

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

**Perfil Metabolômico de Amostras de Pacientes
Diagnosticados com Doença de Wilson**

Thais de Assis Lopes*

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Ciências, área de concentração: Química.

Orientador: Prof^a. Dra. Regina Vincenzi Oliveira

Co-orientador: Prof. Dr. Tiago Venâncio

* bolsista CNPq

**São Carlos - SP
2024**



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Química

Folha de Aprovação

Defesa de Tese de Doutorado da candidata Thais de Assis Lopes, realizada em 27/05/2024.

Comissão Julgadora:

Prof. Dr. Tiago Venâncio (UFSCar)

Prof. Dr. Álvaro José dos Santos Neto (IQSC/USP)

Prof. Dr. Marcone Augusto Leal de Oliveira (UFJF)

Profa. Dra. Gisele André Baptista Canuto (UFBA)

Prof. Dr. Luiz Juliano Neto (UNIFESP)

“Cem vezes todos os dias lembro a mim mesmo que minha vida interior e exterior, depende dos trabalhos de outros homens, vivos ou mortos, e que devo esforçar-me a fim de devolver na mesma medida que recebi”

Albert Einstein

Dedico esta Tese à minha mãe e a minha
irmã, que sempre me incentivaram e
apoiaram, mesmo de longe, a nunca
desistir dos meus sonhos.

AGRADECIMENTOS

À minha orientadora, Prof^a. Dra. Regina Vincenzi Oliveira, gostaria de expressar minha mais profunda gratidão por ter aceitado me orientar. Seu incentivo constante e sua infinita paciência foram inestimáveis em todos os momentos. Agradeço imensamente por compartilhar seus preciosos ensinamentos e conselhos, tanto profissionais quanto pessoais. Não tenho dúvidas de que você contribuiu significativamente para o meu desenvolvimento não apenas como profissional, mas também como pessoa.

À minha mãe e minha irmã, as mulheres da minha vida, que sempre foram fonte constante de inspiração para mim. Agradeço do fundo do coração por todo o incentivo, amor, apoio e dedicação que vocês sempre me proporcionaram. São os pilares que me sustentam, dando-me forças para enfrentar os desafios e jamais desistir. Suas presenças são um constante lembrete do poder do amor e da união familiar.

Ao meu pai e ao meu tio Célio, embora não estejam mais fisicamente presentes, sinto a presença do amor e apoio que sempre me ofereceram. Suas assistências e carinhos deixaram marcas profundas em minha vida, moldando-me de maneiras que eu jamais esquecerei.

A todos que, de alguma forma, estiveram ao meu lado e contribuíram para a realização deste trabalho, especialmente a Juliana Magalhães, Alex Castro e Álex Rosini, que sempre estiveram à disposição para me ajudar. Agradeço profundamente pela dedicação e apoio oferecidos ao longo deste processo. As contribuições de vocês foram fundamentais para a realização deste trabalho, e sou imensamente grata pela generosidade e disponibilidade demonstradas.

Aos meus amigos do Núcleo *Separare*: Julia Lima, Welma Beatriz, Larissa Ramos, Izadora Furlani, Juliana Rodrigues, Cleiton Leandro, e aos agregados do PN: Luciano Pereira, Marília Pires, Marina Zequin e João Gabriel, o apoio, a amizade e o sentimento de família. Vocês foram essenciais em minha trajetória. Agradeço por cada gesto de companheirismo e por tornarem minha experiência tão especial e significativa.

Aos meus eternos amigos de VR, por estarem ao meu lado em todos os momentos, mesmo quando distância física nos separa. Em especial, desejo reconhecer o apoio constante de Juliana Neves, Filipe Lopes, Nilton Mattos e Carolina Possas. A amizade de vocês me fortalece e enriquece minha jornada.

À Sophia e ao tintim, por todo companheirismo, amizade, apoio, palavras de afirmação, presença e lambeijos do tintim nos momentos difíceis e também felizes, vocês estão para sempre no meu coração.

Ao Prof. Dr. Eduardo Luiz R. Cançado e ao Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, pela colaboração e obtenção das amostras clínicas.

À UFSCar, ao Departamento de Química e ao seu corpo docente e administrativo pela competência e contribuição.

À CNPQ pela bolsa de doutorado concedida (Processo: 141435/2020-8 Chamada/Edital: GM/GD – Cotas do Programa de Pós-Graduação) e a FAPESP pelos auxílios financeiros (Processos: 14/50244-6, 16/22215-7 e 20/05965-8).

