

**UNIVERSIDADE ANHEMBI MORUMBI**

**ALEXANDRE SANTOS MELITTO**

**AVALIAÇÃO DO CÂNCER DE MAMA E SEUS SUBTIPOS MOLECULARES  
ATRAVÉS DA ESPECTROSCOPIA RAMAN**

**TESE DE DOUTORADO**

**DOUTORADO EM ENGENHARIA BIOMÉDICA  
PROGRAMA DE PÓS-GRADUAÇÃO STRICTO SENSU**

**São José dos Campos, dezembro/2022**

**UNIVERSIDADE ANHEMBI MORUMBI**

**ALEXANDRE SANTOS MELITTO**

**AVALIAÇÃO DO CÂNCER DE MAMA E SEUS SUBTIPOS MOLECULARES  
ATRAVÉS DA ESPECTROSCOPIA RAMAN**

**TESE DE DOUTORADO**

Tese apresentada ao Programa de Pós-Graduação Stricto Sensu em Engenharia Biomédica – Doutorado, da Universidade Anhembi Morumbi, como requisito parcial para obtenção do título de Doutor em Engenharia Biomédica.

Orientador: Prof. Dr. Landulfo Silveira Junior

**São José dos Campos, dezembro/2022**

# **UNIVERSIDADE ANHEMBI MORUMBI**

**ALEXANDRE SANTOS MELITTO**

## **AVALIAÇÃO DO CÂNCER DE MAMA E SEUS SUBTIPOS MOLECULARES ATRAVÉS DA ESPECTROSCOPIA RAMAN**

Tese apresentada ao Programa de Pós-Graduação Stricto Sensu em Engenharia Biomédica – Doutorado, da Universidade Anhembi Morumbi, como requisito parcial para obtenção do título de Doutor em Engenharia Biomédica. Aprovada pela seguinte Banca Examinadora:

**Prof. Dr. Landulfo Silveira Junior**

Orientador

Doutorado em Engenharia Biomédica

Universidade Anhembi Morumbi

**Prof. Dr. Flávio Aimbire Soares de Carvalho**

UNIFESP

**Prof. Dr. Pedro Luiz de Brito**

Hospital GACC

**Profa. Dra. Adriana Barrinha Fernandes Moretti**

Universidade Anhembi Morumbi

**Prof. Dr. Carlos José de Lima**

Universidade Anhembi Morumbi

**Prof. Dr. Renato Amaro Zângaro**

Universidade Anhembi Morumbi

**Prof. Dr. Leandro Procópio Alves (Suplente)**

Universidade Anhembi Morumbi

**Prof. Dr. Antonio Luiz Barbosa Pinheiro (Suplente)**

Universidade Anhembi Morumbi

**São José dos Campos, dezembro/2022**

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da Universidade, do autor e do orientador.

**ALEXANDRE SANTOS MELITTO**

GRADUAÇÃO EM MEDICINA – UNIFESP / 2002

RESIDÊNCIA MÉDICA EM GINECOLOGIA E OBSTETRÍCIA – UNIFESP / 2005

RESIDÊNCIA MÉDICA EM MASTOLOGIA – UNIFESP / 2007

MESTRADO EM MASTOLOGIA – UNIFESP / 2010

Ficha Catalográfica

M475a Melitto, Alexandre Santos  
Avaliação do câncer de mama e seus subtipos moleculares através da espectroscopia Raman / Alexandre Santos Melitto – 2022.  
83f.: 30 cm.

Orientador: Landulfo Silveira Junior.  
Tese (Doutorado em Engenharia Biomédica) - Universidade Anhembi Morumbi, São José dos Campos, 2022.  
Bibliografia: f. 74-81.

1. Engenharia Biomédica. 2. Espectroscopia Raman. 3. Câncer de Mama. 4. Diagnóstico Óptico. 5. Análise de componente principal.  
I. Título.

CDD 610.28

## **AGRADECIMENTOS**

Agradeço em primeiro lugar a Deus, porque sem Ele nada seria possível.

À minha esposa, por todo companheirismo de uma vida e pelo apoio e compreensão durante essa fase tão importante do meu desenvolvimento profissional.

Ao meu filho, por toda a felicidade que me proporciona e pela mudança de visão do mundo que operou em mim.

Aos meus pais, por terem me criado e formado com tanto amor e carinho.

Em especial, ao Prof. Dr. Landulfo Silveira Junior, pela orientação e amizade.

A todos os professores do Programa de Pós-Graduação em Engenharia Biomédica da Universidade Anhembi Morumbi.

Ao professor Dr. Renato Amaro Zângaro, pela dedicação ao ensino e liderança frente ao Programa de Pós-Graduação da Universidade Anhembi Morumbi.

Ao colega patologista Dr. Victor Eduardo Arrua Arias, pela parceria na obtenção das amostras de tecido mamário.

L. Silveira Jr. agradece à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela aquisição do espectrômetro Raman dispersivo (Processo 2009/01788-5) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de produtividade em pesquisa (Processo No. 306344/2017-3).

## RESUMO

**Introdução:** A espectroscopia Raman tem sido usada para discriminar o câncer de mama humano e seus diferentes subtipos moleculares tumorais (luminal A, luminal B, HER2 e triplo negativo) do tecido de mama normal em amostras cirúrgicas (biópsias). **Objetivos:** Avaliar a discriminação do câncer de mama e seus subtipos moleculares através da espectroscopia Raman. **Materiais e métodos:** Amostras de câncer de mama e fragmentos de tecido normal de 31 pacientes foram obtidos por ressecção cirúrgica e submetidos à espectroscopia Raman com os seguintes parâmetros: laser de excitação de 830 nm, 350 mW, tempo de coleta 30 s. No total, foram obtidos 424 espectros Raman dos 31 fragmentos de câncer de mama e 31 fragmentos de tecido normal. A análise de componente principal (PCA) foi utilizada em uma análise exploratória para identificar diferenças bioquímicas entre os tumores e os tecidos normais. Modelos discriminantes foram desenvolvidos por meio da regressão por mínimos quadrados parciais (PLS) para distinguir os diferentes grupos de subtipos de câncer *versus* normal e entre si. **Resultados:** Os vetores de carregamento da PCA apresentaram características espectrais referentes à constituição bioquímica dos tecidos mamários, tais como lipídios, proteínas, aminoácidos e carotenoides, onde foi encontrada predominância de proteínas nos tumores mamários e lipídios nos tecidos normais. Apesar das pequenas diferenças espectrais entre os diferentes subtipos de tumor e tecidos normais, o modelo discriminante baseado na regressão PLS foi capaz de discriminar os espectros do tumor mamário do tecido normal com acurácia de 97,3% (sensibilidade de 95,4% e especificidade de 99,0%), discriminar entre subtipos luminal, não luminal e normal com acurácia de 89,3% e discriminar cada subtipo molecular individualmente e tecido normal com acurácia de 80,8%. **Conclusão:** A técnica PCA revelou as diferenças bioquímicas entre o câncer de mama e o tecido mamário normal e a regressão PLS discriminou os espectros Raman de tecidos mamários em relação aos subtipos moleculares do câncer, podendo, após estudos adicionais, se tornar uma ferramenta útil para o diagnóstico do câncer de mama.

**PALAVRAS-CHAVE:** Engenharia biomédica, Espectroscopia Raman; câncer de mama; diagnóstico óptico; análise de componente principal; regressão por mínimos quadrados parciais.

## ABSTRACT

### Diagnosing molecular subtypes of breast cancer by means of Raman spectroscopy

**Introduction:** Raman spectroscopy has been used to discriminate human breast cancer and its different tumor molecular subtypes (Luminal A, Luminal B, HER2 and Triple Negative) from normal breast tissue in surgical specimens (biopsies). **Objectives:** To evaluate the discrimination of breast cancer and its molecular subtypes through Raman spectroscopy. **Materials and methods:** Fragments of breast cancer and normal tissue from 31 patients were obtained by surgical resection and submitted to Raman spectroscopy (830 nm, 350 mW excitation, collection time 30 s, laser parameters). In total, 424 Raman spectra were obtained from 31 breast cancer fragments and 31 normal tissue fragments. Principal component analysis (PCA) was used in an exploratory analysis to unveil the biochemical differences between the tumors and normal tissues. Discriminant models based on partial least squares (PLS) regression were developed to distinguish the different cancer subtype groups. **Results:** PCA loading vectors showed spectral features referred to the biochemical constitution of breast tissues, such as lipids, proteins, amino acids, and carotenoids, where proteins were increased in breast tumors. Despite the small spectral differences between the different subtypes of tumor and normal tissues, the discriminant model based on PLS was able to discriminate the spectra of breast tumor and normal tissue with accuracy of 97.3% (95.4% sensitivity and 99.0% specificity), to discriminate the Luminal subtypes, non-Luminal subtype and normal tissue with accuracy of 89.3%, and each molecular subtype and normal tissue with accuracy of 80.8%. **Conclusion:** PCA could reveal the biochemical difference between tumors and normal tissues and PLS could discriminate the Raman spectra of breast tissues regarding the molecular subtypes of cancer, as a useful tool for cancer diagnosis.

**KEYWORDS:** Biomedical engineering, Raman spectroscopy; breast cancer; optical diagnosis; discriminant analysis; principal component analysis; partial least squares regression.

## SUMÁRIO

LISTA DE FIGURAS .....	10
LISTA DE TABELAS .....	11
LISTA DE ABREVIATURAS E SIGLAS.....	12
1. INTRODUÇÃO .....	14
1.1 A MAMA.....	14
1.2 O CÂNCER DE MAMA.....	15
1.3 OS SUBTIPOS MOLECULARES DO CÂNCER DE MAMA .....	17
1.3.1 Os receptores hormonais .....	17
1.3.2 HER2.....	19
1.3.3 Ki-67 .....	22
1.3.4 Subtipos moleculares .....	22
1.4 A ESPECTROSCOPIA RAMAN .....	26
1.4.1 A espectroscopia Raman aplicada ao diagnóstico do câncer de mama.....	31
1.4.2 Métodos de análise multivariada .....	40
1.5 OBJETIVOS: .....	44
1.5.1 Objetivo principal .....	44
1.5.2 Objetivos secundários .....	44
1.6 HIPÓTESE: .....	45
2.MATERIAIS E MÉTODOS.....	45
2.1 FRAGMENTOS DE TECIDO MAMÁRIO .....	46
2.2 ESPECTROSCOPIA RAMAN DAS AMOSTRAS E PRÉ-PROCESSAMENTO DOS ESPECTROS.....	45
2.3 TESTES ESTATÍSTICOS.....	48



2.4 ANÁLISE EXPLORATÓRIA .....	49
2.5 DISCRIMINAÇÃO E CLASSIFICAÇÃO DE TECIDOS NORMAIS E TUMORAIS EM GRUPOS UTILIZANDO A REGRESSÃO PLS.....	51
3.RESULTADOS.....	51
3.1 DESCRIÇÃO DOS CASOS.....	53
3.2 ESPECTROS RAMAN DE TECIDOS NORMAIS E TUMORAIS.....	52
3.3 ANÁLISE EXPLORATÓRIA .....	58
3.4 ANÁLISE DISCRIMINANTE POR PLS.....	62
4. DISCUSSÃO .....	64
4.1 ESTUDOS FUTUROS E PERSPECTIVAS DO TRABALHO.....	71
5. CONCLUSÃO.....	73
REFERÊNCIAS BIBLIOGRÁFICAS .....	74
ANEXO.....	82

## LISTA DE FIGURAS

Figura 1. Anatomia da mama.....	14
Figura 2. Anatomia seccional da mama, vista frontal e lateral.....	15
Figura 3. História natural e fisiopatologia do câncer de mama .....	16
Figura 4. Carcinoma ductal <i>in situ</i> , carcinoma ductal invasivo e carcinoma lobular invasivo. ....	17
Figura 5. Fisiologia dos receptores de estrogênio (RE) e progesterona (RP). ....	18
Figura 6. Estudo imunoistoquímico do câncer de mama. ....	19
Figura 7. HER2 normal e superexpressão do HER2. ....	20
Figura 8. Escore de HER2. ....	21
Figura 9. Pesquisa da superexpressão do HER2. ....	22
Figura 10. Sobrevida das pacientes (em meses) de acordo com a classificação dos subtipos moleculares.....	24
Figura 11. Tipos de espalhamento de luz a partir do decaimento de um estado excitado virtual. ....	28
Figura 12. Espectro Raman de colesterol apresentando bandas vibracionais típicas. ....	29
Figura 13. Diagrama de blocos e foto do espectrômetro Raman.....	31
Figura 14. Média dos espectros Raman normalizados de tecido mamário normal e câncer de mama.....	56
Figura 15. Média dos espectros Raman normalizados de tecido mamário normal e os diferentes subtipos imunoistoquímicos (moleculares) do câncer de mama.....	58
Figura 16. Plotagem dos sete primeiros Escores do PCA, intensidade média e desvio padrão dos PCs dos tecidos mamários normais e dos subtipos moleculares de câncer extraídos do conjunto de dados de Raman.....	59

## LISTA DE TABELAS

Tabela 1. Subtipos moleculares do câncer de mama.....	26
Tabela 2. Características demográficas das pacientes incluídas no estudo.....	53
Tabela 3. Revisão do número de pacientes incluídos no estudo de acordo com o diagnóstico histopatológico e o respectivo diagnóstico imunoistoquímico.....	54
Tabela 4. Principais picos Raman e suas atribuições utilizados para a avaliação do tecido mamário.....	55
Tabela 5. Resultados da aplicação do modelo discriminante baseado em PLS (PLS-DA) nos espectros Raman dos grupos imunoistoquímicos (subtipos moleculares).....	63
Tabela 6. Resultados da aplicação do modelo discriminante baseado em PLS (PLS-DA) nos espectros Raman dos grupos imunoistoquímicos (luminal e não luminal).....	63
Tabela 7. Resultados da aplicação dos modelos discriminantes baseados em PLS (PLS-DA) nos espectros Raman dos grupos tumor e tecido normal.....	63
Tabela 8. Resultados da aplicação do modelo discriminante baseado em PLS (PLS-DA) nos espectros Raman dos grupos imunoistoquímicos (luminal e não luminal excluído o tecido normal).....	64
Tabela 9. Resultados da aplicação do modelo discriminante baseado em PLS (PLS-DA) nos espectros Raman dos grupos imunoistoquímicos (subtipos moleculares excluído tecido normal).....	64

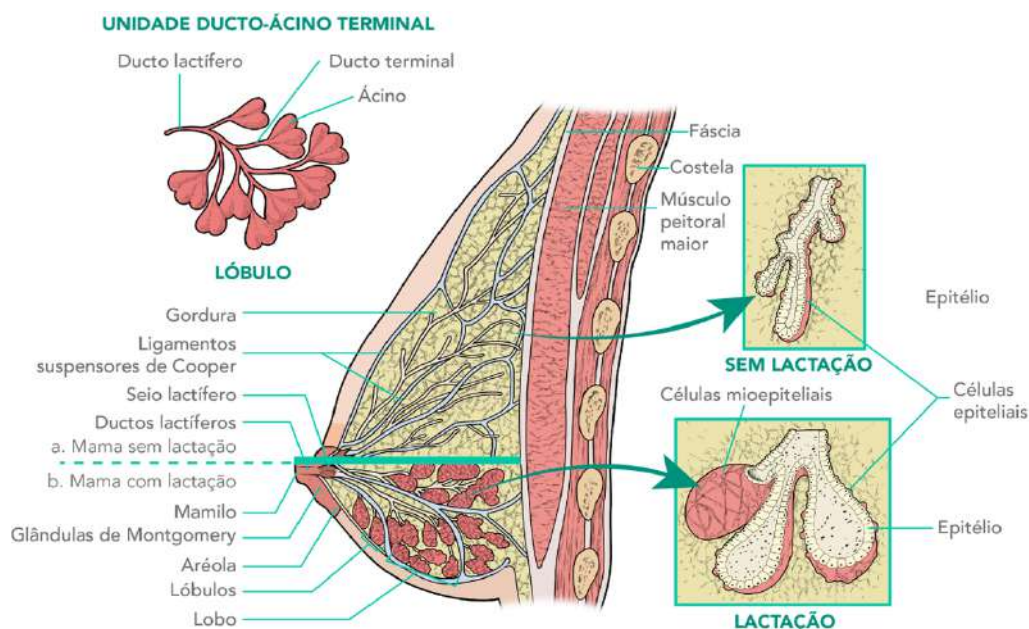
## LISTA DE ABREVIATURAS E SIGLAS

CDI	Carcinoma ductal invasivo
CDIS	Carcinoma ductal <i>in situ</i>
CLI	Carcinoma lobular invasivo
DAID	Isoflavona daidzeína
DNA	Ácido desoxirribonucleico
FISH	Hibridização fluorescente <i>in situ</i>
HDA	Hiperplasia ductal atípica
HER2	Receptor tirosina quinase de membrana do tipo 2
IHC	Imunoistoquímica
INCA	Instituto Nacional do Câncer
Ki-67	Fator de proliferação celular
LA	Luminal A
Laser	Amplificação de luz por emissão estimulada de radiação
LB	Luminal B
NT	Tecido normal
OMS	Organização Mundial da Saúde
PCA	Análise de componente principal
PLS	Regressão por mínimos quadrados parciais
PLS-DA	Análise discriminante por PLS
RE	Receptor de estrogênio
RP	Receptor de progesterona
SVM	Máquina de vetores de suporte
TN	Triplo negativo

# 1. INTRODUÇÃO

## 1.1 A MAMA

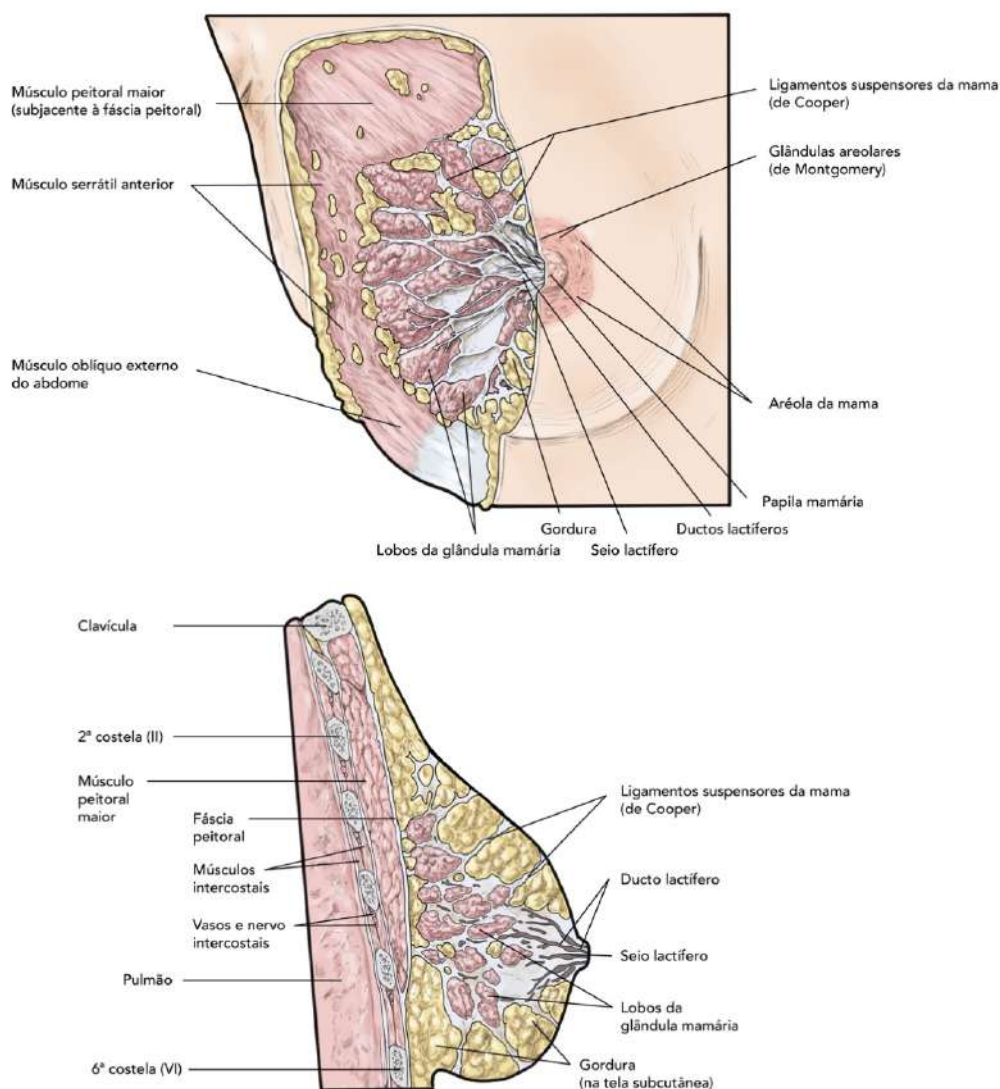
As glândulas mamárias são consideradas glândulas sudoríparas modificadas e altamente especializadas, que são as unidades funcionais mamárias, cuja função é a produção de leite pelos lóbulos. A mama é composta por 12 a 15 sistemas de ductos independentes, sendo que cada um deles drena aproximadamente 40 lóbulos. O tecido mamário se estende até a axila e essa extensão é denominada cauda de Spence. A Figura 1 mostra a anatomia da mama.



**Figura 1.** Anatomia da mama.

Fonte: Adaptado de The Netter Collection of Medical Illustrations, 2a Edição, 2016 (NETTER, 2016).

A mama é recoberta por duas fâscias: anterior (superficial) e posterior (profunda). A fâscia superficial está localizada logo abaixo da pele e tecido celular subcutâneo. A fâscia profunda está localizada atrás da glândula e fica em contato com a fâscia do músculo peitoral maior. Entre essas duas fâscias existem faixas fibrosas (traves) que as conectam; são os ligamentos suspensores de Cooper. Esses ligamentos dão sustentação para a mama e o seu comprometimento por um câncer pode provocar uma retração. A Figura 2 mostra a anatomia seccional da mama.



