

ROBERTA RAYRA MARTINS CHAVES

**LEUCOPLASIAS ORAIS COM E SEM TRANSFORMAÇÃO MALIGNA
SÃO METABOLICAMENTE DIFERENTES**

Universidade Federal de Minas Gerais
Programa de Pós-Graduação em Medicina Molecular
Belo Horizonte – MG
2021

Roberta Rayra Martins Chaves

**LEUCOPLASIAS ORAIS COM E SEM TRANSFORMAÇÃO MALIGNA SÃO
METABOLICAMENTE DIFERENTES**

Tese apresentada ao Colegiado de Pós-Graduação da Faculdade de Medicina da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Doutora em Medicina Molecular.

Orientador: Prof. Dr. Ricardo Santiago Gomez

Belo Horizonte – MG
2021

Chaves, Roberta Rayra Martins
C512I Leucoplasias orais com e sem transformação maligna são metabolicamente diferentes [manuscrito]. / Roberta Rayra Martins Chaves. - - Belo Horizonte: 2021.

131 f.: il.

Orientador (a): Ricardo Santiago Gomez.

Área de concentração: Medicina Molecular.

Tese (doutorado): Universidade Federal de Minas Gerais, Faculdade de Medicina.

1. Leucoplasia Oral. 2. Carcinoma de Células Escamosas de Cabeça e Pescoço. 3. Metabolômica. 4. Cromatografia Líquida. 5. Espectrometria de Massas. 6. Detecção Precoce de Câncer. 7. Dissertação Acadêmica. I. Gomez, Ricardo Santiago. II. Universidade Federal de Minas Gerais, Faculdade de Medicina. III. Título.

NLM: WU 280

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM MEDICINA MOLECULAR

FOLHA DE APROVAÇÃO

LEUCOPLASIAS ORAIS COM E SEM TRANSFORMAÇÃO MALIGNA SÃO METABOLICAMENTE DIFERENTES

ROBERTA RAYRA MARTINS CHAVES

Tese de Doutorado defendida e aprovada, no dia dezois de agosto de dois mil vinte e um, pela Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação Medicina Molecular da Universidade Federal de Minas Gerais constituída pelos seguintes professores doutores:

Ricardo Santiago Gomez - Orientador
Universidade Federal de Minas Gerais (UFMG)

Felipe Paiva Fonseca
Universidade Federal de Minas Gerais (UFMG)

Fabrcio Rezende do Amaral
Faculdade de Ciências Gerenciais Padre Arnaldo Janssen (FAJANSSEN)

Renata Gonçaves de Resende
Faculdade de Ciências Gerenciais Padre Arnaldo Janssen (FAJANSSEN)

Hélder Antônio Rebelo Pontes
Universidade Federal do Pará (UFPA)

Belo Horizonte, 16 de agosto de 2021.



Documento assinado eletronicamente por **Renata Gonçaves de Resende**, Usuário Externo, em 16/08/2021, às 13:23, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ricardo Santiago Gomez**, Professor do Magistério Superior, em 16/08/2021, às 13:29, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Felipe Paiva Fonseca**, Coordenador(a), em 16/08/2021, às 20:07, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Fabrcio Rezende do Amaral**, Usuário Externo, em 17/08/2021, às 07:56, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Helder Antonio Rebelo Pontes**, Usuário Externo, em 17/08/2021, às 13:21, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador 0889673 e o código CRC 940656F7.

Dedico este trabalho à minha amada e saudosa
avó, Josefa Martins de Sousa.

AGRADECIMENTOS

Agradeço pela saúde dos meus familiares e minha, provinda pela proteção Divina e de todos os seres de luz nos quais acredito e deposito minha fé.

À minha saudosa avó, Josefa, a cada dia mais viva no íntimo do meu ser, suscitando os valores de generosidade, idoneidade e amor genuíno que me ensinou em vida. Eu te amo com todas as minhas forças, “Vó”.

Aos meus amados pais, Fátima e Roberto, e à minha admirável irmã, Larissa, pelo amor e apoio incondicional. Tudo o que eu faço é por e para vocês.

Às minhas queridas amigas, Camila, Júlia e Michele, pela presença e torcida.

Ao meu orientador, Prof. Dr. Ricardo Santiago Gomez, pela confiança em meu trabalho e pela imensurável contribuição para o meu desenvolvimento acadêmico e profissional.

Aos integrantes desta banca avaliadora, Prof. Dr. Fabrício Rezende Amaral, Prof. Dr. Felipe Paiva Fonseca, Prof. Dr. Hélder Antônio Rebelo Pontes, Prof.a. Dr.a. Renata Gonçalves Resende, e aos membros suplentes, Prof.a. Dr.a. Flávia Isabela Barbosa e Prof.a. Dr.a. Flávia Sirothêau Corrêa Pontes, pela disponibilidade e presteza em avaliar o nosso trabalho, mesmo meio à situação pandêmica extenuante que estamos vivenciando.

Às Professoras Dra. Liséte Celina Lange e Dra. Lucilaine Valéria de Souza Santos pela disponibilização da infraestrutura do Laboratório de Análise de Água do Departamento de Engenharia Sanitária e Ambiental (DESA) da UFMG durante a execução dos experimentos por LC-MS.

Aos alunos de pós-graduação Yuri Abner Rocha Lebron e Victor Rezende Moreira pela orientação, ensinamentos e presteza durante a execução dos experimentos por LC-MS.

Aos meus queridos colegas da pós-graduação, pela companhia leve e amistosa. Agradeço especialmente ao Victor Coutinho Bastos, pela parceria e contribuição para este estudo.

Aos colegas do Laboratório de Biologia Molecular, Jéssica Gardone Vitório, e Filipe Fidelis Duarte-Andrade, pelos ensinamentos sobre análise metabolômica compartilhados.

Ao Prof. Dr. Wagner Henriques de Castro, e a Prof.a. Dr.a. Flávia Isabela Barbosa, por introduzirem o pensamento científico em minha vida e pelo incentivo de sempre.

Ao Prof. Dr. Luiz Armando de Marco e ao Programa de Pós-Graduação em Medicina Molecular.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), por fomentar a minha bolsa de pós-graduação.

“A realidade biológica não é em preto e branco. Há também áreas cinza importantes.”

(Yuval Noah Harari)

RESUMO

O carcinoma de células escamosas oral é o principal tipo de câncer bucal e também é uma das neoplasias malignas mais comuns em todo o mundo, responsável por uma alta taxa de morbidade e mortalidade. Uma parte dos casos de câncer de boca é precedida por lesões de mucosa denominadas como doenças orais potencialmente malignas, e a leucoplasia oral é a doença oral potencialmente maligna mais frequente. Por isso, diversas pesquisas foram desenvolvidas a fim de elucidar os mecanismos subjacentes à malignização da doença. No entanto, os eventos biológicos que promovem a evolução maligna da leucoplasia oral ainda são pouco compreendidos. Alterações no perfil metabolômico entre amostras de leucoplasias orais, câncer oral, e indivíduos saudáveis já foram descritas na literatura. Entretanto, não há estudos que tenham comparado amostras de leucoplasias orais transformadas e não transformadas. Dessa forma, no presente estudo, objetivamos identificar as alterações metabólicas envolvidas com as etapas iniciais do processo de carcinogênese oral por meio da análise metabolômica *untargeted* de 5 amostras de leucoplasias orais transformadas e 15 amostras de leucoplasias orais não transformadas. As amostras foram obtidas a partir da coleta retrospectiva de blocos de tecido fixado em formol e incluído em parafina. Os blocos de tecido foram cortados, desparafinados e os metabólitos foram extraídos utilizando solução de metanol, clorofórmio e água. Em seguida, os extratos foram analisados utilizando cromatógrafo líquido de alta eficiência acoplado a espectrômetro de massas do tipo *quadrupole time-of-flight* (HPLC-QToF-MS). Após o pré-processamento e pré-tratamento dos dados adquiridos por HPLC-QToF-MS, métodos de análise estatística univariada e multivariada foram combinados para a detecção dos compostos discriminantes entre os grupos. A análise de enriquecimento de vias biológicas foi realizada por meio da associação entre os algoritmos *mummichog* e GSEA para a predição das vias biológicas enriquecidas nas amostras de leucoplasia oral. Dezesete compostos discriminantes foram detectados de modo correspondente em todos os métodos estatísticos aplicados. O agrupamento hierárquico não supervisionado exibiu o potencial destes compostos para a distinção dos grupos em estudo. As vias metabólicas enriquecidas foram majoritariamente relacionadas ao metabolismo lipídico, à síntese de moléculas pró-inflamatórias, ao metabolismo do ácido retinoico e da vitamina E, processos sabidamente associados à carcinogênese. Os resultados deste estudo revelaram, pela primeira vez na literatura, o potencial da metabolômica *untargeted* para a distinção das leucoplasias malignamente transformadas daquelas que não se transformaram. Ainda, foram sugeridas vias metabólicas possivelmente associadas ao processo de carcinogênese oral.

Palavras-chave: Leucoplasia Oral. Carcinoma de Células Escamosas Oral. Metabolômica. Cromatografia Líquida. Espectrometria de Massas. Diagnóstico Precoce do Câncer.

ABSTRACT

Oral squamous cell carcinoma is the main type of oral cancer and is also one of the most common malignancies worldwide, accounting for high morbidity and mortality rates. Some oral cancer cases are preceded by mucosal lesions called oral potentially malignant disorders. Oral leukoplakia is the most frequent oral potentially malignant disorder. Therefore, several studies have been performed to elucidate the mechanisms that underlie the lesion's malignant transformation. However, the biological events that promote the malignant progression of oral leukoplakia are still poorly understood. An altered metabolomic profile between samples of oral leukoplakia, oral cancer, and healthy controls was previously reported. However, no metabolomic study has compared transformed and non-transformed oral leukoplakia samples. Therefore, in the present study we aimed to identify the metabolic alterations at the early stages of the oral carcinogenesis process throughout untargeted metabolomic analysis of 5 transformed and 15 non-transformed oral leukoplakias. Samples were obtained from a retrospective collection of tissue blocks fixed in formalin and embedded in paraffin. Tissue blocks were cut, dewaxed and metabolites were extracted using methanol, chloroform and water solution. Then, the extracts were analyzed using high-performance liquid chromatography coupled to a quadrupole time-of-flight mass spectrometer (HPLC-QToF-MS). After pre-processing and pre-treatment of data acquired by HPLC-QToF-MS, univariate and multivariate statistical analysis methods were combined to detect discriminant compounds between groups. The enrichment analysis was performed through the association between the mummichog and GSEA algorithms for the prediction of enriched biological pathways in the samples of oral leukoplakia. Seventeen discriminant compounds were correspondingly detected in all applied statistical methods. The unsupervised hierarchical clustering revealed the potential of these compounds to distinguish the study groups. Enriched metabolic pathways were mostly related to lipid metabolism, synthesis of pro-inflammatory molecules, retinoic acid, and vitamin E metabolism, which are carcinogenesis-related processes. For the first time in the literature, the present study revealed the potential of untargeted metabolomics to distinguish malignantly transformed from untransformed oral leukoplakia. Additionally, we shed light on the metabolic pathways behind oral carcinogenesis.

Keywords: Oral Leukoplakia. Oral Squamous Cell Carcinoma. Metabolomics. Liquid Chromatography. Mass Spectrometry. Early Detection of Cancer.

LISTA DE FIGURAS

Figura 1	Características clínicas da leucoplasia oral.....	24
Figura 2	Características histopatológicas da leucoplasia.....	27
Figura 3	Fluxograma de pré-processamento.....	51
Figura 4	Esquema de classificação MUVR.....	56
Figura 5	Fluxo de trabalho desenvolvido no presente estudo.....	58
Figura 6	Cromatogramas de íons totais das amostras de QC.....	65
Figura 7	Análise de componente principal – PCA	66
Figura 8	Agrupamento hierárquico supervisionado	73
Figura 9	Análise discriminante de mínimos quadrados parciais – PLS-DA.....	74
Figura 10	Análise SAM.....	75
Figura 11	Agrupamento hierárquico não supervisionado.....	78
Figura 12	Gráficos de Curva ROC e <i>boxplot</i>	82
Figura 13	Gráficos de modelagem ROC multivariada e acurácia.....	83
Figura 14	Gráfico de dispersão para classificação dos grupos.....	84
Figura 15	Gráficos integrados do módulo <i>MS Peaks-to-Pathways</i>	87

LISTA DE TABELAS

Tabela 1	Informações clínicas das amostras do estudo.....	64
Tabela 2	Compostos diferencialmente abundantes.....	71
Tabela 3	Compostos selecionados a partir do critério de correspondência entre as abordagens estatísticas univariadas, análise SAM e classificação MUVR.....	77
Tabela 4	Valores de AUC e <i>p.</i> para análise ROC clássica.....	81

LISTA DE ABREVIATURAS E SIGLAS

μm	Micrômetro
$^1\text{H NMR}$	Ressonância Magnética Nuclear de Prótons
4NQO	1-óxido de 4-nitroquinolina
APCI	Ionização Química por Pressão Atmosférica
BER	Taxa de Erro Balanceada
<i>BRCA1</i>	<i>Breast Cancer 1</i>
<i>BRCA2</i>	<i>Breast Cancer 2</i>
<i>BTBD7</i>	<i>BTB Domain Containing 7</i>
<i>CCDN1</i>	Cyclin D1
CCE	Carcinoma de Células Escamosas
<i>CDKN2A</i>	<i>Cyclin Dependent Kinase Inhibitor 2A</i>
CE-MS	Eletroforese Capilar - Espectrometria de Massa
circ-RNA	RNA circular
<i>COL5A1</i>	<i>Collagen Type V Alpha 1 Chain</i>
CPSI-MS	Espectrometria de Massa de Ionização por Spray de Polímero Condutivo
<i>CXCL</i>	<i>C-X-C Motif Chemokine Ligand 1</i>
CXCR	<i>C-X-C Motif Chemokine Receptor</i>
<i>DCLRE1B</i>	<i>DNA Cross-Link Repair 1B</i>
DESA	Departamento Engenharia Sanitária e Ambiental Desorção de Imagens de Ionização por Eletrospray - Espectrometria de
DESI-MS	Massas
DNA	Ácido desoxirribonucleico
DOK	Queratinócito Displásico Oral
DOPM	Doenças Oraís Potencialmente Malignas
DP	Desvião Padrão
<i>ECT2</i>	<i>Epithelial Cell Transforming 2</i>
<i>EGFR</i>	<i>Epidermal Growth Factor Receptor</i>

<i>EIF5A2</i>	<i>Eukaryotic Translation Initiation Factor 5A2</i>
EMMPRIN	Indutor Extracelular de Metaloproteinase de Matriz
<i>ERBB3</i>	<i>Erb-B2 Receptor Tyrosine Kinase 3</i>
ERK	Quinase Regulada por Sinal Extracelular
ESI	Ionização por Eletrospray
<i>FANCA</i>	<i>FA Complementation Group A</i>
FAO-UFMG	Faculdade de Odontologia da Universidade Federal de Minas Gerais
<i>FDR</i>	Taxa de Descoberta Falsa
FFPE	Tecido fixado em Formol e Incluído em Parafina
<i>FHIT</i>	<i>Fragile Histidine Triad</i>
GC-MS	Cromatografia Gasosa-Espectrometria de Massas
GDI	Inibidor de Dissociação de GDP
GDP	Difosfato de Guanosina
<i>GPR160</i>	<i>G Protein-Coupled Receptor 160</i>
GSEA	Análise de Enriquecimento de Conjuntos de Genes
H&E	Hematoxilina e Eosina
<i>HBEGF</i>	<i>Heparin Binding EGF Like Growth Factor</i>
HILIC	Cromatografia de Interação Hidrofílica
HOEC	Células Humanas Epiteliais Orais
<i>HOXC13</i>	<i>Homeobox protein Hox-C13</i>
<i>HOXC9</i>	<i>Homeobox protein Hox-C9</i>
HPLC	Cromatografia Líquida de Alta Performance
HPV	Papilomavírus Humano
HSP	Proteína de Choque Térmico
HUJBB	Hospital Universitário João de Barros Barreto
IL6	Interleucina 6
ILK	Quinase Ligada à Integrina
IPO	Otimização de Parâmetros Isotopológicos
JNK	Quinase c-Jun de N Terminal

<i>KHDRBS1</i>	<i>KH domain containing, RNA binding, signal transduction associated 1</i>
<i>KLHL6</i>	<i>Kelch Like Family Member 6</i>
<i>KMTC2</i>	<i>Lysine Methyltransferase 2C</i>
kNN	K-ésimo Vizinho mais Próximo
kV	Kilovolt
L/h	Litros por hora
LC-MS	Cromatografia Líquida_Espectrometria de Massas
<i>m/z</i>	Razão Massa/Carga
<i>MAD/Median</i>	Desvio Absoluto da Mediana/Mediana
MAPK	Proteína Quinase Ativada por Mitógeno
MHC-1	Complexo de Histocompatibilidade Maior de Classe 1
mi-RNA	Micro RNA
MMP	Metaloproteinase
mRNA	RNA mensageiro
<i>MUCL1</i>	<i>Mucin-like protein 1</i>
MUVR	Seleção de Variáveis Minimamente Enviesada no R
nl	Nanolitro
NMR	Ressonância Magnética Nuclear
PA28ab	Subunidade Ativadora do Proteassoma 28ab
<i>PAIP2</i>	<i>Binding Protein Interacting Protein 2</i>
PAK	Quinase Ativada P21
<i>PARP1</i>	<i>Poly(ADP-Ribose) Polymerase 1</i>
PCA	Análise de Componentes Principal
PD-1	Proteína de Morte Celular Programada 1
PD-L1	Proteína Ligante de Morte Celular Programada 1
PLS-DA	Análise de Mínimos Quadrados Parciais
ppm	Partes por Milhão
QC	Controle de Qualidade
QTOF	Quadrupolo por Tempo de Vôo

R.S.G.	Ricardo Santiago Gomez
<i>RAB1A</i>	<i>Ras-related protein Rab-1A</i>
<i>RAS</i>	<i>Rat Sarcoma Virus</i>
<i>RASSF1</i>	<i>Ras Association Domain Family Member 1</i>
<i>RB1</i>	<i>RB Transcriptional Corepressor 1</i>
RdCV	Validação Cruzada Dupla Repetida
RF	Floresta Aleatória
<i>Rho</i>	<i>Rhodopsin</i>
RNA	Ácido ribonucleico
ROC	Curva Característica de Operação do Receptor
RSD	Desvio Padrão Relativo
<i>S100A7</i>	<i>S100 calcium-binding protein P</i>
<i>S100P</i>	<i>S100 calcium-binding protein A8</i>
SAM	Análise de Significância de Microarray
<i>STAT3</i>	<i>Signal Transducer And Activator Of Transcription 3</i>
TGF β	Factor Transformador de Crescimento Beta
TGM3	Transglutaminase 3
Th2	Célula T Auxiliar Tipo 2
<i>TIAM1</i>	<i>T-lymphoma invasion and metastasis-inducing protein 1</i>
<i>TP53</i>	Proteína Tumoral p53
tRNA	RNA transportador
UFPA	Universidade Federal do Pará
WNT	Sítio de Integração Relacionado ao Wingless

SUMÁRIO

1. INTRODUÇÃO	18
2. REVISÃO DA LITERATURA	20
2.1 Doenças orais potencialmente malignas	20
2.2 Leucoplasia oral	23
2.3 Abordagens moleculares para avaliação do potencial de progressão maligna da leucoplasia oral	28
2.3.1 Estudos imunohistoquímicos	28
2.3.2 Estudos ômicos: genômica, transcriptômica, proteômica e metabolômica	31
2.3.2.1 Estudos genômicos	32
2.3.2.2 Estudos transcriptômicos e proteômicos	34
2.3.2.3 Estudos metabolômicos	38
3. OBJETIVOS	46
3.1 Objetivo geral	46
3.2 Objetivos específicos	46
4. MATERIAIS E MÉTODOS	46
4.1 Aspectos éticos	46
4.2 Amostras	47
4.3 Reagentes	47
4.4 Análise histopatológica	47
4.5 Preparo das amostras e extração de metabólitos	48
4.6 Análise por cromatografia líquida de alta performance acoplada a espectrômetro de massas	49
4.7 Processamento dos dados	50
4.7.1 Pré-processamento dos dados brutos	50
4.7.2 Pré-tratamento dos dados	52
4.8.2 Análises estatísticas multivariadas	55
4.8.3 Seleção recursiva dos <i>molecular features</i> discriminantes	55
4.9 Avaliação da performance dos <i>molecular features</i> discriminantes	57
5. RESULTADOS E DISCUSSÃO	59
5.1 Caracterização da amostra	59
5.2 Controle de qualidade dos experimentos	65
5.3 Pré-processamento e pré-tratamento dos dados	66
5.4 Seleção de variáveis: Detecção dos compostos discriminantes	68
5.4.1 Análises estatísticas univariadas	69
5.4.2 Análises estatísticas multivariadas	73
5.4.3 Seleção recursiva de variáveis – MUVR	75

5.5	Performance dos <i>molecular features</i> selecionados para a discriminação dos grupos em estudo.....	79
5.6	Análise de enriquecimento de vias metabólicas	84
6.	CONCLUSÃO	91
7.	PERSPECTIVAS	91
	REFERÊNCIAS	92
	APÊNDICE A – Pré-processamento dos dados <i>softwares</i> IPO e XCMS	109
	APÊNDICE B – Normalização dos dados	115
	APÊNDICE C – Seleção recursiva dos <i>molecular features</i> discriminantes.....	117
	APÊNDICE D – Cromatogramas de íons totais da análise por LC-MS	123
	APÊNDICE E – Modelos PCA e PLS-DA	124
	APÊNDICE F – Análise de enriquecimento de vias metabólicas	125
	ANEXO A – Parecer consubstanciado do Comitê de Ética em Pesquisa (CEP)	127
	ANEXO B – Atividades desenvolvidas durante o doutorado	129

1. INTRODUÇÃO

O carcinoma de células escamosas (CCE) da região de cabeça e pescoço é um câncer que pode se desenvolver a partir das células epiteliais da mucosa da cavidade oral, faringe ou laringe e representa a sexta neoplasia maligna mais frequente em todo o mundo (JOHNSON et al., 2020). O carcinoma oral de células escamosas é o subtipo mais comum de câncer de cabeça e pescoço (FERLAY et al., 2019). Segundo o observatório global de câncer da Organização Mundial da Saúde, é previsto um aumento de aproximadamente 42% em sua incidência até o ano de 2040 (GLOBOCAN – IARC, 2021).

Uma parte dos indivíduos afetados por neoplasias malignas de cabeça e pescoço possui um histórico de mudanças locais antecessoras ao desenvolvimento do câncer, também denominadas como lesões potencialmente malignas, que são usualmente representadas pela leucoplasia e eritroplasia (BHURGRI et al., 2006; JOHNSON et al., 2020; VEERAMACHANENI et al., 2019). Na cavidade oral, o espectro de doenças que trazem um risco aumentado para o desenvolvimento do câncer de boca é mais diverso. Esses processos patológicos são referidos como doenças orais potencialmente malignas, e constituem um grupo de desordens com diferentes fatores etiológicos (WARNAKULASURIYA; JOHNSON; VAN DER WAAL, 2007).

A leucoplasia oral é a doença oral potencialmente maligna mais frequente (WARNAKULASURIYA, 2018; WILS et al., 2020). Estima-se que em aproximadamente 10% dos casos a doença pode sofrer transformação maligna em CCE, no entanto, essa taxa pode variar de 1,1 a 40%, dependendo da população estudada. (AGUIRRE-URIZAR; LAFUENTE-IBÁÑEZ DE MENDOZA; WARNAKULASURIYA, 2021). Devido à prevalência e à considerável taxa de transformação maligna da leucoplasia oral, diversos pesquisadores têm se empenhado para elucidar os mecanismos subjacentes ao processo da carcinogênese oral, no intuito de identificar alterações que possam prever o risco de malignização da doença (BARAN et al., 2019; BHOSALE et al., 2017; SIEBERS et al., 2013; SPERANDIO et al., 2013; WEBER et al., 2020; WU et al., 2018).

Embora uma variedade de estudos tenham apontado associações entre alterações moleculares e a transformação da leucoplasia oral em CCE, até o momento, a presença da lesão em indivíduos do sexo feminino, não fumantes, localizadas em língua e soalho bucal, com aparência clínica não homogênea, e quadro histopatológico caracterizado por graus elevados de displasia epitelial oral, são os únicos fatores associados de forma concreta com a evolução maligna da doença (AGUIRRE-URIZAR; LAFUENTE-IBÁÑEZ DE MENDOZA; WARNAKULASURIYA, 2021).

O entendimento dos eventos biológicos subjacentes à malignização da leucoplasia é fundamental para que a elaboração de estratégias para a prevenção do câncer oral se torne possível. Dentre as abordagens laboratoriais, a metabolômica é a que apresenta o maior potencial para a identificação de moléculas biomarcadoras que possibilitem o reconhecimento, por meio de testes clínicos não invasivos, de pacientes sob risco aumentado de desenvolvimento e progressão do câncer (WISHART, 2016). Isso porque a metabolômica foca na detecção dos produtos *downstream* aos processos genômicos e proteômicos e, dessa forma, sintetiza os efeitos produzidos pelos outros níveis ômicos, sendo a abordagem que mais fielmente representa o fenótipo de um sistema biológico. (DINGES et al., 2019) Uma vez que o processo de carcinogênese apresenta uma marcante heterogeneidade de eventos moleculares, a metabolômica, ponto final da cascata ômica, pode refletir o estado atual expresso da mucosa oral analisada.

Diversos estudos metabolômicos propuseram moléculas que podem estar associadas à transformação maligna da leucoplasia oral (BAG et al., 2016; GUPTA; GUPTA; MAHDI, 2015; KONG et al., 2015; SONG et al., 2020; SRIDHARAN et al., 2019; SRIDHARAN; RAMANI; PATANKAR, 2017; WEI et al., 2011; XIE et al., 2012; ZHOU et al., 2009). Entretanto, todos esses estudos compararam a leucoplasia com o CCE e/ou com amostras de pacientes saudáveis. Não há na literatura publicada até a presente data, trabalhos que utilizaram a metabolômica para determinar as diferenças entre espécimes de leucoplasias transformadas e não transformadas. A discriminação entre leucoplasias que sofreram ou não transformação maligna é fundamental para a identificação das alterações envolvidas nos estágios iniciais da carcinogênese oral. Dessa forma, realizamos na presente pesquisa, uma análise metabolômica global de leucoplasias transformadas e não transformadas e demonstramos o potencial dessa técnica para a discriminação entre os dois grupos estudados.

2. REVISÃO DA LITERATURA

2.1 Doenças orais potencialmente malignas

As doenças orais potencialmente malignas (DOPM) constituem um grupo heterogêneo de lesões que oferecem um risco aumentado de transformação maligna em CCE, o principal tipo de câncer bucal (WARNAKULASURIYA; JOHNSON; VAN DER WAAL, 2007). Previamente, as DOPM também podiam ser referidas como “pré-malignas” ou “precursoras”, no entanto, o uso dessas terminologias subentende a ideia de uma inevitável progressão maligna, quando este, na verdade, é um evento raro (SPEIGHT; KHURRAM; KUJAN, 2018; WARNAKULASURIYA, 2018). O uso do termo DOPM sugere que a presença da lesão constitui uma doença de toda a mucosa, o que vai de encontro com a teoria do “campo de cancerização”, que indica um risco aumentado para o desenvolvimento do CCE em áreas de mucosa com aparência clínica normal, e não somente nos sítios afetados pelas lesões (MARKOPOULOS, 2012; WARNAKULASURIYA; JOHNSON; VAN DER WAAL, 2007). Portanto, durante o encontro do ano de 2007, os membros da Organização Mundial da Saúde definiram DOPM como o melhor descritor para esse grupo de processos patológicos (WARNAKULASURIYA; JOHNSON; VAN DER WAAL, 2007).

Em relação à etiopatogenia, as DOPM podem ser subdivididas em geneticamente adquiridas, induzidas pelo tabaco, imuno-mediadas, e geneticamente herdadas (GANESH et al., 2018). A leucoplasia e a eritroplasia são fortemente associadas ao uso crônico de tabaco, álcool, e/ou betel, enquanto a queilite actínica é relacionada com a exposição à radiação ultravioleta emitida pela luz solar (MELLO et al., 2018). Estes fatores carcinogênicos produzem alterações moleculares a nível genômico que culminam no dano à resposta de reparo tecidual, e por isso, as três doenças são classificadas como DOPM geneticamente adquiridas (GANESH et al., 2018). Já a fibrose submucosa oral e a queratose palatina associada ao fumo invertido são induzidas exclusivamente pelo tabaco, enquanto o líquen plano oral e o lúpus eritematoso discoide representam as DOPM imuno-mediadas (GANESH et al., 2018). A disqueratose congênita e a epidermólise bolhosa configuram condições geneticamente herdadas que

aumentam significativamente o risco de desenvolvimento do câncer bucal (BONGIORNO et al., 2017; FINE; MELLERIO, 2009; GANESH et al., 2018).

A candidíase hiperplásica crônica, as lesões orais liquenoides, a hiperplasia exofítica verrucosa e as lesões orais causadas pela doença do enxerto *versus* hospedeiro foram reconhecidas como DOPM durante a última década (WARNAKULASURIYA, 2020a). Cada subtipo desse grupo de doenças apresenta características clínicas definidoras que vêm sendo descritas em diversos estudos de revisão (PARAKH et al., 2020; WARNAKULASURIYA, 2018). Em ordem de salientar os aspectos relevantes para o diagnóstico diferencial das DOPM, Warnakulasuriya, (2020) sintetizou os sinais, sintomas, métodos de diagnóstico e manejo dos pacientes afetados (WARNAKULASURIYA, 2020b). Essa discussão é extensa e foge do escopo deste trabalho e, por isso, encorajamos a leitura das referências supracitadas.

A displasia epitelial oral é o desarranjo da arquitetura epitelial com a presença de atipias celulares, e pode ser observada na análise histopatológica das DOPM (IOCCA et al., 2020a; SPEIGHT; KHURRAM; KUJAN, 2018). Alterações arquiteturais, organizacionais, e citológicas mínimas limitadas a um terço do epitélio são classificadas como displasia epitelial oral leve, enquanto mudanças abrangendo além de dois terços da espessura epitelial são definidas como displasia epitelial oral moderada (SPEIGHT; KHURRAM; KUJAN, 2018). A displasia epitelial oral severa apresenta um nível elevado de alterações por toda a extensão do epitélio, onde um carcinoma *in situ*, ou seja, um CCE restrito à faixa epitelial pode ser encontrado (SPEIGHT; KHURRAM; KUJAN, 2018). Uma associação estatisticamente significativa entre a presença de displasia epitelial oral moderada e/ou severa e o risco aumentado de progressão maligna das DOPM tem sido relatada na literatura (IOCCA et al., 2020a; WARNAKULASURIYA et al., 2011). É válido ressaltar que a gradação da displasia epitelial de acordo com a subdivisão do epitélio da mucosa oral em terços é controversa. Em alguns sítios, como na superfície ventral da língua e assoalho bucal, a espessura do epitélio é biologicamente menor. Nestes casos, as alterações displásicas podem facilmente acometer todo epitélio e, por isso, o tipo e a frequência da modificação celular observada devem se sobrepor ao critério de extensão para a definição do grau de displasia observado.

Existe um subgrupo de displasia epitelial oral com características histopatológicas distintas, que apresenta um número excessivo de células apoptóticas e cariorrexe. Alguns estudos mostraram positividade intensa para HPV (*Human papillomavirus*) e superexpressão da proteína p16 em lesões displásicas marcadas por esse quadro histopatológico, definindo-as como displasias epiteliais orais associadas ao HPV (LERMAN et al., 2017; WOO; CASHMAN; LERMAN, 2013). Embora uma associação causal entre a infecção por HPV e o desenvolvimento das DOPM tenha sido postulada por alguns autores (JAYAPRAKASH et al., 2011; SYRJÄNEN et al., 2011), o papel do vírus no desenvolvimento e progressão dessas doenças permanece incerto (DE LA COUR et al., 2020; WARNAKULASURIYA, 2020a).

O exame de biópsia é considerado padrão ouro para o diagnóstico das DOPM, que é obtido por meio da correlação do quadro histopatológico com os achados clínicos (SPEIGHT; KHURRAM; KUJAN, 2018; WARNAKULASURIYA, 2020a). O maior desafio relacionado a este grupo de desordens é o estabelecimento do plano de tratamento e acompanhamento adequados (IOCCA et al., 2020a), tendo em vista que o risco de progressão maligna varia consideravelmente de acordo com o subtipo de DOPM, com o tipo de alterações moleculares presentes e com o nível de exposição aos fatores carcinogênicos (CHUANG et al., 2018; WARNAKULASURIYA, 2020a).

A classificação das DOPMs em relação ao risco de malignização também é dificultada pela variação na prevalência observada entre populações distintas (MELLO et al., 2018). A taxa geral de transformação de uma DOPM em CCE varia entre 0,13% e 34% (WARNAKULASURIYA; ARIYAWARDANA, 2016). Embora estudos recentes tenham mostrado certa consistência, revelando uma frequência de casos malignamente transformados de aproximadamente 8% (CERQUEIRA et al., 2020; CHUANG et al., 2018; EVREN et al., 2020; IOCCA et al., 2020a; JAYASOORIYA et al., 2020; RUBERT; BAGÁN; BAGÁN, 2020), o número pode não retratar a realidade. Relatos de casos com maior tempo de acompanhamento clínico é fundamental para a compreensão da frequência deste evento. Dentre as DOPM, a leucoplasia proliferativa; eritroplasia; hiperplasia exofítica verrucosa; candidíase hiperplásica; leucoplasia e fibrose submucosa oral são as lesões que mais frequentemente evoluem para o câncer de boca (CHIANG et al., 2020; CHUANG et al., 2018; IOCCA et al., 2020a).

2.2 Leucoplasia oral

A leucoplasia oral é a DOPM mais comum, com prevalência estimada de 4,11% na população global (MELLO et al., 2018; WILS et al., 2020). A doença foi conceituada pela primeira vez no ano de 1978 pela Organização Mundial da Saúde como “uma mancha ou placa branca que não pode ser caracterizada clinicamente ou histopatologicamente como qualquer outra desordem” (KRAMER IR, 1978). O conceito da lesão foi alterado por mais três vezes ao longo dos anos (AXELL et al., 1984; VAN DER WAAL, 2015; WARNAKULASURIYA; JOHNSON; VAN DER WAAL, 2007) e, hoje, a leucoplasia oral é definida como “uma mancha ou placa predominantemente branca que não pode ser caracterizada clinicamente ou histopatologicamente como qualquer outra desordem e que carrega um risco aumentado para o desenvolvimento do câncer tanto na área leucoplásica como em qualquer outro sítio anatômico da cavidade oral” (VAN DER WAAL, 2015).

Clinicamente, a placa branca não removível à raspagem que caracteriza a leucoplasia oral pode apresentar alterações na coloração de sua superfície, em sua espessura e em sua textura, mudanças que distinguem a lesão em dois subtipos clínicos: homogênea e não homogênea. A leucoplasia oral homogênea é usualmente assintomática e corresponde a uma placa uniformemente branca, fina, com superfície lisa, que pode exibir alguns sulcos superficiais. As leucoplasias orais não homogêneas podem ser nodulares, exibindo protuberâncias polipoides de formato arredondado; verrucosas, apresentando superfície mais espessa de aparência corrugada, ou podem se apresentar como eritroleucoplasias, quando há a presença de algumas áreas eritematosas ao longo da placa branca. As áreas avermelhadas podem estar associadas à infecção por cândida. Sintomas como desconforto ou aumento de sensibilidade são raros, mas, podem ser observados, especialmente nas eritroleucoplasias (WARNAKULASURIYA, 2018). A aparência não homogênea foi associada ao risco aumentado de progressão maligna da leucoplasia oral, contudo, os mecanismos biológicos subjacentes a essa associação ainda não foram esclarecidos (CERQUEIRA et al., 2020; JAYASOORIYA et al., 2020; RUBERT; BAGÁN; BAGÁN, 2020; SILVERMAN; GORSKY; LOZADA, 1984; WU; WANG; ZHOU, 2019). Os subtipos clínicos da leucoplasia oral podem ser observados na Figura 1 A-C.



Figura 1. Subtipos clínicos da leucoplasia oral. **(A)** Leucoplasia oral homogênea. Lesão em placa uniformemente branca, aderida à mucosa gengival superior esquerda. Alguns sulcos superficiais podem ser observados na superfície da lesão. **(B)** Leucoplasia oral não homogênea. Lesão leucoplásica nodular, predominantemente branca, em comissura labial entendendo-se discretamente em direção à mucosa jugal direita. A superfície da lesão é irregular com a formação de projeções verrucosas. Um pequeno ponto eritematoso pode ser observado em sua porção superior. **(C)** Eritroleucoplasia. Lesão em placa de espessura variável em borda lateral esquerda de língua. Nota-se a presença de áreas eritematosas entremeadas à placa branca, principalmente na região inferior direita da imagem. A superfície da área leucoplásica é corrugada, na qual sulcos epiteliais mais aprofundados podem ser observados.

Fonte: Imagens cedidas pelo Prof. Dr. Ricardo Santiago Gomez.

A doença afeta usualmente indivíduos do sexo masculino, tabagistas, etilistas, a partir da quarta década de vida. A borda lateral de língua, o soalho bucal, a mucosa gengival e jugal são os sítios orais mais frequentemente acometidos (WARNAKULASURIYA, 2018). Uma tendência de progressão maligna foi observada em lesões desenvolvidas em indivíduos do sexo feminino, não tabagistas e localizadas em língua. No entanto, maiores coortes observacionais são necessárias para o melhor entendimento dessa associação (CERQUEIRA et al., 2020).

As leucoplasias orais geralmente se manifestam como lesões únicas. Lesões leucoplásicas múltiplas ou extensas, localizadas sobre uma ampla área da mucosa oral e que possuem tendência à recidiva após a excisão, são denominadas como leucoplasia verrucosa proliferativa (WARNAKULASURIYA, 2020b). A leucoplasia oral e a leucoplasia verrucosa proliferativa compartilham alterações histopatológicas a nível arquitetural, organizacional e citológico semelhantes (LI et al., 2020). Contudo, apesar de por vezes apresentar quadro histopatológico aparentemente inofensivo, a leucoplasia verrucosa proliferativa exhibe comportamento biológico agressivo, progredindo para CCE em cerca de 70–100% dos casos (LI et al., 2020). Portanto, a doença deve ser considerada no diagnóstico diferencial de lesões leucoplásicas extensas (CERERO-LAPIEDRA et al., 2010; HANSEN; OLSON; SILVERMAN, 1985). Outras lesões que se apresentam clinicamente como placas brancas como a queratose friccional, a queratose de rebordo alveolar, a candidíase pseudomembranosa, o leucoedema, a

leucoplasia pilosa, a doença do enxerto *versus* hospedeiro e a leucoqueratose palatina nicotínica também devem ser incluídas como hipóteses de diagnóstico diferencial clínico da leucoplasia oral (WARNAKULASURIYA, 2020b).