Lista de Abreviatura

AASLD – Associação Americana para o Estudo de Doenças Hepáticas

ACN – Acetonitrila

ANOVA - Análise de Variância

ANCOVA - Análise de Covariância

APCI - Ionização Química à Pressão Atmosférica

APPI - Fotoionização à pressão atmosférica

ALT - Alanina Transaminase

ALD - Doenças Hepáticas Alcoólicas

AST - Aspartato Aminotransferase

ASS - Argininosuccinato Sintase

ATOX1 - Chaperona Antioxidante de Cobre 1

ATP7B - ATPase do Tipo P Transportadora de Cobre

AUC - Área Sob a Curva

CB – Controle Branco

CBP - Colestase Biliar Primária

CCS1 - Chaperona de Cobre para Superóxido Dismutase 1

COX17 - Chaperonas de Cobre para o Citocromo C Oxidase

Cp - Ceruloplasmina

CT – Tomografia Computadorizada

Ctr1 – Proteína Transportadora de Cobre

Cu - Cobre

CuExc – Nível Cu-trocável

CV – Coeficiente de Variação

DART - Análise Direta em Tempo Real

DDA – Aquisição Dependente de Dados

DESI - Dessorção via *electrospray*

DW- Doença de Wilson

EASL - Associação Europeia para Estudo do Fígado

ESI- Ionização por *Electrospray*

ESI(+)- *Electrospray* no Modo Positivo de Ionização

ESI(-)- *Electrospray* no Modo Negativo de Ionização

ESPGHAN - Sociedade Europeia de Pediatria Gastroenterologia, Hepatologia e Nutrição

FC- *Fold-change*

k- Fator de Retenção Cromatográfico

FDR – Taxa de Falsa Descoberta

GC- Grupo controle

GC-MS - Cromatografia Gasosa Acoplada a Espectrometria de Massas

HC-FMUSP – Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo

HILIC - Cromatografia de Interação Hidrofílica

HILIC ESI(+) - Cromatografia de Interação Hidrofílica com Ionização *Elestrospray* no Modo Positivo de Ionização

HILIC ESI(-) - Cromatografia Líquida de Interação Hidrofílica com Ionização *Elestrospray* no Modo Negativo de Ionização

HMDB – Banco de Dados do Metaboloma Humano

HRMS-Espectrometria de Massas de Alta Resolução

IT – Analisador de Massas *Ion trap*

KEGG – Enciclopédia Kyoto de Genes e Genomes

K-F – Anel de Kayser-Fleischer

KNN – *k*-Vizinhos Mais Próximos

LOWESS/Spline - *Locally Weighted Scatterplot Smoothing*

LOESS - *Locally Estimated Scatterplot Smoothing*

LC - Cromatografia Líquida

LC-HRMS - Cromatografia Líquida Acoplada à Espectrometria de Massas de Alta Resolução

LC-MS - Cromatografia Líquida Acoplada à Espectrometria de Massas

LC-MS/MS - Cromatografia Líquida Acoplada à Espectrometria de Massas em *tandem*

LOOCV – Validação Cruzada *leave-one-out*

MALDI - Ionização por Dessorção a Laser Assistida por Matriz

MeOH- Metanol

MoNA - Banco de Massas da América do Norte

m/z- Razão massa/carga

min- Unidade de Tempo (minutos)

mm-Unidade de Distância (milímetro)

MS - Espectrometria de Massas

MSTUS - *MS total useful signal*

NAFLD – Doença Esteatótica Não Alcoólica do Fígado NAFLD

nm-Unidade de Distância (nanômetro)

OPLS-DA - Análise de Componentes Principais por Projeção Ortogonal

PCA - Análise de Componentes Principais

PAH - Fenilalanina Hidroxilase

PLS-DA - Análise Discriminante dos Mínimos Quadrados Parciais

PQN - *probabilistic quotient normalization*

QB - Controle de Qualidade da Amostra Branco (QB)

QC - Controle de Qualidade

QC-RLSC - *Locally Estimated Scatterplot Smoothing*

QqTOF- Analisador de Massas Híbrido Quadrupolo-Tempo de Voo

QTRAP - Analisador de Massas Híbrido Quadrupolo-*Ion Trap*

Q² - Coeficiente de Predição

r - Coeficiente de Correlação de Pearson

R² - Coeficiente de Determinação

REC – Nível Relativo de Cu-trocável

RMN – Ressonância Magnética Nuclear

ROC – Característica de Operação do Receptor

RPLC – Cromatografia Líquida de Fase Reversa

RPLC ESI(+) - Cromatografia Líquida de Fase Reversa com Ionização *Elektrospray* no Modo Positivo de Ionização

RPLC ESI(-) - Cromatografia Líquida de Fase Reversa com Ionização *Elektrospray* no Modo Negativo de Ionização

RSD – Desvio Padrão Relativo

SD – Desvio Padrão

SIG – Significativo

S/N – Razão sinal/ruído

SNP - Polimorfismos de Nucleotídeo Único

SOD1 - Superóxido Dismutase de Cu/Zn

TC – Tomografia Computadorizada

TQ- Analisador de Massas Triplo Quadrupolo

Tr - Tempo de Retenção

TGN - Rede Trans de Golgi

UHPLC - Cromatografia Líquida de Ultra Eficiência

VIP - Importância da Variável na Projeção

\bar{X}_{QC} – Média dos Valores das Amostras de Controle de Qualidade

\bar{X}_{QB} – Média dos Valores das Amostras de Controle de Qualidade das Amostras