**Figura 2.** Anatomia seccional da mama, vista frontal e lateral.

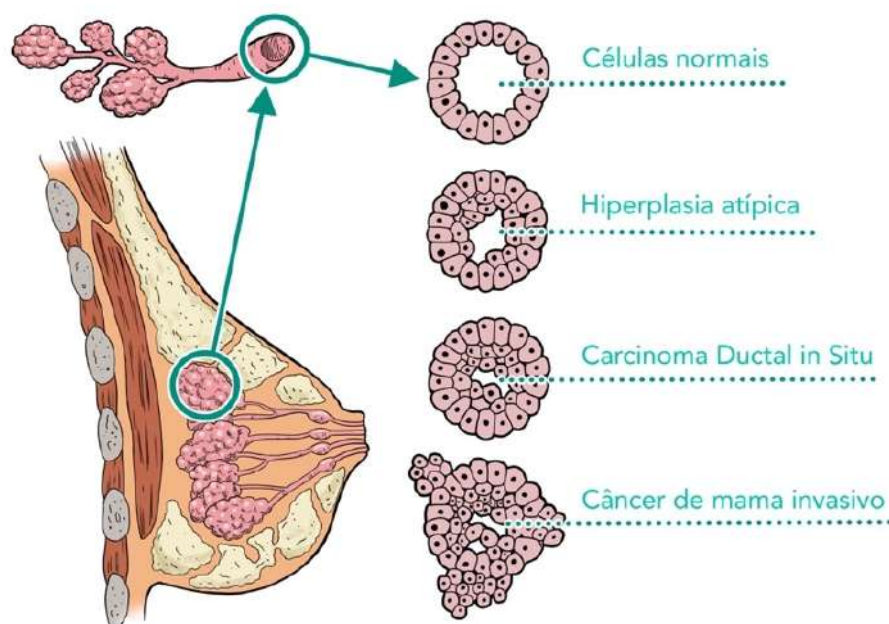
Fonte: Adaptado de The Netter Collection of Medical Illustrations, 2ª Edição, 2016 (NETTER, 2016).

## 1.2 O CÂNCER DE MAMA

Estima-se que a prevalência do câncer de mama em todo o mundo em 2020 seja de 2,3 milhões de mulheres, com uma taxa de mortalidade de 685 mil mulheres. Estima-se que 7,8 milhões de mulheres desenvolveram câncer de mama nos últimos 5 anos, tornando esse câncer o mais prevalente no mundo. Além disso, entre todos os tipos de câncer, o câncer de mama é o principal responsável pela mortalidade das mulheres (BRAY et al., 2018; WHO, 2021). No Brasil, estima-se

66.280 novos casos de câncer de mama a cada ano entre 2020 e 2022. Esse número corresponde a um risco médio de 61,61 novos casos por 100.000 mulheres (INCA, 2016). O diagnóstico do câncer de mama provoca alto impacto psicológico e social nas pacientes, sem mencionar os altos custos envolvidos no tratamento (VINET; ZHEDANOV, 2011).

O tipo histológico mais frequente de câncer de mama é o carcinoma ductal invasivo, representando cerca de 80% dos casos, seguido pelo carcinoma lobular, invasivo que representa cerca de 5 a 15% dos casos (INCA, 2016). A Figura 3 mostra a história natural do carcinoma ductal invasivo.

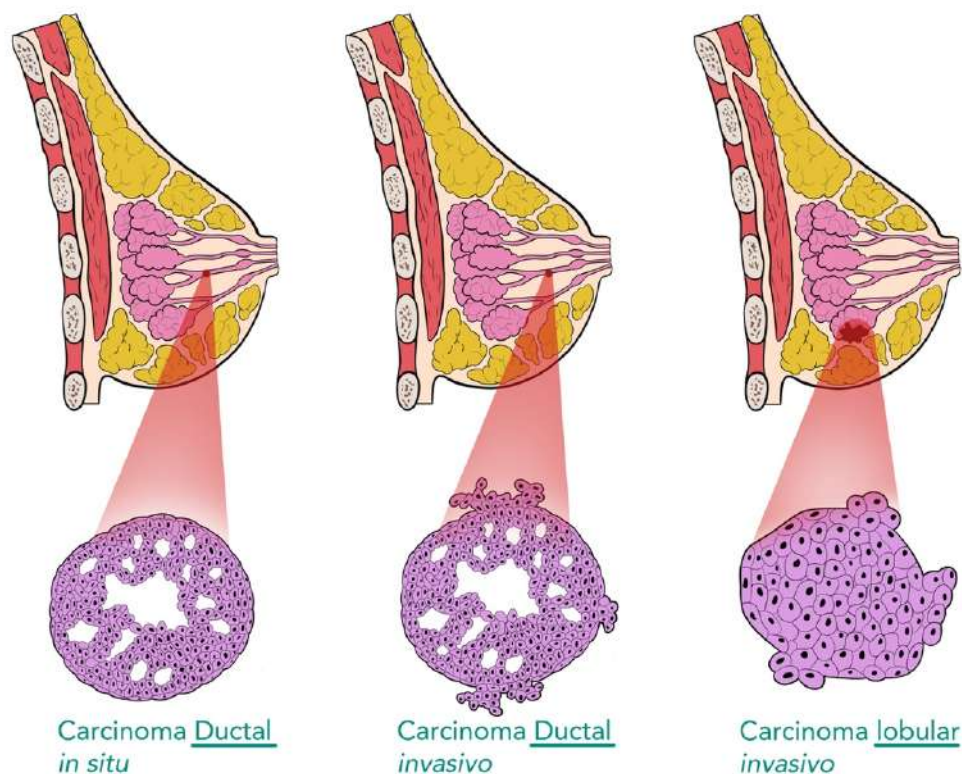


**Figura 3.** História natural e fisiopatologia do câncer de mama

Fonte: Adaptado de The Netter Collection of Medical Illustrations, 2ª Edição, 2016 (NETTER, 2016).

O diagnóstico de câncer de mama em pacientes sem sintomas depende basicamente de estudos de imagem (mamografia, ultrassonografia, tomossíntese e ressonância magnética) que identifiquem e localizem lesões e posteriormente biópsias para obter um diagnóstico patológico anatômico (método padrão ouro) ou avaliação através do método imunoistoquímico (avaliação do subtipo molecular). Tais métodos de diagnóstico são invasivos e às vezes dolorosos, envolvem o uso de radiação ionizante e, em alguns casos, são caros para os serviços públicos de saúde (NOUNOU et al., 2015). O diagnóstico histológico do câncer de mama é feito

através do estudo anatomopatológico. A Figura 4 mostra de forma ilustrada a diferença entre o carcinoma ductal *in situ*, o carcinoma ductal invasivo e o carcinoma lobular invasivo.



**Figura 4.** Carcinoma ductal *in situ*, carcinoma ductal invasivo e carcinoma lobular invasivo.

Fonte: Adaptado de The Netter Collection of Medical Illustrations, 2ª Edição, 2016 (NETTER, 2016).

## 1.3 OS SUBTIPOS MOLECULARES DO CÂNCER DE MAMA

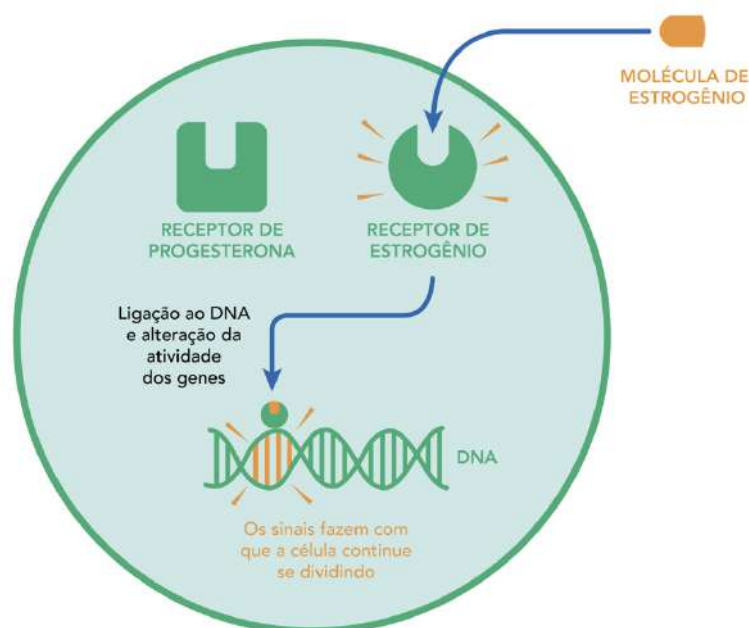
### 1.3.1 Os receptores hormonais

O câncer de mama humano é clinicamente e biologicamente uma doença heterogênea. Esses tumores são múltiplos em sua história natural e resposta aos tratamentos (PRAT et al., 2015). Nos últimos anos, dois grandes alvos moleculares foram identificados na patogênese do câncer de mama, que são os receptores de hormônios estrogênio (RE) e progesterona (RP) e fator de crescimento epidérmico 2 (HER2). O HER2 é um receptor tirosina quinase transmembrana celular que,



quando superexpresso, favorece a proliferação de células tumorais e está associado a um pior prognóstico (WAKS; WINER, 2019).

As células do tecido mamário normal apresentam RE e RP. Alguns cânceres de mama também apresentam RE e RP, como o tecido mamário normal. Por isso, esses tumores são mais diferenciados, menos agressivos e apresentam um prognóstico melhor, com sobrevida maior (GAO; SWAIN, 2018). Além disso, a presença de receptores hormonais no câncer mostra que o estrogênio e a progesterona têm a capacidade de estimular a proliferação desses tumores. Sendo assim, uma das opções de tratamento para esses tipos de tumor é a hormonioterapia, na qual são indicadas medicações que agem inibindo a produção hormonal ou a ação desses hormônios nos receptores (GAO; SWAIN, 2018). A Figura 5 mostra o esquema fisiológico dos RE e RP.

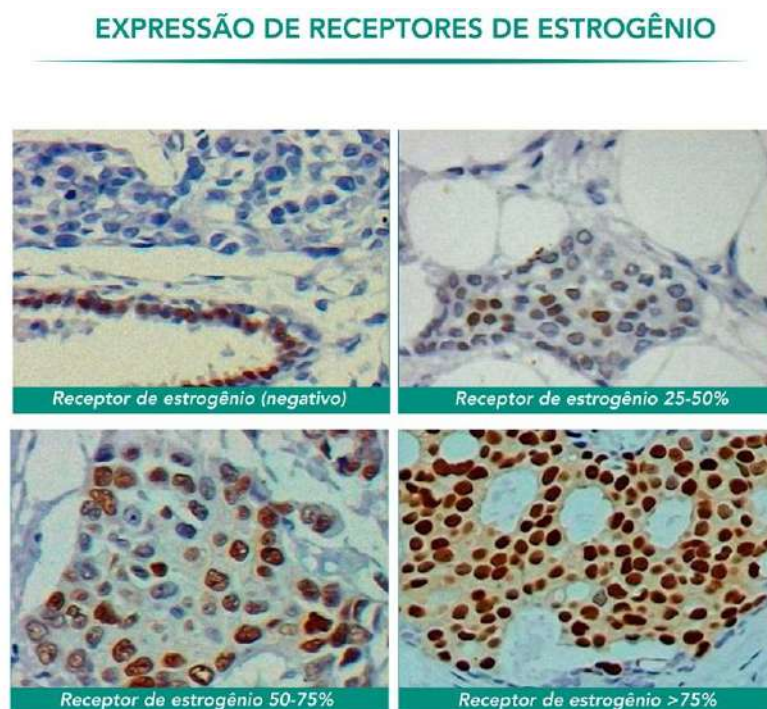


**Figura 5.** Fisiologia dos receptores de estrogênio (RE) e progesterona (RP).

Fonte: Adaptado de GAO et al. (2018).

A avaliação da expressão de receptores hormonais é feita através da aplicação de sondas ao fragmento do tumor. Essas sondas se ligam aos receptores (quando presentes) e mostram se o tumor em análise apresenta receptores ou não. Com isso é possível graduar a quantidade de receptores que o tumor apresenta. Isso é chamado de escore de Alfred. Todo câncer de mama é avaliado quanto à

presença de receptores hormonais através do estudo imunoistoquímico (GOLDHIRSCH et al., 2013). A Figura 6 apresenta o estudo imunoistoquímico dos RE.



**Figura 6.** Estudo imunoistoquímico do câncer de mama.

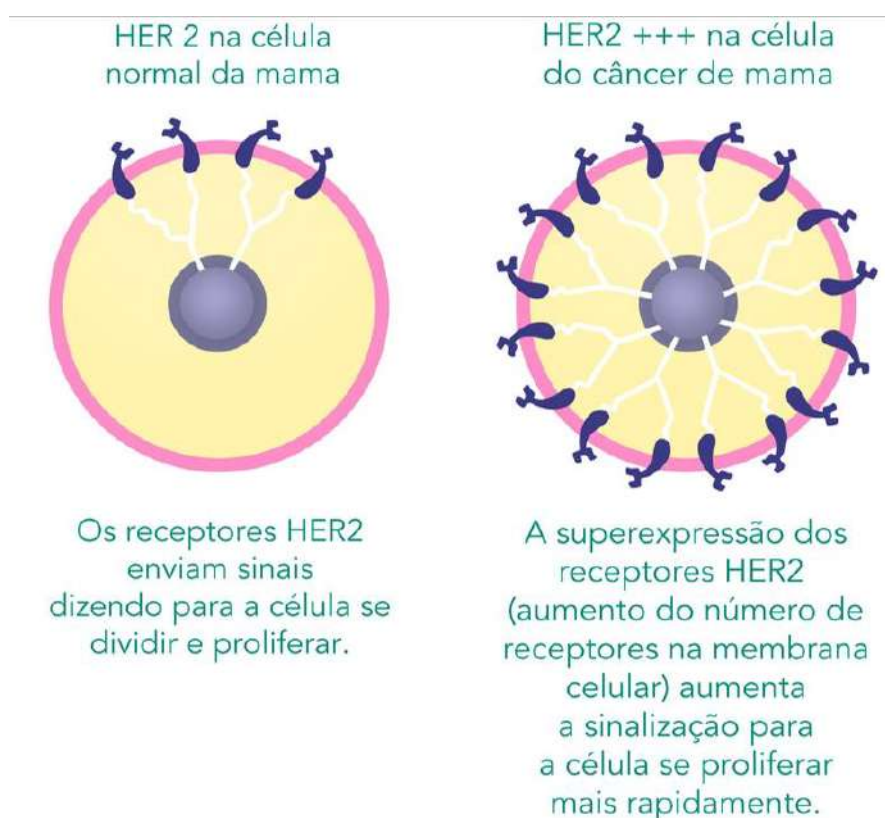
Fonte: adaptado de GOLDHIRSCH et al. (2013).

### 1.3.2 HER2

O receptor do fator de crescimento epidérmico (EGFR) é um receptor da família ErbB. Essa família é formada por quatro genes: ErbB1 ou HER1, ErbB2 ou HER2, ErbB3 ou HER3 e ErbB4 ou HER4 (WOLFF et al., 2018).

O gene HER2 é um proto-oncogene que tem um papel regulador no desenvolvimento e proliferação celular. Um proto-oncogene é um gene que, quando desregulado, pode promover o desenvolvimento de um câncer. O gene HER2 sintetiza a proteína HER2, que é uma proteína que possui um receptor localizado na membrana celular. O receptor HER2 está relacionado com a proliferação das células epiteliais (WOLFF et al., 2018).

A expressão de alguns receptores HER2 nas células é normal, mas a superexpressão ocorre devido a uma mutação no gene HER2, que leva ao aumento do número de receptores HER2 na membrana celular e confere maior capacidade de proliferação. Os tumores que apresentam superexpressão do HER2 são mais agressivos, crescem mais rápido e têm um prognóstico pior (WOLFF et al., 2018). A Figura 7 mostra o esquema de apresentação do HER2 nas células normais e a superexpressão nas células de câncer de mama.

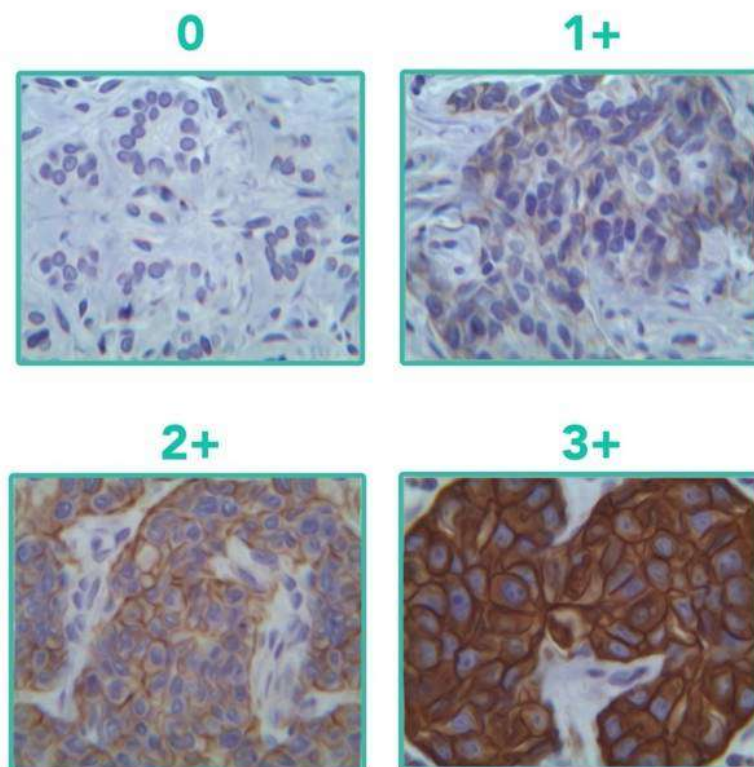


**Figura 7.** HER2 normal e superexpressão do HER2.

Fonte: Adaptado de WOLF et al. (2018).

O estudo imunohistoquímico avalia a expressão do HER2 da mesma forma que avalia a expressão dos RE e RP, através da aplicação de sondas específicas para o receptor HER2 (NASSAR; POOJA; CHIVUKULA, 2020). A Figura 8 apresenta o estudo imunohistoquímico do receptor HER2.

## ESCORES DE REAÇÃO DE TECIDO DE CARCINOMA MAMÁRIO PARA EXPRESSÃO DE HER2



Escores 0 e 1 são considerados negativos;

Escore 3 é considerado positivo;

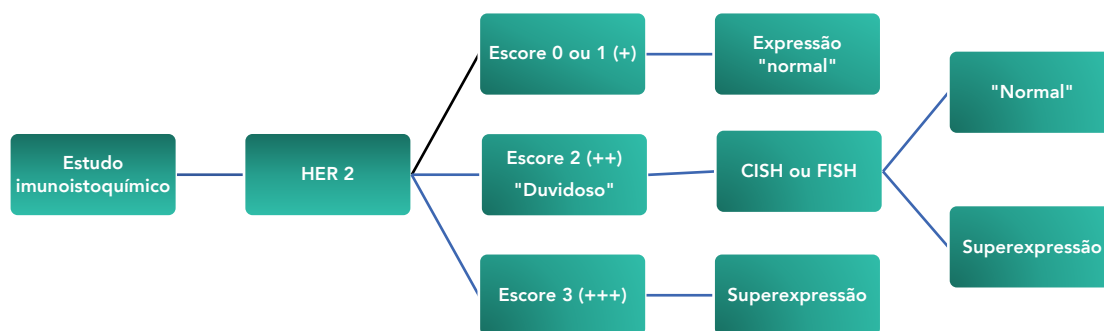
Escore 2 é duvidoso (necessita de confirmação).

**Figura 8.** Escore de HER2.

Fonte: Adaptado de NASSAR et al. (2020).

As diretrizes da CAP/ASCO foram atualizadas em 2013 e 2018. O método atual de pontuação HER2 baseia-se na intensidade e proporção de coloração da membrana com corte de 10%. Tumores com mais de 10% das células com coloração completa e intensa da membrana são considerados Escore 3+ (HER2). Uma coloração de membrana completa fraca a moderada observada em > 10% das células tumorais é considerada Escore 2+ (indeterminada) e posteriormente testada com hibridização *in situ* por fluorescência (FISH – *fluorescence in situ hybridization*) ou a hibridização *in situ* por cromatografia (CISH – *chromatography in situ hybridization*), através da aplicação de fluorescência ou cromatografia para a detecção e quantificação do HER2 nas células tumorais. Tumores com mais de 10% das células com coloração de membrana fraca ou quase imperceptível são considerados Escore 1+ (HER2 negativo) e tumores nos quais não é observada

nenhuma coloração ou a coloração da membrana está incompleta ou fraca em  $\leq 10\%$  das células são considerados Escore 0 (NASSAR; POOJA; CHIVUKULA, 2020). A Figura 9 apresenta o esquema para esclarecimento dos casos duvidosos de expressão do HER2.



**Figura 9.** Pesquisa da superexpressão do HER2.

Fonte: Adaptado de NASSAR et al. (2020).

Somente os tumores que apresentam superexpressão do HER2 têm indicação de tratamento com as drogas alvo específicas para esses receptores (WAKS; WINER, 2019).

### 1.3.3 Ki-67

O Ki-67 é um marcador de proliferação celular. Ou seja, os tumores que apresentam valores altos de Ki-67 estão se proliferando mais e, portanto, são mais agressivos. Em oposição, os tumores com valores baixos de Ki-67 apresentam baixa proliferação celular, sendo menos agressivos. O estudo do Ki-67 também é feito através do perfil imunistoquímico (ŁUKASIEWICZ et al., 2021).

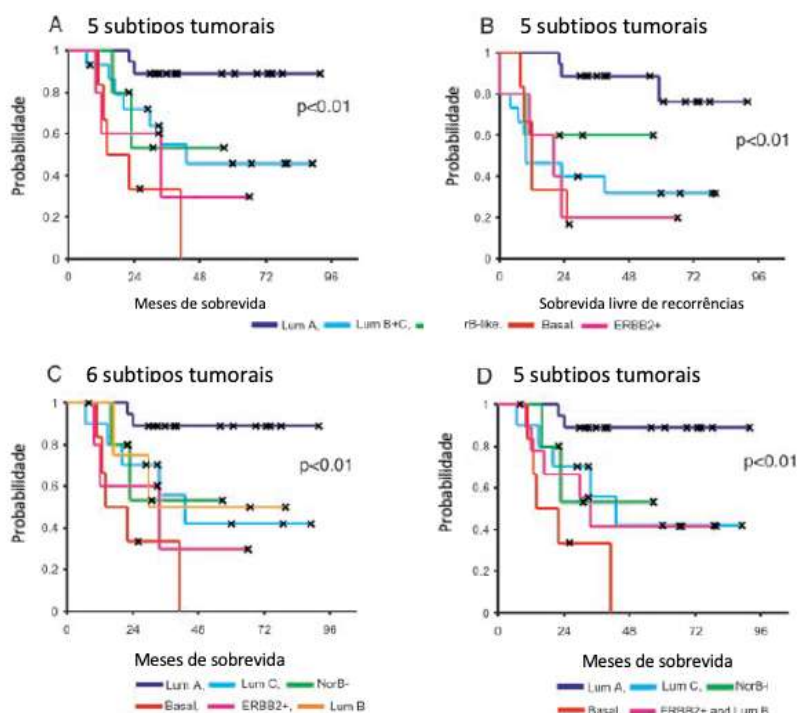
### 1.3.4 Subtipos moleculares

Considerando esses “alvos moleculares” que são investigados por meio do estudo imunistoquímico, utilizando reagentes químicos e o marcador de

proliferação celular Ki-67, quatro subtipos moleculares principais do câncer de mama podem ser caracterizados e tipificados como luminal A, luminal B, HER2 e triplo negativo, apresentando diferenças em relação à sua incidência, fatores de risco, prognóstico e opções de tratamento (PRAT et al., 2015).