A análise anatomohistopatológica de um espécime representativo da lesão leucoplásica é necessária para a definição do diagnóstico de leucoplasia (WARNAKULASURIYA, 2018). As características histopatológicas são geralmente representativas do quadro clínico observado. Leucoplasias homogêneas e delgadas usualmente exibem discreta hiperqueratose, acantose e escasso infiltrado inflamatório subepitelial (NEVILLE et al., 2016).

Lesões não homogêneas de aparência clínica verrucosa são caracterizadas principalmente pela hiperplasia epitelial verrucosa ou papilar, com padrão de crescimento exofítico e/ou endofítico, além de hiperqueratose ou hiperparakeratose acentuada. A displasia epitelial oral pode ou não estar presente. Alguns casos de leucoplasia verrucosa podem exibir atrofia epitelial associada à hiperqueratose ou hiperparakeratose e infiltrado inflamatório linfocitário de interface. Já foi sugerido que a presença desse quadro indica que o desenvolvimento da lesão se deu por uma proliferação clonal das células epiteliais, ao contrário de uma resposta tecidual a alguma injúria local. Isso porque a presença de atrofia epitelial em resposta a irritação local não é um achado usual (LI et al., 2020). Portanto, lesões hiperqueratóticas com atrofia epitelial e banda linfocitária na interface epitélio/lâmina própria devem ser avaliadas com cautela, uma vez que essa morfologia “liquenoide” pode ser um sinal histopatológico precoce de uma leucoplasia proliferativa (GILLIGAN et al., 2020; MCPARLAND; WARNAKULASURIYA, 2020).

As eritroleucoplasias mostram geralmente hiperqueratose ou hiperparakeratose irregular, atrofia epitelial, infiltrado inflamatório denso e displasia epitelial oral moderada ou severa. Ocasionalmente um CCE *in situ* ou invasivo pode ser encontrado nessas lesões (NEVILLE et al., 2016). Li e colaboradores, (2020) propuseram novos critérios para a análise da displasia epitelial oral (LI et al., 2020). Segundo os autores, as mudanças displásicas devem ser subdivididas entre arquiteturais, organizacionais e citológicas. As alterações arquiteturais displásicas são características que podem ser facilmente observadas na microscopia em menor aumento, como a morfologia epitelial corrugada,

papilar ou verrucosa, a hiperqueratose demarcada e/ou segmentada, a atrofia epitelial, a proliferação escamosa de padrão endofítico ou exofítico e a formação de projeções epiteliais em gota (LI et al., 2020). As mudanças organizacionais, previamente descritas como arquiteturais (REIBEL et al., 2017), podem ser visualizadas em aumento intermediário e se referem à relação entre os queratinócitos, como a perda de maturação e coesão celular, disqueratose, formação de pérolas de queratina e a presença de atividade mitótica suprabasal. As alterações citológicas são visualizadas em maior aumento e correspondem ao pleomorfismo celular/nuclear, a relação núcleo/citoplasma aumentada, ao hiperchromatismo nuclear, aos nucléolos evidentes e múltiplos, a cromatina grosseira, a atividade mitótica aumentada, e a presença de mitoses atípicas (LI et al., 2020). O quadro histopatológico da leucoplasia oral é ilustrado na Figura 2 A-C.

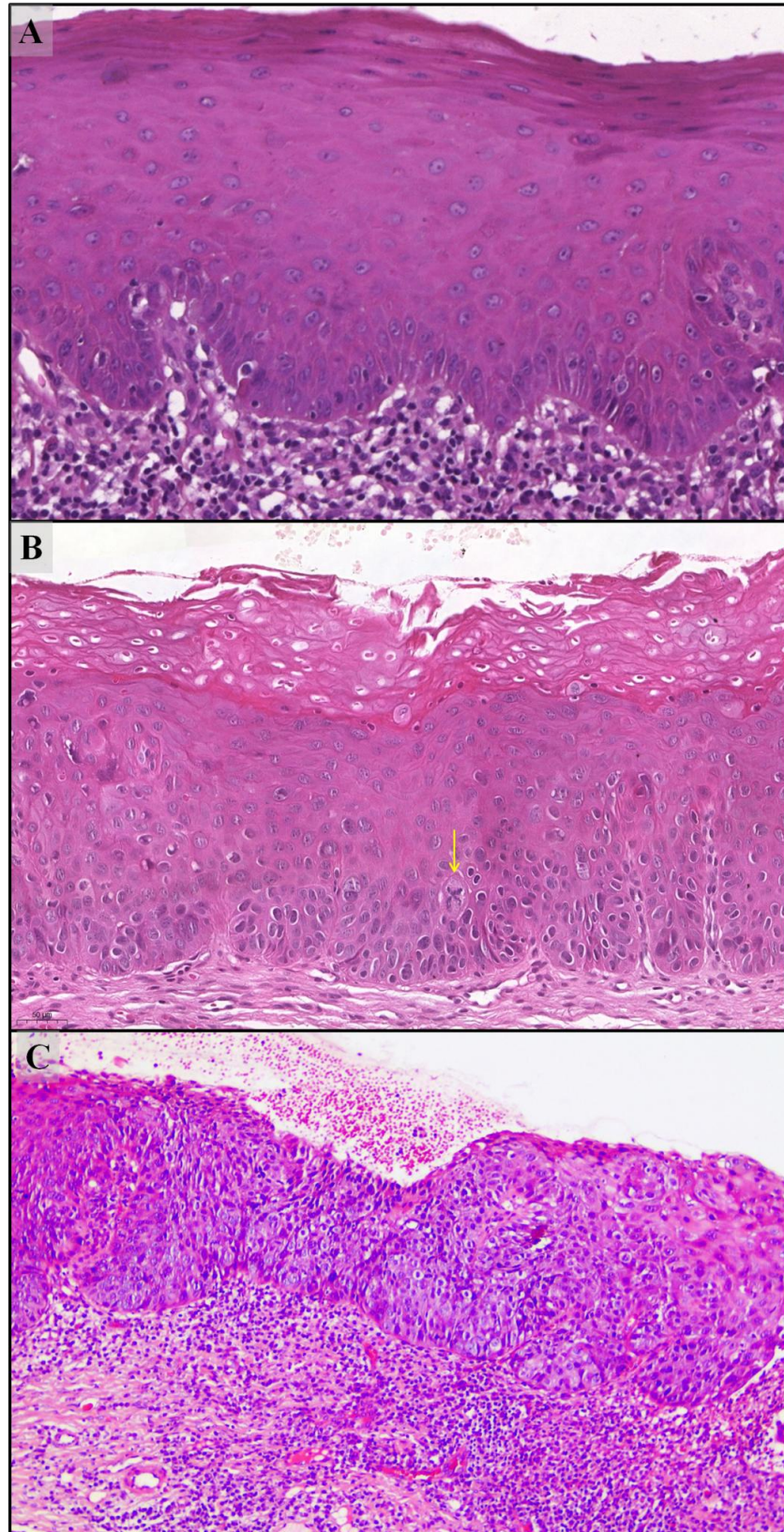


Figura 2. Características histopatológicas da leucoplasia oral. **(A)** Leucoplasia com displasia epitelial oral leve. Na imagem a perda de polaridade e pontos focais de perda de estratificação nas camadas basal e parabasal do epitélio, além de pleomorfismo celular e nuclear discreto e células com núcleos hiper cromáticos e nucléolos evidentes podem ser observados. Na lâmina própria, observa-se infiltrado inflamatório subepitelial predominantemente mononuclear. As alterações displásicas citológicas e arquiteturais estão confinadas ao terço inferior do epitélio. **(B)** Leucoplasia com displasia epitelial oral

moderada. O quadro histopatológico observado revela epitélio hiperparaqueratinizado com perda de estratificação e de polaridade das células das camadas basal e espinhosa. As células apresentam-se pleomórficas, com núcleos hiper cromáticos e nucléolos evidentes. Figuras de mitoses típicas e atípicas (seta amarela) também podem ser observadas. As alterações displásicas citológicas e arquiteturais estendem-se predominantemente à metade da espessura epitelial. (C) Leucoplasia com displasia epitelial oral severa. Nota-se a presença de alterações displásicas como perda da estratificação e polaridade, pleomorfismo celular e nuclear além de disceratose, exocitose e ocasionais figuras de mitose atingindo, em alguns pontos, toda a espessura epitelial. Na lâmina própria, observa-se intenso infiltrado inflamatório de células predominantemente mononucleares.

Screenshots obtidos a partir do aplicativo Case Viewer 2.3.

Fonte: (A) e (B) elaborado pela autora. (C) Cortesia do Prof. Dr. Felipe Paiva Fonseca.

Assim como para as demais DOPM, a gradação do nível de displasia epitelial oral tem sido descrita como um parâmetro importante para predição do risco de progressão da leucoplasia oral para o câncer (EVREN et al., 2020; JAYASOORIYA et al., 2020; RANGANATHAN; KAVITHA, 2019; SPEIGHT; KHURRAM; KUJAN, 2018; WARNAKULASURIYA et al., 2008; WILS et al., 2020). As leucoplasias displásicas sofrem transformação maligna em cerca de 7–34% dos casos, enquanto as que não exibem displasia no seu quadro histopatológico no momento da biópsia apresentam uma taxa de progressão para o câncer entre 0.1–14% (LI et al., 2020). Embora inferior, a porcentagem de lesões que não apresentam displasia epitelial oral e progridem para carcinoma é considerável (LI et al., 2020; SILVERMAN; GORSKY; LOZADA, 1984). Diante disso, diversas técnicas moleculares vêm sendo aplicadas no intuito de elucidar os mecanismos subjacentes ao desenvolvimento da leucoplasia e, a partir disso, definir biomarcadores preditores do seu potencial de malignização (CAMISASCA et al., 2017; DE MORAIS et al., 2020; RAI et al., 2018; WEBER et al., 2020).

2.3 Abordagens moleculares para avaliação do potencial de progressão maligna da leucoplasia oral

Os fatores predisponentes ao câncer de boca são bem reconhecidos, entretanto, o processo de carcinogênese oral ainda não foi completamente elucidado. Nessa sessão, apresentaremos os principais achados obtidos no estudo da leucoplasia oral a partir de abordagens moleculares distintas.

2.3.1 Estudos imunohistoquímicos

A imunohistoquímica é uma técnica de diagnóstico auxiliar extensivamente utilizada para a detecção de biomarcadores preditivos e prognósticos de inúmeras neoplasias malignas (MAGAKI et al., 2019). A expressão aumentada ou diminuída de proteínas e glicoproteínas de membrana associadas à transição epitélio-mesênquima (DE MORAIS et al., 2020; ZHANG et al., 2017a), bem como de espécies oncogênicas (BARAN et al., 2019; LIU et al., 2012a, 2012b; VICENTE et al., 2019; YOSHIDA et al., 2019), angiogênicas (NAYAK et al., 2015), e relacionadas à diferenciação, proliferação e morte celular (FERNÁNDEZ-VALLE et al., 2016; LAMEIRA et al., 2014; NEGI et al., 2015), foi associada ao maior risco de transformação maligna da leucoplasia oral (SARODE et al., 2019).

O desequilíbrio do sistema imunológico é um dos fatores correlacionados ao desenvolvimento de diversos tumores malignos, incluindo o câncer oral (GOMES et al., 2016). O processo de carcinogênese é lento e pode ocorrer ao longo de diversos anos, semelhante a uma infecção crônica. Nesse contexto, a ausência de uma fase aguda modifica a resposta imunológica celular do hospedeiro, criando um microambiente imunossupressivo que suporta o crescimento tumoral (ZOU, 2005). Yagyuu e colaboradores (2017) sugeriram que as DOPM surgem durante a fase de equilíbrio do processo de imunoedição tumoral, na qual o sistema imunológico não consegue eliminar a lesão, mas previne a invasão do estroma pelas células epiteliais displásicas (YAGYUU et al., 2017).

As modificações na resposta imunológica que marcam o estágio inicial da carcinogênese oral vão desde a desregulação da expressão de citocinas até a mudança na densidade e função das células do sistema imune (BABIUCH et al., 2020; GRIGOLATO et al., 2020). As células dendríticas são células apresentadoras de antígeno essenciais para a ativação do sistema autoimune na resposta contra as células tumorais. Uma densidade significativamente reduzida de células dendríticas foi observada em lesões orais displásicas e em leucoplasias orais de pacientes tabagistas (CÉSAR et al., 2020; PELLICIOLI et al., 2017; RIBEIRO et al., 2018).

As leucoplasias que sofreram transformação maligna exibiram uma infiltração macrófagica significativamente maior e uma ativação acentuada de macrófagos tumorais (M2) em comparação com aquelas que mantiveram o seu comportamento benigno (WEBER et al., 2020). Ainda, uma correlação positiva entre a presença de macrófagos do tipo M2 e de um infiltrado linfocitário subepitelial marcado por um predomínio de células Th2 foi observada na leucoplasia oral. De acordo com os autores, o microambiente dominado por linfócitos Th2 polariza os macrófagos para o fenótipo M2, promovendo a supressão da resposta antitumoral (STASIKOWSKA-KANICKA; WĄGROWSKA-DANILEWICZ; DANILEWICZ, 2018b). O predomínio dessas células na região de interface epitélio/mesênquima da leucoplasia sugere uma ação nos estágios iniciais da carcinogênese, e a detecção imunohistoquímica deste infiltrado específico poderia prever o risco de transformação maligna da leucoplasia oral (SHIGEOKA et al., 2019; STASIKOWSKA-KANICKA; WĄGROWSKA-DANILEWICZ; DANILEWICZ, 2018b; WEBER et al., 2020).

Os estudos voltados à compreensão dos mecanismos imunomodulatórios da leucoplasia oral ainda estão em um estágio incipiente, contudo, uma associação consistente entre a imunomodulação por meio da via PD-L1/PD-1 e a transformação maligna da leucoplasia e de outras DOPM tem sido observada (CHEN et al., 2019; DAVE; ALI; MAGALHAES, 2020; STASIKOWSKA-KANICKA; WĄGROWSKA-DANILEWICZ; DANILEWICZ, 2018a, 2018b; YAGYUU et al., 2017). O complexo PD-L1/PD-1 é um ponto de verificação imune inibitório, constituído pela proteína transmembrana PD-1 (*programmed cell death protein – 1*), ligada à membrana citoplasmática das células T, e sua ligante PD-L1 (*programmed cell death ligand – 1*), presente na membrana de algumas células do sistema imunológico, queratinócitos, células endoteliais, pulmonares, placentárias e tumorais (LENOUVEL et al., 2019). Esse complexo regula a ativação da fase efetora das células T.

Em diversas neoplasias malignas, a ligação PD-L1/PD-1 cria mecanismos de escape para as células tumorais e promove a apoptose das células T efetoras (LENOUVEL et al., 2019). Nas DOPM, as células epiteliais displásicas positivas para PD-L1 podem escapar do sistema imunológico do hospedeiro e progredirem

para o câncer (YAGYUU et al., 2017). O uso de anticorpos imunomodulatórios que bloqueiam a proteína PD-1 ou PD-L1 têm se mostrado eficaz no tratamento de inúmeros processos neoplásicos malignos (WU et al., 2019). Dessa forma, o bloqueio dessa via pode ser considerado como um alvo molecular promissor para a prevenção da transformação maligna da leucoplasia oral.

Os achados imunohistoquímicos proporcionaram *insights* importantes tanto para a determinação de marcadores histopatológicos preditores do potencial de malignização da leucoplasia oral (ZHANG et al., 2017b), como para a detecção de alvos moleculares terapêuticos (GRIGOLATO et al., 2020). No entanto, esses alvos não são utilizados no âmbito clínico devido à heterogeneidade dos resultados publicados até a presente data.

2.3.2 Estudos ômicos: genômica, transcriptômica, proteômica e metabolômica

A patogênese molecular do câncer oral é complexa, resultante de diversos mecanismos biológicos interdependentes que envolvem desde alterações genômicas a mudanças na expressão gênica, na síntese e interações proteicas e nos processos metabólicos (VITÓRIO et al., 2020). As ciências ômicas, incluindo a genômica, transcriptômica, proteômica, e a metabolômica, são ferramentas que permitem o estudo de um grande conjunto de moléculas distintas e, por isso, são úteis para a compreensão de processos fisiológicos e patológicos complexos (RAI et al., 2018). A aplicação dessas abordagens moleculares trouxe avanços notáveis para o entendimento do processo de carcinogênese oral.

É importante salientar que a metagenômica, estudo do material genético de microorganismos recuperados de um nicho biológico, representa um novo e promissor horizonte para a pesquisa em câncer (BANERJEE; MISHRA; DHAS, 2015). O microbioma oral é constituído por diversas espécies bacterianas envolvidas com a manutenção de um estado inflamatório crônico e com a eclosão de alterações genéticas nas células epiteliais que desregulam a proliferação e sobrevivência celular e, portanto, podem induzir a progressão da carcinogênese oral (CHATTOPADHYAY; VERMA; PANDA, 2019). Apesar do potencial para o desenvolvimento de biomarcadores preditores e prognósticos de transformação

maligna, os estudos metagenômicos não foram incluídos nesta revisão de literatura, pois fogem do escopo desta pesquisa.

2.3.2.1 Estudos genômicos

A transformação de uma célula normal para uma célula maligna é marcada por diversas alterações gênicas como danos em genes associados ao reparo do DNA. A mensuração da capacidade de reparo do DNA foi aproximadamente 92% sensível e 84% específica para a predição do risco de progressão maligna da leucoplasia oral (THOMAS et al., 2020). Farah e colaboradores (2019) mostraram uma maior desregulação nas vias de reparo do DNA nas amostras de leucoplasias que se transformaram em CCE. Os genes associados com a via de quebra da fita dupla, principalmente *BRCA1*, *BRCA2*, *FANCA* e *DCLRE1B*, foram os que mais contribuíram para a discriminação entre as amostras de leucoplasias transformadas e não transformadas (FARAH et al., 2019).

As alterações cromossômicas também foram correlacionadas com a evolução maligna da leucoplasia oral. Numerosos estudos reportaram que a aneuploidia e graus moderado e severo de displasia epitelial oral são fatores correlacionados que podem prever o risco de transformação maligna da leucoplasia e de outras DOPM (KHANNA et al., 2010; SPERANDIO et al., 2013; VAN ZYL et al., 2012; VIJAYAVEL; ASWATH, 2013). A perda de heterozigossidade envolvendo os cromossomos 3p, 4q, 9p e 17p foi associada à progressão da displasia epitelial oral para o câncer (GUIMARÃES et al., 2020). Lesões com graus distintos de displasia epitelial oral mostraram um perfil semelhante de perda de heterozigossidade (GOMES et al., 2015). Ainda, algumas alterações displásicas foram associadas a perdas específicas. A perda do segmento 22 no braço curto do cromossomo 9 (9p22) foi relacionada a formação de projeções epiteliais em gota, enquanto que as perdas em 3p14.2 e 17p13.1 foram associadas a estratificação epitelial irregular e a disqueratose, respectivamente (FONSECA-SILVA et al., 2016). Esses achados corroboram a ideia de que a leucoplasia oral é uma doença molecularmente heterogênea, na qual cada característica displásica

abriga alterações moleculares distintas e, por isso, podem representar diferentes significados biológicos (FONSECA-SILVA et al., 2016; GOMES et al., 2015).

As neoplasias malignas usualmente apresentam instabilidade cromossomial (WANG et al., 2020) e, da mesma forma, essa alteração parece exercer influência no processo de carcinogênese oral (BHOSALE et al., 2017; SIEBERS et al., 2013). Cervigne et al. (2014), analisaram 32 amostras de leucoplasias transformadas, seus carcinomas correspondentes, e 21 leucoplasias não transformadas. Os genes *KHDRBS1*, *PARP1*, *RAB1A*, *HBEGF*, *PAIP2* e *BTBD7* exibiram perdas e ganhos recorrentes no número de cópias. Interessantemente, essas alterações foram detectadas apenas nas leucoplasias progressivas e em seus carcinomas correspondentes (CERVIGNE et al., 2014). As alterações no número de cópias foram mais evidentes em lesões displásicas, transformadas, não homogêneas e localizadas em sítios anatômicos propensos à malignização (GUIMARÃES et al., 2020). Os genes *EGFR*, *CDKN2A*, *CCDN1*, *RASSF1*, *FHIT* e *RBI*, sabidamente relacionados com funções oncogênicas ou supressoras de tumor, foram os mais frequentemente afetados por essas mudanças, o que sugere a associação entre as aberrações cromossômicas e a transformação maligna de lesões orais displásicas (BAGAN et al., 2012; GUIMARÃES et al., 2020; POH et al., 2012; TSUI et al., 2009). No entanto, é importante ressaltar que alguns desses eventos também foram detectados em casos de leucoplasias não displásicas (WERKMEISTER; BRANDT; JOOS, 1999). Por isso, apesar dos resultados promissores, outros estudos longitudinais são necessários para elucidar o papel da instabilidade cromossômica na progressão maligna da leucoplasia oral (GUIMARÃES et al., 2020).

A elucidação dos eventos iniciais subjacentes ao desenvolvimento das DOPM é igualmente importante à identificação dos fatores que ativam o gatilho para a transformação maligna. Villa e colegas (2019) revelaram que lesões leucoplásicas com e sem displasia epitelial oral abrigam alterações somáticas semelhantes. Mutações *missense* e *frameshift* envolvendo o gene *KMT2C* foram identificadas em 75% dos casos analisados. Os genes *TP53* e

TIAMI se mostraram alterados em 35% e 30% das amostras, respectivamente. O compartilhamento dessas assinaturas moleculares sugere que lesões de leucoplasia sem displasia, representam uma alteração inicial do processo de carcinogênese que possivelmente sofre influência de outras mudanças epigenéticas e/ou alterações locais do microambiente oral que potencializam a progressão maligna (VILLA et al., 2019). Dessa forma, a integração dos achados resultantes de estudos genômicos com aqueles obtidos por meio das outras abordagens ômicas é fundamental para a compreensão da variedade de fenômenos biológicos subjacentes ao desenvolvimento do câncer oral.

2.3.2.2 Estudos transcriptômicos e proteômicos

A transcriptômica e a proteômica são abordagens que permitem o estudo de alterações celulares *downstream*, as quais concernem em mudanças no conteúdo de RNA mensageiro (mRNA) e no perfil de expressão de proteínas, respectivamente. O mRNA é o molde para a síntese de proteínas que ocorre por meio da tradução e, portanto, alterações no transcriptoma celular podem levar à modificações na expressão de proteínas codificadas no genoma. Dessa forma, a correlação entre os achados genômicos, transcriptômicos e proteômicos contribui para a classificação das doenças e para o entendimento de sua progressão (RAI et al., 2018).

Em uma análise integrativa, Bhosale e colaboradores (2017) mostraram que os perfis genômicos e transcriptômicos das amostras de leucoplasia são semelhantes aos observados em amostras de carcinoma de células escamosas. Dezesesseis genes foram super-regulados e 29 foram regulados negativamente de modo consistente entre as leucoplasias e os carcinomas de células escamosas em estágio inicial e tardio de evolução. Essa consistência revela a influência da desregulação gênica na transição de um estado precursor ao de invasividade no processo de carcinogênese oral. A integração dos achados genômicos e transcriptômicos das lesões leucoplásicas e de carcinoma evidenciou uma forte correlação entre a alta expressão dos genes *ECT2*, *EIF5A2*, *KLHL6*, *GPR160* e de *HOXC9*,

HOXC13, *ERBB3*, *MUCL1* com a amplificação em 3q26.31 e em 12q13.2, respectivamente. As similaridades observadas entre as leucoplasias e os carcinomas de células escamosas sugeriram que essas alterações são eventos iniciais essenciais para a transformação e progressão maligna celular. Algumas sobreposições também foram identificadas entre o perfil de expressão gênica das leucoplasias e das amostras de tecido normal, o que vai ao encontro com o *status* transitório da doença (BHOSALE et al., 2017).

Apesar da similaridade no perfil de transcritos expressos, Marakev e colegas (2017) observaram que o cenário de ativação das vias de sinalização celular pelas leucoplasias foi profundamente diferente daquele visto nos carcinomas de células escamosas. O grupo realizou uma análise de bioinformática *in silico* de dados de transcriptômica e mostraram que a ativação das vias relacionadas à sobrevivência e proliferação celular como AKT/mTOR, ERK, JNK, RAS/MAPK, p38, PAK1, integrina/ILK e TGF β , foi substancialmente menor na maioria das amostras de leucoplasia oral em relação ao observado nos carcinomas. Interessantemente, essas amostras revelaram uma ligeira super-regulação de vias pró-apoptóticas, e então, os autores postularam que esse achado associado à baixa regulação das vias mitogênicas e de sobrevivência pode direcionar a regressão espontânea da leucoplasia oral e/ou sua não progressão para o câncer, uma vez que a transformação maligna é um evento raro. Apenas 20 de 86 amostras de leucoplasia analisadas mostraram similaridade ao perfil de ativação de vias de sinalização observado nos casos de carcinoma, exibindo uma notável super-regulação em MAPK, ERK, JNK, IL6-STAT3, WNT, TGF β e na sinalização do receptor de glicocorticoide. Embora apenas sete dessas 20 amostras tenham sofrido evolução maligna, foi sugerido que a desregulação dessas vias pode ser a força condutora inicial da tumorigênese oral (MAKAREV et al., 2017).

A desregulação de proteínas associadas à apoptose foi ligada com o silenciamento da expressão de transglutaminase 3 (TGM3) nas leucoplasias (WU et al., 2018). A TGM3 é uma proteína amplamente expressa na camada suprabasal da mucosa e é essencial para a diferenciação terminal

dos queratinócitos. A supressão dessa proteína tem sido correlacionada com o desenvolvimento de diversos tipos de cânceres humanos. A expressão proteica e de mRNA de TGM3 foi baixamente regulada em linhagens celulares e amostras de leucoplasia, em comparação com o tecido normal. Os baixos níveis de TGM3 foram fortemente associados com o grau de displasia epitelial oral e com a transformação maligna, indicando um papel importante da proteína no processo de carcinogênese oral (WU et al., 2018). Outras modificações a nível transcricional também foram associadas com a transformação maligna da leucoplasia como a expressão elevada de proteases responsáveis pela degradação da membrana basal e invasão do tecido conjuntivo subjacente, como a MMP1 e 9 (JORDAN et al., 2004) e a catepsina L (MACABEO-ONG et al., 2003), e a superexpressão de moléculas indutoras de tumor, como a EMMPRIN (*extracellular matrix metalloproteinase inducer*) (VIGNESWARAN et al., 2006).

O estudo dos micro-RNAs (mi-RNAs), sequencias curtas de RNA compostas por cerca de 19 a 25 nucleotídeos, revelou a influência das mudanças pós-transcricionais sobre a transformação da leucoplasia oral em carcinoma de células escamosas. Chang *et al.*, (2018) analisaram amostras de plasma de três grupos de indivíduos: saudáveis, com leucoplasia oral e com carcinoma de células escamosas, e reportaram que um painel composto por três miRNAs (miR-222-3p, miR-150-5p, e miR-423-5p) foi capaz de discriminar a leucoplasia oral do carcinoma com alta acurácia (AUC = 0,88). Essas moléculas estão envolvidas com a regulação das vias de sinalização Wnt, PI3K-Akt, MAPK, e RAS, sabidamente associadas à progressão do câncer (CHANG et al., 2018). Analisando amostras de tecido de leucoplasias orais transformadas e não transformadas, Maimaiti e colaboradores (2015) associaram três assinaturas de miRNAs (miR-129-5p, miR-339-5p and miR-31*) com a indução dos eventos iniciais e de progressão da leucoplasia para o câncer por meio da regulação de fatores de transcrição envolvidos com o processamento de antígenos e com o controle da resposta imunológica mediada por células T. Os autores sugeriram que a transformação maligna da leucoplasia ocorre em consequência de uma resposta imune desregulada (MAIMAITI et al., 2015).

A análise do perfil de expressão dos RNAs circulares (circ-RNAs), moléculas de RNA não codificantes possivelmente associadas com a regulação dos mi-RNAs, considerou o circHLA-C como um alvo promissor para o diagnóstico (AUC = 0.95) e prognóstico da leucoplasia oral. Curiosamente, essa molécula foi correlacionada com a regulação da citotoxicidade mediada por células T e com a expressão aumentada de MHC-1 (*Major Histocompatibility Complex Class I*) (XU et al., 2020). O envolvimento do sistema imunológico com a progressão maligna da leucoplasia foi ainda proposto por Wang, Zhang e Zhou (2020), que mostraram a super-regulação do eixo CXCL9/10/11-CXCR3A em amostras de leucoplasia e carcinoma de células escamosas. Esse eixo regula a migração, diferenciação e ativação de células do sistema imunológico, e sua super-regulação poderia resultar na desregulação de pontos de verificação imunológicos nas lesões leucoplásicas e de câncer (WANG; ZHANG; ZHOU, 2020). Em uma análise proteômica, as proteínas CXCL11 e CCL11 foram as únicas espécies diferencialmente expressas entre amostras de saliva de pacientes com leucoplasia e indivíduos saudáveis, sendo super-reguladas nas leucoplasias. O estudo da rede de interação proteína-proteína revelou a associação dessas citocinas com a quimiotaxia de linfócitos, resposta imune humoral antimicrobiana e a sinalização de quimiocinas (SCHOLTZ et al., 2020).

Wang e colaboradores (2009) realizaram uma análise pareada do perfil proteômico de lesões de leucoplasias orais displásicas e carcinomas de células escamosas. Ao todo, 85 proteínas foram diferencialmente abundantes entre os grupos, representadas principalmente por espécies relacionadas à apoptose, resposta à estímulos e regulação metabólica. Duas proteínas homólogas do complexo ativador do proteassoma 28 (PA28ab) apresentaram-se 4 a 6 vezes mais expressas nos carcinomas. Essas subunidades são particularmente abundantes em tecidos imunes e contribuem para a apresentação de antígenos de classe I. As peroxiredoxinas (3 e 4), anexinas (A1, A3, A4, A5 e A8), proteínas Rho inibidoras de dissociação de GDP(Rho-GDI 1 e 2), proteínas HSP(70, 71 e

β 1) e proteínas da família S100 (A7-A11 e A16) também foram superexpressas nos carcinomas e, portanto, associadas à transição maligna (WANG et al., 2009). Interessantemente, as proteínas HSP70, anexina 2 e peroxiredoxina 3 e 4 foram correlacionadas à indução de uma resposta imunológica por meio da produção de autoanticorpos em uma análise do soro sanguíneo de pacientes com leucoplasia oral (SHUKLA et al., 2007). Recentemente, foi visto que a nicotina pode promover o desenvolvimento da leucoplasia oral através da desregulação de proteínas que interagem com a peroxiredoxina 1. Os autores postularam que a proteína e suas ligantes poderiam ser consideradas como alvos para o tratamento da leucoplasia e para a quimioprevenção do carcinoma de células escamosas (QI et al., 2019).

Os estudos proteômicos são promissores para o desenvolvimento de métodos não invasivos preditores de malignização, uma vez que a saliva é o material biológico mais comumente utilizado nas análises (RAI et al., 2018). Camisasca *et al.*, (2016) encontraram 22 proteínas altamente abundantes na saliva de pacientes com leucoplasia oral as quais foram pouco presentes ou ausentes no grupo de indivíduos saudáveis. Entre as proteínas encontradas, os autores sugeriram a citoqueratina 10 como alvo mais promissor para o diagnóstico de lesões leucoplásicas (CAMISASCA et al., 2017). Sivadasan *et al.*, (2019) mostraram níveis significativamente aumentados de S100A7, S100P, CD44 e COL5A1 na saliva de pacientes com leucoplasia oral e com carcinoma de células escamosas e, portanto, as propuseram como candidatas a biomarcadoras para detecção precoce de malignidade oral. Ainda, algumas proteínas foram exclusivamente expressas nos casos de leucoplasia oral. No entanto, os autores não realizaram uma análise detalhada dessas espécies (SIVADASAN et al., 2020). Até o momento da finalização desta revisão de literatura, não existe um painel de biomarcadores definido para o diagnóstico precoce e não invasivo do câncer oral.

2.3.2.3 Estudos metabolômicos

O metabólito é uma molécula de baixo peso molecular (<2000Da), que é produto intermediário ou final do metabolismo. Essas moléculas constituem o metaboloma humano, e podem representar a saída *downstream* das mudanças genéticas, das alterações transcricionais e das modificações pós-traducionais das proteínas, ou a entrada *upstream* do ambiente, resultado de fatores exógenos como os hábitos de estilo de vida. Dessa forma, a metabolômica, estudo dos metabólitos e do metabolismo, permite explorar o eixo de interações entre genótipo e fenótipo (CHEN et al., 2012; WILD, 2005; WISHART, 2016). Os ensaios metabolômicos podem ainda ser aplicados para outros fins como na análise química alimentar (TRIMIGNO et al., 2020), no monitoramento ambiental (ZHEN et al., 2018), e nas ciências forenses (LOCCI et al., 2020).

O fluxo de trabalho das análises metabolômicas é constituído por duas etapas. A primeira inclui o delineamento do desenho experimental, fase crucial para a definição da pergunta do estudo e, a partir disso, para a escolha do tipo e número de amostras por grupo amostral, do modo de coleta e de armazenamento, do método a ser utilizado para extração dos metabólitos, e da plataforma analítica a ser empregada para a aquisição dos dados. A segunda engloba o processamento dos dados obtidos, seguido pelas análises estatísticas, pela identificação dos metabólitos (no caso das análises *untargeted*), e pela análise de enriquecimento das vias de sinalização celular (BARNES et al., 2016a, 2016b; GOODACRE et al., 2007; SUMNER et al., 2007). Os estudos metabolômicos podem seguir uma abordagem *targeted*, na qual compostos específicos, previamente conhecidos pelo pesquisador, são avaliados, ou *untargeted*, também chamada de análise global ou de perfil metabolômico. Essa abordagem de descoberta fornece uma visão compreensiva do espectro de modificações metabólicas subjacentes à condição ou doença analisada, permitindo ao investigador postular novas hipóteses que não poderiam ser observadas em uma análise direcionada (BARNES et al., 2016a).

Diversos tipos de materiais biológicos podem ser utilizados para o estudo do metabolismo humano incluindo gases, biofluidos e tecidos (WISHART,

2016). A escolha da plataforma analítica para a análise do metaboloma é diretamente associada com o tipo de amostra usada no estudo. Hoje, a ressonância magnética nuclear e a espectrometria de massas acoplada às técnicas de separação, como a eletroforese capilar e as cromatografias gasosa e líquida são as principais técnicas de análise empregadas para a determinação de metabólitos em uma amostra biológica.

A eletroforese capilar acoplada à espectrometria de massas (CE-MS) é aplicada para a separação de compostos hidrofílicos utilizando pequenos volumes de amostra (2-10nL) (BARNES et al., 2016a). Plasma, suor, fluido cerebrospinal, urina, e diversos outros biofluidos podem ser analisados com sucesso por CE-MS (RAMAUTAR; SOMSEN; DE JONG, 2019).

A cromatografia gasosa acoplada à espectrometria de massas (GC-MS) é indicada para a detecção de compostos voláteis e lábeis. Uma limitação dessa técnica é a necessidade de derivatização dos compostos, método que utiliza reações de silição e/ou oximação para que as moléculas possam se volatilizar nas temperaturas de trabalho do GC-MS. A derivatização pode levar ao aumento da massa molecular do composto derivado, dificultando sua passagem pela coluna de cromatografia gasosa ou fazendo com que ele fique fora da faixa de detecção do espectrômetro de massas utilizado (BARNES et al., 2016a; FIEHN, 2016; HALKET et al., 2005; PHELAN, 2020).

A introdução de fontes de ionização como a ESI (*electrospray ionization*) e a APCI (*atmospheric-pressure chemical ionization*) foram a chave para o sucesso da análise metabolômica por cromatografia líquida acoplada à espectrometria de massas (LC-MS), pois permite a transformação de íons carregados da fase líquida para gasosa sem a necessidade de derivatização (BARNES et al., 2016a; PHELAN, 2020). A técnica de LC-MS possibilita a análise de compostos polares, apolares, hidrofílicos ou hidrofóbicos. Dada a complexa diversidade de metabólitos presentes nas amostras biológicas, a captura dessa ampla gama de moléculas é alcançada com o uso de diferentes tipos de colunas de separação cromatográfica (BARNES et al., 2016a). As

colunas de fase reversa (C₄–C₁₈) proporcionam a separação de compostos apolares e moderadamente polares, enquanto as colunas de fase mista (HILIC – *hydrophilic interaction liquid chromatography*) são empregadas para a separação de moléculas com alto grau de hidrofiliçidade (BARNES et al., 2016a; KOVALCHUK; JENSEN; ROGOWSKA-WRZESINSKA, 2019).

Após a separação, os metabólitos são detectados por espectrometria de massas. O espectrômetro de massas é constituído por uma fonte de ionização, um analisador de massas e um ou mais detectores. Os compostos ionizados são separados pelo analisador de massas, detectados e exibidos no sistema computacional como espectro de massas. Em seguida, técnicas de bioinformáticas são empregadas para o pré-processamento dos dados brutos. A saída do experimento reúne informações sobre o tempo de retenção do composto na coluna cromatográfica, sua razão massa/carga (m/z) e a intensidade ou área do pico em cada amostra analisada (GROSS, 2011).

A espectroscopia por ressonância magnética nuclear (NMR) tem se popularizado como plataforma analítica para estudos de metabolômica devido a algumas vantagens apresentadas em relação à análise por GC-MS ou LC-MS. O método requer pouco ou nenhum tratamento prévio das amostras, não necessita de derivatização, e é automatizado, o que aumenta a reprodutibilidade analítica do estudo. Por ser um procedimento não destrutivo, a NMR pode ser utilizada para a análise metabolômica de tecidos intactos, como órgãos inteiros e até mesmo de amostras vivas. No entanto, a NMR apresenta duas desvantagens importantes, a baixa sensibilidade e a grande quantidade de sinais gerados. Por isso, pode não retornar informações sobre compostos que estejam presentes em baixas concentrações e o processamento dos dados é dificultado, respectivamente. (EMWAS et al., 2019).

Até a presente data, quatro estudos publicados utilizaram abordagens baseadas em NMR para a análise do perfil metabolômico da leucoplasia oral (GUPTA; GUPTA; MAHDI, 2015; KONG et al., 2015; SONG et al., 2020;

ZHOU et al., 2009). Quatro basearam-se em GC-MS (CHEN et al., 2021; MOHAMED ANSER; ASWATH, 2014; MUSHARRAF et al., 2016; XIE et al., 2012), e cinco em LC-MS (BAG et al., 2016; GUTTENPLAN et al., 2017; SRIDHARAN et al., 2019; SRIDHARAN; RAMANI; PATANKAR, 2017; WEI et al., 2011). As abordagens metabolômicas foram aplicadas para análise de amostras de saliva (SONG et al., 2020; SRIDHARAN et al., 2019; WEI et al., 2011), soro (GUPTA; GUPTA; MAHDI, 2015; SRIDHARAN; RAMANI; PATANKAR, 2017), plasma (KONG et al., 2015; ZHOU et al., 2009), urina (MOHAMED ANSER; ASWATH, 2014; XIE et al., 2012) e linhagens celulares (CHEN et al., 2021).