Branco

Zn - Zinco

Lista Tabela

Tabela 1- Sumário dos principais parâmetros e valores de referência no diagnóstico da DW	10
Tabela 2- Classificação do coeficiente de correlação de Pearson ¹²⁰	39
Tabela 3- Informações demográficas das amostras de plasma e urina que integraram o grupo estudo (DW) e o grupo controle (GC).....	41
Tabela 4- Descrição do número de <i>features</i> moleculares em cada modo de eluição e de ionização empregado nas análises por LC-HRMS, detectados e extraídos e após o pré-processamento dos dados.	46
Tabela 5 - Metabólitos discriminantes identificados com $VIP \geq 1,0$, $p \leq 0,05$ e Taxa de Falsa Descoberta (FDR) $< 0,10$	71
Tabela 6- Metabólitos discriminantes dos VIPS scores obtidos dos modelos PLS-DA provenientes das análises das amostras de plasma e urina	82
Tabela A1- Resultado do teste de <i>t-Student</i> para a análise estatística referente à idade dos indivíduos dos dois grupos (DW e HC), nas amostras de plasma	126
Tabela A2- Resultado do teste de Qui-quadrado para a análise estatística referente ao gênero dos indivíduos dos dois grupos (DW e HC), nas amostras de plasma ..	126
Tabela A3- Determinação dos <i>features</i> moleculares significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de plasma no modo RPLC ESI(+/-) de eluição (continuação).	127
Tabela A4- Determinação dos <i>features</i> moleculares significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de plasma no modo HILIC ESI(+/-) de eluição (continuação).	131
Tabela A5- Resultado do teste de <i>t-Student</i> para a análise estatística referente à idade dos indivíduos dos dois grupos (DW e HC), nas amostras de urina.....	139
Tabela A6- Resultado do teste de Qui-quadrado para a análise estatística referente ao gênero dos indivíduos dos dois grupos (DW e HC), nas amostras de urina.....	139
Tabela A7- Determinação dos <i>features</i> significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de urina no modo RPLC ESI(+/-) de eluição (continuação).	140
Tabela A8- Determinação dos <i>features</i> significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de urina no modo HILIC ESI(+/-) de eluição (continuação).	143

Lista de Figuras

Figura 1- Distribuição de $\text{Cu}^+/\text{Cu}^{2+}$ no hepatócito, com acúmulo do metal e radicais hidroxila	5
Figura 2- Características Clínicas da DW - (A) coloração amarelada da pele; (B) coloração amarelada dos olhos; (C) fígado com cirrose; (D) anéis de Kayser-Fleischer. A, B e C: Biorender D: <i>Image Creator from Designer. Powered by DAA E 3</i>	7
Figura 3- Estruturas químicas dos fármacos empregados no tratamento da DW....	12
Figura 4- Evolução do número de artigos publicados na plataforma “ <i>pubmed</i> ” com o termo “ <i>metabolomic</i> ”. Pesquisa realizada em 12/09/2023.....	13
Figura 5- Representação de fatores que podem influenciar o fenótipo de um indivíduo.	14
Figura 6- Classificação das abordagens metabolômicas.....	16
Figura 7- Estratégias metabolômicas na descoberta e validação de biomarcadores	17
Figura 8- Representação de componentes de um Espectrômetro de Massas.	22
Figura 9- Esquema simplificado do preparo das amostras de plasma.	32
Figura 10- Esquema simplificado de extração das amostras de urina.....	33
Figura 11- Cromatogramas de picos extraídos de LC-HRMS representativos das análises da amostra de controle de qualidade (QC) de plasma no (A) modo RPLC ESI(+) e (B) RPLC ESI (-).	53
Figura 12- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de <i>scores plots</i> 2D e 3D baseadas nos modelos PCA, no modo RPLC ESI(+) combinado com RPLC ESI(-).	54
Figura 13- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de <i>score plots</i> 2D do modelo OPLS-DA no modo RPLC ESI(+) combinado com RPLC ESI(-), teste de permutação do modelo OPLS-DA e importância da variável na projeção (VIP).....	56
Figura 14- Cromatogramas de picos extraídos de LC-HRMS representativos das análises da amostra de controle de qualidade (QC) de plasma no (A) modo HILIC ESI(+) e (B) HILIC ESI (-).....	57
Figura 15- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de <i>scores plots</i> 2D e 3D baseadas nos modelos PCA, no modo HILIC ESI(+) combinado com HILIC ESI(-).....	58
Figura 16- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de <i>score plots</i> 2D do modelo OPLS-DA no modo HILIC ESI(+) combinado com HILIC ESI(-), teste de permutação do modelo OPLS-DA e importância da variável na projeção (VIP).....	60
Figura 17- Cromatogramas de picos extraídos de LC-HRMS ilustrativos das análises da amostra de controle de qualidade (QC) de plasma no (A) modo RPLC ESI(+) e (B) RPLC ESI (-).....	61
Figura 18- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de <i>scores plots</i> 2D e 3D baseadas nos modelos PCA, no modo RPLC ESI(+) combinado com RPLC ESI(-).	62