Perou et al. (2000) foram os primeiros a propor uma classificação de subtipos de câncer de mama de acordo com o estudo de múltiplos genes e distinção dos tumores de acordo com o seu comportamento biológico e prognóstico. A observação dos padrões de expressão dos genes mostrou “retratos moleculares” relacionados com uma variação na taxa de crescimento e diferença da composição celular tumoral.

Posteriormente, Sorlie et al. (2001) deram continuidade a esse estudo com o objetivo de propor uma classificação dos carcinomas de mama baseada nas variações dos padrões de expressão gênica derivados do DNA complementar (cDNA) e sua correlação com a sobrevida das pacientes. A Figura 10 mostra a diferença de sobrevida dos diversos subtipos moleculares de câncer de mama, com destaque para o melhor prognóstico dos tumores luminal A, com maior sobrevida geral e livre de recorrências, e pior prognóstico dos tumores triplo negativos, com menor sobrevida geral e livre de recorrências.



**Figura 10.** Sobrevida das pacientes (em meses) de acordo com a classificação dos subtipos moleculares.

Adaptado de SØRLIE et al. (2001).

Em 2013, o Consenso de St. Gallen estabeleceu os quatro subtipos moleculares como descritos acima: Luminal A (LA), Luminal B (LB), HER2 e Triplo Negativo (TN). Os tumores LA foram diferenciados dos tumores LB por serem mais sensíveis ao tratamento endócrino e apresentarem melhor prognóstico. Na época foi conceituado que a diferença entre esses subtipos moleculares estava relacionada com a presença ou ausência de RP e variação da taxa de proliferação celular (Ki-67) (GOLDHIRSCH et al., 2013).

O tumor LA é o de melhor prognóstico, com maior sobrevida. Foi definido pelo Consenso de St. Gallen como o tumor que apresenta receptor positivo de estrogênio (RE+), receptor positivo de progesterona ( $RP \geq 20\%$ ), HER2 negativo (Escore 0 ou 1) e Ki-67 < de 14%. Os estudos clínicos de prognóstico que foram baseados em estudos genéticos e/ou imunoistoquímica confirmaram o melhor prognóstico desses casos em comparação com outros subtipos moleculares de câncer de mama (GAO; SWAIN, 2018; GOLDHIRSCH et al., 2013).

Atualmente os subtipos moleculares são definidos da seguinte forma: os tumores do tipo LA são caracterizados pela presença de RE e/ou RP e ausência de

HER2. Os tumores LB são RE positivos e podem ser RP negativos e/ou HER2, com Ki67 > 20%. Os tumores com superexpressão do HER2 (HER2 puros) são caracterizados pela superexpressão do HER2 (3+), com a ausência de RE e RP. Os tumores TN não apresentam RE, nem RP, nem superexpressão HER2 (ŁUKASIEWICZ et al., 2021). A Tabela 1 mostra os subtipos moleculares de câncer de mama de acordo com o perfil imunoistoquímico.



**Tabela 1.** Subtipos moleculares do câncer de mama.

<b>Subtipos moleculares de acordo com o Perfil Imunoistoquímico</b>	
<b>Luminal A</b>	Receptor de Estrógeno (RE) e Progesterona (RP) positivos; HER2 negativo (Escore 0 ou 1+); Ki-67 < 20%.
<b>Luminal B</b>	RE positivo; HER2 positivo ou negativo, RP < 20% e/ou Ki-67 $\geq$ 20%.
<b>HER 2 (puro)</b>	HER2 positivo (Superexpressão=Escore 3+), RE e RP negativos
<b>Triplo negativo</b>	RE e RP negativos e HER 2 negativo

Fonte: LUKASIEWICZ et al. (2021).

A definição do tipo histológico e do subtipo molecular do câncer de mama é fundamental para a decisão do protocolo de tratamento da paciente. O tratamento do câncer de mama está se tornando cada vez mais personalizado. Cada subtipo molecular apresenta riscos, prognósticos e estratégias de tratamento distintas. A terapêutica ideal para cada caso depende do subtipo do tumor, do estadiamento anatômico do câncer e das preferências da paciente (WAKS; WINER, 2019).

Os subtipos moleculares podem ser definidos a partir do estudo imunoistoquímico que avalia a presença dos receptores RE, RP, a superexpressão do HER2 e a taxa de proliferação celular (Ki-67) na amostra do tumor. Mas a literatura descreve uma discrepância de cerca de 20 a 30% quando o estudo imunoistoquímico do espécime cirúrgico é comparado com os testes genéticos realizados na amostra anatomopatológica (YAO et al., 2015).

## 1.4 A ESPECTROSCOPIA RAMAN

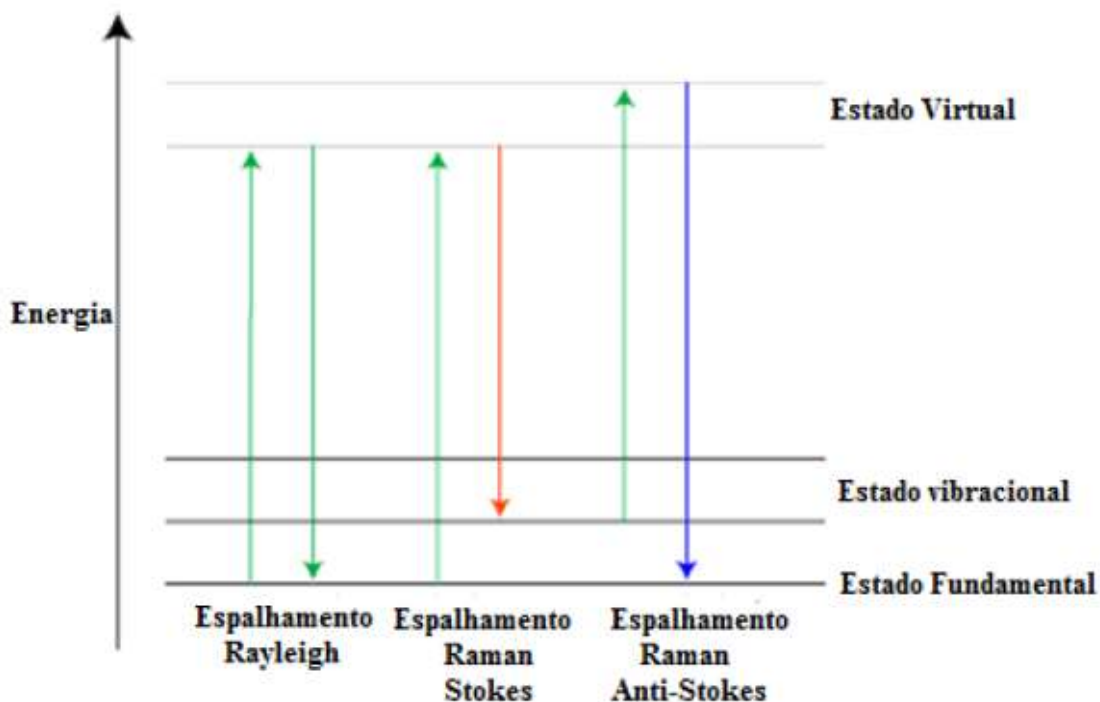
Muitos grupos de pesquisa têm estudado o uso de técnicas ópticas para diagnosticar doenças. Entre essas técnicas, a espectroscopia Raman tem um impacto importante no fornecimento de análise química menos invasiva e não destrutiva de células e tecidos biológicos, pois utiliza luz (radiação não ionizante)

para revelar a composição biomolecular das amostras, fornecendo informações significativas e úteis ao diagnóstico médico (HANLON et al., 2000).

O efeito Raman foi descoberto por Chandrasekhara Venkata Raman em 1928, com sua pesquisa sobre o espalhamento da luz, e seu estudo recebeu o prêmio Nobel dois anos depois. O efeito Raman é um processo fundamental de troca de energia entre a luz e a matéria. Quando a luz incide sobre uma substância, a mesma pode ser absorvida ou espalhada, sendo que a radiação espalhada é composta pela componente elástica (espalhamento Rayleigh) e pela componente inelástica (espalhamento Raman), apresentando mudança na energia da radiação espalhada (deslocamento Raman) devido à modulação na frequência de vibração da nuvem eletrônica induzida pelo laser (luz) causada pela vibração das ligações moleculares.

A maior parte da luz espalhada tem a mesma frequência que a luz incidente (componente elástica – espalhamento Rayleigh). No entanto, uma pequena fração da luz incidente tem a sua frequência alterada, e essa alteração tem relação com a constituição molecular do material. Ocorre um processo de troca de energia entre as moléculas do material e a radiação incidente, com a conseqüente mudança da frequência da radiação espalhada e que traz informação da frequência vibracional das moléculas do material. Esse processo de troca de energia entre as moléculas da matéria e a luz incidente é conhecido como efeito Raman (HANLON et al., 2000; RAMAN; KRISHNAN, 1928).

A Figura 11 mostra os tipos de espalhamento. No espalhamento Rayleigh, a radiação espalhada e a luz incidente têm a mesma frequência. No espalhamento Raman Stokes, a radiação espalhada tem energia menor do que a incidente e no espalhamento Raman anti-Stokes, a radiação espalhada tem energia maior do que a incidente (FRANZEN; WINDBERGS, 2015).



**Figura 11.** Tipos de espalhamento de luz a partir do decaimento de um estado excitado virtual.

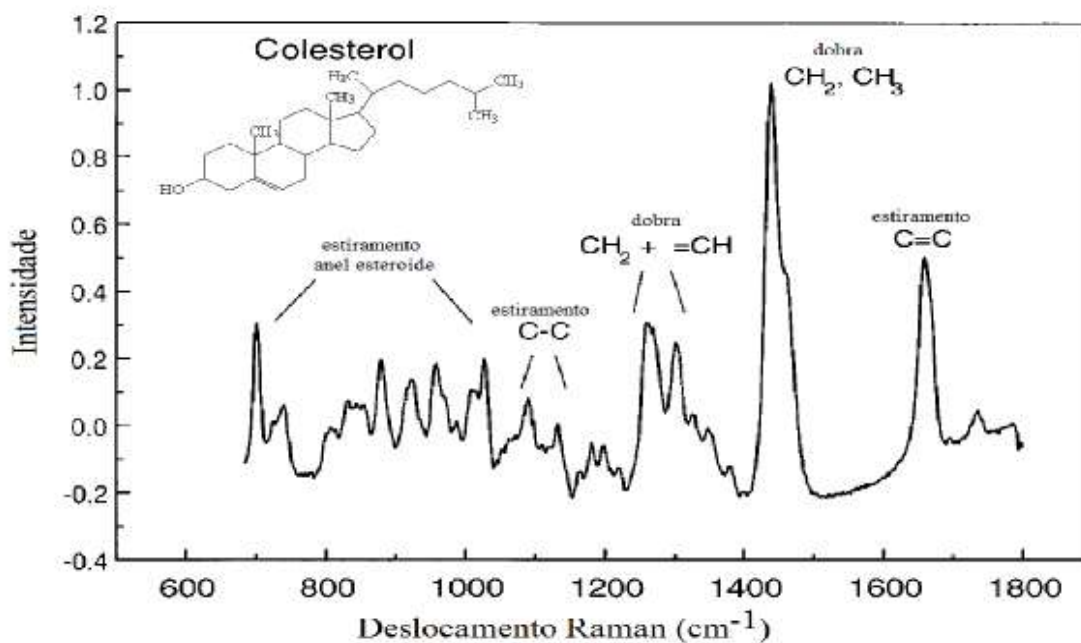
Fonte: SANDEMAN et al. (2015).

A incidência do feixe de luz (radiação eletromagnética) sobre as moléculas da matéria provoca a sua polarização, que é uma transição da molécula do seu estado fundamental para um estado vibracional excitado. O espalhamento Raman ocorre devido a mudanças na polarizabilidade da molécula durante a vibração, acompanhada pela absorção simultânea de um fóton incidente e a emissão de um fóton espalhado inelasticamente (Raman). A luz espalhada Raman pode ser coletada e exibida como um espectro, no qual a intensidade é exibida como uma função da mudança de frequência em relação à frequência de incidência. Como cada substância é formada por diferentes moléculas e possui vibrações moleculares específicas, o espectro Raman mostra uma série de bandas com picos característicos, cada pico deslocado por uma frequência vibracional característica dessa molécula (HANLON et al., 2000; RAMAN; KRISHNAN, 1928).

O espectro Raman é um gráfico da intensidade da luz espalhada *versus* sua mudança em frequência em relação à luz incidente. As mudanças em frequência do espalhamento Raman são convencionalmente medidas em números de onda ( $\text{cm}^{-1}$ ), uma unidade adequada para relacionar a alteração na energia vibracional da molécula com a mudança na frequência da luz espalhada. O valor de

$1 \text{ cm}^{-1}$  é igual a 30 GHz e isso é normalmente 10 mil vezes menor do que a frequência da luz (HANLON et al., 2000).

A Figura 12 mostra um exemplo do espectro do colesterol, que apresenta bandas vibracionais moleculares em  $1440 \text{ cm}^{-1}$ , que representa as vibrações de deformação do metileno e metil ( $\text{CH}_2$  e  $\text{CH}_3$ ) e o pico em  $1670 \text{ cm}^{-1}$ , que representa a vibração de estiramento  $\text{C}=\text{C}$ . Por exemplo, esses picos são encontrados ao se analisar um tecido biológico que contém colesterol em quantidade suficiente para ser detectado pela espectroscopia Raman (HANLON et al., 2000).



**Figura 12.** Espectro Raman de colesterol apresentando bandas vibracionais típicas.

Fonte: Adaptado de HANLON et al. (2000).

Cada tipo de molécula apresenta picos específicos no espectro Raman. Com isso, ao analisar um tecido biológico, pode-se avaliar as substâncias constituintes desse tecido de acordo com os picos presentes no seu espectro Raman. Ou seja, a estrutura molecular e a composição de um material em estudo são codificadas como um conjunto de mudanças de frequência na luz espalhada Raman. Com isso, o espectro Raman pode fornecer a “impressão digital” de uma substância, a partir da qual pode ser determinada a sua composição molecular (HANLON et al., 2000).

A espectroscopia Raman é um método prático, rápido, não destrutivo e específico para avaliação da fisiologia e morfologia celular e tecidual. Com obtenção da assinatura vibracional de moléculas e grupos moleculares, é possível sondar o estado fisiológico de células isoladas para acompanhar as mudanças em sua fisiologia e ampliar esses conceitos para o contexto dos tecidos e sua aplicação para diagnóstico médico (HUSER; CHAN, 2015).

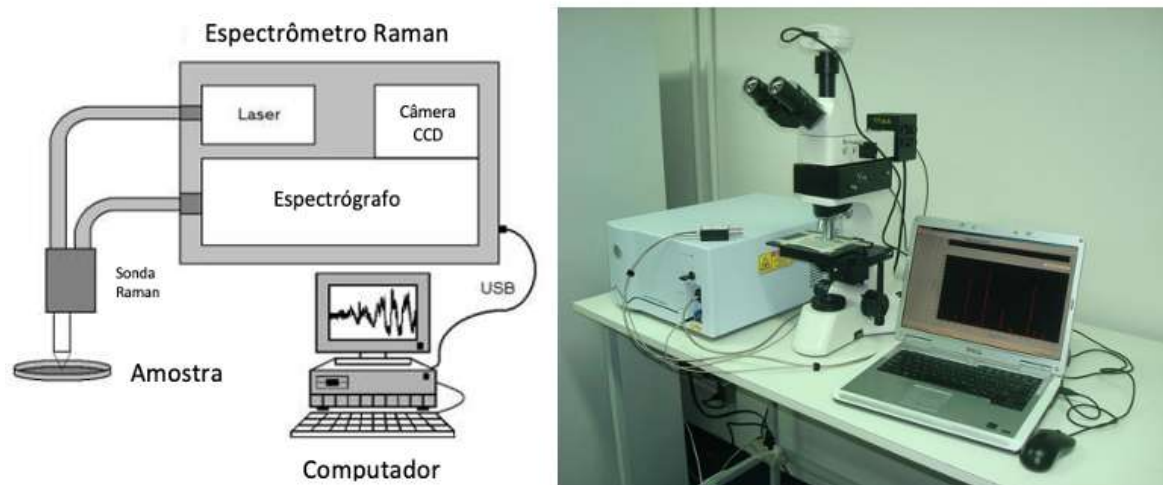
Sendo assim, a constituição molecular do tecido pode ser obtida e a base bioquímica das alterações neoplásicas pode ser evidenciada, auxiliando no diagnóstico dessas patologias. A espectroscopia Raman é preferível em relação a outras técnicas ópticas propostas para o diagnóstico tecidual devido à sua capacidade de gerar uma “impressão digital” espectral que representa a composição bioquímica específica da amostra. A espectroscopia Raman produz uma assinatura espectroscópica que corresponde a aspectos bioquímicos específicos do tecido e pode ser usada para discriminar subtipos de câncer de mama e sua classificação molecular, permitindo, assim, que o método seja introduzido como técnica de diagnóstico molecular (RALBOVSKY; LEDNEV, 2019).

A espectroscopia Raman tem o potencial de identificar marcadores associados a alterações malignas e pode ser usada como ferramenta de diagnóstico para a detecção precoce de lesões pré-neoplásicas e malignas *in vivo*, tendo uma capacidade promissora de ser utilizada como ferramenta auxiliar diagnóstica em tempo real para essas doenças neoplásicas (KALLAWAY et al., 2013).

A espectroscopia Raman possui um alto nível de especificidade bioquímica; no entanto, as diferenças espectrais em diferentes classes teciduais (normal *versus* neoplasia, por exemplo) podem ser muito pequenas e difíceis de serem observadas sem um tratamento estatístico/matemático adequado. Vários algoritmos quimiométricos multivariados foram desenvolvidos para abordar uma ampla gama de questões relacionadas à extração de informações espectrais do conjunto de dados para diagnósticos. Particularmente, as duas categorias de modelos estatísticos multivariados mais importantes que podem ser desenvolvidas incluem técnicas não supervisionadas e supervisionadas. Os algoritmos não supervisionados, como a análise de componente principal (PCA), não usam etiquetas de amostra (como concentração ou grupo a que a amostra pertence por exemplo) ou quaisquer informações definidas pelo usuário quando o modelo está

sendo construído (GELADI, 2003). Esses modelos destinam-se à análise exploratória e à extração de informações e podem ser usados para mostrar diferenças ou semelhanças entre grupos de dados e, em seguida, essas diferenças podem ser usadas para classificação.

A Figura 13 apresenta o diagrama de blocos e uma foto do espectrômetro Raman. O espectrômetro apresenta um elemento dispersor de luz integrado; a coleta do sinal luminoso espalhado é feita por uma sonda Raman (Raman probe), que envia a informação para o espectrógrafo e a uma câmera CCD de alta eficiência. A faixa espectral útil compreende de 400 a 1800  $\text{cm}^{-1}$ . A vantagem da utilização da radiação infravermelha para excitar o material é a minimização da absorção pela excitação de níveis eletrônicos e, conseqüentemente, a minimização da emissão de fluorescência, que pode degradar o sinal Raman, que é fraco, devido à diminuição de fótons disponíveis para o espalhamento Raman e aumento de fótons da fluorescência, que causam ruído.



**Figura 13.** Diagrama de blocos e foto do espectrômetro Raman.

Fonte: PASSONI (2017).

#### 1.4.1 A espectroscopia Raman aplicada ao diagnóstico do câncer de mama

Muitos estudos utilizaram a espectroscopia Raman para o diagnóstico de câncer de mama. Em tais estudos, a espectroscopia Raman foi capaz de

diferenciar lesões malignas de neoplasias benignas e definir o subtipo histológico da lesão (análise morfológica), auxiliando na determinação da base bioquímica de cada tecido.

Alfano et al. (1991) foram os primeiros a fazer uso da espectroscopia Raman por transformada de Fourier (FT-Raman) com excitação em 1064 nm em amostras de tecidos mamários normais, benignos e malignos *ex vivo*. Os autores observaram diferenças espectrais entre os três diferentes tecidos mamários que correspondem a diferenças nos modos de vibração do grupo amida I de proteínas (banda com pico entre 1627 e 1640  $\text{cm}^{-1}$ ).

Haka et al. (2002) estabeleceram um modelo bioquímico e morfológico para determinação da composição de microcalcificações em tecidos mamários humanos por meio de espectroscopia Raman e PCA visando o diagnóstico. As microcalcificações são alterações vistas na mamografia e que podem ser um sinal precoce de malignidade, devendo ser biopsiadas. Mas nem todas as microcalcificações estão relacionadas com o câncer de mama. Microcalcificações do tipo I consistem em oxalato de cálcio e são observadas quase exclusivamente em lesões benignas e muito raramente no carcinoma lobular *in situ*, que é uma lesão pré-neoplásica. Por outro lado, as microcalcificações do tipo II são compostas por fosfatos de cálcio, principalmente hidroxiapatita, com carbonato intercalado, depositada em uma matriz protéica, e estão associadas a lesões proliferativas benignas e malignas, incluindo o carcinoma ductal *in situ* e invasivo. Nesse estudo, foram observadas diferenças bioquímicas nas microcalcificações formadas em ductos benignos e lesões malignas. As microcalcificações benignas apresentaram maior quantidade de carbonato de cálcio e menor quantidade de proteína do que as formadas nos ductos malignos, e foi possível distinguir essas microcalcificações com sensibilidade de 88% e especificidade de 93%.

Haka et al. (2005) usaram a espectroscopia Raman para diagnosticar lesões benignas (tecido normal, doença fibrocística e fibroadenoma) e lesões malignas (carcinoma invasivo) em tecido mamário humano *ex vivo* com base na composição bioquímica. No modelo desenvolvido, foram utilizados coeficientes de ajuste para diferenciar as informações espectrais de lipídios, incluindo colesterol, e colágeno. Observou-se que o tecido normal apresentou abundância de lipídios e as lesões mamárias apresentaram aumento da quantidade de colágeno, sendo o fibroadenoma mais abundante em colágeno. Os autores encontraram sensibilidade

de 94% e especificidade de 96% para o diagnóstico das amostras. Em outro estudo, que também utilizou os mesmos coeficientes de ajuste, os autores avaliaram as margens tumorais durante o procedimento cirúrgico em 31 espectros Raman de 9 pacientes, sendo possível caracterizar se essas margens estavam livres ou não de neoplasia (HAKA et al., 2006). Haka et al. (2009) analisaram espectros Raman adquiridos de espécimes cirúrgicos de câncer recentemente excisados e tecidos normais adjacentes aos locais do tumor de 21 pacientes e, a partir dessa análise, a espectroscopia Raman foi capaz de distinguir o câncer do tecido normal com sensibilidade de 83%, especificidade de 93% e valor preditivo negativo de 99%.