O perfil metabolômico dos biofluidos pode ser alterado por uma variedade de processos fisiológicos em decorrência de um estímulo fisiopatológico. Por isso, a perturbação global desse perfil pode refletir a presença e a progressão das doenças (DUNN; BAILEY; JOHNSON, 2005). Zhou *et al.* (2009) analisaram, pela primeira vez, amostras de plasma de indivíduos saudáveis, com leucoplasia oral e com carcinoma de células escamosas empregando ressonância magnética nuclear de prótons (^1H NMR). Os autores relataram que a abordagem foi eficaz para a diferenciação entre os grupos. A análise por PLS-DA (*Partial Least-Squares Discriminant Analysis*) revelou que os compostos que mais contribuíram para a distinção entre os grupos foram: mio-inositol, taurina, creatinina, glicose, e compostos contendo colina. Essas moléculas podem participar da osmoregulação tumoral, da síntese e degradação da membrana celular, do controle de processos apoptóticos e do metabolismo glicolítico (ZHOU et al., 2009).

Amostras de plasma de ratos foram utilizadas por Kong e colaboradores (2015) para a identificação de mudanças no metabolismo associadas ao tempo durante os vários estágios da carcinogênese oral. O composto tumorigênico 1-óxido de 4-nitroquinolina (4NQO) foi administrado para a indução das lesões orais entre 16 e 24 semanas. O tratamento induziu variados graus de alterações displásicas em todos os modelos animais tratados. As lesões diagnosticadas como leucoplasia se desenvolveram entre as semanas 14 e 16 e os carcinomas de células escamosas foram

diagnosticados entre as semanas 22 e 24. Oito metabólitos discriminantes entre os grupos foram detectados por ^1H NMR: ácido láctico, colina, glicose, prolina, valina, isoleucina, ácido aspártico e ácido 2-hidroxi-butírico. A abundância de ácido láctico, colina e glicose foi aumentada progressivamente, enquanto os níveis de prolina, valina, isoleucina, e ácido aspártico foram reduzidos gradualmente de acordo com a evolução maligna. Essa progressão de mudanças metabólicas tempo-dependentes foi condizente com o processo *multistep* da carcinogênese oral (KONG et al., 2015).

As mudanças do perfil metabolômico associadas com os diferentes estágios do processo carcinogênico oral também foram observadas por Song e colegas (2020). O grupo analisou amostras de saliva de 373 voluntários, sendo 124 indivíduos saudáveis, 124 portando DOPM, e 125 com CCE por CPSI-MS (*Conductive Polymer Spray Ionization Mass Spectrometry*). Dentre os compostos detectados, 58 foram significativamente alterados considerando a progressão dos controles saudáveis para os indivíduos com leucoplasia. Tendo em vista o avanço da leucoplasia para o carcinoma de células escamosas, 116 metabólitos foram detectados significativamente elevados ou reduzidos. As vias metabólicas de biossíntese de aminoacil tRNA, metabolismo de arginina/prolina, biossíntese de arginina, degradação de lisina e metabolismo de histidina foram desreguladas em ambos os estágios de progressão tumoral. Os autores validaram as mudanças metabólicas detectadas usando amostras de tecido de carcinoma de células escamosas por DESI-MS (*Desorption Electrospray Ionization MS Imaging*), nas quais os mesmos metabólitos e vias de sinalização metabólica se mostraram desregulados. O estudo foi ainda validado com a aplicação de um algoritmo de aprendizado de máquina que atingiu acurácia de 95% e 87% nas fases de treino e de teste, respectivamente, para a distinção entre os grupos com base no perfil de alterações metabólicas encontrado (SONG et al., 2020).

O estabelecimento de um painel de biomarcadores para o diagnóstico precoce do câncer oral é uma das principais motivações por trás dos estudos

metabolômicos. A análise de amostras de soro sanguíneo por ^1H NMR mostrou que um painel composto por glutamina, acetona, acetato e colina foi capaz de discriminar a leucoplasia oral do carcinoma de células escamosas com alta acurácia (ROC = 0,96) (GUPTA; GUPTA; MAHDI, 2015). Uma abundância significativamente aumentada de estradiol-17-beta-3-sulfato, L - carnitina, 5-metiltioadenosina, 8 - hidroxadenina, ácido 2-metilcítrico, putrescina e estrona-3-sulfato foi detectada por LC-MS em amostras de soro de pacientes com leucoplasia oral e carcinoma de células escamosas em comparação com o grupo de indivíduos saudáveis. Os metabólitos 5,6-di-hidrouridina, 4-hidroxiipenbutolol glucuronido, 8-hidroxadenina e putrescina foram significativamente mais abundantes nas amostras de carcinoma do que nas de leucoplasia oral. Assim, os autores sugeriram que a detecção desses compostos pode apresentar um potencial promissor para a predição da transformação maligna da doença (SRIDHARAN; RAMANI; PATANKAR, 2017).

Um painel formado por ácido 6-hidroxinicotínico, cisteína, e tirosina foi 85% sensível e 90% específico, apresentando valor preditivo positivo de 92% e ROC = 0,93 para a distinção entre leucoplasia oral e carcinoma de células escamosas em uma análise do perfil metabolômico de amostras de urina por GC-MS. Foi visto que o aumento da concentração do ácido 6-hidroxinicotínico indica uma menor probabilidade para o diagnóstico de carcinoma, enquanto o aumento da concentração de cisteína e tirosina aponta uma chance aumentada para a presença do câncer oral (XIE et al., 2012). A abundância do composto 4-(metilnitrosamino)-1-(3-piridil)-1-butanol (NNAL) foi maior em amostras de urina de pacientes usuários de tabaco sem fumaça do que os que fazem uso do tabaco tradicional. Por isso, a presença de NNAL foi sugerida como potencial biomarcador de risco aumentado para o desenvolvimento de câncer oral nesse grupo de indivíduos (MOHAMED ANSER; ASWATH, 2014).

A análise de amostras de saliva por LC-MS sugeriu outros metabólitos como potenciais preditores de transformação maligna da leucoplasia como um painel formado por valina, ácido láctico e fenilalanina (ROC = 0,97). O

nível de ácido láctico foi aproximadamente 3 vezes maior nos carcinomas, enquanto a valina e a fenilalanina foram cerca de 2 e 3 vezes mais abundantes nas leucoplasias (WEI et al., 2011). A super-regulação dos metabólitos 1-metilhistidina, inositol 1,3,4-trifosfato, d-glicerato-2-fosfato, 4-nitroquinolina-1-óxido, 2-oxoarginina, nitróxido de norcocaína, esfingalina-1-fosfato e pseudouridina e a regulação negativa de l-homocisteico, ubiquinona, ácido neuramínico e valerato de estradiol também foram sugeridos como biomarcadores sugestivos de malignidade. No entanto, estudos adicionais são necessários para a confirmação da utilidade clínica de cada composto para a prevenção da transformação maligna da leucoplasia e para o diagnóstico precoce do câncer oral (SRIDHARAN et al., 2019).

Chen *et al.* (2021) cultivaram linhagens celulares de células epiteliais orais humanas (HOEC) e queratinócitos displásicos orais (DOK) e detectaram, por GC-MS, 4 metabólitos cruciais para a distinção entre os dois subtipos celulares: piruvato, glutamina, metionina e lisina. A desregulação do metabolismo energético, o metabolismo de aminoácidos e da via glicolítica foi sugerida como eventos metabólicos importantes para a diferenciação entre os grupos (CHEN et al., 2021).

Embora a análise dos biofluidos e das linhagens celulares tenham proporcionado informações relevantes sobre a carcinogênese oral, alguns mecanismos fisiopatológicos relacionados à promoção e progressão tumoral podem originar compostos que são mais bem detectados diretamente no tecido afetado. Ainda, as comparações entre tecido tumoral e tecido normal adjacente, bem como à elucidação da heterogeneidade intratumoral só pode ser alcançada a partir do estudo de amostras de tecido. No entanto, as análises comparativas entre DOPM e carcinoma de células escamosas foram, até a presente data, limitadas à fibrose submucosa oral e o líquen plano (VITÓRIO et al., 2020). Ainda, todos os estudos de metabolômica publicados compararam amostras obtidas de indivíduos saudáveis com as de pacientes afetados por leucoplasia oral e carcinoma de células escamosas. Até o momento de finalização dessa revisão de literatura, nenhum estudo

avaliou o perfil metabolômico de amostras de leucoplasia que sofreram ou não transformação maligna.

3. OBJETIVOS

3.1 Objetivo geral

Avaliar o perfil metabolômico de amostras de tecido de leucoplasia oral transformada e não transformada.

3.2 Objetivos específicos

Analisar o metaboloma da leucoplasia oral por abordagem *untargeted* em cromatógrafo líquido acoplado a espectrômetro de massas.

Correlacionar os resultados obtidos por meio de análises estatísticas uni e multivariadas e definir compostos discriminantes.

Investigar os processos biológicos subjacentes à transformação maligna da leucoplasia oral mediante análise por enriquecimento de vias de sinalização celular.

4. MATERIAIS E MÉTODOS

4.1 Aspectos éticos

O presente estudo retrospectivo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais (CAAE: 40027620.0.0000.5149) (ANEXO A) e foi realizado de acordo com os princípios estabelecidos pela Declaração de Helsinki, e pela Resolução CNS 466/12. Foram selecionados casos de leucoplasia oral a partir dos arquivos do Serviço de Patologia Oral e Maxilo Facial da Faculdade de Odontologia da Universidade Federal de Minas Gerais (FAO-UFGM) e do Serviço Integrado de Patologia e Cirurgia Bucomaxilofacial do Hospital Universitário João de Barros Barreto (HUIBB) e da Faculdade de Odontologia da Universidade Federal do Pará (UFPA).

4.2 Amostras

Devido às dificuldades para a obtenção desse tipo de amostra, foi realizada uma amostragem de conveniência. Trabalhamos com o maior número de leucoplasias orais progressivas e não progressivas que obtivemos. A amostragem foi constituída por 6 leucoplasias transformadas e 15 leucoplasias não transformadas, fixadas em formol e incluídas em parafina (FFPE). Para cada um dos casos, foram coletadas informações clínicas de idade, sexo, hábitos (uso de tabaco e álcool), localização anatômica, tempo de evolução, grau de displasia, tratamento, recorrência, progressão maligna, tempo até a progressão maligna, e tempo de acompanhamento. A partir deste material foi conduzida análise histopatológica, extração de metabólitos e análise metabolômica em cromatógrafo líquido acoplado a espectrômetro de massas.

4.3 Reagentes

Todos os reagentes e solventes utilizados foram de grau de cromatografia líquida de alta performance (*High Performance Liquid Chromatography – HPLC*) e espectrofotométrico. Acetonitrila (C_2H_3N), ácido fórmico (CH_2O_2), água deionizada (H_2O), clorofórmio ($CHCl_3$), metanol (CH_3OH) e xileno (C_8H_{10}), foram comprados da empresa Sigma-Aldrich (*Merck KGaA, Missouri, U.S.*).

4.4 Análise histopatológica

Para confecção das lâminas de análise histopatológica foram obtidos dois cortes de $5\mu m$ de cada bloco. Os cortes foram levados ao banho-maria em uma temperatura de $40^\circ C$ para remoção da parafina e de possíveis dobras no tecido. Posteriormente, foram coletados por meio de uma lâmina adesivada. As lâminas foram levadas à estufa aquecida a $60^\circ C$ para remoção do excesso da parafina e adesão do corte à lâmina. Por fim, os tecidos foram corados com hematoxilina e eosina (H&E).

A análise histopatológica foi conduzida por um patologista experiente (R.S.G.) para confirmação do diagnóstico. Ao realizar novos cortes microscópicos para confirmação

do diagnóstico observamos em uma das seis amostras incluída no grupo de leucoplasias progressivas quadro microscópico consistente com o diagnóstico de CCE. Dessa forma, a amostra foi excluída do estudo. A amostragem final foi constituída por 5 leucoplasias progressivas e 15 leucoplasias não-progressivas.

4.5 Preparo das amostras e extração de metabólitos

A quantidade de tecido incluído em cada bloco de parafina variou consideravelmente. Assim, não foi possível padronizar um único número de cortes para a obtenção de 10mg de massa tecidual por amostra. A estimativa da quantidade de cortes necessários para cada amostra foi feita utilizando o cálculo de densidade $(d = \frac{m}{v})$. Para determinar o volume $(v = a \times h)$, a área de tecido incluído em cada bloco foi mensurada utilizando o *software Image J (National Institutes of Health, Maryland, U.S)*. A espessura $(e = 20\mu\text{m})$ e o número de cortes a serem realizados (n) foram considerados para determinar a espessura de cada tecido. Em seguida, o número de cortes necessários foi estimado considerando o valor da massa tecidual desejada por amostra $(m = 10\text{mg})$.

$$\left(n = \frac{m}{dm(a \times e)} \times 10 \right)$$

$$\text{Em que, } dm = \left(\frac{m/v}{n^{\circ} \text{ total de blocos mensurados}} \right)$$

Os blocos de tecido foram seccionados em cortes de 20 μm de espessura utilizando micrótopo manual. Os dois primeiros cortes de cada bloco foram desprezados para eliminar possíveis contaminantes inerentes à exposição ambiental. A quantidade de cortes variou entre 3 e 83, de acordo com o valor estimado pelo cálculo previamente descrito. Os cortes foram acondicionados em microtubos e pesados em balança analítica.

Para remover a parafina dos cortes, 4 banhos em 1 mL de Xileno foram realizados. Em seguida, os tubos foram centrifugados por 5 minutos a 15.000 g em uma temperatura de 4°C e incubados em banho seco a 56°C para a secagem do xileno. Após estarem completamente secos, os tecidos foram novamente pesados em balança analítica de alta

precisão para o ajuste do valor da massa tecidual desejada em 10 mg de tecido por amostra.

A solução para extração de metabólitos utilizada foi constituída por metanol, clorofórmio e água, na proporção de (CH₃HO:CHCl₃:H₂O – 3:1:1). Após a adição de 200 µL da solução de extração, as amostras foram incubadas no gelo por 10 minutos, vortexadas brevemente e inseridas em banho ultrassônico por 10 minutos. Subsequentemente, as amostras foram centrifugadas por 15 minutos a 15.000 g em uma temperatura de 4°C. O sobrenadante foi coletado e filtrado em filtros de 0,22µm e 13mm de diâmetro para remoção de resíduos contaminantes. Após a filtragem, 10µL de cada amostra foram coletados para o preparo de uma amostra de controle de qualidade (*Quality Control – QC*), constituído por uma mistura de todas as amostras em volumes iguais. Finalmente, 100µL de cada amostra foram inseridos em insertos de vidro, que foram adaptados em *vials*. O conjunto foi imediatamente selado com tampa lacrável com septo de silicone. As amostras foram armazenadas a -80°C até o momento da análise por cromatógrafo líquido acoplado a espectrômetro de massas.

4.6 Análise por cromatografia líquida de alta performance acoplada a espectrômetro de massas

O sistema de cromatografia líquida de alto desempenho (*High Performance Liquid Chromatography – HPLC, LC-20A Prominence*, Shimadzu do Brasil, São Paulo, B.R.) foi equipado com uma coluna de fase reversa C18 (TITAN 18 – 5×2.1 cm id, 1.9 µm, *Supelco Discovery HS, Bellefonte, PA, EUA*) para a separação dos analitos a 40°C. As fases móveis que constituíram o sistema binário de eluição por gradiente foram: (A) água deionizada, contendo 0,1% de ácido fórmico e (B) acetonitrila, contendo 0,1% de ácido fórmico para o modo de ionização positivo e (A) água milliQ e (B) acetonitrila para o modo de ionização negativo. Foram injetados no sistema 10 µL de cada amostra em ordem aleatória. O tempo de análise foi de 30 minutos. Os compostos eluíram da coluna de acordo com o gradiente: 5 – 95% B por 2 – 20 minutos (separação das moléculas); 95% B por 3 minutos (limpeza da coluna); 95 – 20% B por 2 minutos (retorno às condições iniciais); 20% B por 5 minutos (condicionamento da coluna) em um fluxo de 0.3 mL/min. KamLeh e colaboradores, (2012) recomendaram a injeção de

uma amostra de QC a cada duas horas de corrida analítica (KAMLEH et al., 2012). Portanto, a cada cinco amostras de leucoplasia analisadas, uma injeção da amostra de QC foi realizada, proporcionando um conjunto de dados a partir do qual a reprodutibilidade do experimento pôde ser verificada.

A análise por MS utilizou um espectrômetro de massas com detector QTOF (*Quadrupole time-of-flight – QTOF - Bruker*). A fonte de ionização de elétrons do tipo eletrospray (*Electro Spray Ionization – ESI – ESI-micro QTOF II, Bruker*) foi operada a uma energia de colisão de 4,5 kV (modo positivo) e 3,5 kV (modo negativo). O nitrogênio foi utilizado como gás de dessolvatação com pressão de 2 bar e vazão de 7,0 L/h. A temperatura da fonte foi de 100 °C e a temperatura de dessolvatação foi de 180 °C. O nitrogênio também foi usado como gás de colisão e foi gerado a partir do gerador de nitrogênio *NM32LA, Peak Scientific, Inchinnan, UK*. Uma varredura completa de 90-1200 *m/z* foi realizada, usando formato de sódio como calibrador.

Toda a etapa de aquisição dos dados por LC-MS foi realizada no Laboratório de Espectrometria de Massas, do Departamento Engenharia Sanitária e Ambiental (DESA), da Escola de Engenharia da UFMG, sob supervisão e orientação da Prof.a. Dr.a. Lisete Celina Lange, Prof.a. D.r.a. Lucilaine Valéria de Souza Santos e dos alunos de pós-graduação Yuri Abner Rocha Lebron e Victor Rezende Moreira.

4.7 Processamento dos dados

4.7.1 Pré-processamento dos dados brutos

Os dados espectrais adquiridos foram convertidos em arquivos com formato *mzXML* utilizando o *software ProteoWizard (ProteoWizard©, California, U.S.)*. Em seguida, os cromatogramas foram analisados por meio do *software Mass ++ (Mass Plus Plus, Laboratory of Advanced Science and Technology, Shimadzu Corporation, Kyoto, Japan)*. O pacote *Isotopologue Parameter Optimization (IPO – Bioconductor©)* foi instalado no *software R (R Project for Statistical Computing, v.3.6.2, x64)* e utilizado para otimizar os parâmetros de pré-processamento dos dados brutos.

Os melhores parâmetros sugeridos após análise dos dados brutos pelo *software* IPO foram utilizados para configuração das etapas de pré-processamento que foram realizadas no *software* XCMS (SMITH et al., 2006). Basicamente, o pré-processamento é constituído por etapas que objetivam a extração de *molecular features*. Os *molecular features* são sinais tridimensionais equivalentes ao tempo de retenção (*retention time – r.t.*) do composto retido na coluna com determinada relação massa/carga (m/z) e área associada. Na etapa final do pré-processamento pelo *software* XCMS, as informações referentes a cada *molecular feature* identificada foram agrupadas em uma tabela que foi exportada no formato *.xlsx*, compatível com o *software Excel (Microsoft Excel, Washington, U.S.)*. As funções de cada etapa de pré-processamento dos dados no *software* XCMS estão descritas na Figura 3. O processo de obtenção de parâmetros por meio do *software* IPO bem com as linhas de comando utilizadas no pré-processamento pelo *software* XCMS foram descritos no APÊNDICE A.

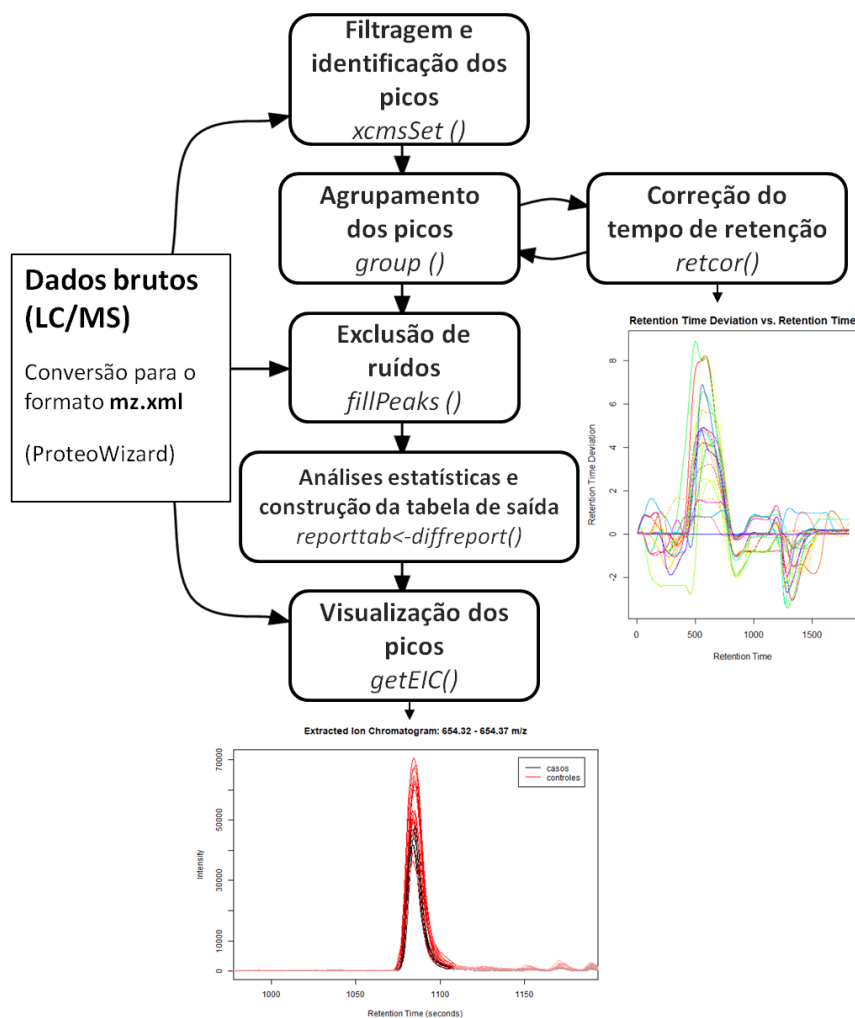


Figura 3. Etapas de pré-processamento dos dados por meio do *software* (XCMS)

Fonte: Adaptado de XiuXia, Du. *Introduction to XCMS in R*, 2017.

4.7.2 Pré-tratamento dos dados

A primeira etapa de pré-tratamento dos dados foi conduzida no *software* Excel. Os compostos detectados durante os primeiros dois minutos e após o vigésimo minuto da corrida foram removidas por se tratarem de compostos eluídos dentro da faixa de tempo morto, tempo de eluição das substâncias que não interagem com a fase estacionária, e compostos eluídos na fase de limpeza da coluna, respectivamente. Em seguida, foi realizada a exclusão dos *molecular features* duplicados. A duplicidade ocorre quando os parâmetros de detecção de pico não são capazes de detectá-lo por inteiro, resultando na divisão de um único pico em dois ou mais *molecular features* detectados com valores de razão massa/carga e tempo de retenção muito próximos. A diferença da razão massa/carga (m/z) entre

cada *molecular feature* foi calculada e os compostos que apresentaram valores menores do que 0.0005 foram considerados como correlacionados. Então, a correlação de Pearson foi realizada para confirmar ou descartar a existência de correlação entre os *molecular features* com diferença de m/z menor do que 0.0005. Os compostos que apresentaram valores de correlação maiores do que 0.7 foram considerados duplicados e, portanto, excluídos do estudo.

A segunda etapa de pré-tratamento dos dados foi realizada na plataforma *online* MetaboAnalyst (*MetaboAnalyst v.4.0, Xia lab, McGill University, Montreal, Canada*). Inicialmente, foi realizada a exclusão e imputação dos *missing values*, sinais de intensidade não detectados. O *cutoff* de no mínimo 80% de presença foi adotado, assim, os *molecular features* com valores de intensidade iônica ausentes em mais do que 20% da amostra (4/20) foram excluídos do estudo. A imputação dos *missing values* restantes foi realizada pelo método *k-nearest neighbors (kNN)*.

A filtragem dos dados foi o segundo tratamento realizado. O objetivo da filtragem é eliminar os dados de baixa qualidade como, *molecular features* com baixa repetibilidade, não informativos e ruídos. O filtro de desvio padrão relativo (*Relative Standard Deviation – RSD*) foi aplicado para a exclusão dos *molecular features* com baixa repetibilidade. Dessa forma, os compostos com *RSD* maior do que 30% dentro do grupo de QCs foram excluídos do estudo. O filtro adicional de *RSD* não paramétrico (Desvio Absoluto da Mediana/Mediana – *MAD/Median*) foi aplicado para a remoção de dados caracterizados como ruídos, que possuem valores de intensidade muito baixos, e dos *molecular features* não informativas, variáveis que permanecem constantes ao decorrer das diferentes condições experimentais.

A terceira etapa de tratamento foi constituída pela normalização, transformação e escalonamento. Nessa fase, foram utilizadas duas abordagens de tratamento. O conjunto de dados detectados no modo de ionização positivo foi normalizado pela mediana e os dados detectados no modo de ionização negativo foram normalizados por quantil. A transformação por *log* foi realizada para os dois conjuntos. O escalonamento não foi realizado em ambos os conjuntos de dados para as análises estatísticas univariadas. Para as análises estatísticas multivariadas,

os dados foram normalizados por quantil, transformados por *log* e escalonados por Pareto. A representação da distribuição dos dados antes e depois da aplicação dos métodos de normalização pode ser visualizada no APÊNDICE B.

4.8 Seleção de variáveis

Embora os métodos de pré-processamento e pré-tratamento dos dados confirmem uma matriz de *molecular features* qualificados, análises estatísticas adicionais são necessárias para a seleção de variáveis relevantes com informações biologicamente significativas dentro do conjunto de dados.

4.8.1 Análises estatísticas univariadas

As análises estatísticas univariadas foram realizadas utilizando o *software* IBM SPSS *Statistics* v.26. O teste de normalidade Shapiro-Wilk foi conduzido e, então, o teste de hipóteses T de Welch (Teste T combinado com o teste de Levene para igualdade de variâncias) foi realizado com os *molecular features* que apresentaram distribuição normal ($p > 0.05$). Os compostos que possuíram distribuição não normal ($p < 0.05$) foram analisadas pelo teste de hipóteses de Mann Whitney U. Os pós-testes de Bonferroni e Tukey foram aplicados, contudo, essas abordagens foram muito rígidas para o conjunto de dados desse estudo. Após as correções pós-teste, nenhum *molecular feature* apresentou diferença estatisticamente significativa entre os grupos estudados. Por isso, os valores de p obtidos nos testes de hipóteses não foram corrigidos. Para contornar o problema de testagem múltipla, a diferença entre os grupos de leucoplasias não progressivas *versus* progressivas foi também avaliada por meio do cálculo do *t-score* e *fold-change*, utilizando o *software* Excel. Esses cálculos foram realizados fazendo uso do conjunto de dados pré-tratados de acordo com os parâmetros aplicados para a análise estatística multivarada.

A contribuição dos *molecular features* discriminantes ($p < 0.05$) para a distinção dos grupos em estudo foi avaliada por meio de um agrupamento hierárquico supervisionado representado em dendograma gerado por meio da plataforma *online* MetaboAnalyst v4.0.

4.8.2 Análises estatísticas multivariadas

Para ilustrar o padrão de distribuição dos grupos de leucoplasia oral estudados foram utilizados o Componente de Análise Principal (*Principal Component Analysis – PCA*) como método não supervisionado e a Análise Discriminante de Mínimos Quadrados Parciais (*Partial Least Squares-Discriminant Analysis – PLS-DA*) como método de análise supervisionado. O método SAM (*Significance Analysis of Microarray*) foi empregado para uma análise mais robusta de todo o conjunto de dados contando com o controle da taxa de descobertas falsas (*False Discovery Rate – FDR*). As análises multivariadas descritas nesta seção foram realizadas na plataforma *online* MetaboAnalyst v4.0.

4.8.3 Seleção recursiva dos *molecular features* discriminantes

O algoritmo de modelagem multivariada MUVR (*Minimally Biased Variable Selection in R – Swedish University of Agricultural Sciences*) foi utilizado como método adicional para a seleção de variáveis dentro de um princípio de validação cruzada dupla repetida (*repeated double cross validation – rdCV*). O pacote MUVR foi instalado no *software* RStudio (*RStudio Desktop– v. 1.4.1717*) e, em seguida, os conjuntos de dados obtidos em modo de ionização positivo e negativo foram pré-processados conforme as etapas descritas no item 4.7.1, normalizados pelo grupo de referência (QC), transformados em *log* e escalonados por Pareto, utilizando a plataforma *online* MetaboAnalyst v5.0 (*MetaboAnalyst v.5.0, Xia lab, McGill University, Montreal, Canada*). Subsequentemente, o *software* Excel foi utilizado para a formatação dos conjuntos de dados normalizados. As observações foram dispostas nas linhas (x) e as variáveis foram dispostas nas colunas (y). O grupo “QC” foi excluído e os arquivos foram salvos no formato .xls. Então, os modelos de classificação foram realizados para cada conjunto de dados utilizando os métodos PLS (*Partial Least Squares*) e RF (*Random Forest*) do algoritmo MUVR.

O esquema de classificação MUVR é baseado na divisão do conjunto de dados em seguimentos externos de teste (*nOuter*) e seguimentos internos de treino (*nInner*).

Neste estudo, a correspondência entre os métodos estatísticos foi o critério final utilizado para a seleção das variáveis discriminantes. Foram selecionados os compostos que se mostraram estatisticamente significativos para a distinção dos grupos nas análises estatísticas univariadas, análise SAM e em pelo menos um dos modelos MUVR, PLS ou RF. Finalmente, um agrupamento hierárquico não supervisionado representado por um mapa de calor foi gerado na plataforma *online* MetaboAnalyst v. 5.0 para ilustrar a contribuição dos *molecular features* selecionados para a distinção dos grupos em estudo.

4.9 Avaliação da performance dos *molecular features* discriminantes

A plataforma *online* MetaboAnalyst v. 5.0 foi utilizada para avaliar o poder preditivo do conjunto de *molecular features* selecionados de modo correspondente entre os métodos estatísticos uni e multivariados previamente aplicados.

A análise foi conduzida utilizando o módulo “*Biomarker Analysis*”. O dado de entrada foi constituído pelos valores de intensidade dos *molecular features* discriminantes. O conjunto de dados foi normalizado pela distância interquantílica, transformados por *log*, e escalonados por Pareto. A curva ROC (*Receiver Operating Characteristics*) foi utilizada para avaliar a contribuição de cada composto para a discriminação dos grupos em estudo de modo individual, usando como métrica o cálculo AUC (*Area Under the Curve*). Em seguida, o algoritmo de análise multivariada SVM (*Support Vector Machine*) foi utilizado para a construção de modelos baseados nas curvas ROC previamente realizadas. Por último, foi construído um gráfico de dispersão para a visualização do poder discriminatório do modelo com maior valor de AUC.

4.10 Enriquecimento de vias metabólicas

O módulo “*Functional Analysis*” disponível na plataforma *online* MetaboAnalyst v.5.0 foi utilizado para o enriquecimento das vias metabólicas por meio da integração entre os algoritmos *Mummichog v.2.0* e *Gene Set Enrichment Analysis – GSEA*. Os dados de entrada foram constituídos pelos valores de *p*, razão

massa/carga (m/z) e t . score dos *molecular features* identificados em ambos os modos de ionização. As análises foram conduzidas considerando uma tolerância de peso molecular de 1, 3, 5 e 10 ppm, em modo analítico misto (modo de ionização positivo e negativo). Foram consideradas válidas apenas as associações de m/z em compostos que apresentaram íons primários. O *cutoff* para o valor de p foi configurado em 0.05. Todos os adutos e isótopos foram considerados como combinações potenciais. O modelo MFN humano (KEGG, BiGG, e modelo de Edinburgh) foi a base de dados interrogada. Essa base é constituída por dados manualmente curados das bases KEGG, BiGG e Edinburgh. O fluxo de trabalho do presente estudo está ilustrado na Figura 5.

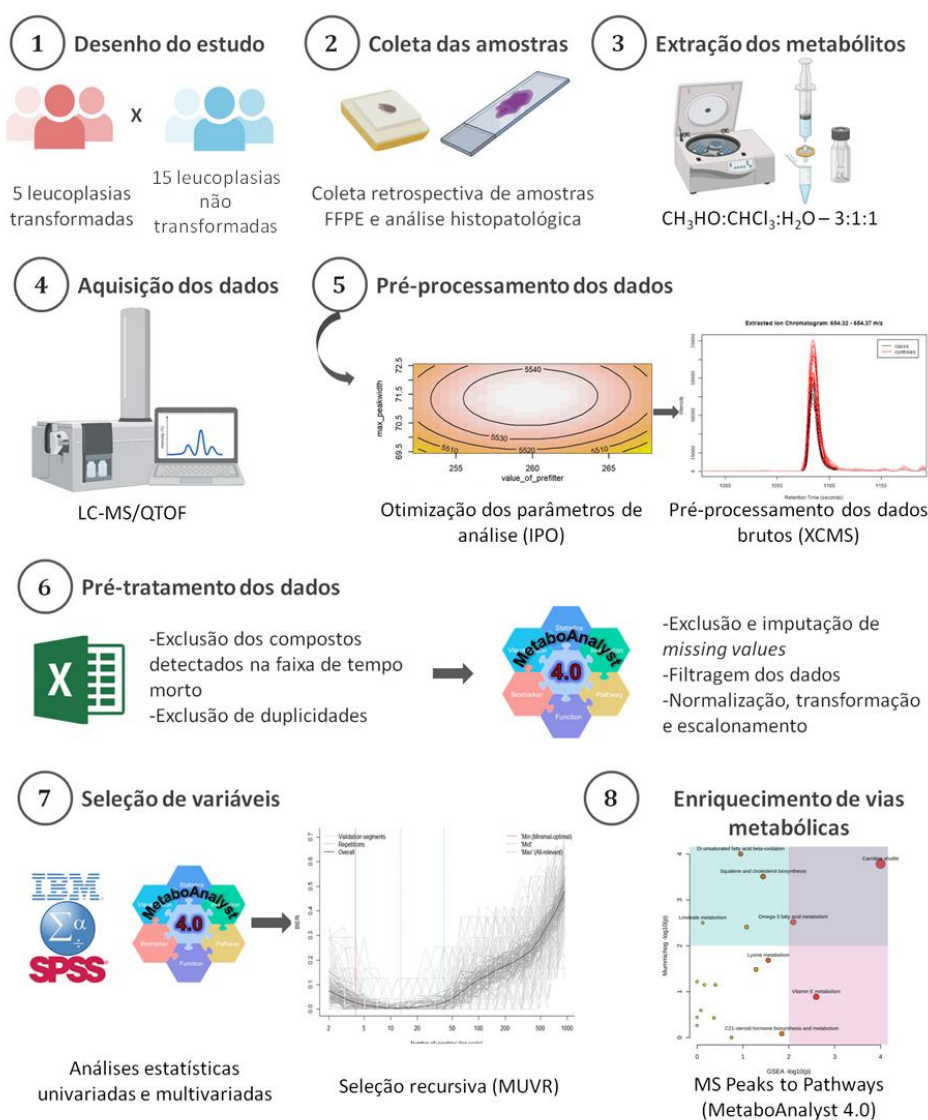


Figura 5. Fluxo de trabalho desenvolvido no presente estudo.
Fonte: Elaborado pela autora. Vetores 1-4: BioRender

5. RESULTADOS E DISCUSSÃO

5.1 Caracterização da amostra

O grupo de estudo foi formado por 5 amostras de leucoplasias transformadas e 15 amostras de leucoplasias não transformadas com acompanhamento mínimo de 12 meses após o diagnóstico da lesão. Os pacientes afetados pertenciam principalmente à terceira (5/20), quarta (4/20) e sétima (4/20) décadas de vida. A idade média dos indivíduos que compuseram o grupo de leucoplasias transformadas foi de $\cong 51$ anos ($DP = \pm 17,8$) e daqueles incluídos no grupo de leucoplasias não transformadas foi de $\cong 56$ anos ($DP = \pm 15,7$). Três pessoas do grupo de leucoplasias transformadas (T1, T4 e T5) tinham entre 31 e 41 anos no momento do diagnóstico inicial. Para esses casos, o tempo até a progressão maligna variou entre 32 e 84 meses, com média de 40 meses ($DP = \pm 22,4$). Curiosamente, essas três pacientes mais jovens eram do sexo feminino, não tabagistas, não etilistas e se encontravam na quarta década de vida quando o câncer oral foi diagnosticado. Os outros dois sujeitos acometidos pela transformação maligna da leucoplasia eram do sexo masculino e pertenciam à sétima década de vida (T2 e T3). O tempo até a malignização para esses dois indivíduos foi de 24 meses. A literatura publicada até a presente data aponta que a evolução maligna ocorre em um período aproximado de 3 anos, variando entre 1,8 e 5,1 anos (AGUIRRE-URIZAR; LAFUENTE-IBÁÑEZ DE MENDOZA; WARNAKULASURIYA, 2021).

O reduzido número de amostras analisadas no estudo dificulta a realização de inferências robustas, no entanto, estes resultados demonstram convergência com algumas observações já publicadas. No geral, a transformação maligna ocorre após os 50 anos, principalmente entre a sétima e oitava década de vida (AGUIRRE-URIZAR; LAFUENTE-IBÁÑEZ DE MENDOZA; WARNAKULASURIYA, 2021; EINHORN; WERSALL, 1967; EVREN et al., 2020; WARNAKULASURIYA; ARIYAWARDANA, 2016). Contudo, em um clássico estudo observacional, Bánóczy (1977) reportou uma proporção considerável de evolução da leucoplasia para o câncer oral durante a quarta década de vida (BÁNÓCZY, 1977). Relatos de casos de leucoplasia oral em pessoas menores do que 40 anos de idade são escassos na literatura. Em uma revisão sistemática incluindo 1246 pacientes com lesões cancerizáveis, apenas 115 foram diagnosticados com leucoplasia oral. Destes, oito sofreram progressão

maligna (ROZA et al., 2021). Diante disso, mais estudos prospectivos observacionais são necessários para a compreensão dos fatores associados ao processo de cancerização em indivíduos jovens.

Enquanto a relação entre a malignização da leucoplasia em jovens é pouco compreendida, o risco aumentado de progressão maligna para lesões diagnosticadas em mulheres tem sido extensivamente reportado (AGUIRRE-URIZAR; LAFUENTE-IBÁÑEZ DE MENDOZA; WARNAKULASURIYA, 2021; CERQUEIRA et al., 2020; EVREN et al., 2020; HO et al., 2012; NAPIER; SPEIGHT, 2008; REIBEL, 2003; SCHEPMAN et al., 1998; SPEIGHT; KHURRAM; KUJAN, 2018; WARNAKULASURIYA; ARIYAWARDANA, 2016). Em uma recente visão sistemática com meta-análise, as mulheres apresentaram um risco aproximadamente duas vezes maior de sofrerem evolução da leucoplasia oral para o carcinoma de células escamosas (AGUIRRE-URIZAR; LAFUENTE-IBÁÑEZ DE MENDOZA; WARNAKULASURIYA, 2021). Ainda, Schepman e colaboradores (1998), observaram que a chance de transformação maligna foi aumentada para as mulheres não fumantes, em relação às tabagistas (SCHEPMAN et al., 1998).