Figura 19- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de <i>score plots</i> 2D do modelo OPLS-DA no modo RPLC ESI(+) combinado com RPLC ESI(-), teste de permutação do modelo OPLS-DA e importância da variável na projeção (VIP).....	64
Figura 20- Cromatogramas de picos extraídos de LC-HRMS ilustrativos das análises da amostra de controle de qualidade (QC) de plasma no (A) modo HILIC ESI(+) e (B) HILIC ESI (-).....	65
Figura 21- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de <i>scores plots</i> 2D e 3D baseadas nos modelos PCA, no modo HILIC ESI(+) combinado com HILIC ESI(-).....	66
Figura 22- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de <i>score plots</i> 2D do modelo PLS-DA no modo HILIC ESI(+) combinado com HILIC ESI(-), teste de permutação do modelo PLS-DA e importância da variável na projeção (VIP). (A) <i>Score plots</i> 2D das amostras de urina dos grupos DW e GC;.....	68
Figura 23- Curvas ROC pelo método estatístico multivariado no ponto de corte ideal definido pela distância mínima para o canto superior esquerdo. (A) Modelo otimizado gerou uma AUC de 0,995 (IC 95%: 0,956–1) a partir de amostras de plasma. (B) Modelo otimizado gerou uma AUC de 0,976 (IC 95%: 0,944–0,976) a partir de amostras de urina. <i>Software: MetaboAnalyst</i>	78
Figura 24- Resultados da Curva ROC método classificatório PLS-DA para a análise das amostras de plasma. A) exatidão preditiva, B) Curva ROC, modelo 3 metabólitos C) predição de classes e D) gráfico de <i>VIP scores</i> . <i>Software: MetaboAnalyst</i>	79
Figura 25- Resultados da Curva ROC método classificatório PLS-DA para a análise das amostras de urina. A) exatidão preditiva; B) Curva ROC, modelo 2 metabólitos; C) predição de classes; D) gráfico de <i>VIP scores</i> . <i>Software: MetaboAnalyst</i>	81
Figura 26- Gráficos circulares descrevendo as superclasses (A) e as subclasses (B) dos metabólitos putativamente anotados que alcançaram valores de <i>fold-change</i> no intervalo de $0,7 \geq FC \geq 1,5$	83
Figura 27- Mapeamento das vias metabólicas impactadas entre os grupos DW e GC em amostras de plasma e urina. O eixo x representa o impacto da via metabólica, enquanto o eixo y representa o $-\log(p\text{-value})$. A cor varia de amarelo claro a vermelho escuro, indicando o nível de significância (<i>p</i>) do enriquecimento e o tamanho do círculo reflete o nível da taxa de enriquecimento (impacto). <i>Software MetaboAnalyst</i>	85
Figura 28- Estrutura química da molécula de Arginina	86
Figura 29- Via Metabólica da Síntese da Arginina.....	89
Figura 30- <i>Boxplots (box-and-whisker)</i> para (A) Glutamato (Glu); (B) Ácido glutâmico (GluT); (C) Fenilacetilglutamina (PAG); (D) Alanil-Hidroxiprolina (Alanil-Hyp); (E) Aspartil-Hidroxiprolina (Aspartil-Hyp); (F) Glicil-Hidroxiprolina (Glicil-Hyp); (G) Citrulina e (H) Ácido aspártico (Asp), em amostras de indivíduos com DW vs. GC. Metabólitos discriminantes identificados com $VIP \geq 1,0$, $p \leq 0,05$ e $FDR < 0,10$	91
Figura 31- Reações de transaminação envolvendo o par α -cetoglutarato/glutamato como um receptor/doador de grupo amino.....	92
Figura 32- Ciclo Glicose-Alanina.	93
Figura 33- Biossíntese do Aspartato.....	94
Figura 34- Estrutura química da fenilalanina, tirosina e triptofano.	96

Figura 35- Conversão de fenilalanina em tirosina.....	97
Figura 36- Reações do catabolismo da tirosina, com produção de fumarato e acetacetato.....	98
Figura 37- Conversão de fenilalanina e tirosina em catecolaminas neurotransmissores.....	99
Figura 38- Metabolismo do Triptofano via Biossíntese da Melatonina e Serotonina	100
Figura 39- Metabolismo do Triptofano via Biossíntese da Quinurenina.....	101
Figura 40- <i>Boxplots (box-and-whisker)</i> para (A) Tirosina (Glu); (B) Ácido quinurênico (KU) em amostras de indivíduos com DW vs. GC. Metabólitos discriminantes identificados com $VIP \geq 1,0$, $p \leq 0,05$ e $FDR < 0,10$	102
Figura 41- Estrutura química do ácido linoleico	103
Figura 42- Metabolismo do Ácido Linoléico (18:3n-3).....	103
Figura 43- <i>Boxplots (box-and-whisker)</i> para (A) Ácido linoleico; (B) Ácido δ -linoleico; (C) Ácido araquidônico em amostras de indivíduos com DW vs. GC. Metabólitos discriminantes identificados com $VIP \geq 1,0$, $p \leq 0,05$ e $FDR < 0,10$	104
Figura 44- Identificação do ácido quinurênico por meio da comparação entre o espectro experimental obtido nas análises LC-HRMS e o espectro de massas teórico disponível na biblioteca do MoNA.	147
Figura 45- Identificação do ácido cinâmico por meio da comparação entre o espectro experimental obtido nas análises LC-HRMS e o espectro de massas teórico disponível na biblioteca do MoNA.	148
Figura 46- Identificação do ácido aspártico por meio da comparação entre o espectro experimental obtido nas análises LC-HRMS e o espectro de massas teórico disponível na biblioteca do MoNA.	149
Figura 47- Identificação da citrulina por meio da comparação entre o espectro experimental obtido nas análises LC-HRMS e o espectro de massas teórico disponível na biblioteca do MoNA.	150
Figura 48- Identificação do ácido linoleico por meio da comparação entre o espectro experimental obtido nas análises LC-HRMS e o espectro de massas teórico disponível na biblioteca do MoNA.	151
Figura 49- Identificação do LPE (18:2) por meio da comparação entre o espectro experimental obtido nas análises LC-HRMS e o espectro de massas teórico disponível na biblioteca do MoNA.	152