Bitar et al. (2006) também usaram espectroscopia Raman para estabelecer diferenças espectrais entre tecido normal e seis subtipos de patologias mamárias: doença fibrocística, carcinoma ductal *in situ*, carcinoma ductal *in situ* com necrose, carcinoma ductal infiltrante colóide, carcinoma ductal invasivo e carcinoma lobular invasivo. Os autores foram capazes de associar as diferenças bioquímicas entre o tecido normal e os subtipos de câncer de mama com as diferenças nas bandas Raman específicas a proteínas, determinando, portanto, a base bioquímica para cada tecido, com predominância de colágeno nos carcinomas.

Rehman et al. (2007) avaliaram 60 casos de carcinoma da mama, incluindo carcinoma ductal *in situ* e carcinoma ductal invasivo, e sete casos de tecidos normais da mama usando a espectroscopia Raman para distinguir entre os graus nucleares de carcinoma *in situ* e carcinoma invasivo. As diferenças espectrais nas regiões  $960-800\text{ cm}^{-1}$  e  $1400-1080\text{ cm}^{-1}$  mostraram o aumento do conteúdo de proteína e diminuição relativa nos lipídios nos tecidos cancerosos comparado aos normais. Foram observados picos em  $1662$ ,  $1442$  e  $1260\text{ cm}^{-1}$ , decorrentes da estrutura de folha da amida I e de vibrações de alongamento C=O de colágeno. O espectro Raman do tecido normal se mostrou com intensidade mais fraca do que o carcinoma invasivo e a intensidade dos picos aumentaram com o aumento do grau tumoral, com o grau III tendo a maior intensidade, grau II intensidade intermediária e grau I a menor intensidade. O aumento na intensidade dos picos OH, NH e CH sugeriu uma mudança na concentração dos lipídios, proteínas e conteúdo de DNA. A intensidade absoluta de cada um dos picos OH, NH e CH na faixa de  $3500$  a  $2700\text{ cm}^{-1}$  varia com o aumento do grau (I, II e III), indicando concentrações variáveis de cadeias de ácidos graxos, fosfolipídios, colesterol, creatina, proteínas e



ácidos nucleicos. Não houve diferença significativa nos espectros entre os graus de carcinoma ductal *in situ*. Os autores concluíram que a espectroscopia Raman pode distinguir objetivamente entre os graus de carcinomas ductais *in situ* e invasivo e que essa diferença pode ser usada como um método diagnóstico para o câncer de mama.

Kong et al. (2014) usaram a espectroscopia Raman para avaliar as margens de ressecção tumoral nas cirurgias conservadoras da mama. Foi usado um modelo espectral que distinguiu o câncer do tecido mamário normal com uma sensibilidade de 95,6% e especificidade de 96,2%. Nesse modelo, foi observado que o câncer apresentou aumento de ácidos nucleicos e redução de colágeno e lipídios. O estudo sugeriu que a microscopia Raman de amostragem seletiva tem o potencial de fornecer um diagnóstico intraoperatório rápido e objetivo para detectar o carcinoma mamário e avaliar as margens de ressecção.

Talari et al. (2015) usaram a espectroscopia Raman para diferenciar duas linhagens de células de câncer de mama (MDA-MB-436 e MCF-7) e uma de tecido mamário normal (MCF-10A) *in vitro*. Nesse estudo foram observadas diferenças na concentração de compostos bioquímicos como lipídios, ácidos nucleicos e proteínas. As concentrações de bases nitrogenadas de ácido nucleico eram diferentes nas linhagens celulares e os espectros de MCF-7 apresentaram uma quantidade relativamente maior de bases de DNA. As proteínas, especialmente nas regiões espectrais de amida II e III, e os lipídios foram considerados diferentes entre as linhagens de células, sugerindo que a concentração de lipídios, proteínas e ácidos nucleicos podem contribuir para a diferenciação. A PCA foi capaz de identificar variações que discriminaram de forma precisa e confiável as três linhagens celulares. Os dados espectrais Raman combinados com a PCA mostraram ser um excelente método que permite a separação de diferentes tipos de linhagens celulares com base em seus teores de lipídios, ácidos nucleicos (DNA) e proteínas. A linhagem celular MCF-7 (câncer) apresentou maior concentração de lipídios em comparação com MDA-MB 436 (câncer) e MCF 10A (tecido normal). Essa linhagem também apresentou maiores concentrações de ácido nucleico/base de DNA em relação às outras duas linhagens celulares. O modelo de análise discriminante linear (LDA) foi capaz de diferenciar as três linhagens celulares com 100% de sensibilidade e 91% de especificidade. Esse estudo mostrou que a combinação de espectroscopia Raman com a quimiometria é

capaz de diferenciar o câncer de mama do tecido normal e também é capaz de diferenciar as diferentes linhagens celulares de câncer de mama, podendo ser útil para identificar diferentes subtipos de câncer de mama.

Li et al. (2015) discriminaram tecidos mamários normais de tecidos cancerosos com o uso da espectroscopia Raman no infravermelho próximo (785 nm) através da aplicação de uma técnica de hiperplano local adaptativo K (ALHK) com uma acurácia de 93,2%. Nesse estudo foram observados, no tecido normal, quatro picos atribuíveis às moléculas de lipídios (1078, 1447, 1653 e 1747  $\text{cm}^{-1}$ ), enquanto o tumor apresentou apenas dois picos lipídicos (1078 e 1750  $\text{cm}^{-1}$ ). Além disso, esses picos se mostraram atenuados no câncer. Por outro lado, no tecido tumoral foram observados picos em 1243, 1308, 1453 e 1663  $\text{cm}^{-1}$ , que representam as moléculas de proteínas, enquanto o tecido normal apresentou um único pico proteico em 1305  $\text{cm}^{-1}$ . Os picos atribuídos às moléculas de ácido nucleico em 815 e 827  $\text{cm}^{-1}$  se apresentaram atenuados no tecido canceroso. As proporções de proteínas foram significativamente aumentadas nos tumores e as proporções de lipídios foram significativamente aumentadas no tecido normal.

Sathyavathi et al. (2015) aplicaram a espectroscopia Raman para analisar as microcalcificações presentes em fragmentos de biópsia de lesões mamárias. Foi usado um modelo quimiométrico para caracterizar a composição química das lesões. Foi observado que o aumento de carbonato intercalado na rede de hidroxiapatita pode ser empregado de forma confiável para diferenciar lesões benignas de lesões malignas. Usando algoritmos baseados apenas no conteúdo de carbonato e proteína citoplasmática, obteve-se um excelente valor preditivo negativo (93 a 98%).

Brozek-Pluska et al. (2016) usaram a espectroscopia Raman para detectar processos de acetilação e metilação em células epiteliais de câncer de mama (carcinoma ductal e lobular) e para identificar e discriminar estruturas em tecidos normais e cancerosos. Demonstraram, em seu estudo, que a vibração do grupo acetil observada em torno de 2938-2942  $\text{cm}^{-1}$  e do grupo metil em 2970  $\text{cm}^{-1}$  é útil no monitoramento de processos moleculares epigenéticos desse câncer. O marcador vibracional Raman se mostrou com desvio para o azul, de 2905  $\text{cm}^{-1}$  (vibrações C–H do grupo funcional não acetilado) para células normais, a 2942  $\text{cm}^{-1}$  (vibrações do grupo funcional acetilado) para células cancerosas. A sensibilidade e especificidade obtidas com o uso da análise discriminante através da regressão por

mínimos quadrados parciais (PLS-DA) foi de 86,1% e 91,3% para calibração e 85,3% e 91,3% para validação cruzada, respectivamente. Com isso, os autores concluíram que a acetilação está aumentada no câncer e que esse processo pode estar envolvido no desenvolvimento da doença.

Han et al. (2017) estudaram a aplicação da espectroscopia Raman em 39 amostras congeladas de pacientes que haviam sido submetidas a ressecção cirúrgica ou biópsia com agulha grossa, sendo nove tecidos mamários normais, sete hiperplasias ductais atípicas (HDA), oito carcinomas ductais *in situ* (CDIS) e quinze carcinomas ductais invasivos (CDI), com o objetivo de encontrar as diferenças e relações entre essas lesões. Foi usada a validação cruzada “deixe-um-de-fora” (*leave-one-out* – LOOCV) e a função de base radial (RBF) para construir um modelo de diagnóstico utilizando máquina de vetores de suporte (SVM). Foram observadas diferenças expressivas entre os espectros Raman médios dos tecidos normais, tecidos HDA, CDIS e CDI, sendo que o mais notável foi o aumento das proteínas e a redução dos níveis de lipídios na HDA em comparação com os tecidos normais. As principais diferenças de espectrais observadas foram um amplo pico de CH<sub>2</sub> (1301 cm<sup>-1</sup>), o pico atribuído aos carotenoides (1526 cm<sup>-1</sup>), uma banda relativamente forte de amida I (1656 cm<sup>-1</sup>) e uma banda de ácidos nucleicos (882 cm<sup>-1</sup>). Considerando a progressão da doença, foi observado que as amostras de CDI apresentaram menores variações de constituintes do que as amostras CDIS e HDA. A acurácia geral observada foi de 74,39%. As sensibilidades observadas do tecido normal, HDA, CDIS e CDI, foram de 62,5%, 50%, 90% e 66,7%, respectivamente. As especificidades observadas do tecido normal, HDA, CDIS e CDI, foram de 100%, 100%, 66,7% e 89,06%, respectivamente. Foi observada diferença significativa entre a HDA e o tecido normal e o câncer nos picos de CH<sub>2</sub> e amida I.

Winnard et al. (2017) estudaram linhagens de células cancerígenas metastáticas específicas da mama, usando um modelo de xenoinxerto ortotópico. A espectroscopia Raman e a PLS-DA foram utilizadas para diferenciar metástases específicas dos órgãos e categorizar as diferentes linhagens celulares, obtendo-se 96,8% de precisão, mostrando que a propagação metastática leva a adaptações específicas das células cancerígenas em nível molecular. Isso sugere a possibilidade de usar a espectroscopia Raman para diagnosticar metástases do câncer de mama.

Li et al. (2017) usaram a espectroscopia Raman para o diagnóstico precoce de câncer de mama e verificar a margem do tumor com precisão e rapidez no estágio inicial do tumor. Nesse estudo, a acurácia da discriminação foi de 93,1%, a sensibilidade foi de 99,2%, a especificidade foi de 79,7%, o valor preditivo positivo foi de 91,6% e o valor preditivo negativo foi de 97,9%. Os picos Raman nos tecidos normais ocorreram em 872, 1078, 1305, 1447, 1659 e 1747  $\text{cm}^{-1}$ , e os picos Raman nos tumores ocorreram em 870, 1093, 1278, 1305, 1447, 1466, 1659 e 1747  $\text{cm}^{-1}$ . De acordo com os autores, os picos nos espectros do tecido normal são atribuídos a moléculas de lipídios, enquanto os picos nos espectros de câncer são atribuídos a moléculas de proteínas.

Zúñiga et al. (2019) também usaram 164 espectros Raman para distinguir tecido normal de câncer de mama durante a mastectomia de 6 pacientes com sensibilidade de 90%, especificidade de 86% e precisão de 88%. Foi observado no câncer de mama picos em 785  $\text{cm}^{-1}$  (nucleotídeos de DNA e RNA), em torno de 840 e 941  $\text{cm}^{-1}$  (colágeno e glicogênio), 1004  $\text{cm}^{-1}$  (fenilalanina), 1650  $\text{cm}^{-1}$  (amida I), 1550  $\text{cm}^{-1}$  (amida II) e 1300  $\text{cm}^{-1}$  (amida III). Também foram observados picos em 1004, 1006, 1152, 1158, 1259 e 1518  $\text{cm}^{-1}$  (carotenoides).

Koya et al. (2020) aplicaram a espectroscopia Raman e examinaram 88 amostras de tecido mamário congeladas, incluindo 44 amostras de tecido normal e 44 fragmentos de tumor usando um algoritmo de aprendizado profundo com rede neural convolucional (CNN) que foi desenvolvido para distinguir características histopatológicas. Houve discriminação do câncer da mama em relação ao tecido normal com acurácia de 90%, sensibilidade de 88,8% e especificidade de 90,8%. Nos fragmentos de tumor foram observados picos Raman nas frequências de 1002  $\text{cm}^{-1}$  (fenilalanina), 869  $\text{cm}^{-1}$  (prolina, estiramento C–C de hidroxiprolina e colágeno), e 1309  $\text{cm}^{-1}$  ( $\text{CH}_3/\text{CH}_2$ ). Foram produzidas “imagens” Raman com o uso do algoritmo, associando cores aos picos, e com isso, foi feita uma correlação entre as imagens e as amostras de tecido coradas com hematoxilina e eosina, que permitem a visualização clara dos limites entre o tecido adiposo normal, tecido conjuntivo e tumor. A espectroscopia Raman foi capaz de fornecer “imagens” compatíveis com as características microscópicas de alta precisão, mostrando que o método pode auxiliar na avaliação das margens da peça durante uma cirurgia conservadora da mama.

A espectroscopia Raman foi capaz de desvendar as diferenças bioquímicas entre os subtipos moleculares. Diferenças na quantidade de lipídios, proteínas e ácidos nucleicos podem ser capazes de identificar os respectivos subtipos moleculares.

Bi et al. (2014) utilizaram espectroscopia Raman para diferenciar tumores HER2 positivos (superexpressos) e HER2 negativos e obtiveram 99% de sensibilidade e especificidade na análise de três linhas celulares diferentes. As células positivas HER2 apresentaram maior teor lipídico e menor teor proteico. Os autores do estudo interpretaram que o aumento de lipídios nas células cancerígenas de mama ocorre devido à síntese exacerbada de ácidos graxos e fosfolipídios e esse aumento na síntese de ácidos graxos estaria associado à superexpressão dos receptores de tirosina quinase HER2 nas células, que é a característica que define o subtipo molecular HER2 positivo.

Medeiros et al. (2016) estudaram o impacto da isoflavona daidzeína (DAID), que é semelhante ao estrogênio, em células cancerosas de mama humana dos tipos MDA-MB-231 (RE-) e MCF-7 (RE+), utilizando espectroscopia Raman e PCA. A DAID pode competir com estrogênio endógeno em casos de câncer de mama dependente de estrogênio (RE+) e pode inibir o crescimento de células cancerígenas ou induzir sua apoptose. A espectroscopia Raman e a análise PCA mostraram um mecanismo de ação diferente da DAID em cada linha celular estudada, de acordo com a expressão dos RE. O PC1 mostrou que nas células MDA-MB-231 (ER-) houve uma interferência predominante da DAID com aminoácidos (fenilalanina a  $1003\text{ cm}^{-1}$ ), glicoproteínas ( $\text{CH}_3$  a  $1375\text{ cm}^{-1}$ ) e amida I (em cerca de  $1660\text{ cm}^{-1}$ ), além de fosfatos de DNA ( $1090\text{ cm}^{-1}$ ) e lipídios ( $1450\text{ cm}^{-1}$ ). A PCA mostrou que, nas células MCF-7 (ER+), os principais componentes celulares afetados pelas isoflavonas foram os ácidos nucleicos de DNA ( $800$ ,  $1090$  e  $1480\text{ cm}^{-1}$ ) e lipídios ( $1450\text{ cm}^{-1}$ ). Isso mostra que, nas células que possuem RE, o mecanismo de ação foi a inibição da estimulação à mitose (que afeta principalmente o DNA) e, em células que não possuem RE, o mecanismo de ação foi a interferência com a síntese proteica, afetando a síntese de fenilalanina, proteínas e lipídios.

Téllez-Plancarte et al. (2018) observaram picos Raman em  $1002$  e  $1030\text{ cm}^{-1}$  (atribuídos a fenilalanina) mais intensos para linhagens celulares de câncer de

mama SKBR3 e MCF-7 (HER2) do que para as linhagens celulares T47D e MDA-MB-231 (não HER2).

Bendau et al. (2020) usaram a espectroscopia Raman para diferenciar tumores TN e não TN, uma vez que os primeiros apresentam pior prognóstico. Eles foram capazes de diferenciar características moleculares relacionadas a colágeno entre tumores metastáticos TN e tumores não metastáticos com RE positivo.

Deng et al. (2022) utilizaram os espectros Raman para diferenciar amostras de câncer de mama de diferentes subtipos moleculares e tecido mamário normal, levando em consideração as contribuições do glicogênio e lipídios (colesterol, fosfatidilserina e ácido esteárico), com uma acurácia > 97% (sensibilidade > 97% e especificidade > 97%). Nas linhagens celulares MCF10A, MDA-MB-231 e SK-BR-3, observou-se pico em  $1436\text{ cm}^{-1}$ , atribuídos a lipídios/proteínas. Na linhagem celular BT-474, observou-se um pico de  $1438\text{ cm}^{-1}$ , que foi atribuído a ácidos graxos saturados e triglicérides. Na linhagem celular MCF-7, observou-se um pico em  $482\text{ cm}^{-1}$  atribuído ao glicogênio; observou-se elevado nível de glicogênio nas linhas de células MCF-7 e BT-474, ER+ e PR+. Maiores concentrações de colesterol foram observadas na linha celular MCF 10A (tecido normal) do que nas linhagens tumorais, e maiores concentrações de colesterol nas linhas MDA-MB-231 e SK-BR-3 (triplo negativo e HER2) em comparação com as linhas MCF-7 e BT-474 (luminal A e luminal B).

Zhang et al. (2022) aplicaram técnicas de aprendizagem de máquina ao espectro Raman para distinguir células normais do câncer de mama com uma acurácia superior a 97% e classificaram subtipos de câncer de mama com uma acurácia superior a 92%, através do uso de dois algoritmos de aprendizagem: PCA análise de função discriminante (DFA) e PCA por máquina de vetores de suporte (SVM), e mostraram que as células cancerígenas de mama tinham níveis mais elevados de glicogênio (pico de  $490\text{ cm}^{-1}$ ) do que o tecido normal. Além disso, as células cancerígenas de mama apresentaram níveis mais elevados de colágeno e lipídios (picos de  $1333$  e  $1304\text{ cm}^{-1}$ , respectivamente).

Santos et al. (2022) aplicaram a microespectroscopia Raman e espectroscopia no infravermelho próximo por transformada de Fourier (FT-IR) e depois de analisar os componentes principais dos dados, foram capazes de discriminar entre o câncer de mama triplo negativo (TNBC) e não triplo negativo

(NTN), bem como câncer triplo negativo mesenquimal e linhagens altamente metastáticas e pouco metastáticas. A PCA permitiu discriminar entre TNBC (MDA-MB-231, HCC-1143 e MDA-MB-468) em comparação com linhagens celulares NTN (MCF-7). Houve maior contribuição do DNA (pico de  $1238\text{ cm}^{-1}$ ) e menor contribuição do RNA/DNA (pico de  $1302\text{ cm}^{-1}$ ) e da amida I (pico de  $1650\text{ cm}^{-1}$ ) para células tumorais TNBC. A diferenciação das linhas TNBC e NTN baseou-se principalmente no DNA e nas proteínas, que predominaram nas linhas TNBC, enquanto os picos referentes a lipídios pareciam ser predominantes em linhas celulares NTN.

#### **1.4.2 Métodos de análise multivariada**

A análise multivariada é um método que tem por objetivo analisar dados que contêm simultaneamente múltiplas medidas. Dentre as técnicas de análise multivariada destacam-se a PCA e a PLS.

A PCA é uma técnica estatística multivariada na qual se transforma linearmente um conjunto de variáveis correlacionadas entre si em um conjunto significativamente menor de variáveis não correlacionadas entre si, mas que apresenta a parte mais representativa do conjunto de variáveis inicial (HONGYU; SANDANIELO; OLIVEIRA JUNIOR, 2015).

A PCA tem como objetivo a redução do volume de dados, mas com a menor perda possível da informação. Isso é feito através da distribuição da variação observada nos eixos originais de tal maneira que se obtenha um conjunto de eixos ortogonais não correlacionados. Assim sendo, na PCA é feita a transformação de um conjunto de variáveis original em um novo conjunto de variáveis chamados de componentes principais, sendo que cada componente principal é uma combinação linear de todas as variáveis originais, independente dos outros componentes principais e tem o objetivo de manter o máximo de informação da variação total contida nas informações originais (HONGYU; SANDANIELO; OLIVEIRA JUNIOR, 2015).

O principal objetivo da PCA é explicar a variância e covariância de um vetor aleatório através de combinações lineares das variáveis originais, os componentes principais, que não se correlacionam entre si. A PCA pode ser utilizada para criar

índices e agrupar objetos de estudo de acordo com a variação contida no grupo, ou seja, os objetos são agrupados de acordo com as suas variâncias dentro da amostra total. Assim sendo, a PCA agrupa os objetos de acordo com a variação das suas características. A PCA é capaz de criar combinações lineares com a máxima variância através do estabelecimento de um índice que possibilita ordenar um conjunto de objetos de acordo com um critério definido por um conjunto de variáveis adequadas (HONGYU; SANDANIELO; OLIVEIRA JUNIOR, 2015).

As técnicas de análise multivariada foram criadas para solucionar problemas específicos, como, por exemplo, definir porque alguns alunos apresentavam notas maiores em comparação com outros alunos, mas podem ser usadas para resolver problemas como a redução da dimensão das variáveis e agrupamento de objetos de estudo através de similaridades, como as necessárias na interpretação dos espectros Raman e as diferenças que ocorrem em grupos distintos (HONGYU; SANDANIELO; OLIVEIRA JUNIOR, 2015).

As principais vantagens da PCA são retirar a multicolinearidade das variáveis, já que permite transformar o conjunto de variáveis originais correlacionadas em um novo conjunto de variáveis não correlacionadas, além de redução das variáveis a eixos ortogonais, explicando a variação dos dados de forma independente e decrescente (HONGYU; SANDANIELO; OLIVEIRA JUNIOR, 2015).

As desvantagens do PCA são a sensibilidade não desejada aos “outliers”, além de não ser indicada nos casos de dados com muitos valores de 0 na matriz ou ausência de dados. Também não é indicada quando o número de variáveis excede o número de amostras, pois, nesses casos, quando se reduz o número de variáveis, ocorre grande perda de informação. Em alguns casos a PCA também não é adequada porque mesmo após a redução, ainda existem muitas variáveis, principalmente quando as variáveis originais são pouco correlacionadas (HONGYU; SANDANIELO; OLIVEIRA JUNIOR, 2015).