Embora o tabaco desempenhe um papel importante para o surgimento da leucoplasia, a progressão maligna tem afetado principalmente indivíduos não fumantes (CERQUEIRA et al., 2020; HO et al., 2012; HOGEWIND; VAN DER KWAST; VAN DER WAAL, 1989; REIBEL, 2003; SCHEPMAN et al., 1998). Em uma análise longitudinal de 455 pacientes, foi visto que as lesões orais displásicas em indivíduos não fumantes tinham cerca de duas vezes mais chance de progredirem para o câncer. Quando essas lesões estavam localizadas em soalho de boca, o risco foi 38 vezes maior (ROCK et al., 2018). O risco aumentado de transformação dessas lesões, também denominadas como leucoplasias idiopáticas, sugere que uma predisposição genética à malignização pode ocorrer em pelo menos uma proporção de indivíduos (SPEIGHT; KHURRAM; KUJAN, 2018).

As lesões leucoplásicas ocorrem com maior frequência na mucosa jugal (SILVERMAN et al., 1976; WARNAKULASURIYA; ARIYAWARDANA, 2016), contudo, a taxa de transformação maligna nesta localização é baixa (DOST et al., 2013; JABER et al., 2003; SPEIGHT; KHURRAM; KUJAN, 2018; WALDRON; SHAFER, 1975). A língua

e o soalho de boca são sítios anatômicos contíguos considerados como áreas altamente susceptíveis ao desenvolvimento e transformação maligna da leucoplasia (CERQUEIRA et al., 2020; RUBERT; BAGÁN; BAGÁN, 2020; SPEIGHT; KHURRAM; KUJAN, 2018; WALDRON; SHAFER, 1975). Neste estudo, 4 das 5 lesões de leucoplasia progressivas (80%) ocorreram em língua (T3,T4 e T5) e soalho (T2). A língua também foi a localização mais afetada pelos casos não progressivos (7/15), seguida pela mucosa jugal (4/15), fundo do saco do vestíbulo (2/15), palato duro (1/15) e mucosa gengival (1/15). Ainda que alguns autores tenham apontado que leucoplasias localizadas em língua e soalho exibem alterações displásicas mais marcantes (CERQUEIRA et al., 2020; DOST et al., 2013; EVREN et al., 2020; GANDARA-VILA et al., 2018; HO et al., 2012; JABER et al., 2003; NAPIER; SPEIGHT, 2008), o reduzido número de amostras avaliadas no presente estudo impede a associação entre o sítio anatômico acometido e a gradação da displasia epitelial oral.

Um dos resultados reportados de forma mais consistente pelas análises clinicopatológicas da leucoplasia é a associação entre a presença de graus elevados de displasia epitelial oral e a evolução maligna da doença. Após a análise histopatológica, as lesões leucoplásicas podem ser subdivididas entre leucoplasias de baixo risco, quando há a presença de hiperplasia e/ou displasia epitelial oral leve, ou de alto risco, quando alterações displásicas severas são observadas (WU; WANG; ZHOU, 2019). A probabilidade de uma leucoplasia de alto risco se transformar em carcinoma de células escamosas é cerca de 5 vezes maior do que aquelas classificadas como lesões de baixo risco (AGUIRRE-URIZAR; LAFUENTE-IBÁÑEZ DE MENDOZA; WARNAKULASURIYA, 2021). Quatro dos cinco casos transformados analisados neste trabalho exibiram displasia epitelial moderada (T2-T5 – 80%) e, um, displasia epitelial severa (T1). Dos casos não transformados, 6 mostraram displasia epitelial leve (C4-C6, C9, C11 e C12 – 40%), 5 displasia epitelial moderada (C1, C7, C8, C10 e C13 – \cong 33%), 3 displasia epitelial severa (C3, C2 e C14 – 20%), e 1 não exibiu alterações displásicas no seu quadro histopatológico (C15 – \cong 7%).

Todas as leucoplasias incluídas nesta pesquisa foram tratadas por excisão cirúrgica à lâmina fria. De acordo com a literatura, a taxa de recorrência da leucoplasia oral varia entre 7% e 38%, independentemente do tratamento realizado (IOCCA et al., 2020b; KURIBAYASHI et al., 2012). Metade dos pacientes deste estudo (10/20) foram

afetados por lesões recidivantes. Curiosamente, a recidiva ocorreu em apenas uma de seis lesões que apresentaram displasia epitelial oral leve (17%), e em nove das treze leucoplasias caracterizadas por um quadro moderado ou severo de displasia epitelial oral (69%). No entanto, a relação entre o grau de displasia e a recorrência não foi observada em um estudo multicêntrico envolvendo 226 pacientes com leucoplasia. O mesmo trabalho mostrou uma associação estatisticamente significativa ($p < 0.001$) entre a recidiva e a progressão maligna (SUNDBERG et al., 2019). Quatro das cinco amostras transformadas exibiram comportamento recorrente (T2-T5 – 80%). Dois destes casos sofreram múltiplas recorrências (T2 – dois episódios e T5 – quatro episódios). Das seis amostras de leucoplasias não transformadas e recidivantes, duas também recidivaram, ambas por duas vezes após a remoção cirúrgica (C1 e C4 – $\cong 33\%$). Em um caso (C9), uma nova lesão de leucoplasia surgiu na mucosa jugal contralateral após 72 meses do diagnóstico inicial. Alguns estudos associaram a recidiva e/ou a formação de novas lesões de leucoplasia à ressecção cirúrgica incompleta, na qual anomalias epiteliais são observadas nas margens do espécime ressecado (KURIBAYASHI et al., 2012; PANDEY et al., 2001). Em um estudo recente, os fatores relacionados à propensão para a recidiva de leucoplasias localizadas em língua foram: o histórico de câncer em região de cabeça e pescoço, a aparência clínica não homogênea e a presença de graus moderado e severo de displasia epitelial oral. (YANG et al., 2021).

O tempo até a recorrência foi de em média $\cong 12$ meses para os dois grupos estudados (casos transformados – DP= $\pm 7,11$; casos não transformados – DP= $\pm 16,46$). O acompanhamento clínico após o diagnóstico inicial dos casos avaliados no presente estudo foi de no mínimo um ano, variando entre 12 a 144 meses, com média de $\cong 50$ meses (DP= $\pm 33,72$). A porcentagem de lesões malignamente transformadas aumenta conforme o período de acompanhamento. Já foi estimado que cerca de 50% dos pacientes com leucoplasia poderiam desenvolver um carcinoma de células escamosas dentro de um período de 200 meses após o primeiro diagnóstico (SCHEPMAN et al., 1998; SPEIGHT; KHURRAM; KUJAN, 2018). Por isso, é importante ressaltar que as leucoplasias orais incluídas no grupo de lesões não transformadas poderiam estar em processo de malignização até o momento da finalização do estudo, uma vez que o período máximo de acompanhamento após o diagnóstico foi de 144 meses. Uma vez que não há, até o momento, ferramentas capazes de predizer e/ou prevenir a evolução da leucoplasia para o câncer oral, longos e regulares períodos de acompanhamento são

fundamentais para que o diagnóstico precoce da progressão maligna seja realizado. As informações clínicas dos casos reportados nesse estudo estão sumarizadas na Tabela 1.

Além da limitação relacionada ao curto período de acompanhamento de algumas amostras do presente estudo, a fragilidade do conceito de leucoplasia é um entrave importante para este e qualquer outro estudo que se propõe a estudar a doença. A lesão foi definida pela primeira vez no ano de 1978 pela Organização Mundial da Saúde como “uma mancha ou placa branca que não pode ser caracterizada clinicamente ou histopatologicamente como qualquer outra desordem” (KRAMER IR, 1978). O conceito da lesão foi alterado por mais três vezes ao longo dos anos (AXELL et al., 1984; VAN DER WAAL, 2015; WARNAKULASURIYA; JOHNSON; VAN DER WAAL, 2007) e, hoje, a leucoplasia é definida como “uma mancha ou placa predominantemente branca que não pode ser caracterizada clinicamente ou histopatologicamente como qualquer outra desordem e que carrega um risco aumentado para o desenvolvimento do câncer tanto na área leucoplásica como em qualquer outro sítio anatômico da cavidade oral” (VAN DER WAAL, 2015). Dessa forma, o diagnóstico de leucoplasia é realizado por exclusão de outras lesões em placa branca que podem afetar a cavidade oral. Ainda, a coleta retrospectiva das amostras limita a obtenção de características que podem fornecer informações importantes sobre risco de transformação maligna da doença, como, por exemplo, a aparência clínica não homogênea da lesão. No presente estudo, não foi possível recuperar a descrição sobre a aparência clínica das lesões de leucoplasia oral avaliadas.

Caso	Sexo	Idade	Etnia	Hábitos	Localização	Evolução (meses)	Grau de displasia	Transformação maligna (meses)	Recidiva (meses)	Acompanhamento (meses)
T1	F	43	N-B	Não	Palato mole e orofaringe	24	Severo	Sim (36)	Não	36
T2	M	72	B	NI	Soalho de boca	12	Moderado	Sim (24)	Sim (7 e 9)	24
T3	M	73	N-B	Não	Borda posterior de língua	NI	Moderado	Sim (24)	Sim (9)	12
T4	F	37	B	Não	Borda posterior de língua	NI	Moderado	Sim (84)	Sim (24)	84
T5	F	31	N-B	Não	Borda posterior de língua	2	Moderado	Sim (32)	Sim (7, 11, 14 e 46)	72
C1	M	54	B	Não	Palato duro	60	Moderado	Não	Sim (2, 4)	48
C2	F	76	N-B	Tabagista (cachimbo)	Mucosa jugal	NI	Severo	Não	Não	12
C3	M	40	B	Não	Borda posterior de língua	2	Severo	Não	Sim (1)	12
C4	M	73	N-B	NI	Borda posterior e ventre de língua	NI	Leve	Não	Sim (17 e 20)	34
C5	M	53	B	Não	Borda ântero-posterior de língua	NI	Leve	Não	Não	24
C6	M	38	B	Não	Ventre de língua	1	Leve	Não	Não	12
C7	M	69	B	Não	Mucosa jugal	36	Moderado	Não	Não	72
C8	F	85	B	NI	Fundo do saco do vestibulo mandibular	12	Moderado	Não	Sim (1)	90
C9	F	34	N-B	NI	Mucosa jugal (lado direito)	5	Leve	Não	Nova lesão em mucosa jugal, lado esquerdo (72)	144
C10	M	49	B	NI	Borda posterior de língua	3	Moderado	Não	Não	60
C11	M	69	N-B	NI	Mucosa jugal (lado direito)	NI	Leve	Não	Não	36
C12	M	59	B	NI	Borda posterior e ventre de língua	1	Leve	Não	Não	24
C13	F	39	N-B	Não	Fundo do saco do vestibulo maxilar	36	Moderado	Não	Sim (6)	84
C14	F	40	N-B	Não	Borda posterior de língua e soalho de boca	NI	Severo	Não	Sim (47)	50
C15	M	68	N-B	NI	Mucosa gengival	60	Ausente	Não	Não	72

Tabela 1. Informações clínicas das amostras de leucoplasia analisadas neste estudo.

T: grupo de casos transformados; C: grupo de casos controle; NB: não branco; B: branco; NI: não informado.

5.2 Controle de qualidade dos experimentos

A qualificação da performance instrumental é uma etapa importante para a confirmação da confiabilidade e reprodutibilidade do método de análise (SUMNER et al., 2007). A linearidade da corrida cromatográfica foi analisada pela visualização dos espectros das amostras de controle de qualidade (QC) e também foi avaliada por meio da análise de componente principal (PCA). Os cromatogramas de íons totais das amostras de QC mostraram uma boa sobreposição, o que indica uma baixa variação decorrente de problemas no instrumento de análise (FIGURA 6 A e B). Uma variação na intensidade foi observada entre as duas primeiras e as duas últimas amostras de QC durante a corrida no modo de ionização negativo (FIGURA 6 A). No entanto, as demais amostras do estudo acompanharam o mesmo padrão de variação (APÊNDICE D). O agrupamento das amostras de QC no gráfico da PCA, observado tanto no modo de ionização positivo como no modo negativo indica que a reprodutibilidade do experimento não foi afetada (FIGURA 7 A e B).

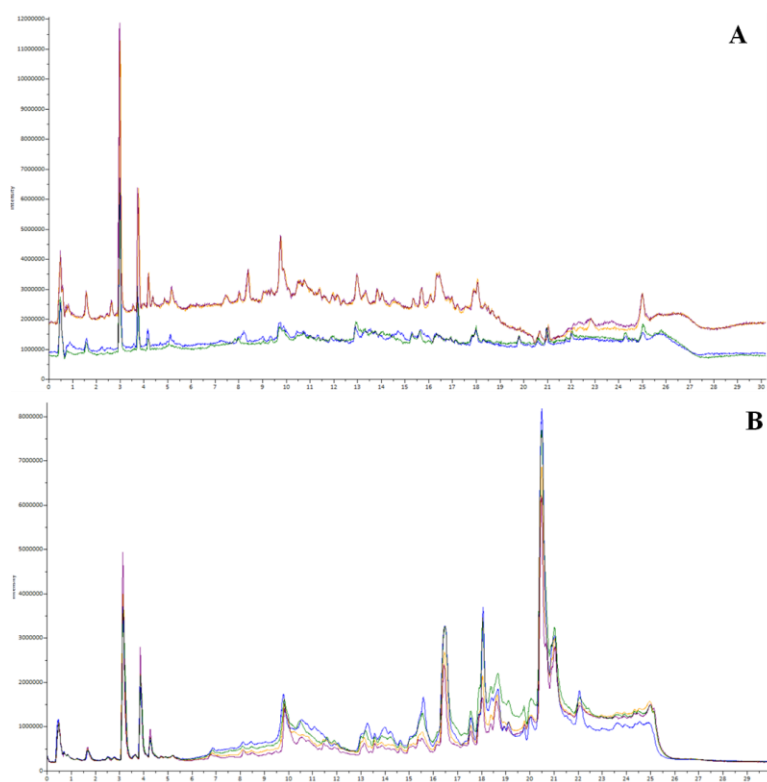


Figura 6. Cromatogramas de íons totais das amostras de QC. O eixo Y corresponde à intensidade de detecção iônica e o eixo X representa o tempo de retenção. **(A)** Modo de ionização negativo. **(B)** Modo de ionização positivo.

Fonte: Capturas de tela obtidas do *software* Mass ++.

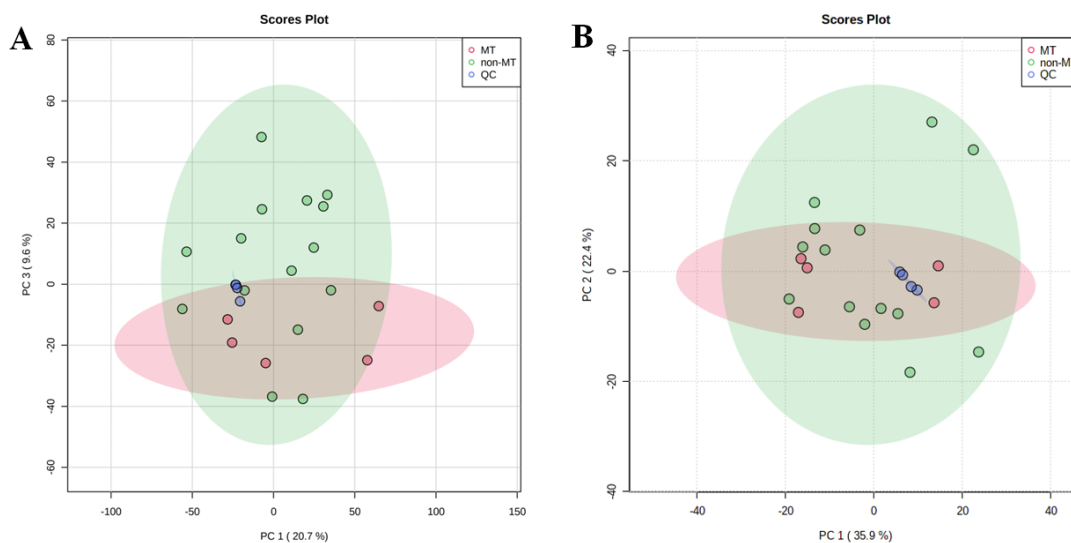


Figura 7. Análise de componentes principal – PCA. O PCA mostra o agrupamento das amostras de QC (pontos azuis). A variância explicada por cada componente está descrita entre parênteses. As áreas em vermelho e em verde delimitam o intervalo de 95% de confiança **(A)** Gráfico de escores entre os componentes principais 3 e 1 representando a distribuição dos dados obtidos em modo de ionização positivo. **(B)** Gráfico de escores entre os componentes principais 2 e 1 representando a distribuição dos dados obtidos em modo de ionização negativo.
PC: principal component. MT: malignant transformed. Non-MT: non-malignant transformed. QC: quality control.

Fonte: Imagens geradas pela plataforma MetaboAnalyst.

5.3 Pré-processamento e pré-tratamento dos dados

Os métodos de pré-processamento dos dados adquiridos por LC-MS requerem o ajuste de diversos parâmetros para a deconvolução e extração de sinais analíticos. A escolha adequada desses parâmetros não é uma tarefa fácil. O pesquisador necessita de um amplo conhecimento sobre processamento de dados uma vez que a definição de parâmetros inadequados pode levar à distorção dos resultados em estudos de metabolômica (LIBISELLER et al., 2015). Por isso, no presente estudo, essa etapa foi otimizada com o uso do *software* IPO, uma ferramenta de otimização automática que possibilita a seleção dos parâmetros ideais por meio da análise interativa de gráficos de superfície de resposta que são gerados durante o desenho dos experimentos (DoEs) realizado pelo algoritmo (ALBÓNIGA et al., 2020). Os gráficos de superfície de resposta e os parâmetros definidos para o pré-processamento dos dados obtidos em ambos os modos de ionização podem ser visualizados no APÊNDICE A. Os parâmetros selecionados foram posteriormente utilizados para o pré-processamento dos dados no *software* XCMS que resultou na detecção de 5.399 *molecular features* no modo de ionização positivo e 287 *molecular features* no modo de ionização negativo.

Após o pré-processamento, os dados devem ser pré-tratados para que se tornem prontos para serem processados nas análises estatísticas subsequentes. (GOODACRE et al., 2007) Conforme detalhado no item 4.7.2 do presente estudo, o pré-tratamento dos dados foi realizado em duas etapas. Na primeira, foram eliminados os compostos detectados na faixa de tempo morto da corrida e as duplicidades. Foram excluídos 2.142 dos 5.399 *molecular features* adquiridos no modo de ionização positivo e 111 dos 287 *molecular features* detectados no modo de ionização negativo. Foram detectados 87 *missing values* no conjunto de dados adquiridos em modo de ionização positivo. Seguindo a regra de 80%, na qual uma variável deve estar presente em pelo menos 80% das amostras analisadas (SMILDE et al., 2005), os *molecular features* que apresentaram mais do que 20% de *missing values* foram excluídos do estudo. O método kNN foi utilizado para imputar valores aos *missing values* restantes. Neste método, a métrica euclidiana é utilizada para encontrar as amostras mais próximas nas quais o valor de interesse pôde ser detectado. Então, o algoritmo calcula a média dos elementos correspondentes dos seus vizinhos e atribui o valor ao *missing value* (TROYANSKAYA et al., 2001). Essa análise tem demonstrado vantagem sobre diversos outros métodos de imputação (DO et al., 2018; FAQUIH et al., 2020) e, por isso, foi a abordagem escolhida no presente estudo. O conjunto de dados adquiridos em modo de ionização negativo não apresentou *missing values*.

Conforme descrito no item 4.7.2 do presente estudo, o filtro de desvio padrão relativo foi aplicado para a exclusão dos *molecular features* com baixa repetitividade utilizando o *cutoff* de 30%. Esse método de filtragem é baseado na mensuração da variação analítica utilizando as amostras de QC, pois elas representam a composição biológica de todas as amostras avaliadas no estudo (KIRWAN et al., 2013). Quanto menor a porcentagem de RSD, maior é a estabilidade analítica (VAN DER KLOET et al., 2009). Embora as diretrizes da FDA recomendem a filtragem por RSD utilizando o *cutoff* de 20% para o estudo de biomarcadores (KIRWAN et al., 2013), a aplicação do *cutoff* de 30% demonstrou bons resultados em diversas análises publicadas (BROADHURST et al., 2018; CHETNIK; PETRICK; PANDEY, 2020; REYES-GARCÉS et al., 2021; WANG et al., 2019). Após a aplicação dos filtros de RSD, 1.684 *molecular features* detectados no modo de ionização positivo e 105 *molecular features* detectados no modo de ionização negativo foram excluídos do estudo.

Depois da filtragem, foram aplicados métodos de normalização para a redução de vieses sistemáticos que poderiam afetar a consistência geral dos dados adquiridos (CHONG; WISHART; XIA, 2019). Os extratos de tecidos podem naturalmente apresentar diferentes quantidades de metabólitos. Essas diferenças podem ser inerentes aos fatores biológicos, ligados à condição ou doença estudada, ou não biológicos, relacionados principalmente aos métodos de coleta, armazenamento e preparo das amostras. Dessa forma, a normalização se faz necessária para determinar que as mudanças no perfil metabolômico sejam causadas apenas por um evento biológico (WU; LI, 2016). Ainda que variados métodos tenham sido propostos para a normalização de dados de metabolômica (DI GUIDA et al., 2016), não existe uma única diretriz para a condução dessa etapa de pré-tratamento dos dados (CRAIG et al., 2006; DI GUIDA et al., 2016; WU; LI, 2016).

Para os dados a serem posteriormente processados por análises estatísticas multivariadas, a normalização, a transformação e o escalonamento foram realizados por meio de testes de tentativa e erro, optando pelos métodos que tornaram a distribuição dos dados o mais próximo possível ao formato gaussiano, conforme orientado por Chong, Wishart e Xia, (2019). Não foram aplicados métodos de escalonamento para os dados a serem posteriormente processados por análises estatísticas univariadas (DI GUIDA et al., 2016). O escalonamento dos dados tem como principal finalidade reduzir grandes discrepâncias de magnitude entre as variáveis. No entanto, a abordagem pode resultar na inflação de valores pequenos, ocasionando um aumento no erro de medição (VAN DEN BERG et al., 2006). Assim, para as análises estatísticas univariadas, o escalonamento dos dados poderia acarretar o aumento de erros do tipo I e II.

Após a aplicação de todos os métodos de pré-processamento e pré-tratamento de dados, 1012 *molecular features*, sendo 941 detectados no modo de ionização positivo e 71 no modo de ionização negativo foram processados por análises estatísticas uni e multivariadas para a identificação dos compostos diferencialmente abundantes e os seus padrões de distribuição entre os grupos de leucoplasias progressivas e não progressivas.

5.4 Seleção de variáveis: Detecção dos compostos discriminantes

Cada *molecular feature* detectado após o pré-processamento e pré-tratamento dos dados não necessariamente se traduz em um metabólito específico. Isso ocorre porque os dados adquiridos por LC-MS são altamente redundantes devido à detecção recorrente de adutos, isótopos ou íons duplamente carregados que inflam acentuadamente o número de picos detectados. Por isso, é essencial que os compostos realmente relevantes dentro do conjunto de dados sejam selecionados a partir de critérios estatísticos. Dessa forma, a combinação entre métodos de análise estatística univariada e multivariada é recomendada para retenção dos *molecular features* potencialmente ricos em informações (VINAIXA et al., 2012)

5.4.1 Análises estatísticas univariadas

Os testes de hipótese são utilizados para a análise estatística univariada de dados de metabolômica a fim de caracterizar se a variação entre os *molecular features* é devido ao fenômeno estudado ou à aleatoriedade. A escolha da abordagem estatística apropriada deve levar em consideração as características do conjunto de dados analisados no que diz respeito ao desenho experimental, à normalidade e à heterocedasticidade (VINAIXA et al., 2012).

O fluxograma de análise estatística univariada aplicado neste estudo seguiu as recomendações realizadas por Vinaixa e colaboradores (2012). O teste de Shapiro-Wilk se comporta melhor para a análise da distribuição dos valores em estudos com um número reduzido de amostras (<50) e, por isso, foi aplicado para a avaliação da normalidade. As variáveis que apresentaram distribuição normal foram posteriormente avaliadas por teste de hipótese paramétrico e aquelas com distribuição não normal, por teste não paramétrico. As amostras analisadas constituíram dois grupos independentes (leucoplasias transformadas *versus* não transformadas). Logo, o teste T de Welch foi a abordagem paramétrica escolhida. Esse teste conta com o Teste de Levene para igualdade de variâncias, que permitiu avaliar a heterocedasticidade dos dados. O teste de Mann Whitney U foi empregado para a análise dos compostos que apresentaram distribuição não normal. Em ambas as análises o *threshold* de 5% para rejeição da hipótese nula foi adotado ($p < 0.05$).

Dos 1012 *molecular features* analisados, 826 apresentaram distribuição normal (Shapiro-Wilk > 0.05) e 186 distribuição não-normal (Sapiro-Wilk < 0.05). Os testes de hipótese revelaram 72 compostos diferencialmente abundantes entre os grupos estudados ($p < 0.05$).

Quanto maior é a quantidade de testes de hipóteses realizados, maior é a probabilidade da ocorrência de resultados falsos positivos. Os falsos positivos são erros do tipo I. Neste tipo de erro, a hipótese nula é rejeitada, indicando que há diferença significativa entre os grupos, quando na verdade não há. Esse “problema de testagem múltipla” representa um dos maiores desafios relacionado à análise univariada de grandes conjuntos de dados. A detecção de falsos positivos pode afetar seriamente os estudos de metabolômica, dificultando a descoberta de compostos com verdadeiro valor biológico, promissores para posterior validação funcional (IOANNIDIS, 2005; VINAIXA et al., 2012).

A correção dos valores de p obtidos dos testes de hipótese é o meio para controlar a detecção de falsos positivos. Os principais métodos de correção pós-teste são a correção de Bonferroni e o cálculo da taxa de descobertas falsas (*False Discovery Rate – FDR*). Embora essas duas correções sejam altamente recomendadas para a análise dos *datasets* de metabolômica, ambas as abordagens foram muito conservadoras para o conjunto de dados deste estudo. Após a aplicação desses métodos de correção, todos os 72 compostos detectados como diferencialmente abundantes ($p < 0.05$), foram considerados falsos positivos. Por isso, o critério de *fold change*, abordagem que representa a magnitude da diferença entre duas populações de estudo (VINAIXA et al., 2012), foi aplicado para detecção das variáveis mais ou menos abundantes no grupo de leucoplasias transformadas em comparação com o grupo de leucoplasias não transformadas. Das 72 variáveis estatisticamente significativas de acordo com os testes de hipótese, 3 foram mais abundantes (M275T1120, M669T1120 e M326T756 – *fold change* > 2) e 10 foram menos abundantes no grupo de leucoplasias transformadas (M308T477, M478T701, M476T636, M479T701, M480T704, M501T644, M480T780, M206T275, M478T735 e M782T592 – *fold change* < -2). A Tabela 2 exhibe os 72 *molecular features* discriminantes de acordo com as análises estatísticas univariadas.

Modo de ionização	Molecular feature	p. valor	m/z*	R.T.**	t.score	fold change
Positivo	M310T1120	0.002	310.3085	1119.5500	17.6269	0.9975
Positivo	M311T1120	0.002	311.3100	1119.5500	9.6889	0.7482
Positivo	M308T477	0.003	308.2998	477.2765	0.2443	-2.6412
Positivo	M275T1120	0.003	275.2658	1119.5550	5.7776	2.8427
Positivo	M669T1120	0.004	668.6128	1119.5500	4.6028	3.1981
Positivo	M325T444	0.005	325.2991	444.0415	1.1243	-0.0317
Positivo	M736T989	0.005	735.5581	988.9240	6.2511	0.6254
Positivo	M280T548	0.005	280.2590	547.9390	0.2256	-0.9180
Positivo	M380T645	0.006	380.3617	645.2285	10.7906	0.6564
Positivo	M293T1120	0.006	293.2782	1119.5400	1.2883	0.4734
Positivo	M324T445	0.007	324.2988	444.8560	3.2590	-0.3550
Positivo	M482T783	0.007	482.3237	783.1750	2.2905	-0.6257
Positivo	M376T671	0.007	376.3650	671.1120	4.9782	0.8245
Positivo	M310T389	0.009	310.2831	388.7310	1.9445	0.6923
Positivo	M266T287	0.010	266.2543	287.2615	1.2443	1.4254
Positivo	M210T265	0.011	210.1813	264.6640	1.2399	-0.0433
Negativo	M478T701	0.013	478.3126	700.5520	33.4965	-3.7919
Positivo	M309T496	0.013	309.2839	496.3915	0.8746	1.0234
Negativo	M476T636	0.014	476.2885	635.6680	29.0667	-3.5066
Positivo	M345T218_2	0.015	345.2616	218.2370	5.3014	-0.0925
Negativo	M479T701	0.015	479.3137	700.8190	53.7392	-2.3204
Positivo	M326T756	0.017	326.3750	755.7100	4.1682	2.2087
Positivo	M369T695	0.018	369.1826	694.7750	3.3534	0.3862
Positivo	M296T305	0.019	296.2665	305.4040	1.5042	0.2051
Positivo	M420T775	0.019	420.3906	774.9380	5.1578	0.3628
Positivo	M502T704	0.019	502.2898	704.4380	3.0951	0.3084
Positivo	M393T631	0.020	393.3643	631.4805	10.4944	1.7841
Positivo	M409T789	0.020	409.3920	788.8010	3.5241	0.0040
Positivo	M480T704	0.021	480.3064	703.6470	3.8792	-3.7299
Positivo	M350T635	0.022	350.3509	634.9625	8.2003	1.6220
Positivo	M503T644	0.023	503.2942	643.5050	3.0356	0.9974
Positivo	M364T691	0.024	364.3638	690.7100	2.1680	0.6860
Positivo	M610T254	0.024	610.4616	253.8030	1.2310	0.5331
Positivo	M524T740	0.025	524.2988	740.4405	3.3033	-1.6675
Positivo	M282T265	0.025	282.2520	264.5460	2.7073	0.6929
Positivo	M387T209_1	0.025	386.5084	209.3025	5.8376	1.0153
Positivo	M381T647	0.026	381.3632	646.6430	7.3122	0.6845
Positivo	M460T218_1	0.026	459.6833	218.1180	19.9925	0.8639
Negativo	M501T644	0.027	501.3004	643.6890	33.0222	-2.4458
Positivo	M309T152_2	0.027	308.9015	152.4020	0.9669	-0.0757
Positivo	M502T644	0.028	502.2917	643.5500	4.8635	0.4163
Positivo	M351T636	0.029	351.3508	636.0410	4.6472	1.3802
Positivo	M690T218	0.030	689.5207	218.0730	6.2525	-0.5128
Positivo	M577T195	0.030	576.9360	194.8930	6.9245	-1.2395
Positivo	M418T665	0.031	418.3759	664.9250	2.1324	-0.8136
Positivo	M492T652	0.032	492.3066	651.5110	1.9702	-0.4296

Positivo	M396T736	0.032	396.3782	736.3315	5.6475	0.7416
Positivo	M259T456	0.033	259.1477	455.9705	0.6856	0.9075
Positivo	M311T389	0.034	311.2827	388.5540	0.8737	-0.1490
Positivo	M392T195_1	0.034	391.6209	194.8925	5.5446	1.4890
Positivo	M526T834	0.035	526.3136	834.2330	3.0977	-0.2602
Positivo	M576T195_1	0.035	575.9350	194.8930	14.8336	-0.1413
Positivo	M339T497	0.035	339.3171	496.6480	1.7990	1.1244
Negativo	M480T780	0.036	480.3277	780.0165	34.3858	-2.4821
Positivo	M781T592	0.037	780.5482	592.2190	4.4063	-0.0800
Positivo	M609T235	0.038	609.4634	235.0100	8.9823	1.3298
Positivo	M467T218_1	0.039	466.6757	218.0730	0.7063	0.4978
Positivo	M392T631	0.039	392.3615	631.1130	13.5795	-1.4551
Positivo	M388T624	0.040	388.3270	624.3700	3.6514	-0.3547
Positivo	M381T837	0.042	381.2140	836.6940	1.1463	0.3809
Positivo	M700T217	0.043	699.5070	217.3770	1.5953	-0.2661
Positivo	M253T265	0.043	253.2226	264.5345	0.2982	0.4298
Positivo	M206T275	0.044	206.1877	275.3175	0.3928	-2.4370
Positivo	M568T740	0.045	568.2678	739.6785	0.8473	0.4830
Positivo	M478T735	0.046	478.3197	734.8950	3.7517	-2.1069
Positivo	M494T722	0.046	494.3220	722.4015	2.8542	1.4696
Positivo	M782T592	0.047	781.5506	592.2220	3.9697	-3.7269
Positivo	M562T805	0.047	562.3285	804.6325	0.4824	1.2105
Positivo	M689T217	0.048	689.0179	217.4750	0.0635	0.0996
Positivo	M500T636	0.048	500.2723	635.9920	0.0804	0.3373
Positivo	M481T704	0.049	481.3066	703.6315	0.4941	1.0075
Positivo	M689T218	0.049	688.5173	217.6900	1.6277	0.7860

Tabela 2. Compostos diferencialmente abundantes classificados de acordo com a ordem crescente dos valores de p . * m/z : relação massa/carga. ** $R.T$: tempo de retenção. **Valores em negrito:** compostos diferencialmente abundantes de acordo com o cálculo de *fold change*.

O agrupamento hierárquico não supervisionado dos 72 compostos discriminantes ($p < 0.05$) mostrou o agrupamento das amostras em dois *clusters* distintos. No dendograma, o primeiro *cluster* formado incluiu todas as amostras de leucoplasia progressivas (T1 – T5) e 4/15 amostras de leucoplasias não progressivas (C2, C4, C11, C15). As demais amostras de leucoplasias não progressivas (11/15) foram agrupadas no segundo *cluster* (FIGURA 8). Não houve associação entre as características clínicas e o agrupamento encontrado.

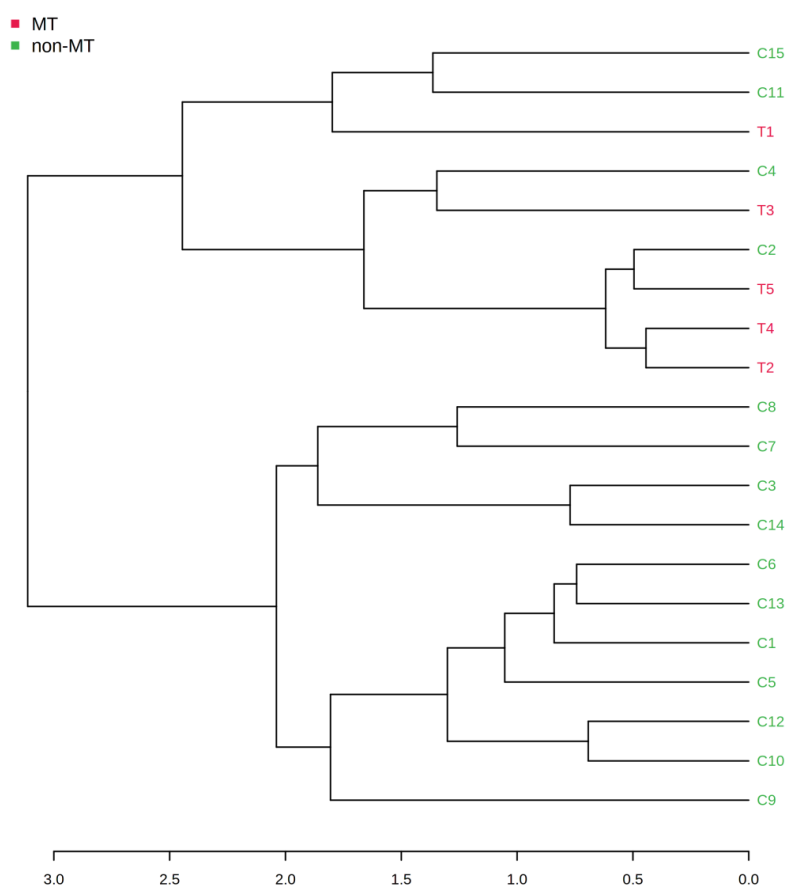


Figura 8. Agrupamento hierárquico não supervisionado exibido como dendograma. Medida de distância: Pearson. Algoritmo de agrupamento: Média. *MT*: *malignant transformed*. *Non-MT*: *non-malignant transformed*.

Fonte: Imagem gerada pela plataforma MetaboAnalyst

5.4.2 Análises estatísticas multivariadas

A análise por PCA foi realizada para avaliar o padrão de distribuição dos compostos dentro de cada grupo do estudo. Apesar de uma sobreposição entre os grupos ter sido observada, a distribuição das amostras ocorreu dentro do intervalo de confiança estabelecido para cada grupo (95%), o que indica a ausência de *outliers* (APÊNDICE E). Para distinguir mais detalhadamente as diferenças entre as amostras de leucoplasia transformadas e não transformadas, o PLS-DA foi utilizado. O modelo PLS-DA apresentou valores de Q^2 negativos nos testes de validação cruzada (*10-fold cross validation*) para separação dos dados em sua primeira componente, tanto para os dados obtidos em modo de ionização positivo como para os dados obtidos em modo de ionização negativo (APÊNDICE E).

Valores baixos de Q^2 podem indicar uma tendência de *overfitting* (CHONG; WISHART; XIA, 2019). Curiosamente, uma tendência de agrupamento foi observada ao realizar a análise por PLS-DA utilizando como *input* todo o conjunto de *molecular features* e subdividindo as amostras em grupos de acordo com o grau de displasia epitelial oral observado (FIGURA 9).

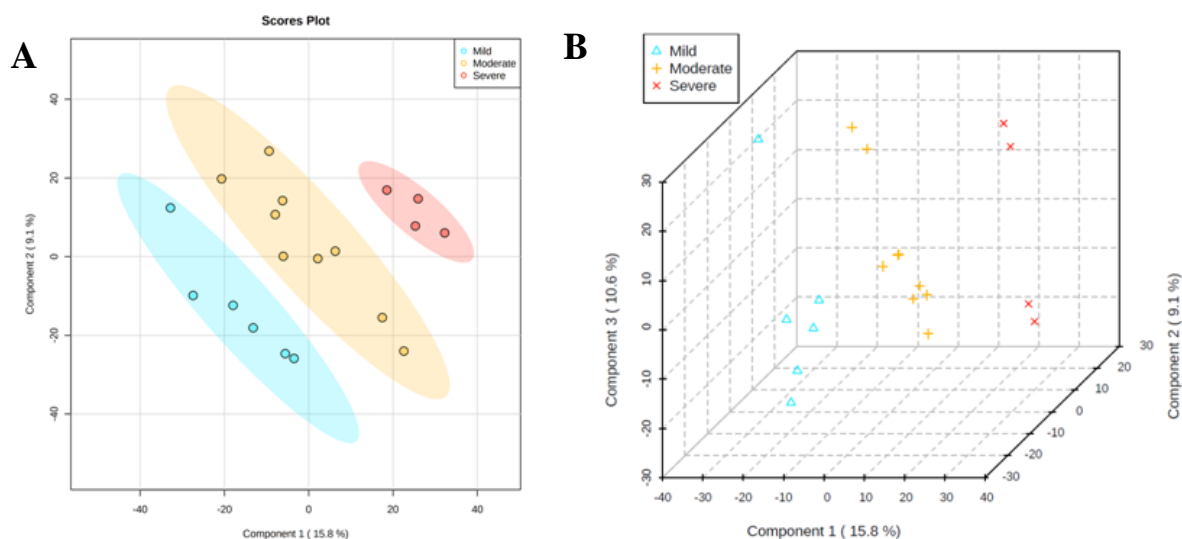


Figura 9. Análise discriminante de mínimos quadrados parciais – PLS-DA. Gráfico de escores 2D (A) e 3D (B) entre os componentes principais 1 e 2 e 1, 2 e 3, respectivamente. A variância explicada por cada componente está descrita entre parênteses. As áreas em vermelho, amarelo e azul possuem 95% de confiança. Os gráficos mostram uma boa separação visual entre as leucoplasias com diferentes graus de displasia. O valor de Q^2 é de 0.5. *Mild*: displasia epitelial oral leve. *Moderate*: displasia epitelial oral moderada. *Severe*: displasia epitelial oral severa.