RESUMO

PERFIL METABOLÔMICO DE PACIENTES DIAGNOSTICADOS COM DOENÇA DE WILSON. A Doença de Wilson (DW) ou degeneração hepatolenticular, é uma doença genética rara, de caráter autossômico recessivo, envolvendo o gene ATP7B. A DW é caracterizada por um distúrbio no metabolismo do cobre (Cu), ocasionando o acúmulo do micronutriente em diversos tecidos do corpo humano, incluindo fígado, cérebro, rins e córneas. Quando não diagnosticada e tratada, a DW progride para doenças hepáticas e disfunção do sistema nervoso central, podendo até resultar em danos fatais. Com isso, o diagnóstico precoce da DW em estágio inicial é fundamental para reduzir as complicações clínicas da doença. Não existem testes bioquímicos específicos para o diagnóstico da DW, mas sim uma combinação de características clínicas que auxiliam no diagnóstico, dentre eles, baixos níveis de ceruloplasmina, aumento da excreção urinária de Cu, níveis elevados de Cu hepático e presença de depósitos de Cu na córnea. Neste trabalho, utilizou-se de uma abordagem metabolômica global com a técnica de cromatografia líquida acoplada à espectrometria de massa de alta resolução (LC-HRMS) para identificar alterações metabólicas em amostras de plasma e urina de indivíduos com DW em comparação a um grupo de indivíduos saudáveis. Os resultados estatísticos mostraram a obtenção de 98 *molecular features* discriminantes nas amostras de urina ($p \leq 0,05$ e $VIP \geq 1$) e 243 *molecular features* discriminantes nas amostras de plasma ($p \leq 0,05$ e $VIP \geq 1$). Desses, um total de 139 metabólitos foram anotados putativamente e avaliados quanto às alterações nas vias bioquímicas, sendo essas relacionadas à biossíntese da arginina, metabolismo do ácido linoleico, metabolismo da alanina, aspartato e glutamato e biossíntese de fenilalanina, tirosina e triptofano. Os resultados obtidos são uma contribuição para o avanço das pesquisas na descoberta de biomarcadores para doenças raras e negligenciadas. Os resultados obtidos auxiliam na compreensão dos potenciais mecanismos patogênicos da DW e promovem direcionamentos a serem empregados em futuros diagnósticos mais precisos para indivíduos. Os resultados deste trabalho têm o potencial de contribuir com o diagnóstico e o tratamento de doenças raras e negligenciadas e abrir caminho para futuras investigações na descoberta de novos biomarcadores.

Palavras-chave: Doença de Wilson, doenças raras, metabolômica, doença hepática, espectrometria de massas de alta resolução

ABSTRACT

METABOLOMIC PROFILE OF PATIENTS DIAGNOSED WITH WILSON'S DISEASE.

Wilson's Disease (WD) or hepatolenticular degeneration, is a rare genetic disorder caused by the ATP7B gene. WD causes the accumulation of copper (Cu) in various body tissues, including the liver, brain, kidneys, and corneas, due to a disorder in Cu metabolism. If left untreated, it can cause liver disease, central nervous system dysfunction, and fatal damage. Early diagnosis of WD is essential to reduce clinical complications. There are no specific biochemical tests for diagnosing WD, but rather a combination of clinical characteristics, including low levels of ceruloplasmin, increased urinary excretion of Cu, high levels of hepatic Cu, and Cu deposits in the cornea. In this study, we used a global metabolomics approach with liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS) to identify changes in metabolic pathways in plasma and urine samples from individuals with WD compared to healthy individuals (control group). The statistical analysis showed that 98 molecular features were significantly different in urine samples ($p \leq 0.05$ and $VIP \geq 1$) and 243 molecular features in plasma samples ($p \leq 0.05$ and $VIP \geq 1$). Out of these, 139 metabolites were putatively annotated and evaluated for changes in biochemical pathways, which are related to arginine biosynthesis, linoleic acid metabolism, alanine, aspartate, and glutamate metabolism, and biosynthesis of phenylalanine, tyrosine, and tryptophan. These results provide valuable insights into potential pathogenic mechanisms of WD and contribute to discovering biomarkers for rare and neglected diseases. The findings of this study have the potential to aid in developing more accurate diagnoses and treatments for rare and neglected diseases. They also could pave the way for further research on discovering new diagnosis biomarkers.