Outras técnicas de análise multivariada são modelos supervisionados, como a regressão por PLS e a regressão por componentes principais (PCR), que consideram rótulos (especificamente classes ou concentrações) definidos pelo usuário para que todas as amostras construam modelos de previsão para novas amostras visando classificação ou regressão (RALBOVSKY; LEDNEV, 2019).



A análise discriminante por PLS (PLS-DA) é um método de análise multivariada supervisionado usado para classificação de objetos dentro de grupos. Através desse método é possível determinar a qual grupo pertence uma amostra desconhecida (validação do modelo ou previsão de novas amostras) a partir de informações previamente fornecidas ao sistema (calibração do modelo) (DE SOUZA et al., 2013).

Nos métodos supervisionados existe a necessidade de conhecimento prévio sobre as amostras e seus grupos (classificação utilizando técnicas clássicas) para que sejam definidas regras a serem usadas para a classificação dessas amostras (DE SOUZA et al., 2013).

Na PLS-DA é utilizada a técnica de regressão multivariada por PLS, que é um método de calibração inversa que utiliza a técnica de PCA para a redução da dimensão dos dados e procura uma relação direta entre uma matriz  $\mathbf{X}$  e uma matriz  $\mathbf{Y}$ , sendo equivalente a calcular a PCA tanto de  $\mathbf{X}$  quanto de  $\mathbf{Y}$  simultaneamente, e que apresente a propriedade a ser pesquisada, que é a covariância entre  $\mathbf{X}$  e  $\mathbf{Y}$ . Caso exista apenas uma propriedade, por exemplo a classe ou grupo a que uma amostra pertence,  $\mathbf{Y}$  é um vetor e, nesse caso, o método é denominado PLS (DE SOUZA et al., 2013).

As matrizes  $\mathbf{X}$  e  $\mathbf{Y}$  são decompostas por PCA por meio das Equações 1 e 2:

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^T + \mathbf{E} \quad (1)$$

$$\mathbf{Y} = \mathbf{U} \cdot \mathbf{Q}^T + \mathbf{F} \quad (2)$$

onde  $\mathbf{P}$  e  $\mathbf{Q}$  são os pesos (carregamentos) de  $\mathbf{X}$  e  $\mathbf{Y}$ , respectivamente; as colunas  $\mathbf{T}$  e  $\mathbf{U}$  são escores de  $\mathbf{X}$  e  $\mathbf{Y}$ ; e as matrizes  $\mathbf{E}$  e  $\mathbf{F}$  são os erros associados à modelagem PLS de  $\mathbf{X}$  e  $\mathbf{Y}$ .

A matriz  $\mathbf{X}$  contém as amostras e a matriz  $\mathbf{Y}$  contém os valores para predição. As variáveis latentes tanto em  $\mathbf{X}$  quanto em  $\mathbf{Y}$  são modeladas pela regressão PLS, para encontrar um espaço em  $\mathbf{X}$  que se adapte à modelagem de  $\mathbf{Y}$ . O modelo PLS utiliza a matriz  $\mathbf{X}$  para estimar os fatores  $\mathbf{Y}$ , e a matriz  $\mathbf{Y}$ , para estimar os fatores  $\mathbf{X}$  (DE SOUZA et al., 2013).

Na PLS os valores dos escores nas matrizes são manipulados para que seja alcançado o melhor modelo linear entre eles, ou seja, obter-se a maior covariância possível entre  $\mathbf{X}$  e  $\mathbf{Y}$ . Isso é realizado através de uma série de etapas que são repetidas para cada componente principal estimada por meio de um processo iterativo (computacional). Dessa forma, o modelamento dos dados via PLS tem o objetivo de explicar a variância da matriz  $\mathbf{X}$ , tal qual a PCA faz, ao mesmo tempo que busca a maior correlação com a matriz  $\mathbf{Y}$  (DE SOUZA et al., 2013).

O procedimento para a elaboração do modelo de classificação na PLS-DA é o mesmo usado na PLS, entretanto, a propriedade a ser investigada nos modelos de classificação é uma variável categórica que descreve a atribuição da classe da amostra, sendo, em geral, o valor 1 atribuído ao grupo de interesse e o valor 2 atribuído ao outro grupo antagônico. Porém, o número de classes não está limitado a dois, podendo-se modelar quantas classes quantas disponíveis, respeitando-se o número de variáveis disponíveis para serem modeladas (DE SOUZA et al., 2013).

Uma das principais etapas na PLS e na PLS-DA é a escolha do número correto de variáveis latentes ( $\mathbf{T}$  e  $\mathbf{U}$ ) que são usadas para melhor correlacionar os dados de  $\mathbf{X}$  e  $\mathbf{Y}$ . Se houver número insuficiente de variáveis latentes, haverá subajuste do modelo, não sendo utilizada toda a informação útil para a sua construção e o modelo pode não corresponder a suas possibilidades de classificação. Por outro lado, se houver número excessivo de variáveis latentes, haverá sobreajuste do modelo, sendo usadas informações que podem estar não correlacionadas com a propriedade investigada na construção do modelo, e ele falhar quando for empregado na investigação de novos dados (DE SANTANA et al., 2020).

A validação cruzada costuma ser usada para a escolha do número de variáveis latentes, em que uma parte das amostras, denominada de conjunto de calibração, é usada em um processo de validação interna (determinação das variáveis de interesse). Após isso, é confeccionado um modelo de treinamento com números variados de variáveis latentes (em geral 1 a 10) e então as amostras de validação interna são previstas em termos de qual classe ou grupo pertencem para cada quantidade de variáveis latentes. Então os erros de classificação obtidos são armazenados e esse processo é repetido até que haja previsão de todas as amostras de treinamento com todos os números de variáveis latentes, sendo que o

modelo mais adequado é aquele que encontra o menor erro com o menor número de variáveis latentes (DE SANTANA et al., 2020).

## 1.5 OBJETIVOS

Este estudo teve como objetivo utilizar a espectroscopia Raman para diferenciar o tecido normal do câncer de mama e seus subtipos moleculares comparados com a análise anatomopatológica.

### 1.5.1 Objetivo principal

- Avaliar a capacidade de discriminação dos subtipos moleculares do câncer de mama utilizando a espectroscopia Raman e a PLS-DA e propor a sua utilização como ferramenta de diagnóstico para as alterações teciduais no nível celular (morfológico) e molecular.

### 1.5.2 Objetivos secundários

- Avaliar as diferenças bioquímicas entre os subtipos moleculares do carcinoma de mama (Luminal A, Luminal B, HER2 e Triplo Negativo) em tecidos mamários entre si e comparativamente ao tecido mamário normal *in vitro* por meio da espectroscopia Raman.

- Realizar a análise exploratória dos espectros Raman por meio da PCA para encontrar as diferenças espectrais entre os subtipos moleculares e correlacionar essas diferenças com os componentes bioquímicos e as estruturas morfológicas responsáveis por essas alterações moleculares.

- Desenvolver modelos de análise discriminante que empregam tanto os modelos de regressão por PCA (PCR) quanto de PLS para classificar os subtipos moleculares do câncer de mama *versus* o tecido mamário normal e os subtipos moleculares de câncer de mama entre si.

## 1.6 HIPÓTESE

A hipótese levantada nesta Tese é a de que a mudança na composição bioquímica do tecido maligno em comparação com o tecido normal e de acordo com o subtipo molecular de câncer de mama pode ser identificada pela espectroscopia Raman, e de que um modelo definido através dos constituintes bioquímicos e de sua contribuição espectral pode ser utilizado para classificar os subtipos moleculares do câncer de mama através da PLS-DA.

## 2. MATERIAIS E MÉTODOS

### 2.1 FRAGMENTOS DE TECIDO MAMÁRIO

O estudo foi aprovado pelo Comitê de Ética em Pesquisa do Hospital Pérola Byington sob N<sup>o</sup>. 2.788.474, seguindo as diretrizes brasileiras para pesquisa com seres humanos.

Foram selecionadas 31 pacientes do sexo feminino do Centro de Referência em Saúde da Mulher (CRSM) do Hospital Pérola Byington, São Paulo, SP, com nódulos mamários palpáveis com diâmetro mínimo de 2 cm (estádio I ou II), com diagnóstico definido de câncer de mama através de biópsia e estudo anatomopatológico com indicação de tratamento cirúrgico (mastectomia, quadrantectomia ou tumorectomia). Após a cirurgia, foram obtidos fragmentos de cerca de 0,5 cm<sup>3</sup> representativos da lesão e fragmentos de cerca de 0,5 cm<sup>3</sup> de tecido mamário normal adjacentes à lesão. Estes fragmentos foram removidos antes de se encaminhar as peças anatômicas para o exame anatomopatológico e armazenados para análise via espectroscopia Raman. As peças anatômicas dos espécimes cirúrgicos foram então submetidas à análise de rotina anatomopatológica no CRSM. Os cortes histológicos foram avaliados por um patologista especialista que não tinha conhecimento dos resultados obtidos pela análise de espectroscopia Raman (cego).

O estudo imunoistoquímico foi utilizado para determinar o subtipo molecular do câncer de mama. Os tumores do tipo LA são caracterizados pela presença de RE e/ou RP e HER2 Escore 1+ ou HER2 Escore 2+. Os tumores LB são RE positivos e podem ser RP negativos e/ou HER2, com Ki67 > 20%. Os tumores HER2 (HER2+++ ) são caracterizados pela superexpressão do HER2 com a ausência de RE e RP. Os tumores TN não têm receptor de estrogênio, nem receptor de progesterona, nem superexpressão de HER2.

Anteriormente à análise pela espectroscopia Raman, os fragmentos de tecido mamário (amostras) foram congelados com nitrogênio líquido e armazenados por 3 meses em ultra freezer (-75 °C). No dia de realização da espectroscopia Raman, esses fragmentos foram armazenados em caixa térmica contendo CO<sub>2</sub> sólido (“gelo seco”, -78 °C) e transportados para o laboratório. Antes

das avaliações espectroscópicas, as amostras foram passivamente descongeladas com soro fisiológico, colocadas em um suporte de alumínio para amostras e submetidas à espectroscopia Raman. As substâncias biológicas residuais foram descartadas de acordo com a lei vigente para o manuseio de tecidos biológicos.

## **2.2 ESPECTROSCOPIA RAMAN DAS AMOSTRAS E PRÉ-PROCESSAMENTO DOS ESPECTROS**

Os espectros Raman foram obtidos por um espectrômetro Raman dispersivo (modelo Dimensão P-1, Lambda Solutions, Inc., MA, EUA) que possui um laser de diodo com 830 nm excitação, ajustável até 450 mW. O espectrômetro é composto por um espectrógrafo de imagem (F/1.8, grade com 1200 linhas/mm) e uma câmera CCD (1320 x 100 pixels,  $-75$  °C de temperatura de trabalho). A resolução espectral é de aproximadamente  $4\text{ cm}^{-1}$  na faixa espectral de 400 a  $1800\text{ cm}^{-1}$ , conforme relatado na literatura (BODANESE et al., 2012; SILVEIRA et al., 2012). O espectrômetro possui uma sonda Raman (modelo Vector probe, Lambda Solutions, Inc.) com cerca de 3 m de comprimento (Figura 13, seção Introdução) para excitação da amostra e coleta do espectro. Cada espectro Raman foi coletado em 30 s (10 acumulações de 3 s cada).

No total, 424 espectros de 31 fragmentos de tecido cancerígeno e 31 fragmentos de tecido mamário normal provenientes de 31 pacientes foram obtidos através da espectroscopia Raman. Para cada fragmento tumoral foram coletados de 6 a 10 espectros, dependendo do tamanho e qualidade do fragmento; além disso, também foram coletados de 6 a 10 espectros de cada tecido normal adjacente correspondente.

A calibração do número de ondas do espectrômetro foi verificada usando os picos Raman conhecidos de naftaleno. A resposta espectral fornecida pelo fabricante foi utilizada para corrigir cada espectro. A linha de base (sinal de fundo, principalmente fluorescência da amostra) foi removida por meio de um filtro passa-altas desenvolvido no software Matlab (versão 7.0.4, The Mathworks, Inc., MA, EUA), aplicando-se uma função polinomial de 5ª ordem em cada espectro na faixa de  $400\text{-}1800\text{ cm}^{-1}$  e depois realizando a sua subtração, sendo evidenciadas, assim,

as bandas Raman de alta frequência. Em seguida, cada espectro foi normalizado pela área sob a curva (1-norm).

Os espectros Raman pré-processados foram então submetidos à identificação de padrões espectrais e sua correspondência com as alterações imunoistoquímicas teciduais, bem como discriminação entre os tipos normais e tumorais. Para tal, os espectros foram separados em cinco grupos diferentes: tecidos normais (NT) e quatro subtipos de câncer de mama de acordo com a imunoistoquímica: Luminal A (LA), Luminal B (LB), HER2 e Triplo Negativo (TN).

## 2.3 TESTES ESTATÍSTICOS

Os testes de hipóteses de Qui-quadrado ( $\chi^2$ ), análise de variância (ANOVA) e Kruskal-Wallis foram utilizados neste estudo.

O teste  $\chi^2$  é um teste de hipóteses não paramétrico que tem o objetivo de avaliar a associação existente entre variáveis qualitativas comparando proporções para encontrar possíveis divergências entre os resultados observados e os resultados esperados. Os pressupostos para a realização desse teste são: independência dos grupos, seleção aleatória dos itens de cada grupo, observações de frequências ou contagens, sendo que cada observação pertence somente a uma categoria (SCHEUNEMAN, 1979). Neste trabalho foi utilizado o teste  $\chi^2$  para avaliar a existência de correlação entre a raça e os subtipos moleculares do câncer de mama.

O teste ANOVA é um método estatístico paramétrico utilizado para comparar duas ou mais médias de diferentes grupos e fazer um teste de hipóteses, avaliando se existe diferença estatisticamente significativa entre os grupos ( $p < 0,05$ ) ou não. São necessários três pressupostos para se aplicar o teste ANOVA: todas as observações têm que ser independentes, as observações devem apresentar uma distribuição gaussiana (normal) e as variâncias de cada grupo devem ser aproximadamente iguais (STHLE; WOLD, 1989). Neste trabalho foi utilizada a ANOVA para comparar as médias de idade entre os grupos dos subtipos moleculares e para a PCA entre os subtipos moleculares quando apresentarem distribuição normal. O teste de Kolmogorov-Smirnov (KS) é utilizado para

determinar se a amostra estudada segue uma distribuição normal (gaussiana), ou seja, quando os valores da variável se distribuem de forma normal no gráfico de distribuição. Este teste avalia a distância entre uma distribuição normal, previamente conhecida, e a distribuição que está sendo estudada. A hipótese nula do KS ( $p > 0,05$ ) é a de que a amostra estudada segue a distribuição normal e a hipótese alternativa ( $p < 0,05$ ) é a de que a amostra estudada não segue a distribuição normal.

O teste de Kruskal-Wallis é um teste não paramétrico que é utilizado como uma alternativa ao teste ANOVA para comparar três ou mais grupos que não apresentam distribuição gaussiana (normal) na variável de análise. Os pressupostos para a aplicação do teste de Kruskal-Wallis são: comparação de três ou mais amostras independentes, dados com nível de mensuração no mínimo ordinal e amostras com tamanho mínimo de 6 dados (OSTERTAGOVÁ; OSTERTAG; KOVÁČ, 2014). Neste trabalho foi utilizado o teste de Kruskal-Wallis para a PCA entre os subtipos moleculares quando não apresentarem distribuição normal.

## 2.4 ANÁLISE EXPLORATÓRIA

Neste estudo foram utilizados dois métodos de análise multivariada: a PCA e a PLS-DA.

A análise exploratória baseada na PCA foi realizada para extração de características espectrais e para correlacionar as diferenças espectrais entre o tecido normal e o tecido com câncer (grupos tumorais) que estão associadas às alterações bioquímicas dos subtipos tumorais e que se relacionam com as estruturas morfológicas e alterações imunoistoquímicas (diferenças bioquímicas entre tecido normal *versus* carcinoma e subtipos imunoistoquímicos).

O primeiro componente principal (PC1) é o mais importante, pois apresenta a maior variância que explica o máximo da variabilidade dos dados originais. O segundo componente principal (PC2) apresenta a segunda maior variância e assim por diante. Os últimos componentes principais apresentam direções que não estão associadas a variabilidade de dados, e sim de ruído (HONGYU; SANDANIELO; OLIVEIRA JUNIOR, 2015).



A PCA decompõe o conjunto de dados em duas matrizes: os escores, que mostram as direções em que a variância dos dados é maximizada, e os PCs, que mostram as projeções (cossenos) dessas direções no eixo original (dados originais). Os escores se assemelham aos espectros Raman e os PCs são as intensidades de cada escore para reconstruir os dados originais. Os testes estatísticos ANOVA (paramétrico) ou Kruskal-Wallis (não paramétrico) (Instat versão 3.0, GraphPad Software, Inc., CA, USA) foram aplicados aos PCs para verificar qual PC seria capaz de diferenciar melhor os tecidos mamários (valor de  $p < 0,05$  para ser considerado significativo). O cálculo das variáveis PCA (ou seja, PCs e Escores) foi determinado pela função princomp.m no MatLab (versão 7.4 do MatLab, The MathWorks Inc., MA, EUA).

A identificação das alterações bioquímicas nos grupos tumorais em relação ao tecido normal foi feita analisando-se as informações espectrais fornecidas pelos Escores da PCA, tendo sido utilizada neste estudo com o objetivo de discriminar as amostras nos grupos de tecido normal, câncer e seus subtipos moleculares (luminal A, luminal B, HER2 e triplo negativo).

Na PCA, a variabilidade total contida nas variáveis originais é igual à variabilidade total contida nos componentes principais. A contribuição de cada componente principal é expressa em porcentagem. O número de componentes principais descritos explica ao menos 80% da variação total presente nos dados.

De modo simplificado, pode-se entender que na PCA procura-se determinar “blocos de construção” do conjunto de dados que está sendo estudado, sendo que esses “blocos de construção” são chamados de variáveis latentes. Cada vetor de forma espectral é chamado então de componente principal (PC ou “loading”). Cada PC descreve um conjunto de informações correlacionadas do conjunto de dados. A PCA organiza os PCs de forma que eles expliquem sequencialmente estas informações em ordem de magnitude e prevalência. Ou seja, os PCs iniciais concentram o maior volume de informações e geralmente se correlacionam fortemente com características químicas. Os PCs posteriores representam informações menos comuns do conjunto de dados e descrevem principalmente o ruído de medição. Os Escores quantificam o quanto cada PC presente no conjunto de dados contribui para as observações originais e pode ser calculada usando a covariância dos espectros. O objetivo da PCA é descobrir a identidade destes

“blocos de construção” (PCs) e usá-los para explicar os dados observados da forma mais sucinta possível (BEATTIE; ESMONDE-WHITE, 2021).

## **2.5 DISCRIMINAÇÃO E CLASSIFICAÇÃO DE TECIDOS NORMAIS E TUMORAIS EM GRUPOS UTILIZANDO A REGRESSÃO PLS**

Os modelos de discriminação foram desenvolvidos com base em técnicas de regressão multivariada, como a regressão pelos mínimos quadrados parciais (PLS). Esses modelos discriminatórios supervisionados não se baseiam na compreensão das características bioquímicas e morfológicas dos tecidos e, ao invés disso, obtêm as características espectrais que variam de acordo com o rótulo do grupo (covariância entre grupos) e, portanto, usam esses recursos para classificar as amostras em um dos grupos via análise discriminante (DA). No entanto, os modelos PLS podem oferecer informações a partir da perspectiva de identificar as diferenças bioquímicas em cada grupo (FERREIRA et al., 1999), já que essas diferenças estão presentes entre os grupos pelo cálculo da covariância.

O modelo de regressão PLS-DA utilizou o método de validação cruzada de omissão (deixe-um-de-fora ou *leave-one-out*), no qual o modelo é construído com um número de  $n-1$  espectros e posteriormente esse espectro omitido é classificado. Em seguida, o modelo é repetido  $n$  vezes, onde cada espectro omitido é classificado um por vez. O modelo PLS-DA foi desenvolvido utilizando-se o número de variáveis latentes (VLs) que apresentaram a melhor capacidade de discriminação, ou seja, a maior acurácia.

O modelo PLS-DA foi utilizado para classificar o espectro Raman dos tecidos normais adjacentes e dos tumores de tecidos mamários em três comparações diferentes: a) cinco grupos: tecido normal (NT) e cada um dos grupos imunoistoquímicos: Luminal A (LA), Luminal B (LB), HER2 e Triplo Negativo (TN); b) três grupos: NT, Luminal (LA e LB) e Não Luminal (HER2 e TN); e c) dois grupos: NT e Tumor (todos os tecidos tumorais agrupados). Foram calculados sensibilidade, especificidade e acurácia para a classificação NT e Tumor e a acurácia foi determinada para as comparações a) e b) conforme cada arranjo imunoistoquímico. O modelo PLS-DA foi desenvolvido utilizando-se o software

Chemoface (<http://www.ufla.br/chemoface>), aplicando a classificação dos espectros ao invés de classificação das amostras (NUNES et al., 2012).

### 3. RESULTADOS

#### 3.1 DESCRIÇÃO DOS CASOS

A Tabela 2 mostra as características demográficas das pacientes incluídas no estudo. A idade variou de 36 a 98 anos, com uma média de 62 anos, com pacientes apenas do sexo feminino atendidas no CRSM – Hospital Pérola Byington, São Paulo, SP. Não houve diferença significativa de idade entre os grupos de subtipos moleculares (ANOVA,  $p = 0,392$ ). Embora 54,8% das pacientes desta amostra tenham sido caucasianas, não houve diferença significativa comparando a raça com os subtipos moleculares ( $\chi^2$ ,  $p = 0,243$ ).

**Tabela 2.** Características demográficas das pacientes incluídas no estudo.

Característica demográfica	Categoria	Valor	%
Idade	–	36-98 anos	
	Média ( $\pm$ Desvio padrão)	62 ( $\pm$ 15 anos)	
Gênero	Feminino	31 pacientes	100%
Raça	Caucasiana	17 pacientes	54,8%
	Negra	3 pacientes	9,7%
	Amarela	1 pacientes	3,2%
	Parda	10 pacientes	32,3%
Subtipo molecular	Luminal A	7 pacientes	22,6%
	Luminal B	12 pacientes	38,7%
	HER2	2 pacientes	6,5%
	Tripló negativo	7 pacientes	22,6%
	Sem subtipagem molecular	3 pacientes	9,7%

Raça *versus* subtipo molecular sem diferença significativa ( $\chi^2$ ,  $p = 0,243$ ).

Comparação de média de idade entre subtipos moleculares sem diferença significativa (ANOVA,  $p = 0,392$ ).