A análise SAM, abordagem primeiramente descrita para a identificação de genes diferencialmente expressos (TUSHER; TIBSHIRANI; CHU, 2001), foi realizada para a detecção dos compostos mais importantes para a discriminação dos grupos deste estudo a partir de um método multivariado. SAM identifica os compostos com diferenças estatisticamente significativas na abundância, assimilando um conjunto de testes t não pareados e definindo uma pontuação específica para cada composto levando em conta a mudança na sua abundância em relação ao desvio padrão das medições repetidas para aquele composto. Na análise, é possível controlar a taxa de descobertas falsas (FDR) a partir da permutação das medições. Neste estudo, foram detectados 18 e 3 compostos discriminantes ($p < 0.05$), para os dados obtidos em modo de ionização positivo e negativo, respectivamente, com $FDR = 0.2$ para ambos os conjuntos de dados (FIGURA 10).

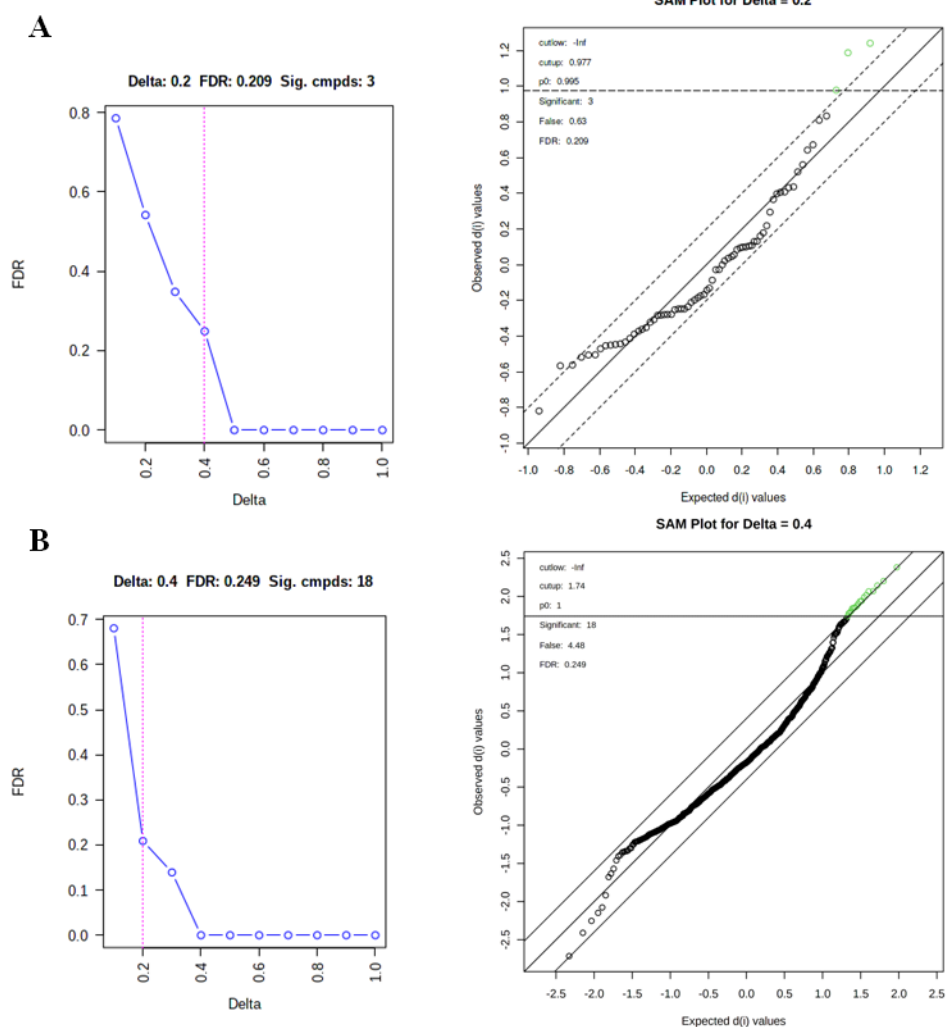


Figura 10. Análise SAM. Os gráficos à esquerda representam o controle de descobertas falsas por meio do ajuste do valor de Delta. Os gráficos de dispersão à direita exibem a diferença relativa observada (eixo y) em relação à diferença relativa esperada pela permutação dos dados (eixo x). A linha diagonal central indica quando essas duas medidas foram iguais. As linhas diagonais externas foram traçadas à distância Delta da linha central. Os pontos verdes correspondem aos compostos discriminantes. **(A)** Dados obtidos em modo de ionização negativo. **(B)** Dados obtidos em modo de ionização positivo.

Fonte: Capturas de tela obtidas da Plataforma MetaboAnalyst

5.4.3 Seleção recursiva de variáveis – MUVR

O objetivo da seleção de variáveis é identificar um subconjunto de compostos dentro de todo o conjunto de dados que sejam relevantes ao estudo. No entanto, os métodos tradicionais para a seleção de variáveis podem sofrer vieses de seleção (KRAWCZUK; ŁUKASZUK, 2016). Para as análises univariadas, o principal gargalo estatístico é o surgimento de erros do tipo I (IOANNIDIS, 2005;

VINAIXA et al., 2012), enquanto as análises multivariadas tem o sobreajuste como a principal limitação para a construção de um modelo confiável (KRAWCZUK; ŁUKASZUK, 2016). O desempenho dos modelos multivariados é estimado por métodos de validação cruzada. O sobreajuste do modelo ocorre quando a seleção de variáveis considera todas as amostras para estimar o erro de predição nesse esquema de validação cruzada (CHRISTIN et al., 2013). No intuito de superar essas limitações, Shi e colaboradores (2019) desenvolveram o MUVR, um algoritmo de modelagem multivariada particularmente útil para dados onde o número de variáveis é maior do que número de observações, assim como os dados do presente estudo (SHI et al., 2019).

O método é baseado no esquema de validação por rdCV, método recursivo que realiza validações internas repetidas do poder preditivo do modelo e que demonstrou previamente estimativas mais realísticas sobre os erros de predição do que as estratégias de validação rotineiramente utilizadas como *k-fold* e *leave-n-out* (FILZMOSEER; LIEBMANN; VARMUZA, 2009; SHI et al., 2019). Uma vez que no presente estudo não foi possível a aplicação de correções pós-teste após as análises univariadas e diante do sobreajuste apresentado pelo modelo PLS-DA exibido no item 5.4.2, o algoritmo MUVR foi utilizado como uma avaliação adicional para a determinação dos compostos discriminantes.

Para os dados obtidos em modo de ionização positivo, a análise classificou como variáveis preditoras 8 compostos pelo método RF e 29 pelo método PLS. Já para os dados obtidos em modo de ionização negativo, foram classificados 5 compostos pelo método RF e 11 pelo método PLS. Os modelos gerados para a classificação dos compostos obtidos em modo de ionização negativo apresentaram valor preditivo estatisticamente significativo de acordo com a análise de permutação. É válido ressaltar que todos os compostos classificados como discriminantes pelo método MUVR também se mostraram estatisticamente significativos para a distinção dos grupos nas análises univariadas, o que suporta o uso das abordagens estatísticas convencionais para dados de metabolômica. O resultado da classificação MUVR pode ser observado na Tabela 4 do APÊNDICE C.

Dezessete *molecular features* foram detectados como discriminantes de modo correspondente entre as análises univariadas, a análise SAM e os modelos MUVR (Tabela 3). Dessa forma, a consistência dos achados estatísticos pôde ser comprovada e os compostos foram selecionados para identificação putativa.

<i>Molecular feature</i>	SAM (p.valor)	SAM (q.valor)	p.valor**	Fold change	MUVR (PLS)	MUVR (RF)	Modo de ionização
M669T1120*	0.000	0.1	0.004	3.2	0.08	0.03*	Positivo
M275T1120*	0.002	0.4	0.003	2.8	0.08	0.03*	Positivo
M736T989*	0.002	0.4	0.005	0.6	0.08	0.03*	Positivo
M376T671	0.003	0.4	0.007	0.8	0.08	-	Positivo
M324T445	0.004	0.4	0.007	-0.4	0.08	-	Positivo
M380T645	0.005	0.4	0.006	0.7	0.08	-	Positivo
M476T636*	0.005	0.2	0.014	-3.5	0.003*	0.002*	Negativo
M325T444	0.007	0.5	0.005	0.0	0.08	-	Positivo
M308T477	0.007	0.5	0.003	-2.6	0.08	-	Positivo
M478T701*	0.007	0.5	0.013	-3.8	0.03*	0.002*	Negativo
M310T389	0.012	0.5	0.009	0.7	0.08	-	Positivo
M420T775*	0.013	0.5	0.019	0.4	0.08	0.03*	Positivo
M280T548	0.013	0.5	0.005	-0.9	0.08	-	Positivo
M309T496	0.016	0.5	0.013	1.0	0.08	-	Positivo
M502T704	0.016	0.5	0.019	0.3	0.08	-	Positivo
M480T704	0.016	0.5	0.021	-3.7	0.08	-	Positivo
M479T701*	0.025	0.5	0.015	-2.3	0.03*	0.002*	Negativo

Tabela 3. Compostos selecionados a partir do critério de correspondência entre as abordagens estatísticas univariadas, análise SAM e classificação MUVR.

* Modelagem MUVR estatisticamente significativa. $nPerm = 500 - p < 0.05$

** Análise estatística univariada - $p < 0.05$

Os 17 *molecular features* selecionados foram utilizados para a construção de um mapa de calor (FIGURA 11). Assim como no dendograma, todas as leucoplasias progressivas foram clusterizadas no mesmo grupo. Apenas as amostras C2, C4 e C11 se agruparam fora do grupo de leucoplasias não transformadas, em conjunto com as leucoplasias transformadas.

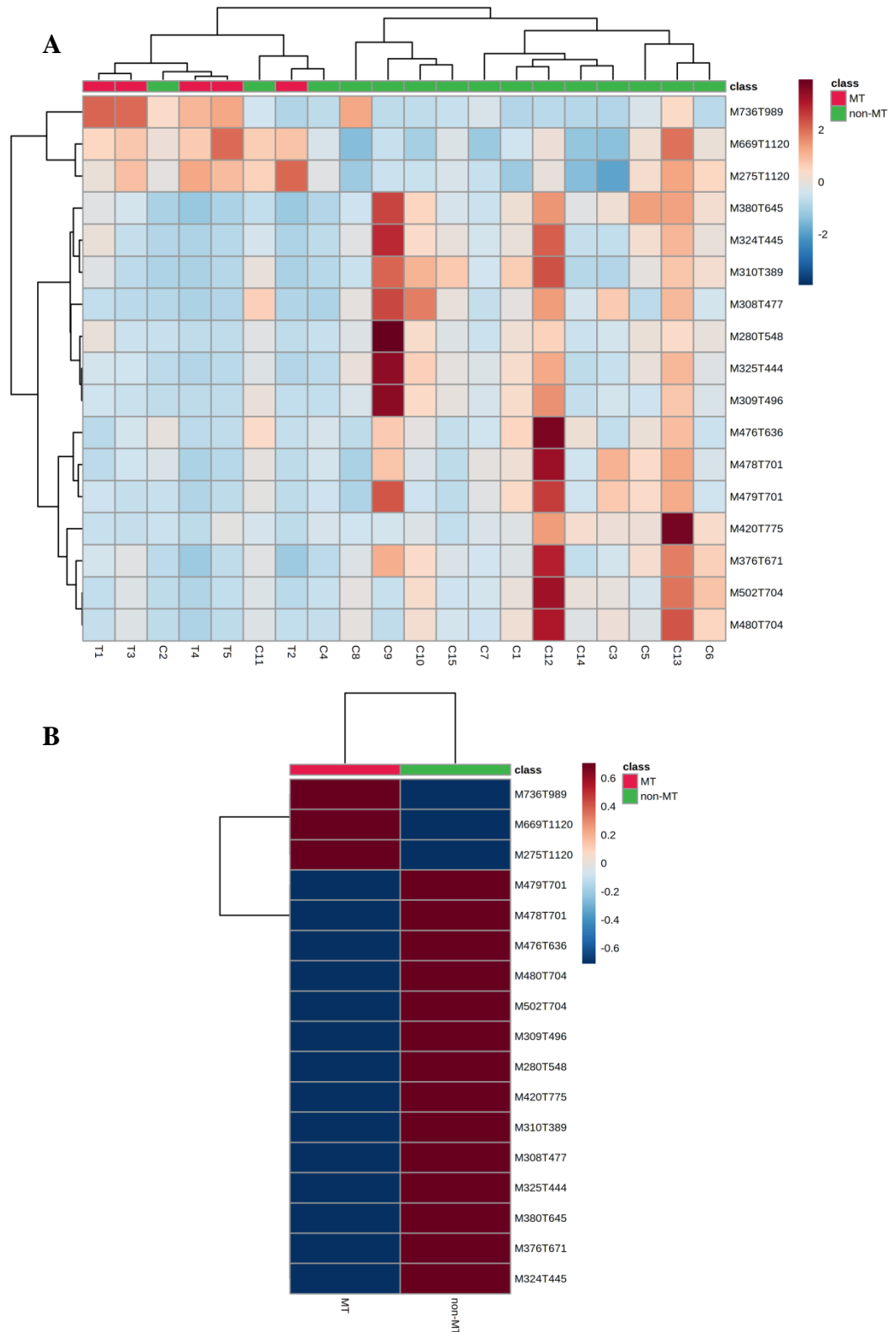


Figura 11. Agrupamento hierárquico não supervisionado exibido como mapa de calor. **A.** O padrão de distribuição das amostras de acordo com a abundância relativa dos compostos selecionados pode ser observado. As amostras de leucoplasias não progressivas C2, C4, e C11 foram clusterizadas junto ao grupo de leucoplasias transformadas. Apesar disso, em uma análise geral, os grupos são visualmente distintos. Note que a abundância representada para os compostos M669T1120, M275T1120, M308T477, M476T636, M478T701 e M479T701 foi correspondente aos seus valores de *fold change* (Tabela 3). **B.** Visão geral da abundância relativa dos compostos selecionados de acordo com os grupos estudados. Medida de distância: Pearson. Algoritmo de

agrupamento: Média. Thresholds: 2 – mais abundante; -2 – menos abundante. MT: malignant transformed. Non-MT: non-malignant transformed.

Os resultados das análises estatísticas sugerem que existem diferenças no perfil metabolômico das amostras de leucoplasia transformadas e não transformadas. A lista de compostos selecionados de modo correspondente por métodos distintos e robustos de análise estatística servirá como base confiável para identificação putativa dos metabólitos equivalentes a estes *molecular features*. Com isso, a definição de um painel de biomarcadores preditores de transformação maligna poderá ser realizada em estudos futuros.

A análise metabolômica tem se mostrado eficiente para a distinção molecular entre o tecido normal, a leucoplasia e o carcinoma de células escamosas (GUPTA; GUPTA; MAHDI, 2015; KONG et al., 2015; SIVADASAN et al., 2020; SONG et al., 2020; SRIDHARAN et al., 2019; SRIDHARAN; RAMANI; PATANKAR, 2017; SRIVASTAVA et al., 2011; WEI et al., 2011; XIE et al., 2012; ZHOU et al., 2009). Da mesma forma, a abordagem parece eficaz para a distinção entre as lesões que sofreram transformação maligna e as que mantiveram o seu comportamento benigno. Esses achados podem representar um avanço promissor para o diagnóstico precoce ou até mesmo para a prevenção do câncer de boca.

5.5 Performance dos *molecular features* selecionados para a discriminação dos grupos em estudo

Uma das vantagens principais das análises ômicas é a possibilidade de definir um conjunto de biomarcadores ábeis a prever ou diagnosticar doenças. Fundamentalmente, o objetivo do desenvolvimento de biomarcadores em metabolômica é criar um modelo preditivo a partir de uma coleção de vários compostos, que podem ser usados para classificar novas amostras em grupos específicos, por exemplo, saudável ou doente, com adequada sensibilidade e especificidade (XIA et al., 2013). Embora os dados obtidos no presente estudo não sejam suficientes para a definição de biomarcadores específicos para a transformação maligna da leucoplasia oral, o raciocínio que permeia a definição de

biomarcadores pode avaliar de forma ainda mais detalhada o poder discriminatório dos *molecular features* selecionados a partir dos métodos descritos anteriormente.

Segundo o ponto de vista analítico, o estudo de biomarcadores envolve três passos, seleção de biomarcadores, avaliação da performance, e desenvolvimento do modelo. A seleção de biomarcadores é o primeiro passo e consiste na identificação de um subconjunto de *molecular features* com poder discriminatório máximo entre os grupos em estudo (XIA et al., 2013). No presente trabalho, 17 compostos discriminantes foram selecionados a partir da correlação entre distintos e robustos métodos de análise estatística.

A avaliação da performance envolve a análise e validação do painel de compostos selecionados pelo primeiro passo. Neste contexto, a curva ROC clássica foi utilizada para avaliar a acurácia de cada um dos 17 compostos para a classificação das leucoplasias orais transformadas e não transformadas. De acordo com Xia e colaboradores (2013), a capacidade de predição dos compostos dos modelos baseados em ROC deve ser avaliada de acordo com os valores de AUC (XIA et al., 2013). Sendo, AUC = 0,9 – 1 excelentes preditores; AUC = 0,8 – 0,9 bons; AUC = 0,7 – 0,8 razoáveis; AUC = 0,6 – 0,7 fracos; e AUC = 0,5 – 0,6 falhos. Três dos dezessete compostos foram detectados como excelentes preditores (AUC > 0,9); um foi considerado bom (AUC > 0,8); seis foram razoáveis (AUC > 0,7); e seis foram fracos discriminadores (AUC > 0,6) (Tabela 4). Nenhum dos 17 compostos foi considerado falho para a discriminação entre as leucoplasias orais transformadas e não transformadas. A Figura 12 A – C exhibe os resultados da análise ROC para os três compostos detectados como excelentes preditores.

<i>Molecular feature</i>	AUC	<i>p.</i> valor
M669T1120	0.96	0
M275T1120	0.96	0
M736T989	0.95	0
M324T445	0.85	0.01
M376T671	0.75	0.04
M325T444	0.72	0.08
M381T647	0.72	0.15
M310T389	0.72	0.15
M309T496	0.72	0.23
M308T477	0.71	0.15
M420T775	0.69	0.23
M476T636	0.68	0.22
M478T701	0.68	0.46
M502T704	0.67	0.26
M280T548	0.64	0.44
M479T701	0.63	0.33
M480T704	0.61	0.99
M380T645	0.6	0.08

Tabela 4. Valores de AUC e *p.* para análise ROC clássica.

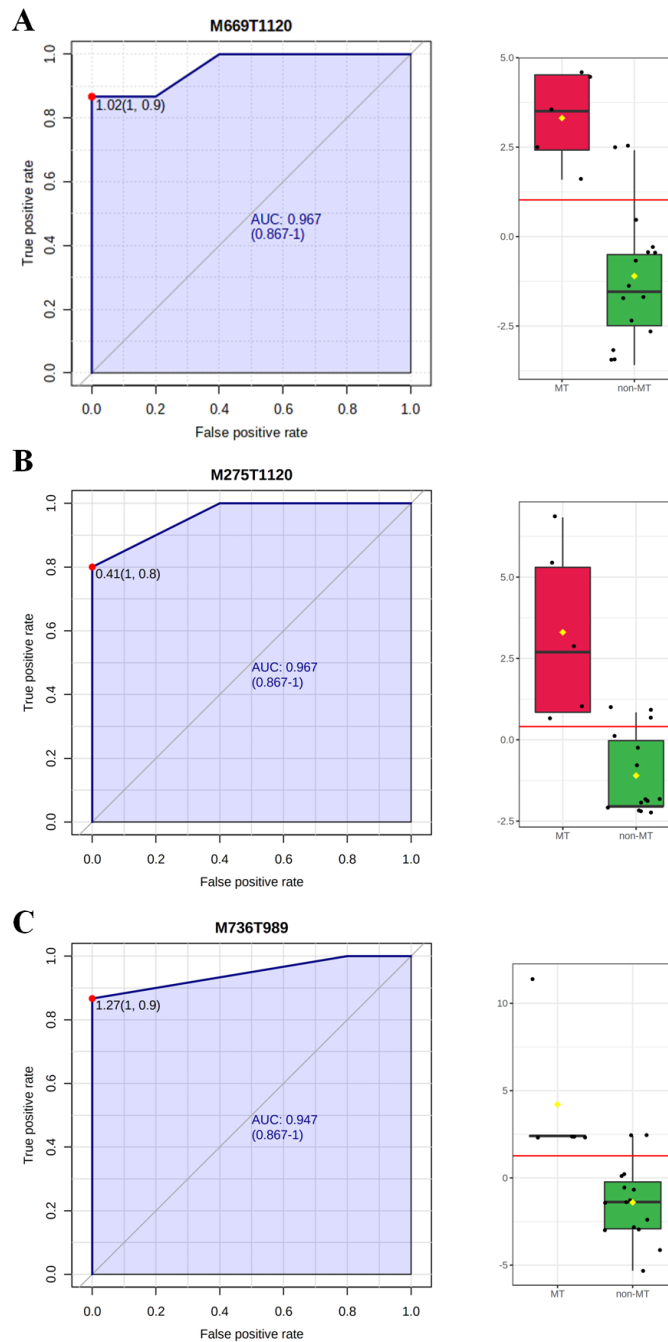


Figura 12. Os gráficos à esquerda exibem a curva ROC para cada composto. A sensibilidade é dada pelo eixo Y, e a especificidade, pelo eixo X. O valor de AUC está destacado em azul. O ponto vermelho representa o ponto de corte ótimo e é caracterizado pela mínima distância até o limite do eixo Y e a máxima distância vertical até a curva diagonal. Os *boxplot* à direita representam a variação da concentração do *molecular feature* selecionado entre os grupos em estudo. A caixa vermelha representa o grupo de leucoplasias transformadas, e a caixa verde, o grupo de leucoplasias não transformadas. A linha horizontal vermelha indica o ponto de corte ótimo. **A.** *Molecular feature*: M669T1120. **B.** *Molecular feature*: M275T1120. **C.** *Molecular feature*: M736T989.

A terceira etapa é constituída pelo desenvolvimento de um modelo final baseado na análise multivariada das curvas ROC criadas na etapa anterior. O painel de compostos selecionados é combinado em um único teste por meio de um algoritmo

multivariado que, basicamente, constrói as curvas ROC aplicando o método de validação cruzada de Monte-Carlo (MCCV), usando subamostragem balanceada. Em cada MCCV, 2/3 das amostras são usadas para avaliar importância do composto e os 1/3 restantes são usados para validar os modelos criados com a primeira etapa. De forma semelhante à descrita para o algoritmo MUVR, o processo é repetido várias vezes para calcular o desempenho e os intervalos de confiança de cada modelo. O valor de predição dos modelos criados no presente estudo variou entre $AUC = 0,79$ e $AUC = 0,89$. O quarto dos seis modelos criados exibiu $AUC = 0,88$ e acurácia de 78% para predição dos grupos em estudo, sendo considerado o melhor modelo preditor (Figura 13 A e B). O potencial do modelo 4 para a predição dos grupos de leucoplasias orais transformadas e não transformadas foi exibido no gráfico de dispersão ilustrado na Figura 14.

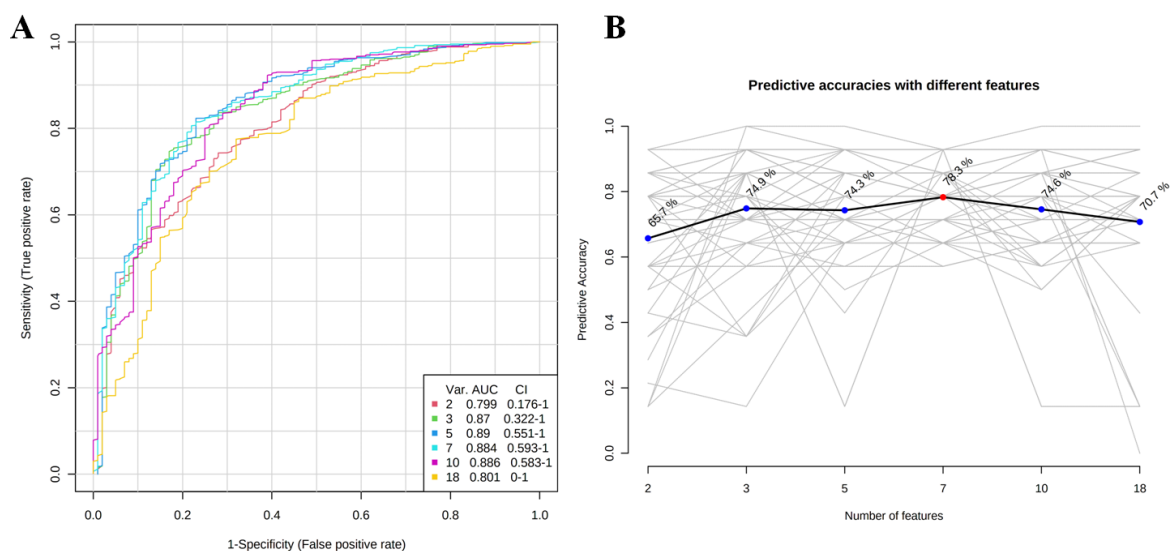


Figura 13. **A.** Gráfico de curvas ROC para os modelos multivariados construídos com base no desempenho médio em todas as execuções MCCV. A caixa no canto inferior direito descreve as cores de cada linha ROC, os valores de AUC, o intervalo de confiança (CI), e a quantidade de compostos incluídos em cada modelo gerado. **B.** Gráfico da precisão preditiva dos modelos multivariados com um número crescente de *molecular features*. O ponto vermelho destaca o modelo mais preciso (Modelo 4 – 7 *molecular features*; $AUC = 0,88$; Acurácia 78,3%).

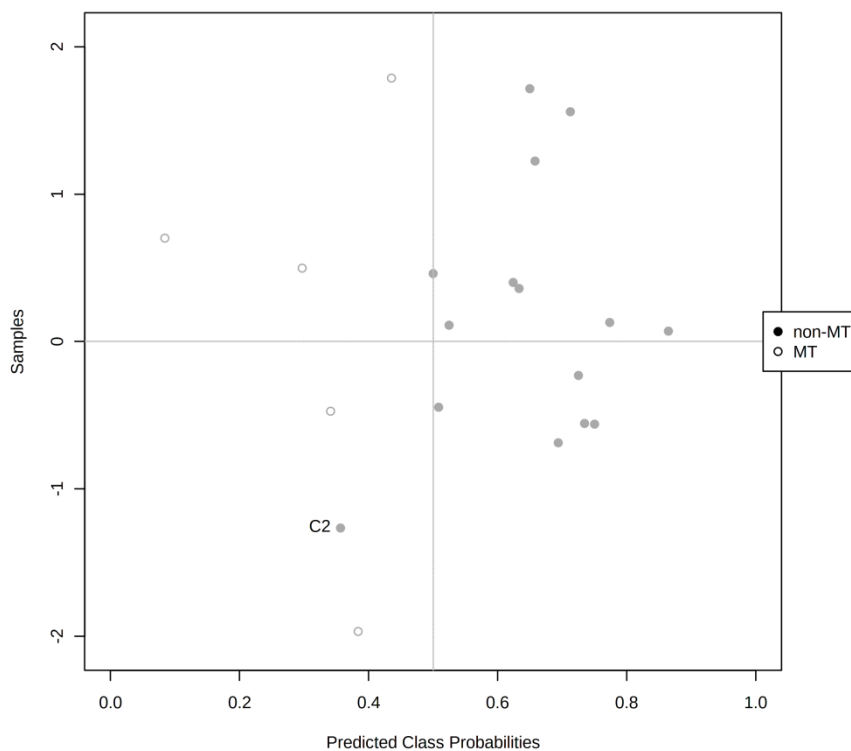


Figura 14. Gráfico de dispersão exibindo as probabilidades de classe previstas para todas as amostras usando o modelo 4. Devido à subamostragem equilibrada, o limite de classificação está no centro ($x = 0,5$ – linha pontilhada). Os pontos vazios representam as amostras de leucoplasia oral transformadas, e os pontos preenchidos, as leucoplasias orais não transformadas. Uma das 15 amostras de leucoplasia oral não transformada (C2) foi erroneamente classificada.

5.6 Análise de enriquecimento de vias metabólicas

Um dos maiores desafios associados aos estudos por metabolômica *untargeted* é a realização de análises funcionais *downstream*. Isso porque a identificação dos compostos em metabólitos é uma etapa complexa usualmente requerida para o estudo das vias de sinalização envolvidas no contexto da doença ou da condição estudada. No entanto, dificilmente um metabólito produz um único pico cromatográfico, mesmo quando instrumentos de alta acurácia são utilizados para a aquisição dos dados (KIND; FIEHN, 2007). Além disso, existem diversos metabólitos ainda não conhecidos que podem apresentar valor de razão massa carga idênticos ao de alguma entidade já identificada (CHONG; WISHART; XIA, 2019). Esses fatores tornam o processo de identificação dos metabólitos extremamente dificultado e laborioso, o que exige do pesquisador um amplo conhecimento sobre química e bioquímica.

No intuito de superar as limitações inerentes ao processo de identificação dos metabólitos, Li e colaboradores (2013), desenvolveram o *software mummichog*. Essa ferramenta baseia-se no uso de um conjunto de algoritmos computacionais que utilizam o poder coletivo das vias metabólicas e de suas redes de interação, para prever a repercussão das alterações metabólicas a um nível funcional diretamente a partir de tabelas de dados espectrais, sem a necessidade da identificação *a priori* dos metabólitos (LI et al., 2013). O algoritmo *mummichog* ganhou uma interface gráfica na plataforma *online* MetaboAnalyst, tornando a análise acessível aos pesquisadores de diversas áreas do conhecimento (CHONG et al., 2018). A ferramenta está disponível no módulo “*Functional Analysis*” e o seu uso pode ser associado ao tradicional método de enriquecimento de conjuntos de genes (*Gene Set Enrichment Analysis – GSEA*) (SUBRAMANIAN et al., 2005).

A análise pelo algoritmo *mummichog* é baseada em super-representação (*Over Representation Analysis – ORA*), na qual o enriquecimento das vias é testado considerando os picos com valores estatisticamente significativos em comparação com modelos nulos baseados em uma coleção de picos do mesmo tamanho que são aleatoriamente retiradas da lista de picos de referência. Já a análise por GSEA avalia as diferenças gerais de duas distribuições com base em testes de Kolmogorov-Smirnov. A base de conhecimento interrogada pelo módulo é formada por dados manualmente curados das fontes KEGG, BiGG e Edinburgh (CHONG; WISHART; XIA, 2019).

Tinnevelt e colegas (2020) compararam os métodos quimiométricos para a seleção de variáveis importantes em conjuntos de dados de metabolômica *untargeted*. Os autores utilizaram uma coorte de pacientes portadores de ataxia-telangiectasia tratados com um tipo de vitamina B3, para comparar o desempenho dos métodos esparsos e globais para a seleção de variáveis importantes. Os métodos esparsos objetivam a detecção do conjunto mínimo de variáveis capazes de discriminar os grupos em estudo, enquanto as abordagens globais visam encontrar todas as variáveis importantes para a distinção das amostras. Os métodos esparsos não selecionaram nenhum metabólito associado à vitamina B3 no grupo estudado, enquanto que a combinação entre métodos de análise univariados e multivariados com o algoritmo *mummichog* foi capaz de selecionar diversos metabólitos associados à vitamina B3 (TINNEVELT et al., 2020). Ao longo dos anos, o algoritmo *mummichog* vem sendo utilizado em estudos de diversas áreas,

proporcionando desde *insights* importantes sobre as vias metabólicas desreguladas em variados tipos de cânceres (DE ARAÚJO et al., 2019; KHAN et al., 2019; LE et al., 2018), até a avaliação do estado nutricional de bovinos na agropecuária (NOVAIS et al., 2019).

No presente estudo, a análise de enriquecimento de vias metabólicas foi realizada utilizando a combinação de dois algoritmos, *mummichog* e GSEA. Embora o equipamento usado para a aquisição dos dados tenha uma acurácia de massa de 10 ppm, as análises foram realizadas considerando toda a variação de acurácia de massa disponibilizada pelo módulo “*Functional Analysis*”, no intuito de observar a influência dos diferentes limiares de acurácia na consistência das vias enriquecidas. Dezesete vias metabólicas foram enriquecidas a uma acurácia de massa de 10ppm, quatorze vias a 5ppm, seis vias a 3ppm e quatro vias a 1ppm. (FIGURA 15 A-D).

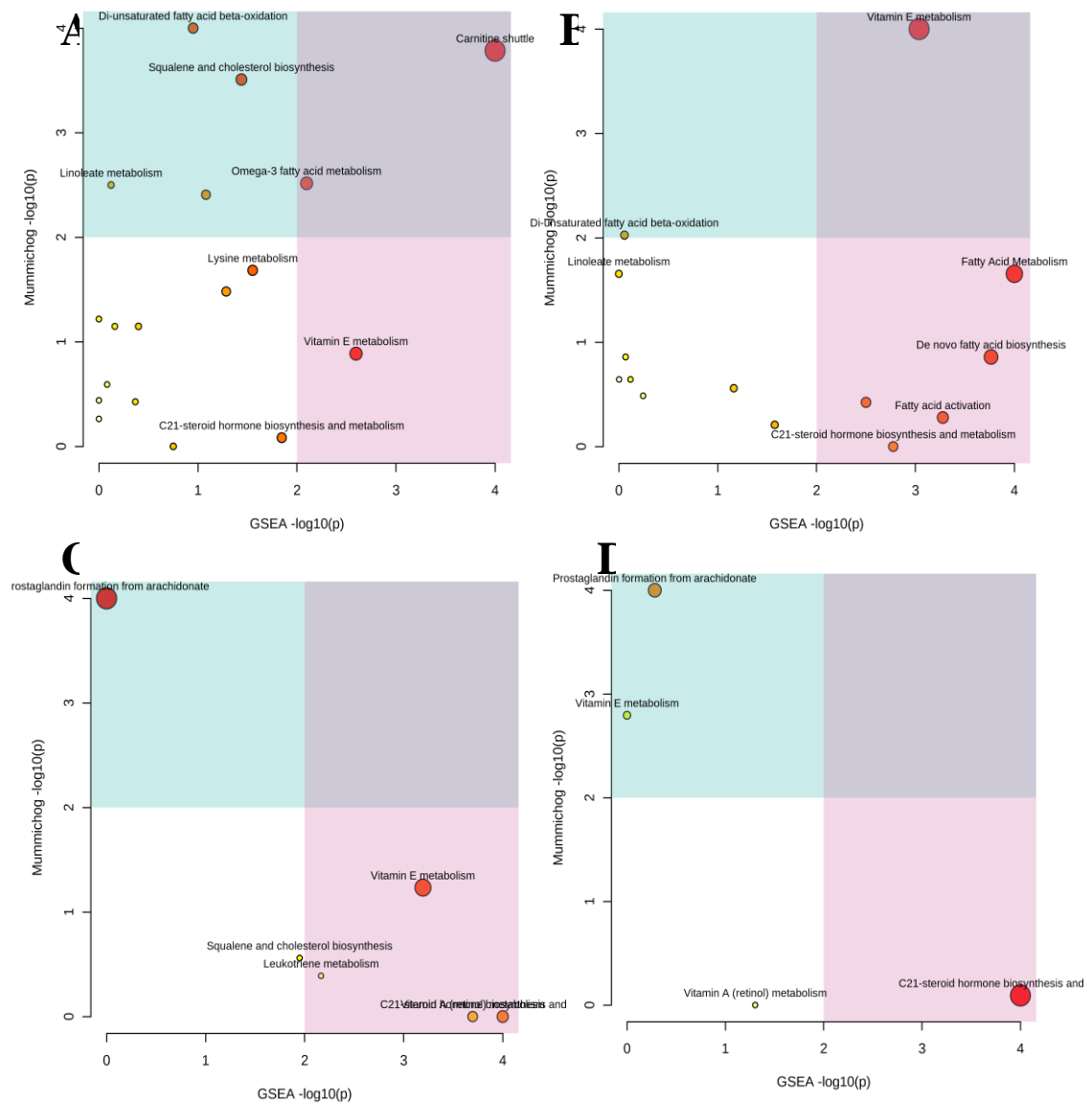


Figura 15. Gráficos integrados do módulo “*Functional Analysis*”. Os gráficos ilustram os resultados do método Fisher para combinar os valores p *mummichog* (eixo Y) e GSEA (eixo X). O tamanho e a cor dos círculos correspondem aos seus valores p combinados transformados. Círculos grandes e vermelhos são considerados os caminhos mais perturbados. As áreas em azul e rosa destacam as vias significativas baseadas em GSEA (rosa) ou *mummichog* (azul), e a área roxa destaca vias significativas identificadas por ambos os algoritmos. (A) Acurácia de massa: 10ppm. (B) Acurácia de massa: 5ppm. (C) Acurácia de massa: 3ppm. (D) Acurácia de massa: 1ppm.

Fonte: Imagem gerada pela plataforma MetaboAnalyst

As vias enriquecidas foram majoritariamente associadas ao metabolismo lipídico, especialmente à síntese e degradação de ácidos graxos, à biossíntese e metabolismo do hormônio esteroide C21, à biossíntese de esqualeno e colesterol e ao metabolismo de glicerofosfolídeos. O desequilíbrio do metabolismo lipídico é uma alteração metabólica extensivamente documentada em inúmeros tipos de cânceres, incluído o

câncer oral (DICKINSON et al., 2020; METGUD; BAJAJ, 2014; MOHANTY et al., 2021; SU et al., 2020; WISHART, 2016).

O microambiente tumoral é marcado pela baixa oxigenação e pela escassez de nutrientes, no qual a síntese das membranas celulares, a modificação de proteínas e a transmissão dos sinais oncogênicos podem depender principalmente de espécies lipídicas (HALCZY-KOWALIK et al., 2018). Nesse contexto, a reprogramação da lipogênese é possivelmente associada à proliferação e progressão tumoral. Um aumento na concentração de ácido salicílico e desse composto ligado à glicolipídeos foi observado em amostras de soro sanguíneo de pacientes com leucoplasia oral e sugerido como um potencial biomarcador para prognóstico da doença (KRISHNAN; BALASUNDARAM, 2017). Neste estudo, o metabolismo do ácido salicílico foi uma das vias enriquecidas pelo o algoritmo GSEA.