Keywords: Wilson disease, rare diseases, metabolomic, liver disease, high resolution mass spectrometry.

SUMÁRIO

1. Introdução.....	1
1.1. Importância e Metabolismo do Cobre	1
1.2. Doença de Wilson: Características e Manifestações Clínicas	3
1.3. Diagnóstico da Doença de Wilson	7
1.4. Tratamento da Doença de Wilson	11
1.5. Metabolômica.....	12
1.5.1. Vantagens e Desvantagens da Metabolômica Global e da Metabolômica Alvo.....	18
1.6. Técnicas de Análise Empregadas em Metabolômica	20
1.6.1. Cromatografia Líquida Acoplada à Espectrometria de Massas...21	
1.7. Processamento de dados	23
1.7.1. Análises Multivariadas e Univariadas.....	23
2. Objetivos.....	28
3. Materiais e Métodos.....	29
3.1. Generalidades	29
3.2. Casuística e Amostragem	29
3.3. Critérios de Inclusão e Exclusão	30
3.3.1. Critérios de Inclusão	30
3.3.2. Critérios de Exclusão	30
3.4. Preparo de amostras para análise metabolômica.....	31
3.4.1. Plasma.....	31
3.4.2. Urina.....	32
3.5. Análise Instrumental por LC-HRMS	33
3.5.1. Separação Cromatográfica no Modo Reverso de Separação.....	34
3.5.2. Separação Cromatográfica por Interação Hidrofílica.....	34
3.5.2.1. Amostras de Plasma.....	34
3.5.2.2. Amostras de Urina.....	35
3.6. Parâmetros para a Aquisição de Dados no Sistema LC-HRMS	35
3.7. Pré-Processamento dos Dados	36
3.7.1. Análises Multivariadas e Univariadas.....	38
3.8. Anotação Putativa dos Metabólitos	40
3.9. Identificação das rotas bioquímicas	40

4. Resultados e Discussão.....	41
4.1. Dados Demográficos	41
4.2. Pré-Processamento dos Dados de Metabolômica Global	41
4.2.1. Detecção, Deconvolução e Alinhamentos de Picos.....	43
4.2.2. Inspeção das amostras branco (QB).....	44
4.2.3. Avaliação da Qualidade dos Dados.....	44
4.2.4. Preenchimento de Lacunas (<i>missing values</i>).....	46
4.2.5. Normalização dos Dados.....	47
4.2.6. Centralização, Escalonamento e Transformação dos Dados.....	49
4.3. Análise Estatísticas	50
4.4. Amostras Biológicas de Plasma.....	53
4.4.1. Plasma - Modo Reverso de Eluição.....	53
4.4.2. Plasma - Modo HILIC de Eluição.....	57
4.5. Amostras Biológicas de Urina.....	61
4.5.1. Urina - Modo Reverso de Eluição.....	61
4.5.2. Urina - Modo HILIC de Eluição.....	64
4.6. Anotação Putativa do Metabólitos e Níveis de Identificação	69
4.7. Curva Característica de Operação do Receptor (ROC)	77
4.7.1. Curva ROC das Análises das Amostras de Plasma e Urina.....	77
4.8. Alterações Vias Bioquímicas	84
4.8.1. Biossíntese da Arginina.....	85
4.8.2. Metabolismo de Alanina, Aspartato e Glutamato.....	92
4.8.3. Biossíntese de fenilalanina, tirosina e triptofano.....	95
4.8.4. Metabolismo do Ácido Linoleico.....	102
5. Conclusão.....	105
6. Limitações da pesquisa.....	106
7. Perspectivas Futuras.....	107
8. Referências bibliográficas.....	108
9. Anexos.....	125

1. Introdução

1.1. Importância e Metabolismo do Cobre

O cobre (Cu) é um micronutriente essencial para o corpo humano sendo obtido através da alimentação. O Cu está presente em diversos alimentos, dentre eles, em produtos integrais, batatas, chocolate, fígado, feijão, nozes, crustáceos e cogumelos.¹ Normalmente, um ser humano consome aproximadamente de 1 a 3 mg de Cu/dia, enquanto o corpo humano precisa apenas de 0,9 mg/dia.^{2,3}

Nos adultos, aproximadamente 20-50% do Cu ingerido é absorvido e essa percentagem varia em função da dieta.³ Destes, aproximadamente 15% são transportados para outros tecidos e o restante (~85%) é excretado. O excesso de Cu é excretado majoritariamente pela bile (~98%) e em baixas concentrações pela urina (~2%).^{4,5} No entanto, se houver uma obstrução do fluxo da bile, o Cu é retido no organismo.⁶

Há evidências científicas de que a idade, gênero e uso de contraceptivos alteram a taxa de absorção de Cu no organismo. Além disso, mulheres apresentam maiores taxas de absorção do metal do que homens, e crianças possuem níveis mais elevados de absorção do que os adultos, que por sua vez, apresentam absorção igual aos idosos.⁷