A Tabela 3 mostra o número de lesões separadas de acordo com os resultados histopatológicos e com a imunistoquímica, e o total de pacientes

representando cada grupo histopatológico e imunoistoquímico; entre parênteses é mostrado o número de espectros em cada grupo histopatológico e imunoistoquímico. O número de amostras no grupo de tecido normal e o número de espectros nesse grupo também são exibidos. A histopatologia das lesões descreveu 28 carcinomas ductais invasivos, 1 carcinoma ductal *in situ*, 1 carcinoma lobular invasivo e 1 carcinoma metaplásico. A classificação das lesões segundo a avaliação imunoistoquímica mostrou 7 tumores luminal A (LA), 12 tumores luminal B (LB), 2 tumores HER2 e 7 tumores triplo negativo (TN) (Tabela 3). Em uma amostra, a imunoistoquímica não foi possível ser realizada e em duas amostras a imunoistoquímica não foi disponibilizada. Neste estudo o subtipo molecular mais frequente foi o LB, mas em geral o subtipo mais frequente é o LA.

**Tabela 3.** Revisão do número de pacientes incluídos no estudo de acordo com o diagnóstico histopatológico e o respectivo diagnóstico imunoistoquímico. O número de espectros Raman está entre parênteses ao lado do número de pacientes.

Diagnóstico histopatológico	Diagnóstico imunoistoquímico					Total
	Luminal A	Luminal B	HER2	Triplo negativo	Indefinido /NA*	
Carcinoma Ductal Invasivo	6 (44)	11 (84)	2 (12)	6 (40)	3 (22)	28 (202)
Carcinoma Ductal <i>in situ</i>	-	1 (6)	-	-	-	1 (6)
Carcinoma Lobular invasivo	1 (6)	-	-	-	-	1 (6)
Carcinoma Metaplásico	-	-	-	1 (6)	-	1 (6)
Tecido Normal	-	-	-	-	31 (204)	31 (204)
Total	7 (50)	12 (90)	2 (12)	7 (46)	34 (226)	62 (424)

\* NA – imunoistoquímica não disponível

### 3.2 ESPECTROS RAMAN DE TECIDOS NORMAIS E TUMORAIS

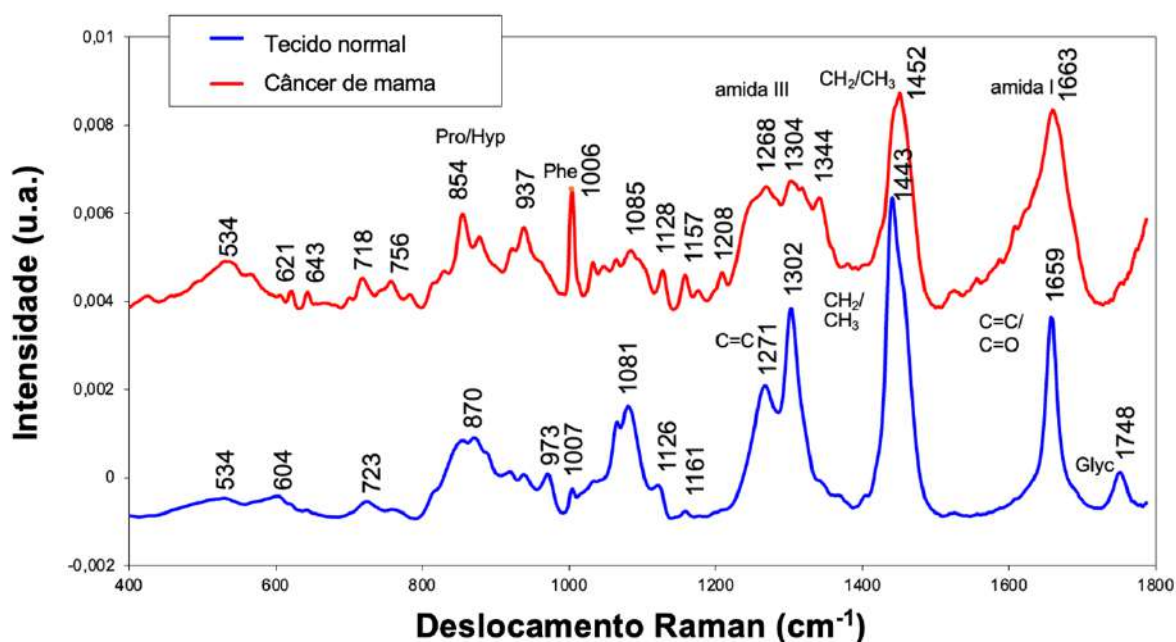
A Tabela 4 mostra a associação entre os valores de deslocamento Raman dos picos e as respectivas substâncias químicas responsáveis por esses picos (BITAR et al., 2006; CHATURVEDI et al., 2016; DE GELDER et al., 2007; KONG et al., 2014; LI; GAO; ZHANG, 2014; MOURANT et al., 2005; MOVASAGHI; REHMAN; REHMAN, 2007; SILVEIRA et al., 2012; ZÚÑIGA et al., 2019).

**Tabela 4.** Principais picos Raman e suas atribuições utilizadas para a avaliação do tecido mamário.

Deslocamento Raman (cm <sup>-1</sup> )	Atribuição
742 – 749	Triptofano, ácidos nucleicos (DNA e RNA)
858 – 882	C–C, colágeno, valina, prolina
938 – 950	Prolina, C–C, colágeno, polissacarídeos C–O–C
1002 – 1004	C–C, fenilalanina
1062 – 1063	C–C (lipídios), C–O, C–N (proteínas), O–P–O (DNA e RNA)
1081 – 1082	Ácidos nucleicos, C–C, C–O (fosfolipídios)
1271 – 1278	Amida III, colágeno
1302 – 1303	Lipídios, ácidos graxos e/ou colágeno
1325 – 1333	DNA, fosfolipídios
1439 – 1442	CH <sub>2</sub> (lipídios)
1448 – 1453	CH <sub>2</sub> (proteínas)
1627 – 1640	Amida I
1653 – 1657	C=C (lipídios)
1662 – 1667	Ácidos nucleicos, DNA, Amida I
1683 – 1697	Amida I, colágeno
1745 – 1750	C=O (fosfolipídios)
2850 – 2875	CH <sub>2</sub> (lipídios)
2885 – 2908	CH <sub>2</sub> (lipídios e proteínas)
2945 – 2957	CH <sub>3</sub> (proteínas, ácidos nucleicos)

Fonte: (BITAR et al., 2006; CHATURVEDI et al., 2016; DE GELDER et al., 2007; KONG et al., 2014; LI; GAO; ZHANG, 2014; MOURANT et al., 2005; MOVASAGHI; REHMAN; REHMAN, 2007; SILVEIRA et al., 2012; ZÚÑIGA et al., 2019).

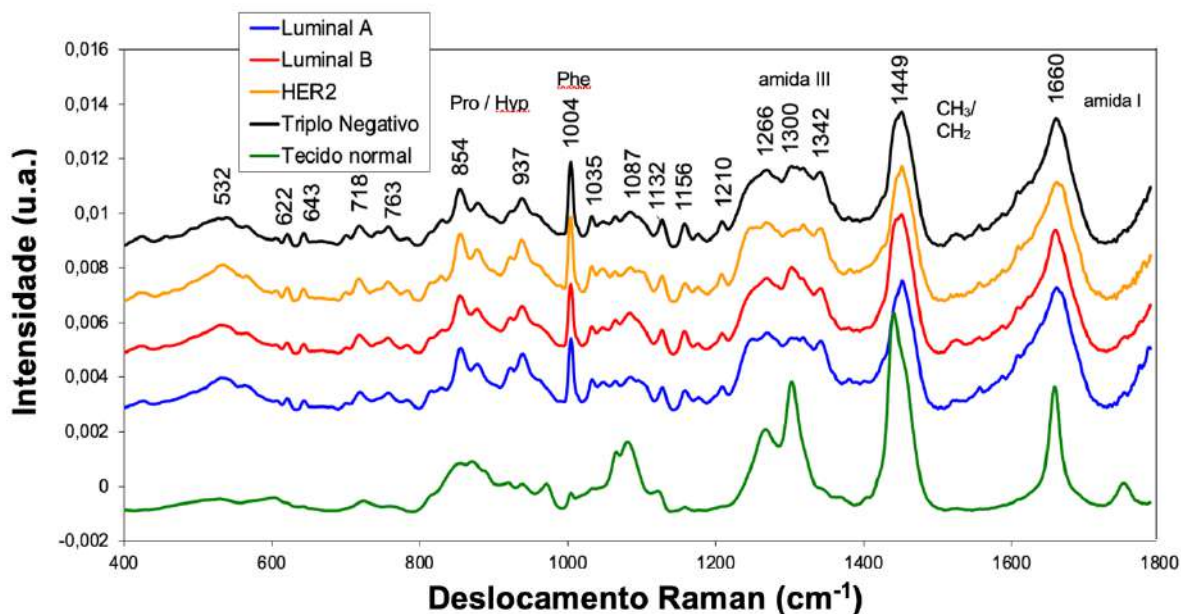
A Figura 14 mostra os espectros Raman normalizados de tecido mamário normal e câncer de mama. Os espectros de tecido normais apresentaram picos atribuídos principalmente aos lipídios (por exemplo, trioleína e ácido esteárico); picos atribuídos a proteínas (por exemplo, colágeno); picos referidos a aminoácidos e ácidos nucleicos também foram observados (HANLON et al., 2000; KONG et al., 2014; LI et al., 2015; SATHYAVATHI et al., 2015; TALARI et al., 2015). Por exemplo, os picos em 870, 1081, 1271, 1302, 1443, 1659 e 1748  $\text{cm}^{-1}$  são atribuídos aos lipídios, enquanto os picos em 854, 937, 1006, 1268, 1304, 1344, 1452 e 1663  $\text{cm}^{-1}$  são atribuídos às proteínas. Observou-se que o tecido normal tem características espectrais com predominância de lipídios e o câncer de mama tem predominância de proteínas. Os picos detalhados para as bandas Raman de tecido mamário podem ser vistos na Tabela 4 (BITAR et al., 2006; CHATURVEDI et al., 2016; DE GELDER et al., 2007; KONG et al., 2014; LI; GAO; ZHANG, 2014; MOURANT et al., 2005; MOVASAGHI; REHMAN; REHMAN, 2007; SILVEIRA et al., 2012; ZÚÑIGA et al., 2019).



**Figura 14.** Média dos espectros Raman normalizados de tecido mamário normal e câncer de mama. As regiões espectrais são apresentadas com as atribuições, segundo a literatura citada (BITAR et al., 2006; CHATURVEDI et al., 2016; DE GELDER et al., 2007; KONG et al., 2014; LI; GAO; ZHANG, 2014; MOURANT et al., 2005; MOVASAGHI; REHMAN; REHMAN, 2007; SILVEIRA et al., 2012; ZÚÑIGA et al., 2019). Os espectros foram deslocados para melhor visualização.

A Figura 15 apresenta os espectros dos diferentes subtipos moleculares de câncer de mama de acordo com o estudo imunohistoquímico, como luminal A (LA), luminal B (LB), HER2 e triplo negativo (TN). Esses espectros mostram perfis semelhantes entre si. No entanto, algumas regiões de espectro têm características particulares devido a diferentes constituições bioquímicas derivadas do processo neoplásico. Intensidades maiores de picos têm sido observadas para os tecidos malignos de 800 a 1010  $\text{cm}^{-1}$  e por volta de 1280, 1350, 1455 e 1665  $\text{cm}^{-1}$ , que são atribuídos às proteínas. Curiosamente, as amostras de LA e LB apresentaram perfil espectral semelhante, sugerindo constituição semelhante, com picos de intensidade maior em praticamente toda a faixa espectral em comparação com o espectro de tecido normal (predominância de lipídios), bem como picos com intensidades maiores nas posições atribuídas a proteínas, enquanto as amostras de HER2 e TN apresentaram intensidade maior nos picos encontrados aproximadamente em 1000, 1250 e 1450  $\text{cm}^{-1}$ , atribuídos às proteínas, e em 1100, 1300 e 1440  $\text{cm}^{-1}$ , atribuídos a lipídios. As principais diferenças na constituição bioquímica dos espectros relativos aos subgrupos imunohistoquímicos são detalhadas na análise exploratória a seguir.





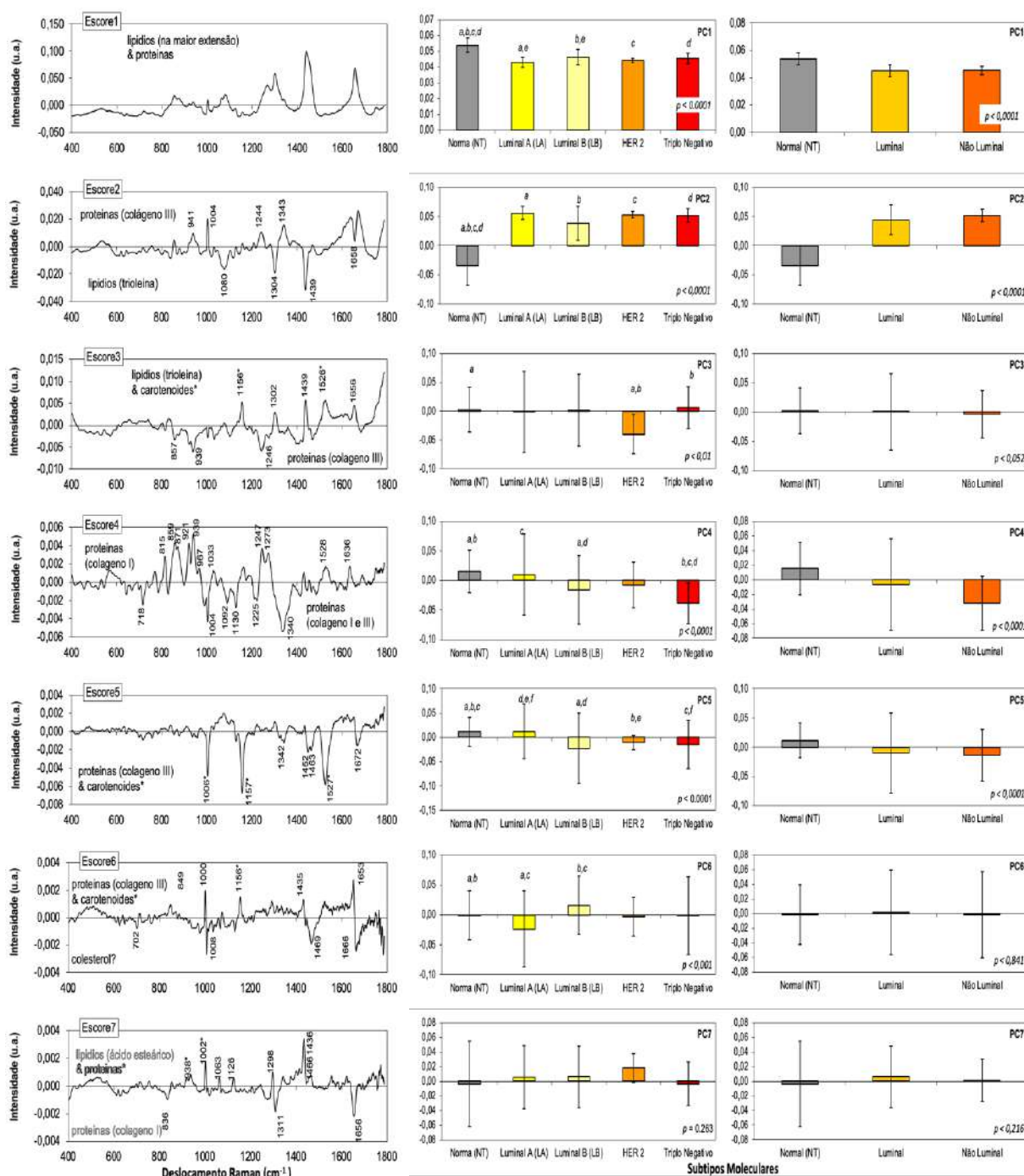
**Figura 15.** Média dos espectros Raman normalizados de tecido mamário normal e os diferentes subtipos imunoistoquímicos (moleculares) do câncer de mama. As regiões espectrais são apresentadas com as atribuições, segundo a literatura citada (BITAR et al., 2006; CHATURVEDI et al., 2016; DE GELDER et al., 2007; KONG et al., 2014; LI; GAO; ZHANG, 2014; MOURANT et al., 2005; MOVASAGHI; REHMAN; REHMAN, 2007; SILVEIRA et al., 2012; ZÚÑIGA et al., 2019). Os espectros foram deslocados para melhor visualização.

### 3.3 ANÁLISE EXPLORATÓRIA

A análise exploratória tem sido empregada como tentativa de desvendar as diferenças espectrais entre os subtipos moleculares do câncer de mama e correlacionar essas diferenças com a possível composição bioquímica dos tecidos (BENDAU et al., 2020; BI et al., 2014; DENG et al., 2022; MEDEIROS et al., 2016; SANTOS et al., 2022; TÉLLEZ-PLANCARTE et al., 2018; ZHANG et al., 2022).

A Figura 16 mostra os sete primeiros vetores dos componentes principais (Escores e PCs) extraídos dos conjuntos de dados. Esses componentes principais apresentaram mais de 99,3% da variância apresentada no conjunto de dados (Escore1/PC1 = 85,4%, Escore2/PC2 = 11,9%, Escore3/PC3 = 1,3%, Escore4/PC4 = 0,4%, Escore5/PC5 = 0,3%, Escore6/PC6 = 0,1%) e são os que mostraram diferenças estatisticamente significativas entre os subtipos moleculares tumorais e o tecido normal (teste ANOVA ou Kruskal-Wallis,  $p < 0,05$ ). Apesar de não ser

significativa ( $p = 0,263$ ), a presença de características espectrais até o Escore7/PC7 revela a natureza complexa das diferenças bioquímicas entre os subtipos moleculares do câncer de mama.



**Figura 16.** Plotagem dos sete primeiros Escores do PCA, intensidade média e desvio padrão dos PCs dos tecidos mamários normais e dos subtipos moleculares de câncer extraídos do conjunto de dados de Raman. O valor  $p$  da ANOVA é mostrado em cada PC. As letras (a, b, c, d, e, f) nos PCs indicam comparação dos PCs com diferenças estatisticamente significativas (ANOVA,  $p < 0,05$ ).

O Escore 1 mostrou características espectrais dos constituintes do tecido mamário, especialmente lipídios vistos nos picos em 1082, 1268, 1304, 1442, 1659 e 1752  $\text{cm}^{-1}$  e proteínas vistas nos picos em 854, 877, 939, 1004, 1342 e 1455  $\text{cm}^{-1}$  (SILVEIRA et al., 2012; TALARI et al., 2015). O PC1 apresentou diferença estatisticamente significativa entre os grupos (Kruskal-Wallis,  $p < 0,0001$ ); diferenças significativas foram observadas entre NT e todos os subtipos tumorais ( $p < 0,001$ ), e entre LA e LB (Kruskal-Wallis,  $p < 0,05$ ). O valor de PC1 maior para o grupo NT em comparação com todos os subtipos tumorais indica maior teor lipídico nos tecidos mamários normais.

O Escore 2 mostrou características espectrais positivas atribuídas a proteínas (principalmente colágeno da matriz extracelular), com picos em 941 e 1004  $\text{cm}^{-1}$  e picos em 1244 e 1343  $\text{cm}^{-1}$ , e características espectrais negativas atribuídas aos lipídios (principalmente trioleína de adipócitos), com picos em 1080, 1304 e 1439  $\text{cm}^{-1}$ . O PC2 apresentou diferença estatisticamente significativa entre os grupos (Kruskal-Wallis,  $p < 0,0001$ ); foram observadas diferenças significativas entre NT e todos os subtipos tumorais (Kruskal-Wallis,  $p < 0,001$ ). O valor de PC2 negativo para o NT e positivo para os subtipos tumorais sugere que os tumores apresentam proteínas (por exemplo, colágeno III) em maiores quantidades e lipídios (por exemplo, trioleína) em menor quantidade do que os tecidos mamários normais, corroborando os achados em Escore1/PC1.

O Escore 3 apresentou características espectrais positivas atribuídas à trioleína, com picos em 1302, 1439 e 1656  $\text{cm}^{-1}$ , e carotenoides, com picos de 1156 e 1526  $\text{cm}^{-1}$ , e características negativas atribuídas a proteínas (por exemplo, colágeno III), com picos de 857, 939 e 1246  $\text{cm}^{-1}$ . O PC3 apresentou diferença estatisticamente significativa entre os grupos (Kruskal-Wallis,  $p < 0,01$ ); diferenças significativas observadas entre HER2 e NT (Kruskal-Wallis,  $p < 0,01$ ) e entre HER2 e TN (Kruskal-Wallis,  $p < 0,05$ ). O valor de PC3 negativo sugere que o HER2, em particular, apresenta maior teor de proteína e menor teor lipídico do que os outros subtipos tumorais e NT.

O Escore 4 apresentou características espectrais muito distintas na região de 815-967  $\text{cm}^{-1}$  e picos em 1247, 1273, 1528 e 1636  $\text{cm}^{-1}$ , todos atribuídos a proteínas (por exemplo, colágeno I), e picos em 718, 1004, 1092, 1130, 1225 e 1340  $\text{cm}^{-1}$ , que podem ser atribuídos ao colágeno III. O PC4 apresentou diferença estatisticamente significativa entre os grupos (Kruskal-Wallis,  $p < 0,0001$ ); foram

observadas diferenças significativas entre NT e LB (Kruskal-Wallis,  $p < 0,01$ ) e NT e TN (Kruskal-Wallis,  $p < 0,001$ ), entre LA e TN e entre LB e TN (ambas Kruskal-Wallis,  $p < 0,001$ ). Essas diferenças sugerem diferença no teor de colágeno entre tecido normal, luminal B e triplo negativo, sendo que o luminal B e o triplo negativo exibem maior teor de proteínas do que os outros subtipos e o tecido normal que exibem mais características de colágeno I.

O Escore 5 apresentou picos em 1342, 1452, 1463 e 1672  $\text{cm}^{-1}$ , atribuídos às proteínas (por exemplo, colágeno), e picos em 1006, 1157 e 1527  $\text{cm}^{-1}$ , que são atribuídos a carotenóides. PC5 apresentou diferença estatisticamente significativa entre os grupos (Kruskal-Wallis,  $p < 0,001$ ); diferenças significativas observadas entre NT e LB (Kruskal-Wallis,  $p < 0,001$ ), NT e HER2 (Kruskal-Wallis,  $p < 0,05$ ), NT e TN (Kruskal-Wallis,  $p < 0,01$ ), e entre LA e LB (Kruskal-Wallis,  $p < 0,001$ ), LA e HER2 (Kruskal-Wallis,  $p < 0,05$ ) e LA e TN (Kruskal-Wallis,  $p < 0,05$ ). O valor de PC5 confirma que os grupos LB e TN apresentam proteínas mais elevadas como visto no PC4; os grupos LB, HER2 e TN apresentam carotenóides mais elevados do que o tecido normal.