Um dos eventos que marcam o processo de carcinogênese é a peroxidação de lipídeos induzida por espécies reativas em oxigênio (METGUD; BAJAJ, 2014). Níveis aumentados de enzimas associadas aos processos de oxidação e diminuídos de espécies antioxidantes foram detectados em amostras de leucoplasia por diversos estudos (GANESAN; GAUTHAM KUMAR, 2014; METGUD; BAJAJ, 2014; SENGHORE et al., 2018; SRIVASTAVA, 2019; SRIVASTAVA; SHRIVASTAVA, 2016). A via de beta oxidação de ácidos graxos insaturados foi enriquecida pela associação dos algoritmos *mumichog* e GSEA a 10ppm e 5ppm, no presente estudo. No entanto, estes enriquecimentos não apresentaram valor estatisticamente significativo.

A via de transporte por carnitina foi significativamente enriquecida na análise combinada pelos dois algoritmos de enriquecimento utilizados neste estudo, a 10ppm ($p= 0.01$). Essa via atua na oxidação lipídica por meio do transporte dos ácidos graxos de cadeia longa para o interior da mitocôndria, onde eles são oxidados gerando ATP. A carnitina palmitoiltransferase-1 e outras enzimas relacionadas à oxidação de ácidos graxos foram superexpressas em amostras de carcinoma de células escamosas de cabeça e pescoço (SU et al., 2020). Amostras de tecido, plasma e urina de pacientes com carcinoma oral de células escamosas também apresentaram níveis elevados de carnitina. Esse metabólito foi significativamente mais abundante no tecido tumoral em comparação com o tecido saudável adjacente. Ainda, foi observado um aumento gradual

em sua concentração conforme a gravidade do estadiamento tumoral (TSAI et al., 2020). Wang e colaboradores (2014) propuseram que a colina, betaína, ácido pipercolínico e L-carnitina poderiam ser utilizadas como um painel de biomarcadores para o diagnóstico precoce do câncer oral em amostras de saliva (WANG et al., 2014).

A super-regulação de L-carnitina também foi observada em amostras de soro sanguíneo de pacientes com leucoplasia oral em relação aos indivíduos saudáveis (SRIDHARAN; RAMANI; PATANKAR, 2017). A concentração aumentada desse composto também pode estar relacionada à maior ativação do ciclo do ácido tricarboxílico por meio do acúmulo de lactato, reflexo da alta demanda energética requerida pelo tumor durante o seu desenvolvimento (SRIDHARAN et al., 2019; SRIDHARAN; RAMANI; PATANKAR, 2017). No presente estudo, a via de síntese de aspartato e asparagina, intermediários do ciclo do ácido tricarboxílico, foi enriquecida na análise combinada entre *mummichog* e GSEA a 10ppm. O ciclo do ácido tricarboxílico também foi uma das vias metabólicas enriquecidas pelo algoritmo GSEA.

O metabolismo de vitamina E foi enriquecido com valor de p estatisticamente significativo ($p= 0.04$) na análise combinada pelos dois algoritmos, a uma acurácia de massa de 5 ppm. Já foi visto que a vitamina E possui um papel protetor contra os danos oxidativos ao DNA ocasionados pelos radicais livres resultantes da peroxidação dos lipídeos de membrana (RAI et al., 2007). Kaur e colaboradores (2016) reportaram a maior concentração de enzimas relacionadas ao estresse oxidativo e níveis reduzidos das vitaminas C e E na saliva de pacientes com leucoplasia oral e com outras doenças cancerizáveis em comparação com as amostras de controles saudáveis. A diminuição desses compostos sugere que a peroxidação lipídica mediada por radicais livres, associada ao consumo de moléculas antioxidantes, pode ser um dos mecanismos patofisiológicos subjacentes ao desenvolvimento das DOPM e do carcinoma de células escamosas (KAUR; POLITIS; JACOBS, 2016; RAI et al., 2007). Ainda, o painel formado pelos compostos associados aos processos oxidativos, 8-hidroxi-2-deoxiguanosina e malondialdeído, em conjunto com as vitaminas C e E foi 80% específico e sensível para a discriminação entre as amostras de saliva entre indivíduos com DOPM e com carcinoma de células escamosas, sendo, portanto, potenciais biomarcadores para o diagnóstico precoce e não invasivo do câncer oral (KAUR; POLITIS; JACOBS, 2016).

A via do metabolismo retinoico bem como a via de síntese de prostaglandina a partir da bioconversão do ácido araquidônico foram enriquecidas de forma consistente em todos os níveis de acurácia de massa. Embora os enriquecimentos não tenham apresentado valor estatisticamente significativo, é interessante observar o possível envolvimento dessas vias com as amostras deste estudo. O efeito quimiopreventivo e terapêutico de variadas espécies vitamínicas para a leucoplasia oral e outras DOPM foi inicialmente postulado na literatura na década de 60 (KOCH, 1978; SILVERMAN; EISENBERG; RENSTRUP, 1965). A ação protetora desses compostos foi associada principalmente à inibição de moléculas pró-inflamatórias como as prostaglandinas. Essas moléculas são sintetizadas por meio das vias da ciclooxygenase e lipooxygenase no metabolismo do ácido araquidônico e desempenham um papel no crescimento, proliferação e metástase tumoral a partir da indução do lançamento de enzimas proteolíticas por células como macrófagos e fibroblastos (ELATTAR; LIN, 1991). Em uma série de estudos, Elattar e Lin relataram que as vitaminas A, C e E poderiam inibir o lançamento de ácido araquidônico e de prostaglandina e, conseqüentemente, a proliferação celular no carcinoma de células escamosas (ELATTAR; LIN, 1991, 1992, 1993, 1995).

A vitamina A modula o processo de diferenciação das células epiteliais e, em alguns estudos, mostrou-se eficaz para a supressão da carcinogênese (HONG et al., 1995; LOTAN, 1994; PUNG et al., 1988; SCHER et al., 1998). As pesquisas clínicas com o uso tópico e/ou sistêmico de retinoides e outros análogos da vitamina A em casos de leucoplasia oral ganharam força durante a década de noventa e no início dos anos 2000. Nagao e colaboradores (2000) reportaram uma associação entre a maior concentração sorológica de betacaroteno com o menor risco de desenvolvimento da leucoplasia (NAGAO et al., 2000). A remissão da leucoplasia oral com a administração de vitamina A foi reportada em alguns estudos (LEUNIG et al., 2000; SANKARANARAYANAN et al., 1997). Entretanto, relapsos da doença afetaram uma proporção aproximada de 50% dos indivíduos após a interrupção do tratamento (EPSTEIN; GORSKY, 1999; GORSKY; EPSTEIN, 2002; KLAASSEN; BRAAKHUIS, 2002; SANKARANARAYANAN et al., 1997). Todos os resultados da análise de enriquecimento de vias realizadas no presente estudo podem ser consultados no APÊNDICE F.

Embora a análise de enriquecimento realizada tenha retornado resultados condizentes com diversos processos biológicos relacionados tanto com a carcinogênese como com o câncer oral, outras análises são necessárias para a validação do envolvimento dessas vias com as amostras de leucoplasia oral analisadas neste estudo.

6. CONCLUSÃO

As leucoplasias orais que sofreram transformação maligna apresentaram um perfil metabolômico diferente daquelas que mantiveram o seu comportamento benigno. Ainda, foi observada a superativação de vias de sinalização celular associadas a processos carcinogênicos nas amostras de leucoplasias orais transformadas. A combinação de métodos de análise estatística univariada e multivariada se mostrou consistente para a identificação dos compostos diferencialmente abundantes entre os grupos estudados. Os resultados do presente estudo suportam o potencial da análise metabolômica *untargeted* para a distinção das leucoplasias orais transformadas e não transformadas. Estes resultados podem servir como base para a definição de biomarcadores de transformação maligna em estudos futuros.

7. PERSPECTIVAS

Identificar os compostos potencialmente discriminantes entre os grupos do estudo e a partir disso conduzir experimentos e análises que possibilitem a definição de um painel de biomarcadores para a triagem e estimativa do risco aumentado de desenvolvimento do câncer oral utilizando métodos não invasivos.

Obter novas amostras biológicas de casos de leucoplasia progressivas e não progressivas de modo prospectivo para possíveis estudos funcionais futuros.

REFERÊNCIAS

- AGUIRRE-URIZAR, J. M.; LAFUENTE-IBÁÑEZ DE MENDOZA, I.; WARNAKULASURIYA, S. Malignant transformation of oral leukoplakia: systematic review and meta-analysis of the last 5 years. **Oral diseases**, p. 0–3, 2021.
- ALBÓNIGA, O. E. et al. Optimization of XCMS parameters for LC–MS metabolomics: an assessment of automated versus manual tuning and its effect on the final results. **Metabolomics**, v. 16, n. 1, p. 1–12, 2020.
- AXELL, T. et al. International seminar on oral leukoplakia and associated lesions related to tobacco habits. **Community Dentistry and Oral Epidemiology**, v. 12, n. 3, p. 145–154, 1984.
- BABIUCH, K. et al. Evaluation of Proinflammatory, NF-kappaB Dependent Cytokines: IL-1 α , IL-6, IL-8, and TNF- α in Tissue Specimens and Saliva of Patients with Oral Squamous Cell Carcinoma and Oral Potentially Malignant Disorders. **Journal of Clinical Medicine**, v. 9, n. 3, p. 867, 2020.
- BAG, S. et al. NanoLC MALDI MS/MS based quantitative metabolomics reveals the alteration of membrane biogenesis in oral cancer. **RSC Advances**, v. 6, n. 67, p. 62420–62433, 2016.
- BAGAN, J. V. et al. Epidermal growth factor receptor copy number in potentially malignant oral disorders and oral squamous cell carcinoma: A short communication and preliminary study. **Journal of Oral Pathology and Medicine**, v. 41, n. 9, p. 662–666, 2012.
- BANERJEE, J.; MISHRA, N.; DHAS, Y. Metagenomics: A new horizon in cancer research. **Meta Gene**, v. 5, p. 84–89, 2015.
- BÁNÓCZY, J. Follow-up studies in oral leukoplakia. **Journal of Maxillofacial Surgery**, v. 5, n. C, p. 69–75, 1977.
- BARAN, C. A. et al. MAGE-A expression in oral and laryngeal leukoplakia predicts malignant transformation. **Modern Pathology**, p. 1068–1081, 2019.
- BARNES, S. et al. Training in metabolomics research. I. Designing the experiment, collecting and extracting samples and generating metabolomics data. **Journal of Mass Spectrometry**, v. 51, n. 7, p. 461–475, jul. 2016a.
- BARNES, S. et al. Training in metabolomics research. II. Processing and statistical analysis of metabolomics data, metabolite identification, pathway analysis, applications of metabolomics and its future. **Journal of Mass Spectrometry**, v. 51, n. 8, p. 535–548, ago. 2016b.
- BHOSALE, P. G. et al. Chromosomal Alterations and Gene Expression Changes Associated with the Progression of Leukoplakia to Advanced Gingivobuccal Cancer. **Translational Oncology**, v. 10, n. 3, p. 396–409, 2017.

BHURGRI, Y. et al. Epidemiological review of head and neck cancers in Karachi. **Asian Pacific Journal of Cancer Prevention**, v. 7, n. 2, p. 195–200, 2006.

BONGIORNO, M. et al. Malignant transformation of oral leukoplakia in a patient with dyskeratosis congenita. **Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology**, v. 124, n. 4, p. e239–e242, 2017.

BROADHURST, D. et al. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. **Metabolomics**, v. 14, n. 6, p. 0, 2018.

CAMISASCA, D. R. et al. A proteomic approach to compare saliva from individuals with and without oral leukoplakia. **Journal of Proteomics**, v. 151, p. 43–52, 2017.

CERERO-LAPIEDRA, R. et al. Proliferative verrucous leukoplakia: A proposal for diagnostic criteria. **Medicina Oral, Patología Oral y Cirugía Bucal**, v. 15, n. 6, 2010.

CERQUEIRA, J. et al. Malignant transformation of oral leukoplakia: a multicentric retrospective study in Brazilian population. **Medicina Oral Patología Oral y Cirugía Bucal**, p. 0–0, 2020.

CERVIGNE, N. K. et al. Recurrent genomic alterations in sequential progressive leukoplakia and oral cancer: Drivers of oral tumorigenesis? **Human Molecular Genetics**, v. 23, n. 10, p. 2618–2628, 2014.

CÉSAR, L. et al. CD1a + and CD207 + cells are reduced in oral submucous fibrosis and oral squamous cell carcinoma. v. 25, n. 1, 2020.

CHANG, Y. A. et al. A three–MicroRNA signature as a potential biomarker for the early detection of oral cancer. **International Journal of Molecular Sciences**, v. 19, n. 3, 2018.

CHATTOPADHYAY, I.; VERMA, M.; PANDA, M. Role of Oral Microbiome Signatures in Diagnosis and Prognosis of Oral Cancer. **Technology in cancer research & treatment**, v. 18, p. 1–19, 2019.

CHEN, R. et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. **Cell**, v. 148, n. 6, p. 1293–1307, 2012.

CHEN, X. et al. Metabolomics study reveals the potential evidence of metabolic reprogramming towards the Warburg effect in precancerous lesions. **Journal of Cancer**, v. 12, n. 5, p. 1563–1574, 2021.

CHEN, X. J. et al. Expression of programmed cell death-ligand 1 in oral squamous cell carcinoma and oral leukoplakia is associated with disease progress and CD8+ tumor-infiltrating lymphocytes. **Pathology Research and Practice**, v. 215, n. 6, p. 152418, 2019.

CHETNIK, K.; PETRICK, L.; PANDEY, G. MetaClean: a machine learning-based classifier for reduced false positive peak detection in untargeted LC–MS metabolomics data. **Metabolomics**, v. 16, n. 11, p. 117, 21 nov. 2020.

CHIANG, W. F. et al. Malignant development in patients with oral potentially malignant

disorders detected through nationwide screening: Outcomes of 5-year follow-up at a single hospital. **Head and Neck**, v. 42, n. 1, p. 67–76, 2020.

CHONG, J. et al. MetaboAnalyst 4.0 : towards more transparent and integrative metabolomics analysis. v. 46, n. May, p. 486–494, 2018.

CHONG, J.; WISHART, D. S.; XIA, J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. **Current Protocols in Bioinformatics**, v. 68, n. 1, p. 1–128, 2019.

CHRISTIN, C. et al. A critical assessment of feature selection methods for biomarker discovery in clinical proteomics. **Molecular and Cellular Proteomics**, v. 12, n. 1, p. 263–276, 2013.

CHUANG, S. L. et al. Malignant transformation to oral cancer by subtype of oral potentially malignant disorder: A prospective cohort study of Taiwanese nationwide oral cancer screening program. **Oral Oncology**, v. 87, n. July, p. 58–63, 2018.

CRAIG, A. et al. Scaling and normalization effects in NMR spectroscopic metabolomic data sets. **Analytical Chemistry**, v. 78, n. 7, p. 2262–2267, 2006.

DAVE, K.; ALI, A.; MAGALHAES, M. Increased expression of PD-1 and PD-L1 in oral lesions progressing to oral squamous cell carcinoma: a pilot study. **Scientific Reports**, v. 10, n. 1, p. 1–11, 2020.

DE ARAÚJO, R. P. et al. Defining metabolic rewiring in lung squamous cell carcinoma. **Metabolites**, v. 9, n. 3, p. 1–14, 2019.

DE LA COUR, C. D. et al. Human papillomavirus prevalence in oral potentially malignant disorders: systematic review and meta-analysis. **Oral Diseases**, p. 0–2, 2020.

DE MORAIS, E. F. et al. Prognostic value of the immunohistochemical detection of epithelial-mesenchymal transition biomarkers in oral epithelial dysplasia: A systematic review. **Medicina Oral Patologia Oral y Cirugia Bucal**, v. 25, n. 2, p. e205–e216, 2020.

DI GUIDA, R. et al. Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. **Metabolomics**, v. 12, n. 5, 2016.

DICKINSON, A. et al. Mass Spectrometry–Based Lipidomics of Oral Squamous Cell Carcinoma Tissue Reveals Aberrant Cholesterol and Glycerophospholipid Metabolism — A Pilot Study. **Translational Oncology**, v. 13, n. 10, p. 100807, 2020.

DINGES, S. S. et al. Cancer metabolomic markers in urine: evidence, techniques and recommendations. **Nature Reviews Urology**, v. 16, n. 6, p. 339–362, 2019.

DO, K. T. et al. Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. **Metabolomics**, v. 14, n. 10, p. 1–18, 2018.

- DOST, F. et al. A retrospective analysis of clinical features of oral malignant and potentially malignant disorders with and without oral epithelial dysplasia. **Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology**, v. 116, n. 6, p. 725–733, 2013.
- DUNN, W. B.; BAILEY, N. J. C.; JOHNSON, H. E. Measuring the metabolome: Current analytical technologies. **Analyst**, v. 130, n. 5, p. 606–625, 2005.
- EINHORN, J.; WERSALL, J. Incidence of oral carcinoma in patients with leukoplakia of the oral mucosa. **Cancer**, v. 20, n. 12, p. 2189–93, dez. 1967.
- ELATTAR, T. M. A.; LIN, H. S. Effect of retinoids and carotenoids on prostaglandin formation by oral squamous carcinoma cells. **Prostaglandins, Leukotrienes and Essential Fatty Acids**, v. 43, n. 3, p. 175–178, 1991.
- ELATTAR, T. M. A.; LIN, H. S. Effect of vitamin C and vitamin E on prostaglandin synthesis by fibroblasts and squamous carcinoma cells. **Prostaglandins, Leukotrienes and Essential Fatty Acids**, v. 47, n. 4, p. 253–257, 1992.
- ELATTAR, T. M. A.; LIN, H. S. Inhibition of human oral squamous carcinoma cell (SCC-25) proliferation by prostaglandin E2 and vitamin E succinate. **Journal of Oral Pathology & Medicine**, v. 22, n. 9, p. 425–427, 1993.
- ELATTAR, T. M. A.; LIN, H. S. Vitamin e succinate potentiates the inhibitory effect of prostaglandins on oral squamous carcinoma cell proliferation. **Prostaglandins, Leukotrienes and Essential Fatty Acids**, v. 52, n. 1, p. 69–73, 1995.
- EMWAS, A. H. et al. Nmr spectroscopy for metabolomics research. **Metabolites**, v. 9, n. 7, 2019.
- EPSTEIN, J. B.; GORSKY, M. Topical application of vitamin A to oral leukoplakia: A clinical case series. **Cancer**, v. 86, n. 6, p. 921–927, 1999.
- EVREN, I. et al. Annual malignant transformation rate of oral leukoplakia remains consistent: A long-term follow-up study. **Oral Oncology**, v. 110, n. July, p. 105014, 2020.
- FAQUIH, T. et al. A workflow for missing values imputation of untargeted metabolomics data. **Metabolites**, v. 10, n. 12, p. 1–23, 2020.
- FARAH, C. S. et al. Exome sequencing of oral leukoplakia and oral squamous cell carcinoma implicates DNA damage repair gene defects in malignant transformation. **Oral Oncology**, v. 96, n. April, p. 42–50, 2019.
- FERLAY, J. et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. **International Journal of Cancer**, v. 144, n. 8, p. 1941–1953, 2019.
- FERNÁNDEZ-VALLE, Á. et al. HERG1 potassium channel expression in potentially malignant disorders of the oral mucosa and prognostic relevance in oral squamous cell carcinoma. **Head & Neck**, v. 38, n. 11, p. 1672–1678, nov. 2016.

- FIEHN, O. Metabolomics by Gas Chromatography–Mass Spectrometry: Combined Targeted and Untargeted Profiling. **Current Protocols in Molecular Biology**, v. 114, n. 1, p. 232–235, abr. 2016.
- FILZMOSER, P.; LIEBMANN, B.; VARMUZA, K. Repeated double cross validation. **Journal of Chemometrics**, v. 23, n. 4, p. 160–171, 2009.
- FINE, J. D.; MELLERIO, J. E. Extracutaneous manifestations and complications of inherited epidermolysis bullosa. Part II. Other organs. **Journal of the American Academy of Dermatology**, v. 61, n. 3, p. 387–402, 2009.
- FONSECA-SILVA, T. et al. Association between histopathological features of dysplasia in oral leukoplakia and loss of heterozygosity. **Histopathology**, v. 68, n. 3, p. 456–460, 2016.
- GANDARA-VILA, P. et al. Survival study of leukoplakia malignant transformation in a region of northern Spain. **Medicina oral, patologia oral y cirugia bucal**, v. 23, n. 4, p. e413–e420, 2018.
- GANESAN, A.; GAUTHAM KUMAR, N. Assessment of lipid peroxides in multiple biofluids of leukoplakia and oral squamous cell carcinoma patients - A clinico- biochemical study. **Journal of Clinical and Diagnostic Research**, v. 8, n. 8, p. 10–13, 2014.
- GANESH, D. et al. Potentially malignant oral disorders and cancer transformation. **Anticancer Research**, v. 38, n. 6, p. 3223–3229, 2018.
- GILLIGAN, G. et al. Lichenoid Proliferative Leukoplakia, Lichenoid lesions with evolution to Proliferative Leukoplakia or a continuum of the same precancerous condition? A revised hypothesis. **Journal of Oral Pathology & Medicine**, p. 0–2, 2020.
- GOMES, C. C. et al. Inter- and intra-lesional molecular heterogeneity of oral leukoplakia. **Oral Oncology**, v. 51, n. 2, p. 178–181, 2015.
- GOMES, J. O. et al. CD1a+ and CD83+ Langerhans cells are reduced in lower lip squamous cell carcinoma. **Journal of Oral Pathology and Medicine**, v. 45, n. 6, p. 433–439, 2016.
- GOODACRE, R. et al. Proposed minimum reporting standards for data analysis in metabolomics. **Metabolomics**, v. 3, n. 3, p. 231–241, 2007.
- GORSKY, M.; EPSTEIN, J. B. The effect of retinoids on premalignant oral lesions: Focus on topical therapy. **Cancer**, v. 95, n. 6, p. 1258–1264, 2002.
- GRIGOLATO, R. et al. Leukoplakia and immunology: New chemoprevention landscapes? **International Journal of Molecular Sciences**, v. 21, n. 18, p. 1–20, 2020.
- GROSS, J. H. **Mass Spectrometry**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
- GUIMARÃES, L. M. et al. The genetic basis of oral leukoplakia and its key role in understanding oral carcinogenesis. **Journal of Oral Pathology & Medicine**, p. 0–3, 2020.
- GUPTA, A.; GUPTA, S.; MAHDI, A. A. ¹H NMR-derived serum metabolomics of leukoplakia and squamous cell carcinoma. **Clinica Chimica Acta**, v. 441, p. 47–55, 2015.

- GUTTENPLAN, J. B. et al. Effects of Black Raspberry Extract and Berry Compounds on Repair of DNA Damage and Mutagenesis Induced by Chemical and Physical Agents in Human Oral Leukoplakia and Rat Oral Fibroblasts. **Chemical Research in Toxicology**, v. 30, n. 12, p. 2159–2164, 18 dez. 2017.
- HALCZY-KOWALIK, L. et al. Fatty acids distribution and content in oral squamous cell carcinoma tissue and its adjacent microenvironment. **PLoS ONE**, v. 14, n. 6, p. 1–20, 2018.
- HALKET, J. M. et al. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. **Journal of Experimental Botany**, v. 56, n. 410, p. 219–243, 2005.
- HANSEN, L. S.; OLSON, J. A.; SILVERMAN, S. Proliferative verrucous leukoplakia. A long-term study of thirty patients. **Oral Surgery, Oral Medicine, Oral Pathology**, v. 60, n. 3, p. 285–298, 1985.
- HO, M. W. et al. The clinical determinants of malignant transformation in oral epithelial dysplasia. **Oral Oncology**, v. 48, n. 10, p. 969–976, 2012.
- HOGEWIND, W. F. C.; VAN DER KWAST, W. A. M.; VAN DER WAAL, I. Oral leukoplakia, with emphasis on malignant transformation. **Journal of Cranio-Maxillofacial Surgery**, v. 17, n. 3, p. 128–133, 1989.
- HONG, W. K. et al. Retinoid Chemoprevention of Aerodigestive Cancer: From Basic Research to the Clinic. **Clinical Cancer Research**, v. 1, n. 7, p. 677–686, 1995.
- IOANNIDIS, J. P. A. Why Most Published Research Findings Are False. **PLoS medicine**, v. 2, n. 8, 2005.
- IOCCA, O. et al. Potentially malignant disorders of the oral cavity and oral dysplasia: A systematic review and meta-analysis of malignant transformation rate by subtype. **Head and Neck**, v. 42, n. 3, p. 539–555, 2020a.
- IOCCA, O. et al. Potentially malignant disorders of the oral cavity and oral dysplasia: A systematic review and meta-analysis of malignant transformation rate by subtype. **Head and Neck**, v. 42, n. 3, p. 539–555, 2020b.
- JABER, M. A. et al. Oral epithelial dysplasia: Clinical characteristics of western European residents. **Oral Oncology**, v. 39, n. 6, p. 589–596, 2003.
- JAYAPRAKASH, V. et al. Human papillomavirus types 16 and 18 in epithelial dysplasia of oral cavity and oropharynx: A meta-analysis, 1985–2010. **Oral Oncology**, v. 47, n. 11, p. 1048–1054, nov. 2011.
- JAYASOORIYA, P. R. et al. Malignant transformation of oral leukoplakia: a follow-up study. **Clinical Oral Investigations**, v. 24, n. 12, p. 4563–4569, 2020.
- JOHNSON, D. E. et al. Head and neck squamous cell carcinoma. **Nature Reviews Disease Primers**, v. 6, n. 1, 2020.

JORDAN, R. C. K. et al. Overexpression of Matrix Metalloproteinase-1 and -9 mRNA Is Associated with Progression of Oral Dysplasia to Cancer. **Clinical Cancer Research**, v. 10, n. 19, p. 6460–6465, 1 out. 2004.

KAMLEH, M. A. et al. Optimizing the use of quality control samples for signal drift correction in large-scale urine metabolic profiling studies. **Analytical Chemistry**, v. 84, n. 6, p. 2670–2677, 2012.

KAUR, J.; POLITIS, C.; JACOBS, R. Salivary 8-hydroxy-2-deoxyguanosine, malondialdehyde, vitamin C, and vitamin E in oral pre-cancer and cancer: diagnostic value and free radical mechanism of action. **Clinical Oral Investigations**, v. 20, n. 2, p. 315–319, 2016.

KHAN, A. et al. Noninvasive Serum Metabolomic Profiling Reveals Elevated Kynurenine Pathway's Metabolites in Humans with Prostate Cancer. **Journal of Proteome Research**, v. 18, p. 1532–1541, 2019.

KHANNA, R. et al. S-phase fraction and DNA ploidy in oral leukoplakia. **ANZ Journal of Surgery**, v. 80, n. 7–8, p. 548–551, 2010.

KIND, T.; FIEHN, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. **BMC Bioinformatics**, v. 20, p. 1–20, 2007.

KIRWAN, J. A. et al. Characterising and correcting batch variation in an automated direct infusion mass spectrometry (DIMS) metabolomics workflow. **Analytical and Bioanalytical Chemistry**, v. 405, n. 15, p. 5147–5157, 2013.

KLAASSEN, I.; BRAAKHUIS, B. J. M. Anticancer activity and mechanism of action of retinoids in oral and pharyngeal cancer. **Oral Oncology**, v. 38, n. 6, p. 532–542, 2002.

KOCH, H. F. Biochemical treatment of precancerous oral lesions: The effectiveness of various analogues of retinoic acid. **Journal of Maxillofacial Surgery**, v. 6, n. C, p. 59–63, 1978.

KONG, X. et al. Analysis of plasma metabolic biomarkers in the development of 4-nitroquinoline-1-oxide-induced oral carcinogenesis in rats. **Oncology Letters**, v. 9, n. 1, p. 283–289, 2015.

KOVALCHUK, S. I.; JENSEN, O. N.; ROGOWSKA-WRZESINSKA, A. FlashPack: Fast and Simple Preparation of Ultrahigh-performance Capillary Columns for LC-MS*. **Molecular & Cellular Proteomics**, v. 18, n. 2, p. 383–390, fev. 2019.

KRAMER IR. World health Organization Collaborating centre for Oral Precancerous Lesions. Definition of leukoplakia and related lesions: An aid to studies on oral precancer. **Oral Surg Oral Med Oral Pathol**, v. 46, p. 518–39, 1978.

KRAWCZUK, J.; ŁUKASZUK, T. The feature selection bias problem in relation to high-dimensional gene data. **Artificial Intelligence in Medicine**, v. 66, p. 63–71, 2016.

KRISHNAN, K.; BALASUNDARAM, S. Estimation of total and lipid bound sialic acid in serum in oral leukoplakia. **Journal of Clinical and Diagnostic Research**, v. 11, n. 3, p. ZC25–ZC27, 2017.

KURIBAYASHI, Y. et al. Recurrence patterns of oral leukoplakia after curative surgical resection: Important factors that predict the risk of recurrence and malignancy. **Journal of Oral Pathology and Medicine**, v. 41, n. 9, p. 682–688, 2012.

LAMEIRA, A. G. et al. MCM3 could be a better marker than Ki-67 for evaluation of dysplastic oral lesions: An immunohistochemical study. **Journal of Oral Pathology and Medicine**, v. 43, n. 6, p. 427–434, 2014.

LE, Y. et al. Accelerated, untargeted metabolomics analysis of cutaneous T-cell lymphoma reveals metabolic shifts in plasma and tumor adjacent skins of xenograft mice. **Journal of Mass Spectrometry**, v. 53, n. 2, p. 172–182, 2018.

LENOUVEL, D. et al. An update of knowledge on PD - L1 in head and neck cancers : Physiologic , prognostic and therapeutic perspectives. v. 1, n. January, p. 1–16, 2019.
LERMAN, M. A. et al. HPV-16 in a distinct subset of oral epithelial dysplasia. **Modern Pathology**, v. 30, n. 12, p. 1646–1654, 2017.

LEUNIG, A. et al. Initial experience in the treatment of oral leukoplakia with high-dose vitamin A and follow-up 5-aminolevulinic acid induced protoporphyrin IX fluorescence. **European Archives of Oto-Rhino-Laryngology**, v. 257, n. 6, p. 327–331, 2000.

LI, C. C. et al. Architectural Alterations in Oral Epithelial Dysplasia are Similar in Unifocal and Proliferative Leukoplakia. **Head and Neck Pathology**, n. 0123456789, 2020.

LI, S. et al. Predicting Network Activity from High Throughput Metabolomics. **PLoS computational biology**, v. 9, n. 7, 2013.

LIBISELLER, G. et al. IPO: A tool for automated optimization of XCMS parameters. **BMC Bioinformatics**, v. 16, n. 1, p. 1–10, 2015.

LIU, W. et al. Two stem cell markers, ATP-binding cassette, G2 subfamily (ABCG2) and BMI-1, predict the transformation of oral leukoplakia to cancer: A long-term follow-up study. **Cancer**, v. 118, n. 6, p. 1693–1700, 2012a.

LIU, W. et al. Expression patterns of cancer stem cell markers ALDH1 and CD133 correlate with a high risk of malignant transformation of oral leukoplakia. **International Journal of Cancer**, v. 132, n. 4, p. 868–874, 2012b.

LOCCI, E. et al. Forensic NMR metabolomics: one more arrow in the quiver. **Metabolomics**, v. 16, n. 11, p. 1–16, 2020.

LOTAN, R. Suppression of squamous cell carcinoma growth and differentiation by retinoids. **Cancer Research**, v. 54, n. 7 SUPPL., p. 1987–1991, 1994.

MACABEO-ONG, M. et al. Quantitative analysis of cathepsin L mRNA and protein expression during oral cancer progression. **Oral Oncology**, v. 39, n. 7, p. 638–647, 2003.

MAGAKI, S. et al. An Introduction to the Performance of Immunohistochemistry. In: **The Scattered Pearls**. [s.l: s.n.]. v. 1897p. 289–298.

MAIMAITI, A. et al. MicroRNA expression profiling and functional annotation analysis of their targets associated with the malignant transformation of oral leukoplakia. **Gene**, v. 558, n. 2, p. 271–277, 2015.

MAKAREV, E. et al. In silico analysis of pathways activation landscape in oral squamous cell carcinoma and oral leukoplakia. **Cell Death Discovery**, v. 3, n. 1, 2017.

MARKOPOULOS, A. K. Current Aspects on Oral Squamous Cell Carcinoma. **The Open Dentistry Journal**, v. 6, n. 1, p. 126–130, 2012.

MCPARLAND, H.; WARNAKULASURIYA, S. Lichenoid morphology could be an early feature of oral proliferative verrucous leukoplakia. **Journal of Oral Pathology & Medicine**, p. 0–3, 2020.

MELLO, F. W. et al. Prevalence of oral potentially malignant disorders: A systematic review and meta-analysis. **Journal of Oral Pathology and Medicine**, v. 47, n. 7, p. 633–640, 2018.

METGUD, R.; BAJAJ, S. Evaluation of salivary and serum lipid peroxidation, and glutathione in oral leukoplakia and oral squamous cell carcinoma. **Journal of oral science**, v. 56, n. 2, p. 135–142, 2014.

MOHAMED ANSER, S.; ASWATH, N. Comparison of the carcinogenic potential of smokeless tobacco and smoked tobacco by quantifying the excretion of nicotine metabolite NNAL in patients with oral leukoplakia. **Indian Journal of Clinical Biochemistry**, v. 29, n. 2, p. 246–249, 2014.

MOHANTY, V. et al. Molecular alterations in oral cancer using high - throughput proteomic analysis of formalin - fixed paraffin - embedded tissue. **Journal of Cell Communication and Signaling**, n. 0123456789, 2021.

MUSHARRAF, S. G. et al. Metabolite Profiling of Preneoplastic and Neoplastic Lesions of Oral Cavity Tissue Samples Revealed a Biomarker Pattern. **Scientific Reports**, v. 6, n. December, p. 1–8, 2016.

NAGAO, T. et al. Serum antioxidant micronutrients and the risk of oral leukoplakia among Japanese. **Oral Oncology**, v. 36, n. 5, p. 466–470, 2000.

NAPIER, S. S.; SPEIGHT, P. M. Natural history of potentially malignant oral lesions and conditions: An overview of the literature. **Journal of Oral Pathology and Medicine**, v. 37, n. 1, p. 1–10, 2008.

NAYAK, S. et al. Fibroblast growth factor (FGF-2) and its receptors FGFR-2 and FGFR-3 may be putative biomarkers of malignant transformation of potentially malignant oral lesions into oral squamous cell carcinoma. **PLoS ONE**, v. 10, n. 10, 2015.

NEGI, A. et al. Comparison of immunohistochemical expression of antiapoptotic protein survivin in normal oral mucosa, oral leukoplakia, and oral squamous cell carcinoma.

Pathology Research International, v. 2015, 2015.

NEVILLE, B.W. **Patologia Oral e Maxilofacial. 3ª edição.** USA: ELSEVIER, 2016.

NOVAIS, F. J. et al. Identification of a metabolomic signature associated with feed efficiency in beef cattle. p. 1–10, 2019.

PANDEY, M. et al. Evaluation of surgical excision of non-homogeneous oral leukoplakia in a screening intervention trial, Kerala, India. **Oral Oncology**, v. 37, n. 1, p. 103–109, 2001.

PARAKH, M. K. et al. Oral potentially malignant disorders: clinical diagnosis and current screening aids: a narrative review. **European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP)**, v. 29, n. 1, p. 65–72, 2020.

PELLICIOLI, A. C. A. et al. Immunosurveillance profile of oral squamous cell carcinoma and oral epithelial dysplasia through dendritic and T-cell analysis. **Journal of Oral Pathology & Medicine**, v. 45, n. 6, p. 433–439, 2 jul. 2017.

PHELAN, V. V. **Computational Methods and Data Analysis for Metabolomics.** New York, NY: Springer US, 2020. v. 2104

POH, C. F. et al. Unique FISH patterns associated with cancer progression of oral dysplasia. **Journal of Dental Research**, v. 91, n. 1, p. 52–57, 2012.

PUNG, A. et al. β -Carotene and canthaxanthin inhibit chemically- and physically- induced neoplastic transformation in 10T1/2 cells. **Carcinogenesis**, v. 9, n. 9, p. 1533–1539, 1988.

QI, M. et al. Proteome profiling to identify peroxiredoxin 1 interacting protein partners in nicotine-associated oral leukoplakia. **Archives of Oral Biology**, v. 108, n. April, p. 104537, 2019.

RAI, B. et al. Salivary vitamins E and C in oral cancer. **Redox Report**, v. 12, n. 3, p. 163–164, 2007.

RAI, V. et al. “Omics” in oral cancer: New approaches for biomarker discovery. **Archives of Oral Biology**, v. 87, n. February 2017, p. 15–34, 2018.

RAMAUTAR, R.; SOMSEN, G. W.; DE JONG, G. J. CE-MS for metabolomics: Developments and applications in the period 2016–2018. **Electrophoresis**, v. 40, n. 1, p. 165–179, 2019.

RANGANATHAN, K.; KAVITHA, L. Oral epithelial dysplasia: Classifications and clinical relevance in risk assessment of oral potentially malignant disorders. **Journal of Oral and Maxillofacial Pathology**, v. 23, n. 1, p. 19–27, 2019.

REIBEL, J. Prognosis of oral pre-malignant lesions: Significance of clinical, histopathological, and molecular biological characteristics. **Critical Reviews in Oral Biology and Medicine**, v. 14, n. 1, p. 47–62, 2003.

REYES-GARCÉS, N. et al. Assessment of solid phase microextraction as a sample

preparation tool for untargeted analysis of brain tissue using liquid chromatography-mass spectrometry. **Journal of Chromatography A**, v. 1638, p. 11–13, 2021.

RIBEIRO, G. et al. Archives of Oral Biology Mature dendritic cell density is affected by smoking habit, lesion size, and epithelial dysplasia in oral leukoplakia samples. **Archives of Oral Biology**, v. 95, n. May, p. 51–57, 2018.

ROCK, L. D. et al. Characterization of epithelial oral dysplasia in non-smokers: First steps towards precision medicine. **Oral oncology**, v. 78, p. 119–125, 2018.

ROZA, A. L. O. C. et al. Oral leukoplakia and erythroplakia in young patients: a systematic review. **Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology**, v. 131, n. 1, p. 73–84, 2021.

RUBERT, A.; BAGÁN, L.; BAGÁN, J. V. Oral leukoplakia, a clinical-histopathological study in 412 patients. **Journal of Clinical and Experimental Dentistry**, v. 12, n. 6, p. e540–e546, 2020.

SANKARANARAYANAN, R. et al. Chemoprevention of oral leukoplakia with vitamin A and beta carotene: An assessment. **European Journal of Cancer Part B: Oral Oncology**, v. 33, n. 4, p. 231–236, 1997.

SARODE, G. S. et al. Recent trends in predictive biomarkers for determining malignant potential of oral potentially malignant disorders. **Oncology Reviews**, v. 13, n. 2, 2019.