No sistema biológico íons Cu existem em dois estados de oxidação, Cu^{+1} (forma reduzida) e Cu^{+2} (forma oxidada), o que fazem do metal um micronutriente essencial e ao mesmo tempo tóxico para as células. O Cu fornecido na dieta, na maioria das vezes encontra-se na forma oxidada (Cu^{+2}) e esses são reduzidos a Cu^{+} por metalorreductases para serem absorvidos. A absorção do Cu ocorre primariamente na mucosa intestinal, os enterócitos. Dos enterócitos, o micronutriente é distribuído para o fígado via albumina e para os rins via transcuprinas. Dos hepatócitos, o Cu pode ser distribuído via ceruloplasmina (Cp) para demais órgãos como cérebro, coração e pulmão.⁸

O processo celular de absorção, distribuição, eliminação e regulação do Cu pelo organismo é bastante complexo e envolve diferentes proteínas transportadoras e chaperonas* de Cu. Resumidamente, nos hepatócitos, o Cu^{+1} alcança o interior das células (citossol) via a proteína transportadora de Cu^{+1} (Ctr1).

*chaperonas – são moléculas que auxiliam as proteínas recém-formadas a assumirem a configuração apropriada (enovelamento) ao desempenho de suas tarefas, evitando agregação que podem ocasionar doenças conformacionais, tais como Alzheimer e Parkinson.

A partir daí, o metal é distribuído intracelularmente por três vias: **1)** chaperona antioxidante 1 de Cu (Atox1), que transfere o Cu para a ATPase beta transportadora de Cu (ATP7B) do tipo P; **2)** chaperonas de Cu para o citocromo C oxidase (COX17), transportando o metal para a mitocôndria e **3)** chaperona de Cu para superóxido dismutase 1 (CCS1), translocando o Cu para o superóxido dismutase de Cu/Zn (SOD1).^{4,5,8}

A etapa 1 de ligação do Cu com a proteína ATP7B é muito importante, pois essa possibilita uma série de funções vitais relacionadas a esse metal. A ATP7B é responsável por transportar o Cu para a rede trans de Golgi (TGN, do inglês: *trans-Golgi network*) para que o micronutriente seja incorporado com a Cp e, então, excretado na corrente sanguínea (plasma). Com isso, o Cu alcança diversos tecidos e órgãos do corpo, nos quais desempenha papéis essenciais em várias reações bioquímicas. A Cp é uma enzima majoritariamente produzida nos hepatócitos. Adicionalmente, a ATP7B é também responsável por transportar o Cu para a periferia celular para ser redistribuído em compartimentos vesiculares, com a excreção de Cu pela bile, a maior via de excreção de Cu.⁸

A liberação controlada de Cu é fundamental para manter a homeostase desse micronutriente no organismo. Esse mecanismo regulatório assegura que o Cu seja distribuído adequadamente para cumprir funções biológicas vitais, ao mesmo tempo em que evita acúmulos tóxicos que poderiam ocorrer sem uma regulação cuidadosa. Distúrbios genéticos, como a Doença de Wilson (DW), podem impactar esse processo, resultando em um acúmulo tóxico de Cu no organismo.⁵

Uma das mais importantes funções da presença de Cu no metabolismo humano está relacionada, principalmente, com suas propriedades como cofator na atividade de inúmeras enzimas, designadas como cuproenzimas.⁹ São muitos os processos fisiológicos no qual o Cu está envolvido, incluindo o sistema cerebral e doenças neurodegenerativas¹⁰, oncologia¹¹, angiogênese¹², sistema endócrino¹³, expressão gênica¹⁴, proteção da pele¹⁵, problemas cardiovasculares¹⁶, depressão¹⁷ e função imunológica¹⁸. Além disso, na defesa contra danos oxidativos dependente de enzimas antioxidantes (superóxido dismutases).¹⁹ Portanto, sua deficiência pode causar impactos negativos no funcionamento enzimático e prejudicar à saúde dos seres humanos, mas em altas concentrações, o micronutriente também apresenta toxicidade e promove danos oxidativos e morte celular.²⁰

1.2. Doença de Wilson: Características e Manifestações Clínicas

A DW, também conhecida por degeneração hepatolenticular, é uma doença rara, autossômica recessiva, na qual ocorre uma falha na excreção do Cu do corpo humano devido a uma mutação específica na proteína ATP7B, localizada no cromossomo 13, mais especificamente na região 13q14.3.^{21,22} Até o momento, o gene ATP7B é o único responsável por ocasionar a DW.