O Escore 6 apresentou picos positivos em 849, 1000, 1156, 1435 e 1653  $\text{cm}^{-1}$ , atribuídos ao colágeno III e proteínas/aminoácidos. Além disso, os picos negativos em 702, 1008, 1469 e 1666  $\text{cm}^{-1}$  podem ser atribuídos ao colesterol. O PC6 demonstrou diferença estatisticamente significativa entre os grupos (ANOVA,  $p < 0,001$ ); diferenças significativas observadas entre NT e LA e entre NT e LB (ANOVA,  $p < 0,05$ ), e entre LA e LB (ANOVA,  $p < 0,001$ ). O valor de PC6, positivo para o luminal B e negativo para o luminal A, sugere que o luminal B é rico em proteínas e o luminal A é rico em colesterol.

O Escore 7 apresentou picos positivos em 1063, 1126, 1298, 1436 e 1466  $\text{cm}^{-1}$ , atribuídos aos ácidos graxos saturados (ácido esteárico); picos em 938 e 1002  $\text{cm}^{-1}$ , atribuídos a proteínas e aminoácidos, e picos negativos em 836, 1311 e 1656  $\text{cm}^{-1}$ , que podem ser atribuídos ao colágeno I. O PC7 não demonstrou diferença estatisticamente significativa entre os grupos (ANOVA,  $p = 0,263$ ); apesar de não significativo, o HER2 apresentou maior teor de proteínas/aminoácidos e teor lipídico e menos colágeno I do que os outros subtipos, corroborando os achados no PC3 de colágeno III mais elevado.

### 3.4 ANÁLISE DISCRIMINANTE POR PLS

O modelo discriminante baseado na PLS (PLS-DA) foi aplicado ao conjunto de dados obtido dos espectros Raman para classificar os subtipos imunoistoquímicos do câncer de mama (LA, LB, HER2 e TN) e o tecido mamário normal (NT). A classificação PLS-DA considerou diferentes agrupamentos dos tumores para comparações: normal e cada subtipo separado, normal, luminal e não luminal, e normal *versus* tumor.

O algoritmo discriminante considerou a classificação dos espectros para distribuí-los em grupos da seguinte forma: (a) classificação contemplando grupos separados (NT x LA x LB x HER2 x TN); (b) classificação considerando as lesões luminiais como um grupo e as lesões não luminiais como outro grupo *versus* tecido normal (NT x [LA + LB] x [HER2 + TN]); (c) classificação considerando todos os tumores agrupados *versus* tecido normal (NT x [LA + LB + HER2 + TN]).

Os números de classificações corretas considerando os grupos imunoistoquímicos acima mencionados e a respectiva precisão são apresentados nas Tabelas 5 a 9. A classificação que considerou grupos individuais, incluindo o tecido normal, apresentou acurácia de 80,8% (Tabela 5); a classificação que considerou os grupos luminal (LA e LB) e não luminal (HER2 e TN), incluindo o tecido normal apresentou acurácia de 89,3% (Tabela 6). A classificação que considerou todas as lesões cancerígenas agrupadas *versus* tecido normal apresentou acurácia de 97,3%, com valores de sensibilidade e especificidade de 95,9% e 99,0%, respectivamente (2 dos 202 tecidos normais foram falsos positivos e 9 das 189 lesões foram falsos negativos) (Tabela 7). A classificação em luminal e não luminal, excluído o tecido normal, apresentou acurácia de 89,9% (Tabela 8), e a classificação dos subtipos moleculares separados, excluído o tecido normal, apresentou acurácia de 73,0% (Tabela 9).

**Tabela 5.** Resultados da aplicação do modelo discriminante baseado em PLS (PLS-DA) nos espectros Raman dos grupos imunoistoquímicos (subtipos moleculares). O total de espectros em cada grupo de subtipos moleculares está entre parênteses.

Grupos de acordo com a imunoistoquímica (subtipos moleculares)	Discriminação/diagnóstico Raman				
	Tecido Normal	Luminal A	Luminal B	HER2	Triplo Negativo
Tecido Normal (204)	202	2	0	0	0
Luminal A (50)	5	31	12	0	2
Luminal B (90)	11	13	60	0	6
HER2 (12)	0	0	6	0	6
Triplo Negativo (46)	0	2	12	0	32
Acurácia			80,8%		

**Tabela 6.** Resultados da aplicação do modelo discriminante baseado em PLS (PLS-DA) nos espectros Raman dos grupos imunoistoquímicos (luminal e não luminal). O total de espectros em cada grupo de subtipos moleculares está entre parênteses.

Grupos de acordo com a imunoistoquímica (subtipos moleculares)	Discriminação/diagnóstico de Raman		
	Tecido normal	Luminal (LA + LB)	Não Luminal (HER2 + TN)
Tecido Normal (204)	202	2	0
Luminal (140)	12	119	9
Não Luminal (58)	1	19	38
Acurácia		89,3%	

LA – Luminal A, LB – Luminal B, TN – Triplo Negativo

**Tabela 7.** Resultados da aplicação dos modelos discriminantes baseados em PLS (PLS-DA) nos espectros Raman dos grupos tumor e tecido normal. O total de espectros em cada grupo de subtipos moleculares é visto entre parênteses.

Grupos de acordo com a histologia	Discriminação/diagnóstico Raman	
	Tecido Normal	Tumor
Tecido Normal (204)	202	2
Tumor (198)	9	189
Acurácia		97,3%
Sensibilidade		95,4%
Especificidade		99,0%

**Tabela 8.** Resultados da aplicação do modelo discriminante baseado em PLS (PLS-DA) nos espectros Raman dos grupos imunoistoquímicos (luminal e não luminal excluído o tecido normal). O total de espectros em cada grupo de subtipos moleculares está entre parênteses.

Grupos de acordo com a imunoistoquímica (subtipos moleculares)	Discriminação/diagnóstico de Raman	
	Luminal (LA + LB)	Não Luminal (HER2 + TN)
Luminal (131)	125	6
Não Luminal (58)	13	45
Acurácia	89,9%	
Sensibilidade	77,5%	
Especificidade	95,4%	

LA – Luminal A, LB – Luminal B, TN – Triplo Negativo

**Tabela 9.** Resultados da aplicação do modelo discriminante baseado em PLS (PLS-DA) nos espectros Raman dos grupos imunoistoquímicos (subtipos moleculares excluído tecido normal). O total de espectros em cada grupo de subtipos moleculares está entre parênteses.

Grupos de acordo com a imunoistoquímica (subtipos moleculares)	Discriminação/diagnóstico Raman			
	Luminal A	Luminal B	HER2	Triplo Negativo
Luminal A (48)	34	13	0	1
Luminal B (83)	12	66	0	5
HER2 (12)	0	5	3	4
Triplo Negativo (46)	5	6	0	35
Acurácia	73,0%			

## 4. DISCUSSÃO

Muitas tentativas foram dedicadas ao uso da espectroscopia Raman para o diagnóstico de câncer de mama (ALFANO et al., 1991; HAKA et al., 2009; HAN et al., 2017; KONG et al., 2014; KOYA et al., 2020; LI et al., 2015; LI; HAO; XU, 2017; REHMAN et al., 2007; WINNARD et al., 2017; ZÚÑIGA et al., 2019).

Neste estudo, espectros Raman de tecido normal apresentaram picos atribuídos principalmente a lipídios (por exemplo, trioleína e ácido esteárico) e picos atribuídos a proteínas (por exemplo, colágeno), aminoácidos e ácidos nucleicos (ABRAMCZYK et al., 2015; BI et al., 2014; HAN et al., 2017; KONG et al., 2014; MARRO et al., 2014; ZÚÑIGA et al., 2019). Foram observados picos de 870, 1081, 1271, 1302, 1443, 1659 e 1748  $\text{cm}^{-1}$  atribuídos aos lipídios, enquanto picos em 854, 937, 1006, 1268, 1304, 1344, 1452 e 1663  $\text{cm}^{-1}$  são atribuídos a proteínas (CHATURVEDI et al., 2016; DE GELDER et al., 2007; HAKA et al., 2005; LI; GAO; ZHANG, 2014; MOURANT et al., 2005; MOVASAGHI; REHMAN; REHMAN, 2007; REHMAN et al., 2007; SILVEIRA et al., 2012; TALARI et al., 2015a; ZÚÑIGA et al., 2019). Observou-se que o tecido normal de mama tem características espectrais com predominância de lipídios e o câncer de mama tem predominância de proteínas, aminoácidos e ácidos nucleicos. Os picos Raman correspondentes à prolina (854  $\text{cm}^{-1}$ ), hidroxiprolina (937  $\text{cm}^{-1}$ ) e fenilalanina (1006  $\text{cm}^{-1}$ ) apresentaram-se entre os maiores discriminadores entre tecido normal e câncer de mama, com picos mais intensos nos tumores, conforme descrição na literatura (CHATURVEDI et al., 2016; HAKA et al., 2005; HAKA et al., 2006; HAKA et al., 2009; KOYA et al., 2020; LI et al., 2015; LI; HAO; XU, 2017; REHMAN et al., 2007; SURMACKI et al., 2013; TALARI et al., 2015a; TÉLLEZ-PLANCARTE et al., 2018; ZÚÑIGA et al., 2019).

Os tumores têm uma ampla expressão de seu genoma que se manifesta através de diferentes comportamentos biológicos, alguns tumores sendo mais agressivos que outros. Diferentes expressões genéticas estão associadas a diferentes comportamentos clínicos e prognósticos. A classificação tumoral de acordo com padrões de expressão genética pode determinar marcadores como a expressão de receptores de estrogênio e progesterona, valores de Ki-67 (relacionados à proliferação celular) e superexpressão de HER2. A expressão



desses marcadores é atualmente essencial para prever o prognóstico e decidir sobre a forma de tratamento.

Os tumores luminais A e B diferem em termos de intensidade de proliferação celular (Ki-67) e regulação hormonal. Tumores luminal B têm maior proliferação celular e menor expressão do receptor de progesterona. O receptor de estrogênio tem expressão semelhante nos subtipos luminal A e luminal B e distingue os tumores luminais dos não luminais (HER2 e triplo negativo). Tumores do tipo luminal A são os tumores com melhor prognóstico e têm menos mutações e alterações cromossômicas em comparação com os tumores luminal B. Os tumores HER2 têm um número maior de mutações do que tumores luminais e são caracterizados pela expressão exacerbada de genes e proteínas relacionadas ao HER2 e pela alta proliferação celular, uma vez que a superexpressão desse gene e seu receptor conferem maior capacidade de proliferação das células tumorais, com maior agressividade e pior prognóstico. O subtipo triplo negativo não tem receptores de estrogênio e progesterona e nenhuma superexpressão do HER2. Esse subtipo possui altos níveis de RNA e proteínas (principalmente colágeno) devido à alta expressão de genes relacionados à proliferação do tumor. Esses tumores também têm um grande número de mutações em nível do DNA. Essas características tornam esse subtipo o mais agressivo e com o pior prognóstico entre os quatro subtipos moleculares (GAO; SWAIN, 2018; GOLDHIRSCH et al., 2013; NASSAR; POOJA; CHIVUKULA, 2020; PEROU et al., 2000; PRAT et al., 2015; SØRLIE et al., 2001; WAKS; WINER, 2019; WOLFF et al., 2018; YAO et al., 2015).

Os espectros Raman dos subtipos moleculares de câncer de mama luminal A, luminal B, HER2 e triplo negativo mostram perfis semelhantes entre si. No entanto, algumas regiões de espectro têm características particulares devido a diferentes constituições bioquímicas do processo neoplásico. Os tecidos cancerígenos apresentaram picos mais intensos entre 800 e 1010  $\text{cm}^{-1}$  e em torno de 1280, 1350, 1455 e 1665  $\text{cm}^{-1}$ , sendo esses picos atribuídos a proteínas. Curiosamente, as amostras de luminal A e luminal B apresentaram perfil espectral semelhante entre si, sugerindo constituição semelhante, já que possuem expressão genética semelhante e, conseqüentemente, constituição bioquímica também semelhante, mas com maior intensidade em praticamente todos os picos em comparação com o espectro de tecido normal. Os tumores luminais apresentaram picos de lipídios com intensidades maiores em comparação com as amostras de

triplo negativo, que apresentaram intensidades maiores nos picos em torno de 1100, 1300 e 1440  $\text{cm}^{-1}$ , atribuídos a proteínas (colágeno). Esses resultados são consistentes com Bendau et al. (2020), que usaram espectroscopia Raman para diferenciar entre tumores triplos negativos e não triplo negativos. Os autores mostraram que a maioria das diferenças espectrais entre tumores metastáticos triplo negativos e tumores não metastáticos não triplo negativos são decorrentes da presença de picos de colágeno e lipídios.

Bi et al. (2014) usaram a espectroscopia Raman para diferenciar tumores HER2 e não HER2 na análise de três diferentes linhagens celulares: BT474 (célula de câncer de mama HER2+), MCF-10A (HER2 negativo) e MCF-10A (HER2 positivo). As células HER2+ apresentaram maior teor lipídico e menor teor proteico. Esse aumento de lipídios nas células malignas de mama ocorre devido à síntese fosfolipídica e à maior síntese de ácidos graxos, que está associada à superexpressão dos receptores HER2 tirosina quinase nas células, e que caracteriza o subtipo molecular HER2.

Téllez-Plancarte et al. (2018) mostraram que os picos Raman localizados em 1002 e 1030  $\text{cm}^{-1}$  atribuídos à fenilalanina são mais intensos em linhagens celulares com superexpressão de HER2; no entanto, não foi possível observar essa diferença no presente estudo. Observou-se que os tumores HER2 apresentavam muitos erros de classificação, provavelmente devido ao número limitado de amostras desse subtipo (2 amostras e 12 espectros). Isso evidenciou que o modelo não foi capaz de construir uma boa representação espectral para diferenciar melhor esse subtipo de câncer dos outros subtipos.

Os estudos citados acima mostraram que a espectroscopia Raman é capaz de discriminar tumores triplo negativos e não triplo negativos e as diferentes expressões de HER2, enquanto no presente trabalho foi demonstrado que a espectroscopia Raman pode discriminar os quatro subtipos moleculares do câncer de mama, com maior diferenciação entre tumores luminais e não luminais, que apresentou uma acurácia de 89,3% (Tabela 3) (TÉLLEZ-PLANCARTE et al., 2018).

O Escore 1/PC1 mostrou as características espectrais dos constituintes do tecido mamário, principalmente lipídios e proteínas. O PC1 indicou maior teor lipídico no tecido normal em comparação com todos os subtipos de câncer e no subtipo luminal A comparado ao luminal B. O Escore 2/PC2 negativo para NT e

positivo para todos os subtipos tumorais sugere que os tumores possuem proteínas (colágeno III) em quantidades maiores e lipídios (trioleína) em quantidades menores do que os tecidos mamários normais, corroborando os achados do PC1. Essas diferenças bioquímicas estão de acordo com a literatura (COTRAN, R. S. KUMAR, V. AND ROBBINS., 1989; ELSTON; ELLIS, 1991; ROSEN et al., 2014).

O Escore 3/PC3 sugere que o HER2, em particular, tem maior teor de proteínas e menor teor de lipídios do que os outros subtipos tumorais e NT, confirmando a maior capacidade de proliferação celular tumoral conferida pela superexpressão do HER2, com maior quantidade de colágeno III constituinte da matriz extracelular, mas não está de acordo com o que foi encontrado no estudo por Bi et al. (2014), que apresentaram maiores quantidades de lipídios e menores quantidades de proteínas no tumor HER2.

Abramczyk et al. (2015) também realizaram espectroscopia Raman e PCA nas células MCF-7 (ER+) e MDA-MB-231 (ER-), bem como em células normais do tecido mamário, e encontraram diferenças na concentração de lipídios dessas linhagens celulares e também em sua composição bioquímica. O número de gotículas lipídicas citoplasmáticas na linhagem celular MCF10A (tecido normal da mama) é duas vezes menor nas células da linhagem MCF-7 (ER+) e quatro vezes maior em comparação com as células da linhagem MDA-MB-231 (ER-), ou seja, o aumento da quantidade de gotículas lipídicas está correlacionado com o aumento da agressividade do câncer devido ao aumento da taxa de síntese lipídica nesses tecidos cancerígenos, sugerindo aumento da lipogênese relacionada à superexpressão dos receptores HER1/HER2. A composição lipídica de adipócitos no câncer de mama é predominantemente composta de triglicérides de ácidos oleicos e linoleicos, em oposição às gotículas lipídicas em células epiteliais de câncer que são compostas de triglicérides e ácidos graxos dominados com predominância do ácido araquidônico. A maior capacidade de proliferação celular pode estar associada a uma maior lipogênese, mostrando tanto um aumento de proteínas (colágeno) quanto de lipídios nos tumores não luminais (HER2 e triplo negativo).

O Escore 4 apresentou características espectrais muito diferentes em relação às proteínas (colágeno I e III) e o PC4 sugere uma diferença no teor de colágeno entre tecidos normais, tumores luminais e triplos negativos, com tecido normal apresentando características de mais colágeno I e tumores luminal B e triplo

negativo exibindo mais proteínas que os outros subtipos. Observou-se maior expressão de colágeno III em tumores triplo negativos. Portanto, a diferença na quantidade de proteínas em tumores luminais em comparação com tumores triplos negativos pode estar associada ao tipo de colágeno. O estudo de Bendau et al. (2020) apresentou maior quantidade de proteínas (colágeno) em tumores triplos negativos em comparação com tumores luminais. Maior teor de colágeno em matriz extracelular tem sido encontrado em carcinomas lobulares e tubulares ductais clássicos (NATAL et al., 2019), o que pode confirmar os achados no Escore 4/PC4, uma vez que a maioria dos tumores dos grupos luminal B e triplo negativo no presente estudo foram ductais invasivos (Tabela 2). Além disso, verificou-se que as linhagens de células malignas de mama especialmente agressivas e metastáticas (não luminais) foram capazes de produzir colágeno dos tipos I, III, IV, V e VI em modelos de xenoenxerto de câncer de mama (NABA et al., 2014), e assinaturas de colágeno associadas ao tumor (*tumor associated collagen signature* – TACS) também parecem estar relacionadas à invasão de células cancerígenas e ao desenvolvimento metastático, provavelmente porque as células tumorais preferem invadir ao longo do alinhamento das fibras de colágeno (FRANCHI et al., 2019). O TACS-3 mostrou-se um indicador prognóstico independente do tamanho do tumor, subtipo e grau histológico, expressão de receptores hormonais, superexpressão de HER2, e envolvimento dos linfonodos (CONKLIN et al., 2011).

O Escore 5 apresentou características espectrais atribuídas às proteínas (colágeno), carotenoides e o PC5 sugeriu que os grupos LB e TN possuem maior teor de proteínas, como observado no PC4; os grupos LB, HER2 e TN apresentam mais carotenoides que o tecido normal, em desacordo com a literatura, que mostra maior concentração de carotenos no tecido normal em comparação com o câncer. Como os carotenos são precursores do ácido retinóico e este age inibindo a proliferação celular por ligação aos receptores nucleares, espera-se que sua concentração seja maior no tecido normal. Além disso, alguns estudos mostram que a ingestão de carotenos pode reduzir o risco de desenvolver câncer de mama, especialmente em pacientes de alto risco e até mesmo melhorar a sobrevivência de pacientes com a doença (HE; GU; ZHANG, 2018; HU et al., 2012; PENG et al., 2021; PRAKASH; KRINSKY; RUSSELL, 2009; SURMACKI et al., 2015). Bendau et al. (2020) não encontraram diferença na concentração de carotenoides entre tumores luminais e tumores triplos negativos através da espectroscopia Raman.

O Escore 6 apresentou características espectrais positivas atribuídas ao colágeno III e a aminoácidos. Além disso, apresentou características negativas que podem ser atribuídas a colesterol. O PC6, positivo para luminal B e negativo para luminal A, sugere que o luminal B é rico em proteínas se comparado ao luminal A, que é rico em colesterol, o que está em concordância com o maior índice de proliferação celular (Ki-67) característico de tumores luminal B, sugerindo aumento da matriz de colágeno nestes (ŁUKASIEWICZ et al., 2021).

Deng et al. (2022) mostraram que o colesterol é o fator mais importante para a classificação do RE e do RP, com maior concentração nas linhagens celulares MCF10A, MDA-MB-231 e SK-BR-3 RE e RP negativos (tecido normal, triplo negativo e HER2, respectivamente). Além disso, esse estudo também mostrou maior concentração de colesterol na linhagem MCF-7 (luminal A) do que na linhagem BT-474 (luminal B).

Medeiros et al. (2016) mostraram que o DAID causa uma diminuição na viabilidade celular e na inibição da migração nas linhagens de células cancerígenas de mama humana MDA-MB-231 (RE-) e MCF-7 (ER+) utilizando espectroscopia Raman e PCA. Nas células que possuem receptores de estrogênio (RE+), o mecanismo de ação foi a inibição da estimulação da mitose (que afeta principalmente o DNA), e em células que não possuem receptores de estrogênio (ER-) o mecanismo de ação foi a interferência com a síntese proteica, afetando fenilalanina, proteínas e lipídios. Isso está de acordo com os resultados observados no Escore 3/PC3, mostrando a maior contribuição bioquímica e metabólica da síntese proteica em subtipos não luminais.

Neste estudo foi observado que a PLS-DA apresentou uma acurácia maior que a PCA para a discriminação dos subtipos moleculares do câncer de mama. Esses resultados são importantes, na medida em que a PCA é um método não supervisionado em que cada componente principal está relacionado com a direção de maior variância naquela matriz de dados, e a PLS-DA é um método supervisionado, em que a correlação entre as variáveis independentes (espectros) e dependentes (grupo a que cada espectro pertence) fornece dados para identificação do grupo que cada amostra pertence, contribuindo para a separação eficiente entre os diferentes subtipos, sendo realizada a identificação dos componentes principais com o objetivo de conseguir a maior discriminação entre os grupos.

Observou-se que a análise das informações espectrais através do modelo discriminante baseado na PLS foi capaz de diferenciar os subtipos imunoistoquímicos (Luminal A, Luminal B, HER2 e Triplo Negativo) com acurácia de 80,8%, entre os subtipos luminal (LA e LB) e não luminal (HER2 e TN) com acurácia de 89,3% e diferenciar todos os subtipos de câncer em relação ao tecido mamário normal com acurácia de 97,3%. Assim, foi demonstrado que a espectroscopia Raman e o modelo de discriminação baseado em regressão PLS podem ser úteis para diferenciar lesões malignas e não malignas, assim como diferenciar os subtipos moleculares, e a técnica pode ser melhorada para ser aplicada na prática clínica como adjunta no diagnóstico de lesões e na decisão terapêutica, já que o protocolo de tratamento é diferente para cada subtipo molecular.