SCHEPMAN, K. P. et al. Malignant transformation of oral leukoplakia: A follow-up study of a hospital-based population of 166 patients with oral leukoplakia from The Netherlands. **Oral Oncology**, v. 34, n. 4, p. 270–275, 1998.

SCHER, R. L. et al. Fenretinide-induced apoptosis of human head and neck squamous carcinoma cell lines. **Otolaryngology - Head and Neck Surgery**, v. 118, n. 4, p. 464–471, 1998.

SCHOLTZ, B. et al. Examination of oral squamous cell carcinoma and precancerous lesions using proximity extension assay and salivary rna quantification. **Biomedicines**, v. 8, n. 12, p. 1–16, 2020.

SENGHORE, T. et al. Biomarkers of oxidative stress associated with the risk of potentially malignant oral disorders. **Anticancer Research**, v. 38, n. 9, p. 5211–5216, 2018.

SHI, L. et al. Variable selection and validation in multivariate modelling. **Bioinformatics**, v. 35, n. 6, p. 972–980, 2019.

SHIGEOKA, M. et al. CD163+ macrophages infiltration correlates with the immunosuppressive cytokine interleukin 10 expression in tongue leukoplakia. **Clinical and Experimental Dental Research**, v. 5, n. 6, p. 627–637, 2019.

SHUKLA, S. et al. Tumor antigens eliciting autoantibody response in cancer of gingivo-buccal complex. **Proteomics - Clinical Applications**, v. 1, n. 12, p. 1592–1604, 2007.

SIEBERS, T. J. H. et al. Chromosome instability predicts the progression of premalignant oral lesions. **Oral Oncology**, v. 49, n. 12, p. 1121–1128, 2013.

SILVERMAN, S. et al. Malignant transformation and natural history of oral leukoplakia in 57,518 industrial workers of gujarat, india. **Cancer**, v. 38, n. 4, p. 1790–1795, 1976.

SILVERMAN, S.; GORSKY, M.; LOZADA, F. Oral Leukoplakia and Malignant Transformation. **Cancer**, v. 53, p. 563–568, 1984.

SIVADASAN, P. et al. Salivary proteins from dysplastic leukoplakia and oral squamous cell carcinoma and their potential for early detection. **Journal of Proteomics**, v.212, 2020.

SMILDE, A. K. et al. Fusion of mass spectrometry-based metabolomics data. **Analytical Chemistry**, v. 77, n. 20, p. 6729–6736, 2005.

SMITH, C. A. et al. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. **Analytical Chemistry**, v. 78, n. 3, p. 779–787, 2006.

SONG, X. et al. Oral squamous cell carcinoma diagnosed from saliva metabolic profiling. **Proceedings of the National Academy of Sciences of the United States of America**, v. 117, n. 28, p. 16167–16173, 2020.

SPEIGHT, P. M.; KHURRAM, S. A.; KUJAN, O. Oral potentially malignant disorders: risk of progression to malignancy. **Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology**, v. 125, n. 6, p. 612–627, 2018.

SPERANDIO, M. et al. Predictive value of dysplasia grading and DNA ploidy in malignant transformation of oral potentially malignant disorders. **Cancer Prevention Research**, v. 6, n. 8, p. 822–831, 2013.

SRIDHARAN, G. et al. Evaluation of salivary metabolomics in oral leukoplakia and oral squamous cell carcinoma. **Journal of Oral Pathology and Medicine**, v. 48, n. 4, p. 299–306, 2019.

SRIDHARAN, G.; RAMANI, P.; PATANKAR, S. Serum metabolomics in oral leukoplakia and oral squamous cell carcinoma. **Journal of Cancer Research and Therapeutics**, v. 14, n. 7, p. 1525–1534, 2017.

SRIVASTAVA, K. C. Comparative evaluation of saliva's oxidant-antioxidant status in patients with different clinicopathological types of oral leukoplakia. **Journal of International Society of Preventive and Community Dentistry**, v. 9, n. 4, p. 396–402, 2019.

SRIVASTAVA, K. C.; SHRIVASTAVA, D. Analysis of plasma lipid peroxidation and antioxidant enzymes status in patients of oral leukoplakia: A case control study. **Journal of International Society of Preventive and Community Dentistry**, v. 6, n. 9, p. S213–S218, 2016.

SRIVASTAVA, S. et al. Proton HR-MAS MR spectroscopy of oral squamous cell carcinoma tissues: An ex vivo study to identify malignancy induced metabolic fingerprints.

Metabolomics, v. 7, n. 2, p. 278–288, 2011.

STASIKOWSKA-KANICKA, O.; WĄGROWSKA-DANILEWICZ, M.; DANILEWICZ, M. CD8+ and CD163+ infiltrating cells and PD-L1 immunoexpression in oral leukoplakia and oral carcinoma. **Apmis**, v. 126, n. 9, p. 732–738, 2018a.

STASIKOWSKA-KANICKA, O.; WĄGROWSKA-DANILEWICZ, M.; DANILEWICZ, M. T cells are involved in the induction of macrophage phenotypes in oral leukoplakia and squamous cell carcinoma—a preliminary report. **Journal of Oral Pathology & Medicine**, v. 47, n. 2, p. 136–143, fev. 2018b.

SU, Y.-W. et al. Prognostic Value of the Overexpression of Fatty Acid Metabolism-Related Enzymes in Squamous Cell Carcinoma of the Head and Neck. **International Journal of Molecular Sciences**, v. 21, n. 18, p. 6851, 18 set. 2020.

SUBRAMANIAN, A. et al. Gene set enrichment analysis : A knowledge-based approach for interpreting genome-wide. **Proceedings of the National Academy of Sciences of the United States of America**, 2005.

SUMNER, L. W. et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). **Metabolomics**, v. 3, n. 3, p. 211–221, 2007.

SUNDBERG, J. et al. Recurrence rates after surgical removal of oral leukoplakia—A prospective longitudinal multicentre study. **PLoS ONE**, v. 14, n. 12, p. 1–13, 2019.

SYRJÄNEN, S. et al. Human papillomaviruses in oral carcinoma and oral potentially malignant disorders: A systematic review. **Oral Diseases**, v. 17, n. SUPPL. 1, p. 58–72, 2011.

THOMAS, G. et al. Prognostic Implications of DNA Repair, Ploidy and Telomerase in the Malignant Transformation Risk Assessment of Leukoplakia. **Asian Pacific journal of cancer prevention : APJCP**, v. 21, n. 2, p. 309–316, 2020.

TINNEVELT, G. H. et al. Variable selection in untargeted metabolomics and the danger of sparsity. **Metabolites**, v. 10, n. 11, p. 1–10, 2020.

TRIMIGNO, A. et al. An NMR metabolomics approach to investigate factors affecting the yoghurt fermentation process and quality. **Metabolites**, v. 10, n. 7, p. 1–16, 2020.

TROYANSKAYA, O. et al. Missing value estimation methods for DNA microarrays. **Bioinformatics**, v. 17, n. 6, p. 520–525, 2001.

TSAI, C.-K. et al. Nuclear Magnetic Resonance Metabolomics Biomarkers for Identifying High Risk Patients with Extranodal Extension in Oral Squamous Cell Carcinoma. **Journal of Clinical Medicine**, v. 9, n. 4, p. 951, 2020.

TSUI, I. F. L. et al. Multiple pathways in the FGF signaling network are frequently deregulated by gene amplification in oral dysplasias. **International Journal of Cancer**, v. 125, n. 9, p. 2219–2228, 2009.

TUSHER, V. G.; TIBSHIRANI, R.; CHU, G. Significance analysis of microarrays applied to the ionizing radiation response. v. 98, n. 9, 2001.

VAN DEN BERG, R. A. et al. Centering, scaling, and transformations: Improving the biological information content of metabolomics data. **BMC Genomics**, v. 7, p. 1–15, 2006.

VAN DER KLOET, F. M. et al. Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. **Journal of Proteome Research**, v. 8, n. 11, p. 5132–5141, 2009.

VAN DER WAAL, I. Oral leukoplakia, the ongoing discussion on definition and terminology. **Medicina Oral, Patologia Oral y Cirugia Bucal**, v. 20, n. 6, p. e685–e692, 2015.

VAN ZYL, A. W. et al. Correlation Between Dysplasia and Ploidy Status in Oral Leukoplakia. **Head and Neck Pathology**, v. 6, n. 3, p. 322–327, 2012.

VEERAMACHANENI, R. et al. Analysis of head and neck carcinoma progression reveals novel and relevant stage-specific changes associated with immortalisation and malignancy. **Scientific Reports**, v. 9, n. 1, p. 1–17, 2019.

VICENTE, J. C. DE et al. SOX2 Expression Is an Independent Predictor of Oral Cancer Progression. 2019.

VIGNESWARAN, N. et al. Increased EMMPRIN (CD 147) expression during oral carcinogenesis. **Experimental and Molecular Pathology**, v. 80, n. 2, p. 147–159, 2006.

VIJAYAVEL, T.; ASWATH, N. Correlation between histological grading and ploidy status in potentially malignant disorders of the oral mucosa: A flow cytometric analysis. **Journal of Oral and Maxillofacial Pathology**, v. 17, n. 2, p. 169–175, 2013.

VILLA, A. et al. Oral keratosis of unknown significance shares genomic overlap with oral dysplasia. **Oral Diseases**, v. 25, n. 7, p. 1707–1714, 2019.

VINAIXA, M. et al. A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data. **Metabolites**, v. 2, n. 4, p. 775–795, 2012.

VITÓRIO, J. G. et al. Metabolic landscape of oral squamous cell carcinoma. **Metabolomics**, v. 16, n. 10, 2020.

WALDRON, C. A.; SHAFER, W. G. Leukoplakia revisited. A clinicopathologic study 3256 oral leukoplakias. **Cancer**, v. 36, n. 4, p. 1386–1392, 1975.

WANG, P. et al. Circulating tumor cell detection and single-cell analysis using an integrated workflow based on ChimeraX® -i120 Platform: A prospective study. **Molecular Oncology**, v. 58, n. 1, p. 1878–0261.12876, 10 dez. 2020.

WANG, Q. et al. Investigation and identification of potential biomarkers in human saliva for the early diagnosis of oral squamous cell carcinoma. **Clinica Chimica Acta**, v. 427, p. 79–85, 2014.

WANG, X.; ZHANG, J.; ZHOU, G. The CXCL11-CXCR3A axis influences the infiltration of CD274 and IDO1 in oral squamous cell carcinoma. **Journal of Oral Pathology and Medicine**, p. 0–3, 2020.

WANG, Y. et al. A Simple Method for Peak Alignment Using Relative Retention Time Related to an Inherent Peak in Liquid Chromatography-Mass Spectrometry-Based Metabolomics. **Journal of Chromatographic Science**, v. 57, n. 1, p. 9–16, 2019.

WANG, Z. et al. Involvement of potential pathways in malignant transformation from oral leukoplakia to oral squamous cell carcinoma revealed by proteomic analysis. **BMC Genomics**, v. 10, p. 383, 2009.

WARNAKULASURIYA, S. et al. Oral epithelial dysplasia classification systems: Predictive value, utility, weaknesses and scope for improvement. **Journal of Oral Pathology and Medicine**, v. 37, n. 3, p. 127–133, 2008.

WARNAKULASURIYA, S. et al. Factors predicting malignant transformation in oral potentially malignant disorders among patients accrued over a 10-year period in South East England. **Journal of Oral Pathology and Medicine**, v. 40, n. 9, p. 677–683, 2011.

WARNAKULASURIYA, S. Clinical features and presentation of oral potentially malignant disorders. **Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology**, v. 125, n. 6, p. 582–590, 2018.

WARNAKULASURIYA, S. Oral potentially malignant disorders: A comprehensive review on clinical aspects and management. **Oral Oncology**, v. 102, n. August 2019, p. 104550, 2020a.

WARNAKULASURIYA, S. Oral potentially malignant disorders: A comprehensive review on clinical aspects and management. **Oral Oncology**, v. 102, n. May 2019, p. 104550, 2020b.

WARNAKULASURIYA, S.; ARIYAWARDANA, A. Malignant transformation of oral leukoplakia: A systematic review of observational studies. **Journal of Oral Pathology and Medicine**, v. 45, n. 3, p. 155–166, 2016.

WARNAKULASURIYA, S.; JOHNSON, N. W.; VAN DER WAAL, I. Nomenclature and classification of potentially malignant disorders of the oral mucosa. **Journal of Oral Pathology and Medicine**, v. 36, n. 10, p. 575–580, 2007.

WEBER, M. et al. Malignant transformation of oral leukoplakia is associated with macrophage polarization. **Journal of Translational Medicine**, v. 18, n. 1, p. 1–18, 2020.

WEI, J. et al. Salivary metabolite signatures of oral cancer and leukoplakia. **International Journal of Cancer**, v. 129, n. 9, p. 2207–2217, 2011.

WERKMEISTER, R.; BRANDT, B.; JOOS, U. Aberrations of erbB-1 and erbB-2 oncogenes in non-dysplastic leukoplakias of the oral cavity. **British Journal of Oral and Maxillofacial Surgery**, v. 37, n. 6, p. 477–480, 1999.

- WILD, C. P. Complementing the genome with an “exposome”: The outstanding challenge of environmental exposure measurement in molecular epidemiology. **Cancer Epidemiology Biomarkers and Prevention**, v. 14, n. 8, p. 1847–1850, 2005.
- WILS, L. J. et al. Incorporation of differentiated dysplasia improves prediction of oral leukoplakia at increased risk of malignant progression. **Modern Pathology**, 2020.
- WISHART, D. S. Emerging applications of metabolomics in drug discovery and precision medicine. **Nature Reviews Drug Discovery**, v. 15, n. 7, p. 473–484, 2016.
- WOO, S. BIN; CASHMAN, E. C.; LERMAN, M. A. Human papillomavirus-associated oral intraepithelial neoplasia. **Modern Pathology**, v. 26, n. 10, p. 1288–1297, 2013.
- WU, W.; WANG, Z.; ZHOU, Z. Risk Factors Associated With Malignant Transformation in Patients With Oral Leukoplakia in a Chinese Population: A Retrospective Study. **Journal of Oral and Maxillofacial Surgery**, v. 77, n. 12, p. 2483–2493, 2019.
- WU, X. et al. Transglutaminase 3 contributes to malignant transformation of oral leukoplakia to cancer. **International Journal of Biochemistry and Cell Biology**, v. 104, p. 34–42, 2018.
- WU, X. et al. Application of PD-1 Blockade in Cancer Immunotherapy. **Computational and Structural Biotechnology Journal**, v. 17, p. 661–674, 2019.
- WU, Y.; LI, L. Sample normalization methods in quantitative metabolomics. **Journal of Chromatography A**, v. 1430, p. 80–95, 2016.
- XIA, J. et al. Translational biomarker discovery in clinical metabolomics: An introductory tutorial. **Metabolomics**, v. 9, n. 2, p. 280–299, 2013.
- XIE, G. X. et al. Urine metabolite profiling offers potential early diagnosis of oral cancer. **Metabolomics**, v. 8, n. 2, p. 220–231, 2012.
- XU, S. et al. Comprehensive analysis of circular RNA in oral leukoplakia: upregulated circHLA-C as a potential biomarker for diagnosis and prognosis. **Annals of Translational Medicine**, v. 8, n. 21, p. 1375–1375, 2020.
- YAGYUU, T. et al. Programmed death ligand 1 (PD-L1) expression and tumor microenvironment: Implications for patients with oral precancerous lesions. **Oral Oncology**, v. 68, p. 36–43, 2017.
- YANG, S. W. et al. Oral tongue leukoplakia: analysis of clinicopathological characteristics, treatment outcomes, and factors related to recurrence and malignant transformation. **Clinical Oral Investigations**, 2021.
- YOSHIDA, T. et al. Association between p62 expression and clinicopathological characteristics in oral leukoplakia. **Clinical and Experimental Dental Research**, v. 5, n. 4, p. 389–397, 2019.
- ZHANG, X. et al. Snail and Axin2 expression predict the malignant transformation of oral leukoplakia. **Oral Oncology**, v. 73, p. 48–55, 2017a.

ZHANG, X. et al. Nomogram for risk prediction of malignant transformation in oral leukoplakia patients using combined biomarkers. **Oral Oncology**, v. 72, p. 132–139, 2017b.

ZHEN, H. et al. Assessing the impact of wastewater treatment plant effluent on downstream drinking water-source quality using a zebrafish (*Danio Rerio*) liver cell-based metabolomics approach. **Water Research**, v. 145, p. 198–209, nov. 2018.

ZHOU, J. et al. ¹H NMR-based metabonomic and pattern recognition analysis for detection of oral squamous cell carcinoma. **Clinica Chimica Acta**, v. 401, n. 1–2, p. 8–13, 2009.

ZOU, W. Immunosuppressive networks in the tumour environment and their therapeutic relevance. **Nature Reviews Cancer**, v. 5, n. 4, p. 263–274, 2005.

APÊNDICE A – Pré-processamento dos dados *softwares* IPO e XCMS

Etapa de Análise	Parâmetro	Intervalo de otimização	Valor definido
Detecção de picos	ppm	35-36	35.1
	min_peakwidth	22.5-24	23.25
	max_peakwidth	68-71	71.3
	mzdiff	0.00140-0.00150	0.00152
	snthresh	5.5-6	5.6
	noise	0-80	1
	value_of_prefilter	245-260	8.3
Correção do tempo de retenção e agrupamento	prefilter	7.5-8.5	260
	gapInIt	0.8-1.2	0.728
	gapExtend	1.9-2.7	2.7
	profStep	0.7-1	0.754
	bw	0-15	2.25
	minfrac	0.7-0.9	0.89
	mzwid	0.020-0.030	0.026

Tabela 1. Valores para otimização dos parâmetros dos dados obtidos em modo de ionização positivo.

Etapa de Análise	Parâmetro	Intervalo de otimização	Valor definido
Detecção de picos	ppm	49-52	52.3
	min_peakwidth	18-20	21
	max_peakwidth	123-127	128.2
	mzdiff	0.0052-0.0056	0.0054
	snthresh	08-10*	9
	noise	250-300	292
	value_of_prefilter	0-150	2.5
Correção do tempo de retenção e agrupamento	prefilter	1-4*	75
	gapInIt	0.3-0.4	0.4
	gapExtend	1.8-2.4	2.1
	profStep	0.55-0.85	0.85
	bw	40-50	40
	minfrac	0.8-0.9	0.85
	mzwid	0.025-0.040	0.025

Tabela 1. Valores para otimização dos parâmetros dos dados obtidos em modo de ionização negativo.

APÊNDICE A – Pré-processamento dos dados *softwares* IPO e XCMS

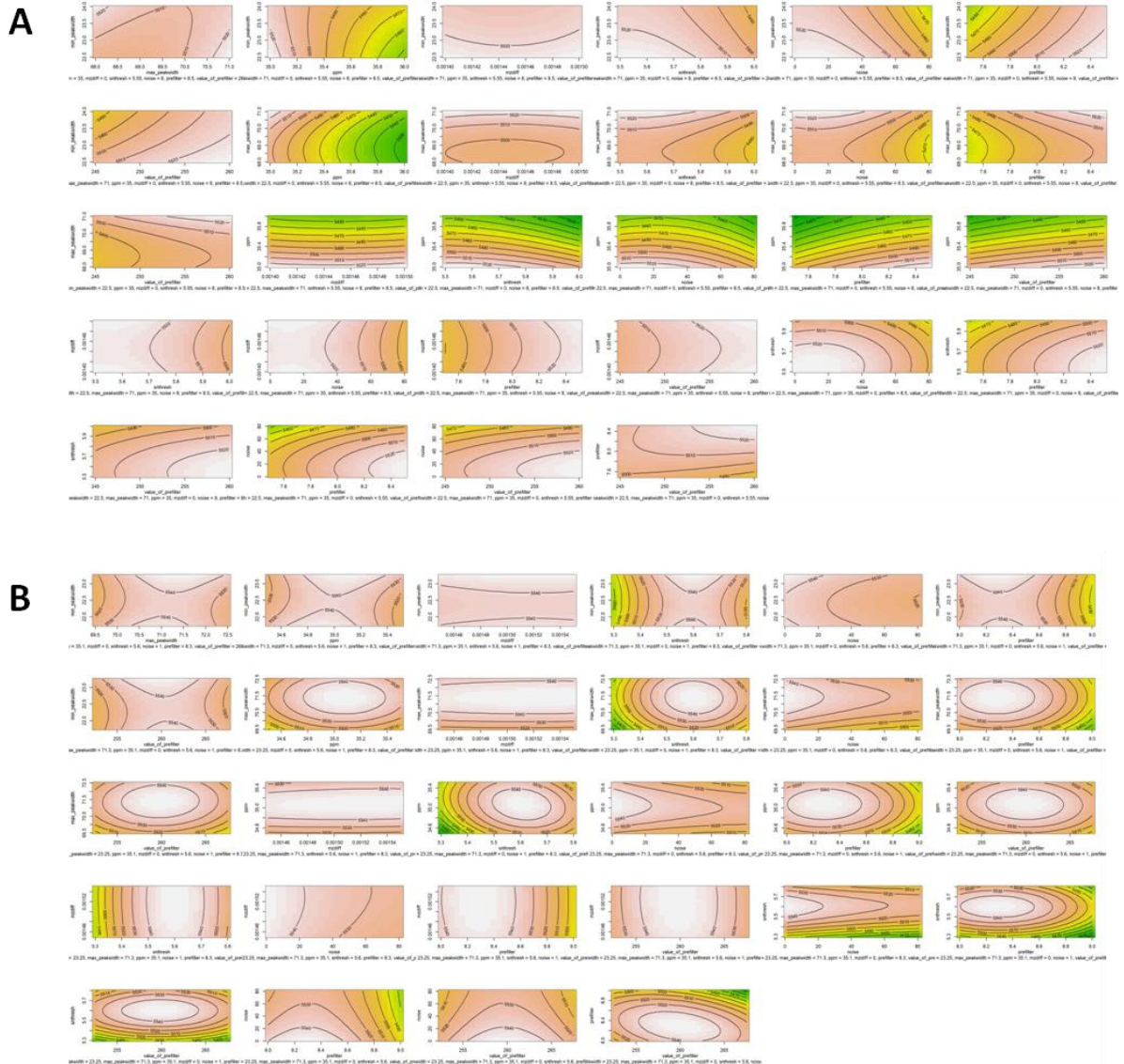


Figura 1: Gráficos de superfície de resposta para otimização dos parâmetros de detecção de picos para os dados obtidos em modo de ionização positivo. **(A)** Gráficos obtidos utilizando os intervalos de otimização padrão do *software*. **(B)** Gráficos obtidos utilizando intervalos definidos a partir da primeira otimização. Os valores considerados para a escolha dos parâmetros são equivalentes às regiões rosa claro do gráfico.

APÊNDICE A – Pré-processamento dos dados *softwares* IPO e XCMS

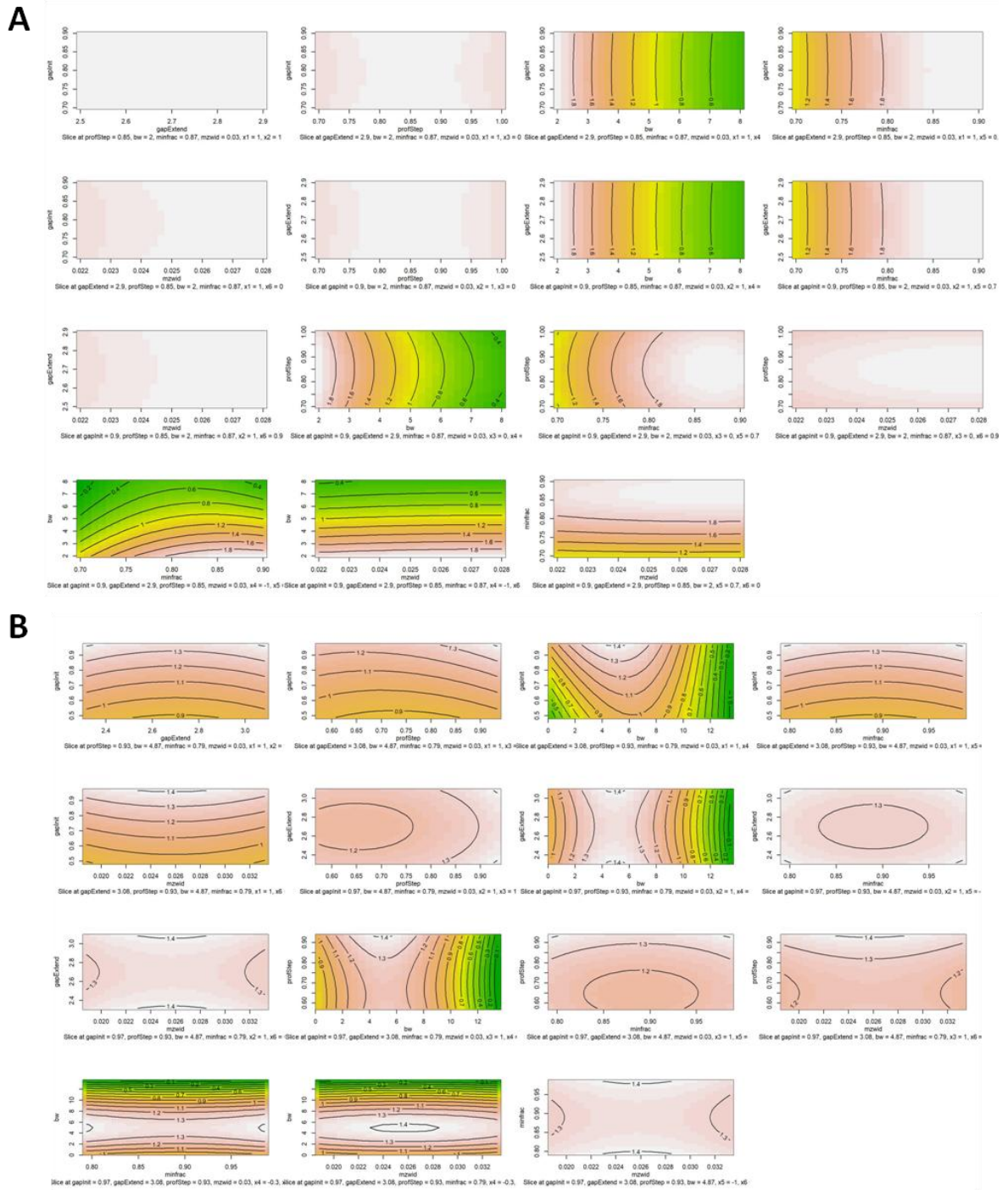


Figura 2: Gráficos de superfície de resposta para otimização dos parâmetros de correção do tempo de retenção e agrupamento para os dados obtidos em modo de ionização positivo. **(A)** Gráficos obtidos utilizando os intervalos de otimização padrão do *software*. **(B)** Gráficos obtidos utilizando intervalos definidos a partir da primeira otimização. Os valores considerados como indicados para a escolha dos parâmetros são equivalentes às regiões rosa claro do gráfico.

APÊNDICE A – Pré-processamento dos dados *softwares* IPO e XCMS

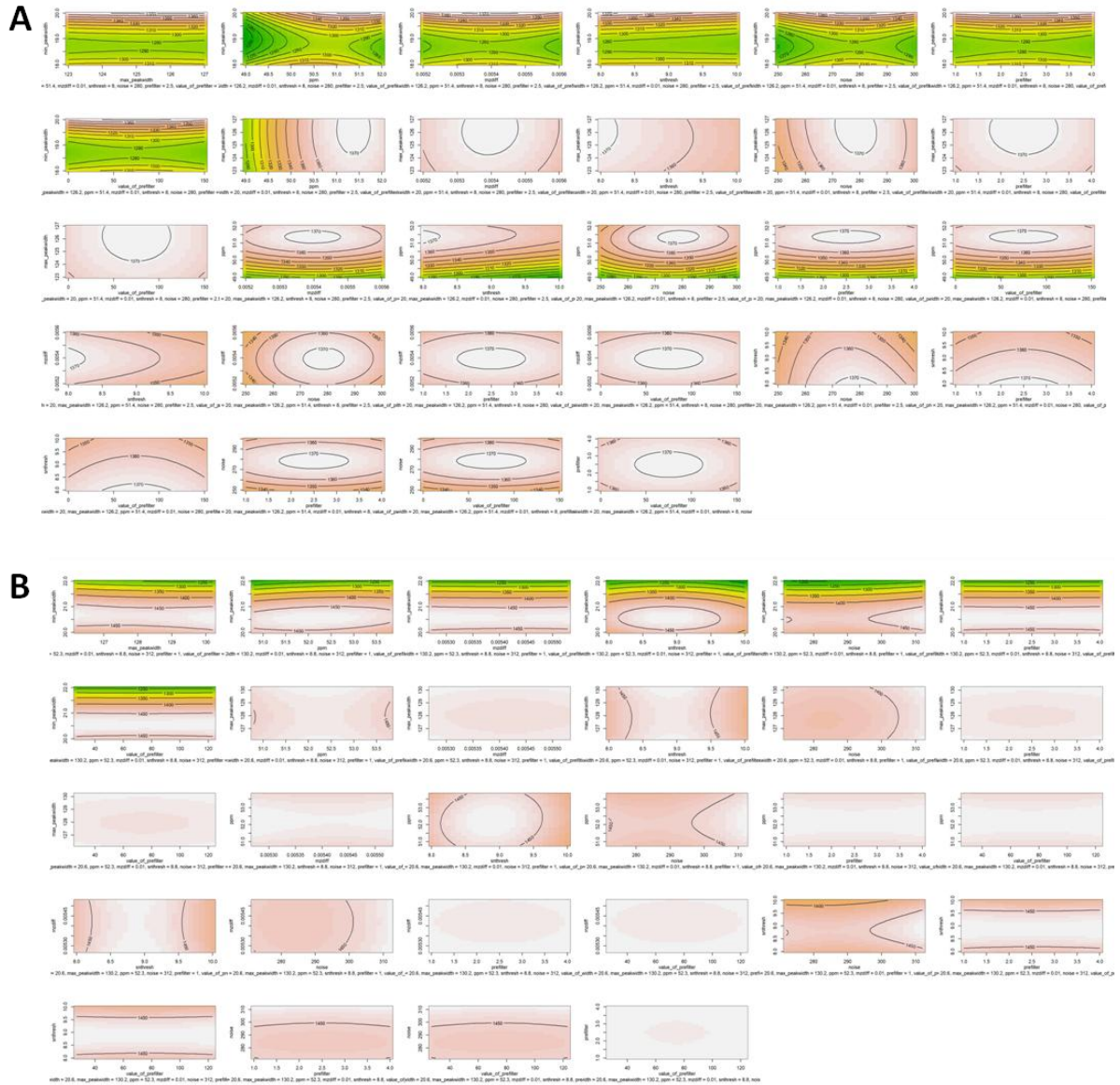


Figura 3: Gráficos de superfície de resposta para otimização dos parâmetros de detecção de picos para os dados obtidos em modo de ionização negativo. (A) Gráficos obtidos utilizando os intervalos de otimização padrão do *software*. (B) Gráficos obtidos utilizando intervalos definidos a partir da primeira otimização. Os valores considerados como indicados para a escolha dos parâmetros são equivalentes às regiões rosa claro do gráfico.

APÊNDICE A – Pré-processamento dos dados *softwares* IPO e XCMS

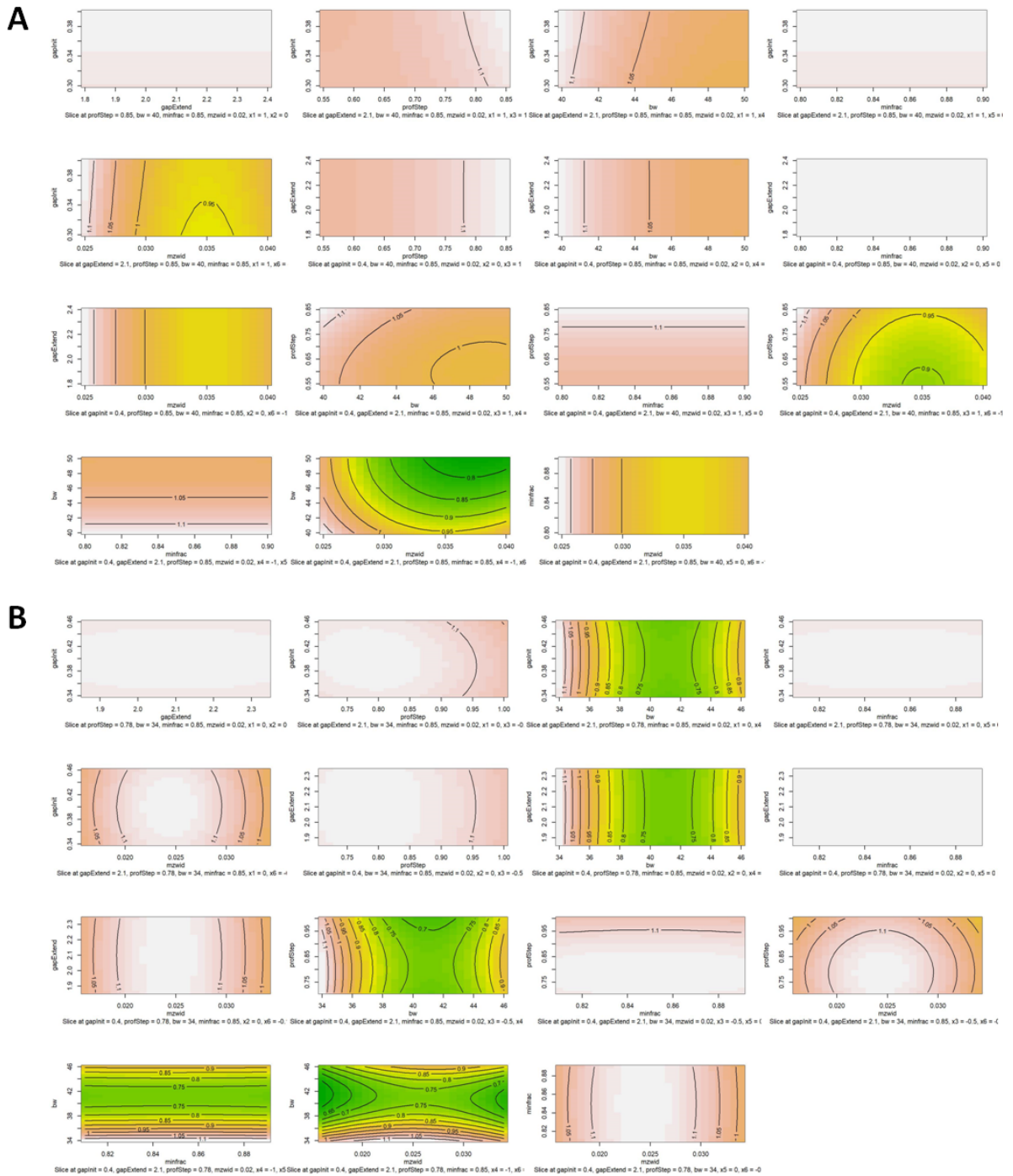


Figura 4: Gráficos de superfície de resposta para otimização dos parâmetros de correção do tempo de retenção e agrupamento para os dados obtidos em modo de ionização negativo. **(A)** Gráficos obtidos utilizando os intervalos de otimização padrão do *software*. **(B)** Gráficos obtidos utilizando intervalos definidos a partir da primeira otimização. Os valores considerados como indicados para a escolha dos parâmetros são equivalentes às regiões rosa claro do gráfico.