A espectroscopia Raman pode ser associada ao estudo histopatológico e imunoistoquímico para confirmar o subtipo molecular, uma vez que a literatura descreve uma discrepância de cerca de 20 a 30% nos resultados da análise imunoistoquímica quando comparada com os testes genéticos (padrão ouro) e a análise anatomopatológica do espécime cirúrgico (GAO; SWAIN, 2018; PRAT et al., 2015; YAO et al., 2015).

O estudo mostrou que a espectroscopia Raman é capaz de diferenciar os subtipos moleculares do câncer de mama, podendo ser uma técnica de diagnóstico útil para a decisão terapêutica, sugerindo, assim, a utilização de uma metodologia (ou protocolo) baseada em técnicas ópticas (biópsia através de fibra óptica) para uma futura adoção na prática clínica.

#### **4.1 ESTUDOS FUTUROS E PERSPECTIVAS DO TRABALHO**

No futuro, espera-se que a espectroscopia Raman possa ser incorporada à prática clínica, podendo ser utilizada durante procedimentos cirúrgicos para avaliação das margens tumorais e para confirmação do subtipo molecular apresentado na biópsia pré-operatória.

Pretende-se realizar novo estudo avaliando os fragmentos de biópsia por agulha grossa através da espectroscopia Raman, para diagnóstico do câncer de mama e diferenciação dos subtipos moleculares. Ao se analisar o espectro Raman

das amostras de biópsia por agulha grossa pode-se aproximar a técnica para um futuro diagnóstico *in vivo*, com o uso de uma sonda Raman no interior da agulha grossa, adaptada para ser aplicada no diagnóstico em tempo real.

## 5. CONCLUSÃO

Este estudo mostrou que a espectroscopia Raman pode ser usada para identificar as alterações bioquímicas associadas aos subtipos moleculares dos tecidos do câncer de mama e para classificar os carcinomas luminais e não luminais com elevada acurácia.

O estudo mostrou que as diferenças bioquímicas entre os subtipos moleculares do carcinoma no tecido mamário *in vitro versus* tecido normal estão associadas principalmente a proteínas (colágeno) e lipídios. A análise exploratória por PCA mostrou aumento do colágeno em tumores luminal B, HER2 e triplo negativo. Tumores luminal B e HER2 apresentaram maior quantidade de colágeno III constituinte da matriz extracelular. Observou-se aumento de lipídios em tumores não luminais (HER2 e triplo negativo), enquanto o grupo de tumores luminal A apresentou aumento no colesterol.

Os modelos discriminantes que empregam a regressão PLS mostraram acurácia de 80,8% para classificação entre cada subtipo molecular e tecidos normais, acurácia de 89,3% para classificação entre subtipo luminal, subtipo não luminal e tecido normal, e acurácia de 97,2% para classificação entre câncer de mama e tecido normal (sensibilidade de 95,4% e especificidade de 99,0%).

O estudo realizado com a espectroscopia Raman para diferenciação dos subtipos moleculares do câncer de mama obteve resultados importantes, incentivando a realização de novos trabalhos com o objetivo de implantação da técnica na prática clínica para a avaliação dos subtipos moleculares *in vivo*.



## REFERÊNCIAS BIBLIOGRÁFICAS

ABRAMCZYK, H. et al. The role of lipid droplets and adipocytes in cancer. Raman imaging of cell cultures: MCF10A, MCF7, and MDA-MB-231 compared to adipocytes in cancerous human breast tissue. **Analyst**, v. 140, n. 7, p. 2224-2235, 2015.

ALFANO, R. R. et al. Human breast tissues studied by IR Fourier transform Raman spectroscopy. **Lasers in the Life Sciences**, v. 4, n. 1, p. 23-28, 1991.

BEATTIE, J. R.; ESMONDE-WHITE, F. W. L. Exploration of principal component analysis: deriving principal component analysis visually using spectra. **Applied Spectroscopy**, v. 75, n. 4, p. 361-375, 2021.

BENDAU, E. et al. Distinguishing metastatic triple-negative breast cancer from nonmetastatic breast cancer using second harmonic generation imaging and resonance Raman spectroscopy. **Journal of Biophotonics**, v. 13, n. 7, p. e202000005, 2020.

BI, X. et al. Evaluating HER2 amplification status and acquired drug resistance in breast cancer cells using Raman spectroscopy. **Journal of Biomedical Optics**, v. 19, n. 2, p. 025001, 2014.

BITAR, R. A. et al. Biochemical analysis of human breast tissues using Fourier-transform Raman spectroscopy. **Journal of Biomedical Optics**, v. 11, n. 5, p. 054001, 2006.

BODANESE, B. et al. Discrimination of basal cell carcinoma and melanoma from normal skin biopsies in vitro through Raman spectroscopy and principal component analysis. **Photomedicine and Laser Surgery**, v. 30, n. 7, p. 381-387, 2012.

BRAY, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: A Cancer Journal for Clinicians**, v. 68, n. 6, p. 394-424, 2018.

BROZEK-PLUSKA, B.; KOPEĆ, M.; ABRAMCZYK, H. Development of a new diagnostic Raman method for monitoring epigenetic modifications in the cancer cells of human breast tissue. **Analytical Methods**, v. 8, n. 48, p. 8542-8553, 2016.

CHATURVEDI, D. et al. Different Phases of Breast Cancer Cells: Raman Study of

Immortalized, Transformed, and Invasive Cells. **Biosensors**, v. 6, n. 4, p. 57, 2016.

CONKLIN, M. W. et al. Aligned collagen is a prognostic signature for survival in human breast carcinoma. **American Journal of Pathology**, v. 178, n. 3, p. 1221-1232, 2011.

COTRAN, R. S. KUMAR, V. AND ROBBINS., S. L. **Robbins' Pathologic Basis of Disease**. 4th. ed. Philadelphia: Wiley, p. 160, 1989.

DE GELDER, J. et al. Reference database of Raman spectra of biological molecules. **Journal of Raman Spectroscopy**, v. 38, n. 9, p. 1133-1147, 2007.

DE SANTANA, F. B. et al. Experimento didático de quimiometria para classificação de óleos vegetais comestíveis por espectroscopia no infravermelho médio combinado com análise discriminante por mínimos quadrados parciais: um tutorial, parte V. **Química Nova**, v. 43, n. 3, p. 371-381, 2020.

DE SOUZA, A. M. et al. Experimento didático de quimiometria para calibração multivariada na determinação de paracetamol em comprimidos comerciais utilizando espectroscopia no infravermelho próximo: Um tutorial, parte II. **Química Nova**, v. 36, n. 7, p. 271-281, 2013.

DENG, X. et al. Group and basis restricted non-negative matrix factorization and random forest for molecular histotype classification and Raman biomarker monitoring in breast cancer. **Applied Spectroscopy**, v. 76, n. 4, p. 462-474, 2022.

ELSTON, C. W.; ELLIS, I. O. Pathological prognostic factors in breast cancer. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. **Histopathology**, v. 19, n. 5, p. 403-410, 1991.

FERREIRA, M. M. C. et al. Chemometrics I: Multivariate calibration, a tutorial. **Química Nova**, v. 22, n. 5, p. 724-731, 1999.

FRANCHI, M. et al. Collagen fiber array of peritumoral stroma influences epithelial-to-mesenchymal transition and invasive potential of mammary cancer cells. **Journal of Clinical Medicine**, v. 8, n. 2, p. 213, 2019.

FRANZEN, L.; WINDBERGS, M. Applications of Raman spectroscopy in skin research - From skin physiology and diagnosis up to risk assessment and dermal drug delivery. **Advanced Drug Delivery Reviews**, v. 89, p. 91-104, 2015.

GAO, J. J.; SWAIN, S. M. Luminal A breast cancer and molecular assays: A review.

**The Oncologist**, v. 23, n. 5, p. 556-565, 2018.

GELADI, P. Chemometrics in spectroscopy. Part 1. Classical chemometrics **Spectrochimica Acta - Part B**, v. 58, n. 5, p. 767-782, 2003.

GOLDHIRSCH, A. et al. Personalizing the treatment of women with early breast cancer: Highlights of the St. Gallen international expert consensus on the primary therapy of early breast cancer 2013. **Annals of Oncology**, v. 24, n. 9, p. 2206-2223, 2013.

HAKA, A. S. et al. Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using Raman spectroscopy. **Cancer Research**, v. 62, n. 18, p. 5375-5380, 2002.

HAKA, A. S. et al. Diagnosing breast cancer by using Raman spectroscopy. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 35, p. 12371-12376, 2005.

HAKA, A. S. et al. In vivo margin assessment during partial mastectomy breast surgery using Raman spectroscopy. **Cancer Research**, v. 66, n. 6, p. 3317-3322, 2006.

HAKA, A. S. et al. Diagnosing breast cancer using Raman spectroscopy: prospective analysis. **Journal of Biomedical Optics**, v. 14, n. 5, p. 054023, 2009.

HAN, B. et al. Differences and relationships between normal and atypical ductal hyperplasia, ductal carcinoma in situ, and invasive ductal carcinoma tissues in the breast based on Raman spectroscopy. **Applied Spectroscopy**, v. 71, n. 2, p. 300-307, 2017.

HANLON, E. B. et al. Prospects for in vivo Raman spectroscopy. **Physics in Medicine and Biology**, v. 45, n. 2, p. R1-R59, 2000.

HE, J.; GU, Y.; ZHANG, S. Vitamin A and breast cancer survival: A systematic review and meta-analysis. **Clinical Breast Cancer**, v. 18, n. 6, p. e1389-e1400, 2018.

HONGYU, K; SANDANIELO, V. L. M.; OLIVEIRA JUNIOR, G. J. O. Análise de componentes principais: resumo teórico, aplicação e interpretação. **Engineering and Science**, v. 5, p. 1, 2015.

HU, F. et al. Carotenoids and breast cancer risk: a meta-analysis and meta-

regression. **Breast Cancer Research and Treatment**, v. 131, n. 1, p. 239-253, 2012.

HUSER, T.; CHAN, J. Raman spectroscopy for physiological investigations of tissues and cells. **Advanced Drug Delivery Reviews**, v. 89, p. 57-70, 2015.

INCA. INSTITUTO NACIONAL DE CANCER JOSÉ ALENCAR GOMES DA SILVA. **Estimativa 2016: incidência de câncer no Brasil**, 2016. Disponível em: <<https://www.inca.gov.br/estimativa/2016>>. Acesso em: 17 de mar. 2021.

KALLAWAY, C. et al. Advances in the clinical application of Raman spectroscopy for cancer diagnostics. **Photodiagnosis and Photodynamic Therapy**, v.10, p. 207-219, 2013.

KONG, K. et al. Towards intra-operative diagnosis of tumours during breast conserving surgery by selective-sampling Raman micro-spectroscopy. **Physics in Medicine and Biology**, v. 59, n. 20, p. 6141-6152, 2014.

KOYA, S. K. et al. Accurate identification of breast cancer margins in microenvironments of ex-vivo basal and luminal breast cancer tissues using Raman spectroscopy. **Prostaglandins and Other Lipid Mediators**, v. 151, p. 106475, 2020.

LI, Q. B. et al. Discrimination of breast cancer from normal tissue with Raman spectroscopy and chemometrics. **Journal of Applied Spectroscopy**, v. 82, n. 3, p. 450-455, 2015.

LI, Q.; GAO, Q.; ZHANG, G. Classification for breast cancer diagnosis with Raman spectroscopy. **Biomedical Optics Express**, v. 5, n. 7, p. 2435, 2014.

LI, Q.; HAO, C.; XU, Z. Diagnosis of breast cancer tissues using 785 nm miniature Raman spectrometer and pattern regression. **Sensors**, v. 17, n. 3, 2017.

ŁUKASIEWICZ, S. et al. Breast cancer - epidemiology, risk factors, classification, prognostic markers, and current treatment strategies - An updated review. **Cancers**, v. 13, n. 17, p. 4287, 2021.

MARRO, M. et al. Molecular monitoring of epithelial-to-mesenchymal transition in breast cancer cells by means of Raman spectroscopy. **Biochimica et Biophysica Acta - Molecular Cell Research**, v. 1843, n. 9, p. 1785-1795, 2014.

MEDEIROS, P. S. C. et al. Raman microspectroscopy for probing the impact of a

dietary antioxidant on human breast cancer cells. **Food and Function**, v. 7, n. 6, p. 2800-2810, 2016.

MOURANT, J. R. et al. Biochemical differences in tumorigenic and nontumorigenic cells measured by Raman and infrared spectroscopy. **Journal of Biomedical Optics**, v. 10, n. 3, p. 031106, 2005.

MOVASAGHI, Z.; REHMAN, S.; REHMAN, I. U. Raman spectroscopy of biological tissues. **Applied Spectroscopy Reviews**, v. 42, p. 493-541, 2007.

NABA, A. et al. Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters. **eLife Sciences**, v. 3, p. 1308, 2014.

NASSAR, A.; POOJA, A.; CHIVUKULA, M. HER2 testing across practices - Have we come to a consensus on the ideal method of testing? A systematic literature review. **Breast Journal**, v. 26, n. 6, p. 1185-1188, 2020.

NATAL, R. DE A. et al. Exploring collagen parameters in pure special types of invasive breast cancer. **Scientific Reports**, v. 9, n. 1, 2019.

NETTER, F. **The Netter collection of medical illustrations**. 2nd ed. Amsterdam: Elsevier, 2016. 3000 p.

NOUNOU, M. I. et al. Breast cancer: Conventional diagnosis and treatment modalities and recent patents and technologies supplementary issue: Targeted therapies in breast cancer treatment. **Breast Cancer: Basic and Clinical Research**, v. 9, p. 17-34, 2015.

NUNES, C. A. et al. Chemoface: A novel free user-friendly interface for chemometrics. **Journal of the Brazilian Chemical Society**, v. 23, n. 11, p. 2003-2010, 2012.

OSTERTAGOVÁ, E., OSTERTAG, O., & KOVÁČ, J. Methodology and application of the Kruskal-Wallis Test. **Applied Mechanics and Materials**, v. 611, p. 115-120, 2014.

PASSONI, D. J. **Análise de óleos lubrificantes utilizando dados do espectro Raman**. 2017. 64 f. Dissertação (Mestrado em Engenharia Mecânica). Universidade Santa Cecília, Santos, 2017. Disponível em: [chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://unisanta.br/arquivos/mestrado/mecanica/dissertacoes/Dissertacao\\_Douglas.pdf](chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://unisanta.br/arquivos/mestrado/mecanica/dissertacoes/Dissertacao_Douglas.pdf). Acesso em: 19 de abr. 2022.

PENG, C. et al. Circulating carotenoids and breast cancer among high-risk individuals. **American Journal of Clinical Nutrition**, v. 113, n. 3, p. 525-533, 2021.

PEROU, C. M. et al. Molecular portraits of human breast tumours. **Nature**, v. 406, n. 6797, p. 747-752, 2000.

PRAKASH, P.; KRINSKY, N. I.; RUSSELL, R. M. Retinoids, carotenoids, and human breast cancer cell cultures: A review of differential effects. **Nutrition Reviews**, v. 58, n. 6, p. 170-176, 2009.

PRAT, A. et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. **Breast**, v. 24, p. S26-S35, 2015.

RALBOVSKY, N. M.; LEDNEV, I. K. Raman spectroscopy and chemometrics: A potential universal method for diagnosing cancer. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 219, p. 463-487, 2019.

RAMAN, C. V.; KRISHNAN, K. S. A new type of secondary radiation. **Nature**, v. 121, p. 501-502, 1928.

REHMAN, S. et al. Raman spectroscopic analysis of breast cancer tissues: Identifying differences between normal, invasive ductal carcinoma and ductal carcinoma in situ of the breast tissue. **Journal of Raman Spectroscopy**, v. 38, n. 10, p. 1345-1351, 2007.

ROSEN, P. P. et al. **Rosen's breast pathology**. 5 ed. Philadelphia: Editora Wolters Kluwer/Lippincott Williams, 2020. 1028 p.

SANTOS, I. P. et al. Who's who? Discrimination of human breast cancer cell lines by Raman and FTIR. **Microspectroscopy**, v. 14, n. 2, p. 452, 2022.

SATHYAVATHI, R. et al. Raman spectroscopic sensing of carbonate intercalation in breast microcalcifications at stereotactic biopsy. **Scientific Reports**, v. 5, n. 1, p. 9907-9919, 2015.

SCHEUNEMAN, J. A method of assessing bias in test items. **Journal of Educational Measurement**, v. 16, n. 3, p. 143-152, 1979.

SILVEIRA, L. et al. Discriminating model for diagnosis of basal cell carcinoma and melanoma in vitro based on the Raman spectra of selected biochemicals. **Journal of Biomedical Optics**, v. 17, n. 7, p. 077003, 2012.

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**, v. 98, n. 19, p. 10869-10874, 2001.

STHLE, L.; WOLD, S. Analysis of variance (ANOVA). **Chemometrics and Intelligent Laboratory Systems**, v. 6, n. 4, p. 259-272, 1989.

SURMACKI, J. et al. Raman imaging at biological interfaces: Applications in breast cancer diagnosis. **Molecular Cancer**, v. 12, n. 1, p. 48, 2013.

SURMACKI, J. et al. The lipid-reactive oxygen species phenotype of breast cancer. Raman spectroscopy and mapping, PCA and PLSDA for invasive ductal carcinoma and invasive lobular carcinoma. Molecular tumorigenic mechanisms beyond Warburg effect. **Analyst**, v. 140, n. 7, p. 2121-2133, 2015.

TALARI, A. C. S. et al. Raman spectroscopic analysis differentiates between breast cancer cell lines. **Journal of Raman Spectroscopy**, v. 46, n. 5, p. 421-427, 2015.

TALARI, A. C. S. et al. Raman spectroscopy of biological tissues. **Applied Spectroscopy Reviews**. v. 50, n. 1, p. 46-111, 2015.

TÉLLEZ-PLANCARTE, A. et al. Development of a nanostructured platform for identifying HER2-heterogeneity of breast cancer cells by surface-enhanced Raman scattering. **Nanomaterials**, v. 8, n. 7, p. 549, 2018.

VINET, L.; ZHEDANOV, A. A “missing” family of classical orthogonal polynomials. **Journal of Physics A: Mathematical and Theoretical**, v. 44, n. 8, 2011.

WAKS, A. G.; WINER, E. P. Breast cancer treatment: A review, **Journal of the American Medical Association**, v. 321, n. 3, p. 288-300, 2019.

WORLD HEALTH ORGANIZATION. **Breast cancer**, 2021. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/breast-cancer>>. Acesso em: 22 de abr. 2022.

WINNARD, P. T. et al. Organ-specific isogenic metastatic breast cancer cell lines exhibit distinct Raman spectral signatures and metabolomes. **Oncotarget**, v. 8, n. 12, p. 20266-20287, 2017.

WOLFF, A. C. et al. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline focused update. **Archives of Pathology and Laboratory Medicine**, v. 142, n. 11, p. 1364-1382, 2018.

YAO, K. et al. Molecular subtyping improves diagnostic stratification of patients with primary breast cancer into prognostically defined risk groups. **Breast Cancer Research and Treatment**, v. 154, n. 1, p. 81-88, 2015.

ZHANG, L. et al. Raman spectroscopy and machine learning for the classification of breast cancers. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 264, p. 120300, 2022.

ZÚÑIGA, W. C. et al. Raman spectroscopy for rapid evaluation of surgical margins during breast cancer lumpectomy. **Scientific Reports**, v. 9, n. 1, p. 14639, 2019.



## ANEXO

CENTRO DE REFERÊNCIA DA  
SAÚDE DA MULHER - CRSM



## PARECER CONSUBSTANCIADO DO CEP

## DADOS DO PROJETO DE PESQUISA

**Título da Pesquisa:** Diagnóstico de alterações celulares e teciduais na mama humana através de técnicas ópticas (espectroscopias Raman, fluorescência)

**Pesquisador:** Alexandre Santos Melitto

**Área Temática:**

**Versão:** 2

**CAAE:** 92135518.3.0000.0069

**Instituição Proponente:** Centro de Referência da Saúde da Mulher

**Patrocinador Principal:** Financiamento Próprio

## DADOS DO PARECER

**Número do Parecer:** 2.788.474

**Apresentação do Projeto:**

Idem ao parecer 2.785.659 de 26/07/2018

**Objetivo da Pesquisa:**

Idem ao parecer 2.785.659 de 26/07/2018

**Avaliação dos Riscos e Benefícios:**

Idem ao parecer 2.785.659 de 26/07/2018

**Comentários e Considerações sobre a Pesquisa:**

Idem ao parecer 2.785.659 de 26/07/2018

**Considerações sobre os Termos de apresentação obrigatória:**

Idem ao parecer 2.785.659 de 26/07/2018

**Recomendações:**

Foi corrigido o item cronograma do qual havia sido emitido pendência.

**Conclusões ou Pendências e Lista de Inadequações:**

N.D.N>

**Considerações Finais a critério do CEP:**

De acordo com o parecer do relator.

**Endereço:** Av. Brigadeiro Luiz Antonio, 683

**Bairro:** Bela Vista

**CEP:** 01.317-010

**UF:** SP

**Município:** SAO PAULO

**Telefone:** (11)3248-8087

**Fax:** (11)3248-8080

**E-mail:** dircient@gmail.com

**CENTRO DE REFERÊNCIA DA  
SAÚDE DA MULHER - CRSM**



Continuação do Parecer: 2.788.474

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1150074.pdf	26/07/2018 08:49:28		Aceito
Projeto Detalhado / Brochura Investigador	Brochura_Investigador_v3.docx	24/06/2018 09:52:44	Alexandre Santos Melitto	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_Raman_V4.docx	24/06/2018 09:33:32	Alexandre Santos Melitto	Aceito
Outros	carta_consentimento.pdf	03/06/2018 11:35:51	Alexandre Santos Melitto	Aceito
Folha de Rosto	folha_rosto_.pdf	03/06/2018 11:34:08	Alexandre Santos Melitto	Aceito

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

SAO PAULO, 27 de Julho de 2018

---

**Assinado por:  
Roberto Euzebio dos Santos  
(Coordenador)**

**Endereço:** Av. Brigadeiro Luiz Antonio, 683

**Bairro:** Bela Vista

**CEP:** 01.317-010

**UF:** SP

**Município:** SAO PAULO

**Telefone:** (11)3248-8087

**Fax:** (11)3248-8080

**E-mail:** dircient@gmail.com