APÊNDICE A – Pré-processamento dos dados *softwares* IPO e XCMS

Etapa	Dados positivos	Dados negativos
Chamada do pacote	<code>library(xcms)</code>	<code>library(xcms)</code>
Deteção dos picos	<code>xsetpos1 <- xcmsSet(method="centWave",peakwidth=c(23.25,71.3),ppm=35.1,noise=1,snthresh=5.6,mzdiff=0.00152,prefilter = c(8.3,260),mzCenterFun="wMean",integrate=1,fitgauss = FALSE,verbose.columns = FALSE)</code>	<code>xsetneg1 <- xcmsSet(method="centWave",peakwidth=c(21,128.2),ppm=52.3,noise=292,snthresh=9,mzdiff=0.0054,prefilter = c(2.5,75),mzCenterFun="wMean",integrate=1,fitgauss = FALSE,verbose.columns = FALSE)</code>
Agrupamento dos picos	<code>xsetpos2<- group(xsetpos1,method="density",bw=2.25,mzwid=0.026,minfrac=0.7,minsamp=1,max =50)</code>	<code>xsetneg2<- group(xsetneg1,method="density",bw=40,mzwid=0.025,minfrac=0.85,minsamp=1,max =50)</code>
Correção do tempo de retenção	<code>xsetpos3 <- retcor(xsetpos2,method="obiwarp",plotype="deviation",distFunc="cor_opt",profStep=0.754,center=22,response=1,gaplnit=0.728,gapExtend= 2.7,factorDiag=2,factorGap= 1,localAlignment=0)</code>	<code>xsetneg3<- retcor(xsetneg2,method="obiwarp",plotype="deviation",distFunc="cor_opt",profStep=0.85,center=23,response=1,gaplnit=0.4,gapExtend= 2.1,factorDiag=2,factorGap= 1,localAlignment=0)</code>
Agrupamento dos picos	<code>xsetpos4<- group(xsetpos3,method="density",bw=2.25,mzwid=0.026,minfrac=0.7,minsamp=1,max =50)</code>	<code>xsetneg4<- group(xsetneg3,method="density",bw=40,mzwid=0.025,minfrac=0.85,minsamp=1,max =50)</code>
Exclusão de ruídos	<code>xsetpos5 <- fillPeaks(xsetpos4)</code>	<code>xsetneg5 <- fillPeaks(xsetneg4)</code>
Análises estatísticas e construção da tabela	<code>reporttab<- diffreport(xsetpos5,"casos","controles","qcs_2",100)</code>	<code>reporttab<- diffreport(xsetneg5,"casos","controles","qcs",100)</code>

Tabela 2. Linhas de comando utilizadas para o pré-processamento dos dados no *software* XCMS

APÊNDICE B – Normalização dos dados

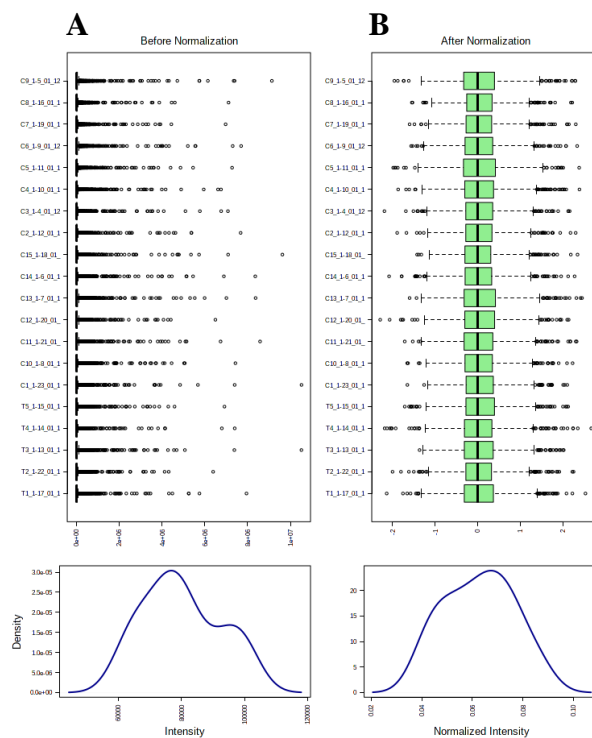


Figura 1 – Normalização para as análises univariadas dos dados obtidos em modo de ionização positivo. Método de normalização: mediana. Transformação: escala logarítmica. O escalonamento não foi aplicado. (A) Antes da normalização. (B) Após a normalização

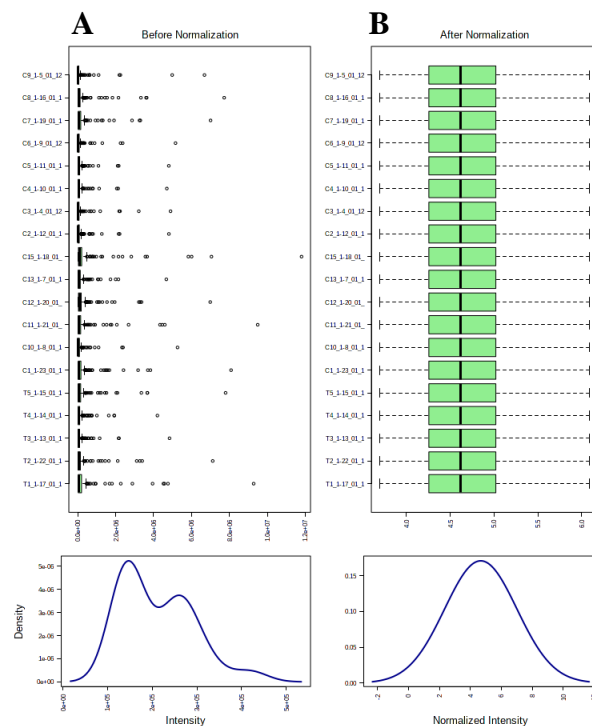


Figura 2 – Normalização para as análises univariadas dos dados obtidos em modo de ionização negativo. Método de normalização: quantil. Transformação: escala logarítmica. O escalonamento não foi aplicado. (A) Antes da normalização. (B) Após a normalização

APÊNDICE B – Normalização dos dados

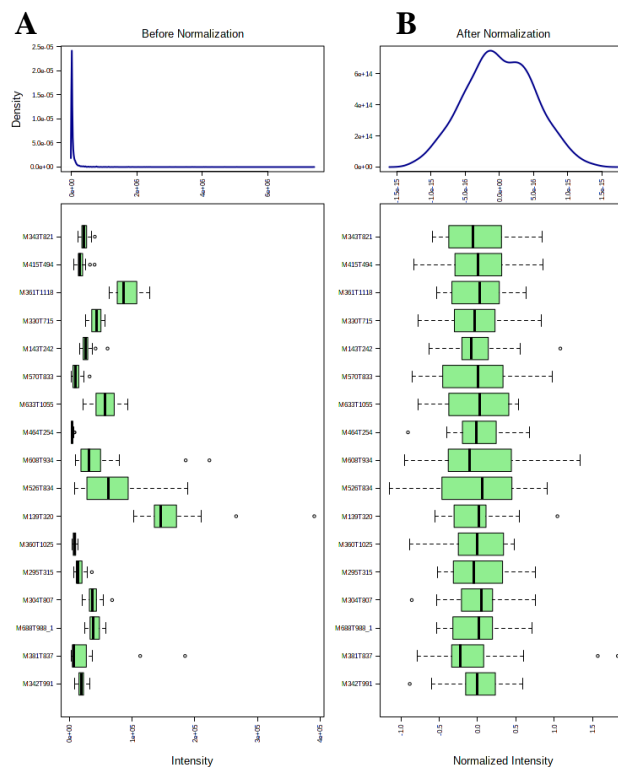


Figura 3 – Normalização para as análises multivariadas dos dados obtidos em modo de ionização positivo. Método de normalização: quantil. Transformação: log. Escalonamento: Pareto.

(A) Antes da normalização. (B) Após a normalização

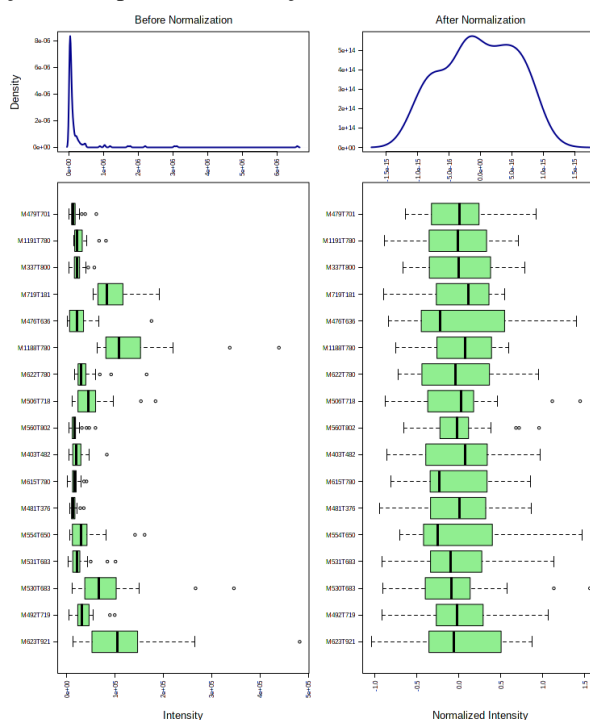


Figura 4 – Normalização para as análises multivariadas dos dados obtidos em modo de ionização negativo. Método de normalização: quantil. Transformação: log. Escalonamento: Pareto.

(A) Antes da normalização. (B) Após a normalização

APÊNDICE C – Seleção recursiva dos *molecular features* discriminantes

Método de classificação	Parâmetro	Modelagem inicial	Modelagem final
PLS	nOuter	5	5
	nInner	4	4
	nRep	15	90
	varRatio	0.5	0.8
	method	PLS	PLS
	fitness	BER	BER
	scale	TRUE	TRUE
RF	nOuter	5	5
	nInner	4	4
	nRep	15	100
	varRatio	0.5	0.8
	method	PLS	RF
	fitness	BER	BER
	scale	TRUE	TRUE

Tabela 1. Valores para modelagem dos dados obtidos em modo de ionização positivo.

Método de classificação	Parâmetro	Modelagem inicial	Modelagem final
PLS	nOuter	5	5
	nInner	4	4
	nRep	15	80
	varRatio	0.5	0.85
	method	PLS	PLS
	fitness	BER	BER
	scale	TRUE	TRUE
RF	nOuter	5	5
	nInner	4	4
	nRep	15	90
	varRatio	0.5	0.8
	method	PLS	RF
	fitness	BER	BER
	scale	TRUE	TRUE

Tabela 2. Valores para modelagem dos dados obtidos em modo de ionização negativo.

APÊNDICE C – Seleção recursiva dos *molecular features* discriminantes

Dados de entrada	Comandos comuns	PLS	RF
Positivo	<pre>filePath <- file.choose(input_muvr_pos) input_muvr_pos<- read_excel(filePath) y <- input_muvr_pos [,1][[1]] x <- input_muvr_pos [,-1] library(doParallel) library(MUVR) nCore <- detectCores()-1 cl<- makeCluster(nCore) registerDoParallel(cl) stopCluster(cl)</pre>	<pre>muvrModel <- MUVR(X = x, Y = y, nOuter = 5, nInner = 4, nRep = 90, varRatio = 0.8, method = 'PLS', fitness= 'BER', scale= 'TRUE')</pre>	<pre>muvrModel <- MUVR(X = x, Y = y, nOuter = 5, nInner = 4, nRep = 100, varRatio = 0.8, method = 'RF', fitness= 'BER', scale= 'TRUE')</pre>
Negativo	<pre>filePath <- file.choose(input_muvr_neg) input_muvr_neg<- read_excel(filePath) y <- input_muvr_neg [,1][[1]] x <- input_muvr_neg [,-1] library(doParallel) library(MUVR) nCore <- detectCores()-1 cl<- makeCluster(nCore) registerDoParallel(cl) stopCluster(cl)</pre>	<pre>muvrModel <- MUVR(X = x, Y = y, nOuter = 5, nInner = 4, nRep = 80, varRatio = 0.85, method = 'PLS', fitness= 'BER', scale= 'TRUE')</pre>	<pre>muvrModel <- MUVR(X = x, Y = y, nOuter = 5, nInner = 4, nRep = 90, varRatio = 0.8, method = 'RF', fitness= 'BER', scale= 'TRUE')</pre>

Tabela 3. Linhas de comando para a modelagem de classificação de variáveis dos dados obtidos em modo de ionização positivo e negativo.

APÊNDICE C – Seleção recursiva dos *molecular features* discriminantes

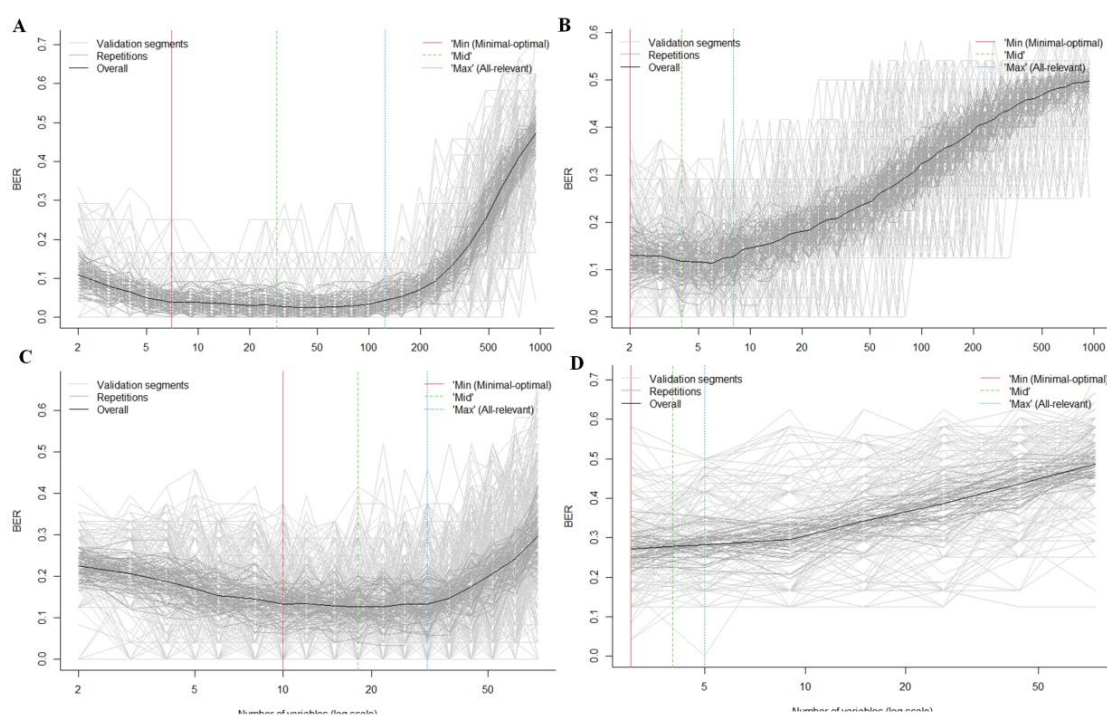


Figura 1. Gráficos de validação para modelagem de classificação MUVR para os dados obtidos em modo de ionização positivo (**A e B**) e negativo (**C e D**). As linhas cinza claro representam o desempenho de validação para os segmentos internos individuais, enquanto as linhas cinza escuro representam as curvas de validação do segmento interno em média sobre as repetições. As linhas pontilhadas em vermelho, verde e azul representam o número mínimo, médio e máximo de variáveis selecionadas, respectivamente. A taxa de erro do modelo (BER) é representada pelo eixo Y. O eixo X representa o número de variáveis incluídas para a modelagem interna. **A e C.** Método PLS. **B e D.** Método RF.

APÊNDICE C – Seleção recursiva dos *molecular features* discriminantes

Método de classificação	<i>Molecular feature</i> (+)	Ranking	<i>Molecular feature</i> (-)	Ranking
RF	m669t1120	1	m478t701	1
	m275t1120	2	m479t701	2
	m539t235	3	m476t636	3
	m311t1120	4	m423t563	4
	m609t235	5	m501t644	5
	m736t989	6	-	-
	m478t636	7	-	-
	m420t775	8	-	-
PLS	m669t1120	1	m476t636	1
	m310t1120	2	m478t701	2
	m311t1120	3	m479t701	3
	m275t1120	4	m501t644	4
	m308t477	5	m480t780	5
	m736t989	6	m1187t780	6
	m325t444	7	m1189t780	7
	m293t1120	8	m492t719	8
	m539t235	9	m615t780	9
	m482t783	10	m432t564	10
	m266t287	11	m558t802	11
	m376t671	12	-	-
	m380t645	13	-	-
	m478t636	14	-	-
	m420t775	15	-	-
	m280t548	16	-	-
	m502t704	17	-	-
	m792t1078	18	-	-
	m309t496	19	-	-
	m369t695	20	-	-
	m503t644	21	-	-
	m539t234	22	-	-
	m324t445	23	-	-
	m191t818	24	-	-
	m310t389	25	-	-
	m838t192	26	-	-
	m210t265	27	-	-
	m524t740	28	-	-
	m480t704	29	-	-

Tabela 4. Resultado da seleção de variáveis para os dados obtidos em modo de ionização positivo e negativo.

APÊNDICE C – Seleção recursiva dos *molecular features* discriminantes

Dados de entrada	Parâmetro	PLS	RF
Positivo	nOuter	5	5
	nInner	4	4
	nRep	90	80
	varRatio	0.8	0.8
	method	PLS	RF
	fitness	BER	BER
	scale	TRUE	TRUE
	model	1	3
Negativo	nPerm	100	100
	nOuter	5	5
	nInner	4	4
	nRep	80	90
	varRatio	0.85	0.8
	method	PLS	RF
	fitness	BER	BER
	scale	TRUE	TRUE
model	1	3	
nPerm	100	100	

Tabela 5. Valores para teste de permutação dos modelos gerados para classificação de variáveis dos dados obtidos em modo de ionização positivo e negativo.

APÊNDICE C – Seleção recursiva dos *molecular features* discriminantes

Dados de entrada	PLS	RF
Positivo	actual=MUVR(X = x, Y = y, nOuter = 5, nInner = 4, nRep = 90, varRatio = 0.8, method= 'PLS', fitness= 'BER', scale= 'TRUE')	actual=MUVR(X = x, Y = y, nOuter = 5, nInner = 4, nRep = 80, varRatio = 0.8, method= 'RF', fitness= 'BER', scale= 'TRUE')
	actua1Fit=actual\$miss[model] for (p in 1:nPerm) { cat('\n Permutation',p,'of',nPerm) YPerm=sample(y) perm= MUVR(X = x, Y = y, nOuter = 5, nInner = 4, nRep = 90, varRatio = 0.8, method= 'PLS', fitness= 'BER', scale= 'TRUE')	actua1Fit=actual\$miss[model] for (p in 1:nPerm) { cat('\n Permutation',p,'of',nPerm) YPerm=sample(y) perm= MUVR (X = x, Y = y, nOuter = 5, nInner = 4, nRep = 80, varRatio = 0.8, method= 'RF', fitness= 'BER', scale= 'TRUE')
Negativo	actual=MUVR(X = x, Y = y, nOuter = 5, nInner = 4, nRep = 80, varRatio = 0.85, method= 'PLS', fitness= 'BER', scale= 'TRUE')	actual=MUVR(X = x, Y = y, nOuter = 5, nInner = 4, nRep = 90, varRatio = 0.8, method= 'RF', fitness= 'BER', scale= 'TRUE')
	actua1Fit=actual\$miss[model] for (p in 1:nPerm) { cat('\n Permutation',p,'of',nPerm) YPerm=sample(y) perm= MUVR(X = x, Y = y, nOuter = 5, nInner = 4, nRep = 80, varRatio = 0.85, method= 'PLS', fitness= 'BER', scale= 'TRUE')	actua1Fit=actual\$miss[model] for (p in 1:nPerm) { cat('\n Permutation',p,'of',nPerm) YPerm=sample(y) perm= MUVR(X = x, Y = y, nOuter = 5, nInner = 4, nRep = 90, varRatio = 0.8, method= 'RF', fitness= 'BER', scale= 'TRUE')

Tabela 6. Linhas de comando para teste de permutação dos modelos gerados para classificação de variáveis dos dados obtidos em modo de ionização positivo e negativo.

Dados de entrada	PLS	RF
Positivo	0.08	0.03*
Negativo	0.003*	0.002*

Tabela 7. Resultado do teste de permutação dos modelos gerados para classificação de variáveis dos dados obtidos em modo de ionização positivo e negativo.

APÊNDICE D – Cromatogramas de íons totais da análise por LC-MS

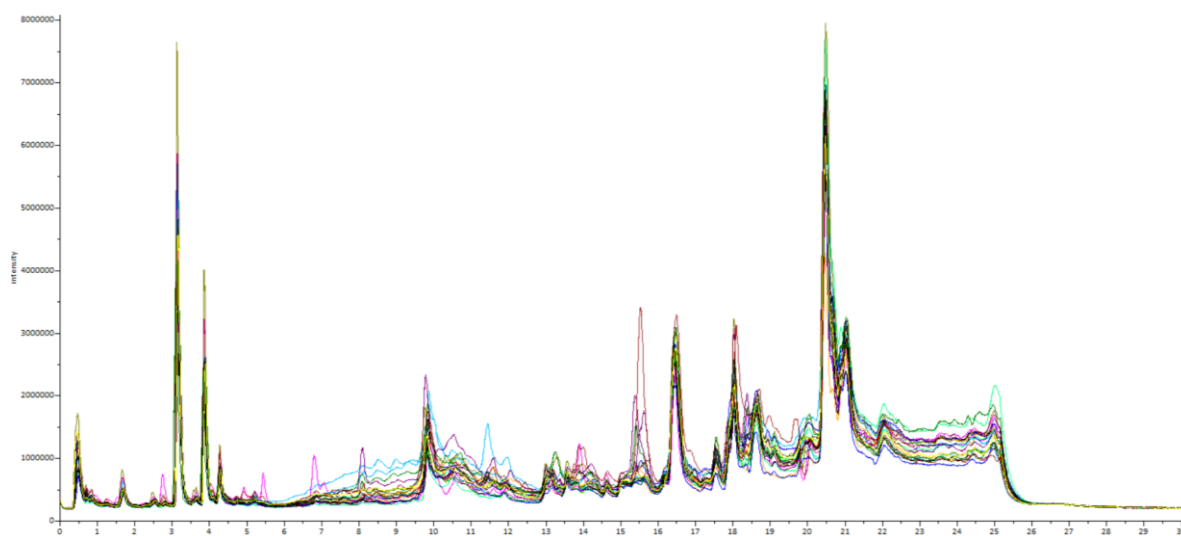


Figura 1. Cromatogramas de íons totais dos dados adquiridos em modo de ionização positivo. A sobreposição dos picos cromatográficos evidencia a estabilidade do instrumento de análise. Cada linha colorida representa uma amostra do estudo. Número total de amostras processadas: 20.

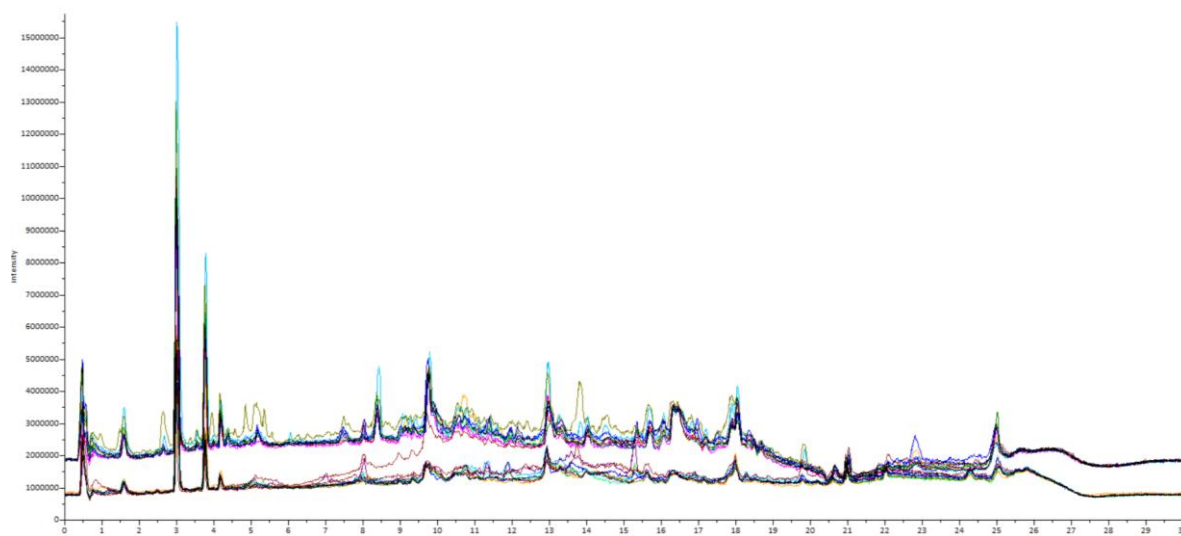


Figura 2. Cromatogramas de íons totais dos dados adquiridos em modo de ionização negativo. Uma variação de intensidade dos sinais adquiridos foi observada a partir da décima amostra analisada (linha roxa). No entanto, as amostras subsequentes acompanharam o mesmo padrão de variação. Por isso, a confiabilidade do método não foi alterada. Cada linha colorida representa uma amostra do estudo. Número total de amostras processadas: 20.

APÊNDICE E – Modelos PCA e PLS-DA

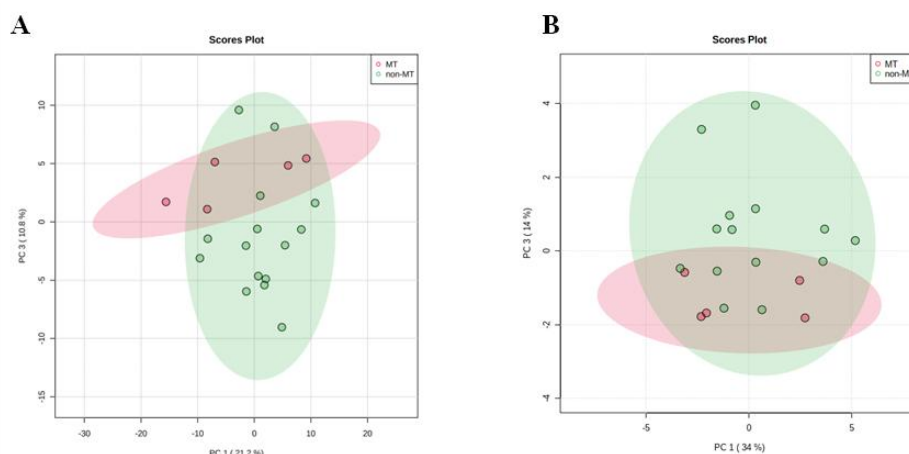


Figura 1. Análise de componente principal – PCA. Gráfico de escores entre os componentes principais 3 e 1 representando a distribuição das amostras dos dados obtidos em modo de ionização positivo (A) e negativo (B). Os pontos vermelhos representam as amostras de leucoplasia transformadas e os pontos verdes representam as amostras de leucoplasia não transformadas. O PCA mostra uma tendência para o agrupamento entre as amostras de mesmo grupo. A variância explicada por cada componente está descrita entre parênteses. As áreas em vermelho e verde possuem 95% de confiança *PC*: *principal component*. *MT*: *malignant transformed*. *non-MT*: *non-malignant transformed*. Fonte: Imagem gerada pela plataforma MetaboAnalyst

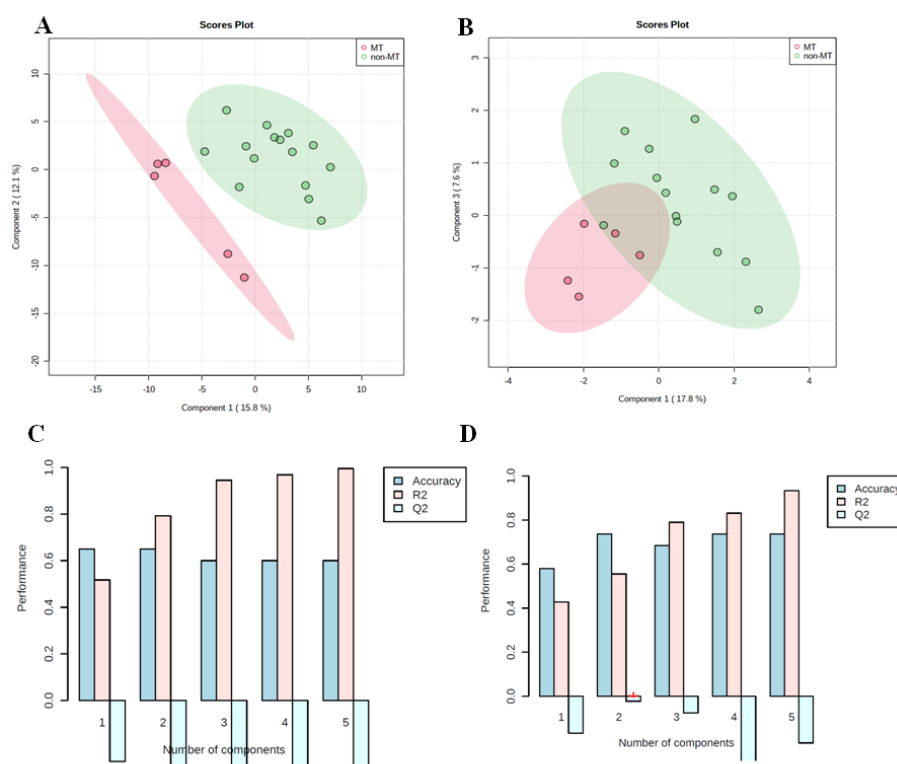


Figura 2. Análise discriminante de mínimos quadrados parciais – PLS-DA. dados obtidos em modo de ionização positivo (A) e negativo (B). A variância explicada por cada componente está descrita entre parênteses. Os pontos vermelhos representam as amostras de leucoplasia transformadas e os pontos verdes representam as amostras de leucoplasia não transformadas. As áreas em vermelho e verde possuem 95% de confiança. Os gráficos mostram uma boa separação visual entre as leucoplasias progressivas e não progressivas. No entanto, *MT*: *malignant transformed*. *non-MT*: *non-malignant transformed*. (C) e (D) Os valores de Q^2 para cada componente são representados pelas barras verdes. Para ambos os modelos, os valores foram negativos, o que sugere sobreajuste.

Fonte: Imagem gerada pela plataforma MetaboAnalyst

APÊNDICE F – Análise de enriquecimento de vias metabólicas

Mummichog + GSEA 10ppm						
Vias enriquecidas	<i>Total_Size</i>	<i>Hits</i>	<i>Sig_Hits</i>	<i>Mummichog_Pvals</i>	<i>GSEA_Pvals</i>	<i>Combined_Pvals</i>
Carnitine shuttle	72	25	3	0.2866	0.0099	0.0195
Vitamin E metabolism	54	15	1	0.6935	0.0500	0.1512
Omega-3 fatty acid metabolism	39	7	1	0.4220	0.0889	0.1607
Squalene and cholesterol biosynthesis	55	15	2	0.3118	0.1900	0.2266
Di-unsaturated fatty acid beta-oxidation	26	4	1	0.2684	0.3333	0.3054
Lysine metabolism	52	10	1	0.5439	0.1667	0.3083
C21-steroid hormone biosynthesis and metabolism	112	47	2	0.8862	0.1188	0.3423
Fatty acid activation	74	32	3	0.4363	0.2871	0.3855
Glycosphingolipid metabolism	67	11	1	0.5786	0.2268	0.3978
Linoleate metabolism	46	19	2	0.4241	0.8673	0.7357
Vitamin A (retinol) metabolism	67	30	1	0.9089	0.4200	0.7494
Fatty Acid Metabolism	63	13	1	0.6405	0.6300	0.7697
Bile acid biosynthesis	82	38	2	0.7980	0.6535	0.8610
De novo fatty acid biosynthesis	106	13	1	0.6405	0.8300	0.8675
Prostaglandin formation from arachidonate	78	42	3	0.6270	0.9990	0.9194
Glycerophospholipid metabolism	156	18	1	0.7589	0.9082	0.9458
Aspartate and asparagine metabolism	114	20	1	0.7947	0.9990	0.9771
Leukotriene metabolism	92	23	1	0.8389	0.9990	0.9861
Mummichog + GSEA – 5 ppm						
Vitamin E metabolism	54	5	2	0.0702	0.1163	0.0474
Fatty Acid Metabolism	63	4	1	0.3253	0.0610	0.0976
De novo fatty acid biosynthesis	106	8	1	0.5476	0.0714	0.1659
Fatty acid activation	74	16	1	0.8003	0.0990	0.2801
Squalene and cholesterol biosynthesis	55	13	1	0.7278	0.1667	0.3772
C21-steroid hormone biosynthesis and metabolism	112	31	1	0.9600	0.1386	0.4014
Di-unsaturated fatty acid beta-oxidation	26	3	1	0.2552	0.8571	0.5511
Vitamin A (retinol) metabolism	67	18	1	0.8379	0.3095	0.6093
Carnitine shuttle	72	11	1	0.6661	0.4086	0.6263

Linoleate metabolism	46	4	1	0.3253	0.8902	0.6485
Prostaglandin formation from arachidonate	78	8	1	0.5476	0.8506	0.8216
Glycerophospholipid metabolism	156	10	1	0.6304	0.8229	0.8592
Bile acid biosynthesis	82	12	1	0.6985	0.7551	0.8648
Leukotriene metabolism	92	10	1	0.6304	0.8889	0.8849
Mummichog + GSEA – 3ppm						
Prostaglandin formation from arachidonate	78	8	3	0.0577	0.7204	0.1738
Vitamin E metabolism	54	4	1	0.4044	0.1512	0.2320
C21-steroid hormone biosynthesis and metabolism	112	24	1	0.9628	0.1020	0.3261
Vitamin A (retinol) metabolism	67	24	1	0.9628	0.1183	0.3614
Squalene and cholesterol biosynthesis	55	8	1	0.6494	0.2778	0.4893
Leukotriene metabolism	92	10	1	0.7321	0.2500	0.4938
Mummichog + GSEA – 1ppm						
C21-steroid hormone biosynthesis and metabolism	112	10	1	0.9133	0.0267	0.1148
Prostaglandin formation from arachidonate	78	3	2	0.1051	0.6800	0.2600
Vitamin E etabolism	54	1	1	0.2048	0.8696	0.4854
Vitamin A (retinol) metabolism	67	13	1	0.9613	0.2794	0.6217

Tabela 1. Vias metabólicas enriquecidas na análise combinada pelos algoritmos *mummichog* e GSEA. *Total_Size*: quantidade de compostos da via. *Hits*: Número de compostos constituintes da via possivelmente presentes nas amostras de leucoplasia analisadas. *Sig_Hits*: número de compostos da via potencialmente presentes nas amostras de leucoplasia analisadas. *Mummichog_Pvals*: valores de *p* resultantes da análise pelo algoritmo *mummichog*. *GSEA_Pval*: valores de *p* resultantes da análise pelo algoritmo GSEA. *Combined_Pvals*: valores de *p* ajustados aos dados do estudo a partir da combinação dos dois algoritmos em uma distribuição Gama.

ANEXO A – Parecer consubstanciado do Comitê de Ética em Pesquisa (CEP)

UNIVERSIDADE FEDERAL DE
MINAS GERAIS



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: PERFIL METABOLÔMICO DA LEUCOPLASIA ORAL COM E SEM TRANSFORMAÇÃO MALIGNA

Pesquisador: Ricardo Santiago Gomez

Área Temática:

Versão: 2

CAAE: 40027620.0.0000.5149

Instituição Proponente: UNIVERSIDADE FEDERAL DE MINAS GERAIS

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 4.572.167

Apresentação do Projeto:

Este projeto de pesquisa visa avaliar o perfil metabólico de amostras de leucoplasia oral progressiva e não progressiva com o intuito de determinar as principais alterações no metabolismo da leucoplasia e definir possíveis biomarcadores para o diagnóstico precoce e predição do seu potencial de transformação maligna. Trata-se de estudo retrospectivo, com população-alvo composta por 15 indivíduos com diagnóstico histopatológico de leucoplasia oral e acompanhamento clínico de dois a doze anos. Para cada um dos casos, serão coletadas informações clínicas de idade, sexo, localização anatômica, grau de displasia, tratamento, recorrência e progressão maligna. A partir de blocos de parafina contendo tecido oriundos de biópsias realizadas na clínica de Estomatologia, Patologia e Radiologia Oral da Faculdade de Odontologia da UFMG serão realizados: análise histopatológica, extração de metabólitos e análise metabólica em cromatógrafo líquido acoplado a espectrômetro de massas.

Hipótese:

Existem alterações metabólicas entre as leucoplasias orais progressivas e não progressivas que podem ajudar na compreensão do processo de carcinogênese oral.

Endereço: Av. Presidente Antônio Carlos, 6627 2º Ad Sl 2005

Bairro: Unidade Administrativa II **CEP:** 31.270-901

UF: MG **Município:** BELO HORIZONTE

Telefone: (31)3409-4592

E-mail: coep@prpq.ufmg.br

ANEXO A – Parecer consubstanciado do Comitê de Ética em Pesquisa (CEP)

UNIVERSIDADE FEDERAL DE
MINAS GERAIS

Continuação do Parecer: 4.572.167

Assentimento / Justificativa de Ausência	TCLE_new.pdf	14:41:31	Gomez	Aceito
Declaração de Manuseio Material Biológico / Biorepositório / Biobanco	constituicao_biorrepositorio.pdf	03/02/2021 14:41:17	Ricardo Santiago Gomez	Aceito
Folha de Rosto	folha_de_rosto.pdf	09/11/2020 08:00:41	Ricardo Santiago Gomez	Aceito
Outros	parecer_departamental.pdf	09/11/2020 08:00:11	Ricardo Santiago Gomez	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

BELO HORIZONTE, 03 de Março de 2021

Assinado por:
Críssia Carem Paiva Fontainha
(Coordenador(a))

Endereço: Av. Presidente Antônio Carlos, 6627 2º Ad SI 2005
 Bairro: Unidade Administrativa II CEP: 31.270-901
 UF: MG Município: BELO HORIZONTE
 Telefone: (31)3409-4592 E-mail: coep@prpq.ufmg.br

ANEXO B – Atividades desenvolvidas durante o doutorado

FORMAÇÃO COMPLEMENTAR

1. Curso de curta duração em Curso para Usuários de GC e GC-MS: fundamentos e práticas avançadas. (Carga horária: 8h). Agilent Technologies Brasil, AGILENT, Barueri, Brasil.
2. Curso de curta duração em Curso de Verão em Bioinformática. (Carga horária: 24h). Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil.
3. Curso de curta duração em I Curso Básico de Análises Metabolômicas. (Carga horária: 16h). Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil.
4. Curso de curta duração em curso de extensão Software Livre R project: Princípios Básicos. (Carga horária: 8h). Instituto Federal Do Ceará, IFCE, Farias Brito, Brasil
5. Curso de curta duração em I Curso Avançado de Análises Metabolômicas. (Carga horária: 40h). Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil

PARTICIPAÇÃO EM EVENTOS

1. Conferencista na I Semana das Especialidades Odontológicas, 2019. (Encontro) Alterações Patológicas da Cavidade Oral: Estomatologia Aplicada à Prática Clínica
2. Apresentação de Poster / Paineis no(a) 45º Congresso Brasileiro de Estomatologia e Patologia Oral, 2019. (Congresso). CONSERVATIVE MANAGEMENT OF CENTRAL GIANT CELL LESION OF THE JAW: A CASE REPORT.
3. 1st Symposium On Mass Spectrometry Applied To Proteomics and Structural Biology, 2019. (Simpósio)
4. 46º Congresso Brasileiro de Estomatologia e Patologia Oral, 2020. (Congresso Online).
5. Apresentação Oral no El VIII Simposio Internacional Cáncer Oral – Online, 2020. (Congresso). BRAFV600E mutation assessment as an auxiliary tool for odontogenic lesions diagnosis.
6. Congress on Oral Pathology and Oral Medicine Online meeting, IAOP, 2021. (Congresso Online)
7. 47º Congresso Brasileiro de Estomatologia e Patologia Oral, 2021. (Congresso Online).

8. Apresentação de Poster / Painel no 47º Congresso Brasileiro de Estomatologia e Patologia Oral, 2021. (Congresso). UNICYSTIC VARIANT OF ADENOID AMELOBLASTOMA WITH DENTINOID?
9. Apresentação de Poster / Painel no 47º Congresso Brasileiro de Estomatologia e Patologia Oral, 2021. (Congresso). ORAL LEUKOPLAKIA WITH AND WITHOUT MALIGNANT TRANSFORMATION ARE METABOLICALLY DISTINCT

REVISOR DE PERIÓDICO

1. Journal of Oral Pathology and Medicine – mar/2020, abr/2020, set/2020, out/2020
2. Oral Diseases – Set/2020
3. Frontiers in Oral Health – Ago/2020

PROJETOS DE PESQUISA

1. COVID-19 e mudanças de Biossegurança em Odontologia
2. Detecção do SARS-COV-2 em espécimes de biópsias de tecidos por RT-qPCR – Nova tecnologia para monitoramento e identificação de surtos de infecções e disseminação do vírus. Financiador: Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG

PRODUÇÃO BIBLIOGRÁFICA

1. LEITE-LIMA, Flávia, et al. Unveiling metabolic changes in marsupialized odontogenic keratocyst: A pilot study. Oral Diseases. Epub ahead of print, 2021.
2. BASTOS, Victor Coutinho, et al. Age-related metabolic pathways changes in dental follicles: a pilot study. Frontiers in Oral Health, v.2, p.31, 2021.
3. MARTINS-CHAVES, Roberta Rayra et al. Glandular Odontogenic Cyst—A Case Series. **Journal of Oral and Maxillofacial Surgery**, 2020.
4. SANT’ANA, Maria Sissa Pereira et al. BRAF p. V600E status in epithelial areas of ameloblastoma with different histological aspects: implications to the clinical practice. **Journal of Oral Pathology & Medicine**, 2020.

5. MARTINS-CHAVES, Roberta Rayra; GOMES, Carolina Cavaliéri; GOMEZ, Ricardo Santiago. Immunocompromised patients and coronavirus disease 2019: a review and recommendations for dental health care. **Brazilian oral research**, v. 34, 2020.
6. VITÓRIO, Jéssica Gardone et al. Metabolic landscape of oral squamous cell carcinoma. **Metabolomics**, v. 16, n. 10, p. 1-18, 2020.
7. MARTINS-CHAVES, Roberta Rayra et al. KRAS mutations in implant-associated peripheral giant cell granuloma. **Oral diseases**, v. 26, n. 2, p. 334-340, 2020.
8. SCHUCH, Lauren Frenzel et al. A Brazilian multicentre study of 2,497 isolated cases of odontogenic keratocysts. **Oral diseases**, v. 26, n. 3, p. 711-715, 2020.