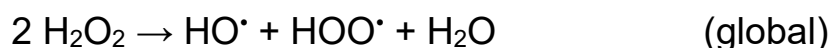
Dados epidemiológicos apontam que a DW afeta indivíduos com idade entre 4 e 40 anos.²³ E embora seja raro diagnosticar a DW em crianças menores de 3 anos ou em adultos acima de 40 anos de idade, há relatos na literatura da DW em bebê com nove meses de vida²⁴ e de casos raros da DW em pacientes com 72 anos de idade.²⁵ Segundo Wang *et al.* (2023) a prevalência mundial da DW é de aproximadamente 1 a 3 indivíduos em 100.000, sendo mais evidenciada no Leste Asiático.²⁶

Em estudo realizado com 371 pacientes diagnosticados com a DW de janeiro de 2005 a dezembro de 2020,²⁷ os resultados apontam que homens indicam ter maior incidência à doença. Em crianças, no entanto, não há notável diferença de gênero ou idade que impacta na prevalência da DW. O início dos sintomas apresenta-se mais tardiamente nos homens do que nas mulheres. Os sintomas mais comuns são fatores neuropsiquiátricos, em crianças e adultos, com predomínio do gênero masculino. Uma elevada porcentagem dos pacientes (82%), apresentaram mais de dois sintomas neurológicos. Aqueles indivíduos com fatores hepáticos apresentaram sintomas da DW mais jovens do que os com sintomas neurológicos. Nas mulheres observou-se maior incidência de fatores hepáticos e nos homens neuropsiquiátricos. A idade de início dos sintomas foi um fator decisivo para prever o fenótipo da DW e os autores recomendaram estudos adicionais com foco no efeito dos estrogênios na patologia da DW.

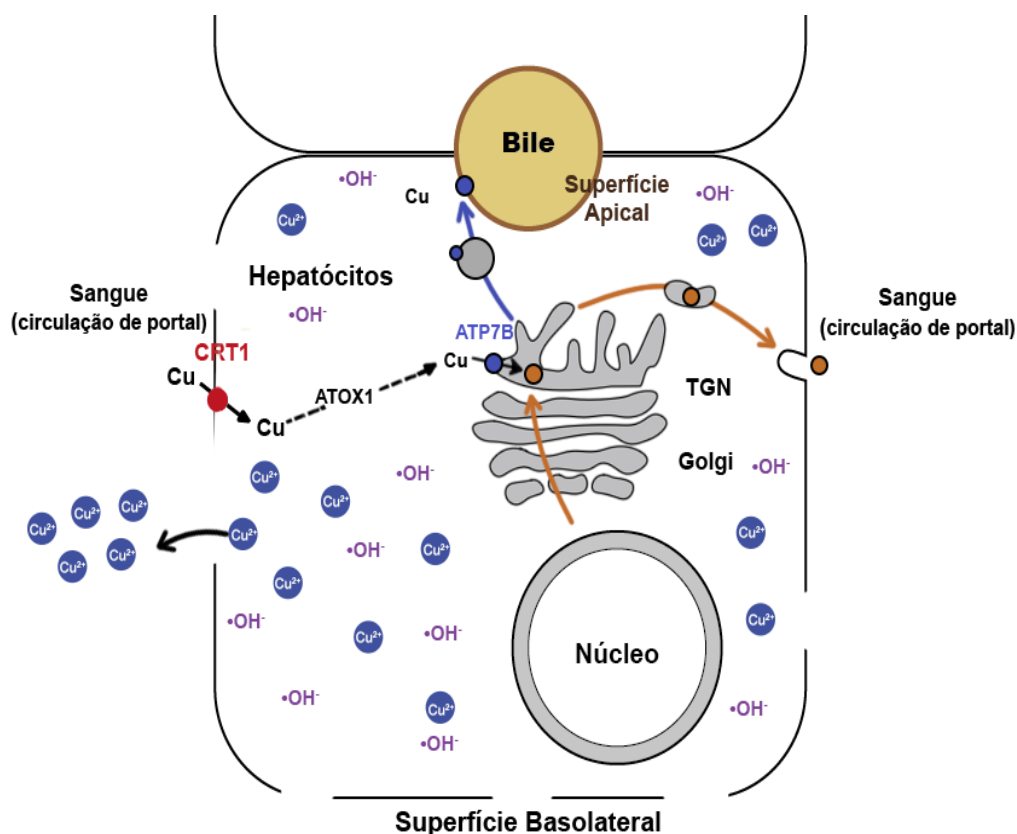
Alguns estudos, relatam que a DW é conhecida desde 1883, sendo inicialmente denominada pseudoesclerose por Westphall em decorrência dos tremores e rigidez apresentados pelos pacientes.²⁸ Em 1912, Kinnier Wilson descreveu a forma clássica da doença, fazendo sua correlação à doença do fígado, sendo denominada por ele como degeneração lenticular progressiva. Em 1921, Hall descreveu a patologia como uma degeneração hepatolenticular, ou seja, mais uma vez associada à doença do fígado.²⁸

Somente em 1993 o gene que codifica a *beta* ATPase transportador de Cu do tipo P foi definido como o responsável pela DW. Esse gene codifica 1.465 aminoácidos com um domínio fosfatase, outro de fosforilação e um domínio de ligação a nucleotídeos, além de seis domínios de ligação metálicas do tipo *N*-terminal e por fim oito canais iônicos transmembranares.^{21,22,29} O número de mutações do gene ATP7B já identificadas ultrapassa 1.000 diferentes mutações e aproximadamente 800 polimorfismos de nucleotídeo único (SNP, do inglês: *single nucleotide polymorphism*), o que dificulta consideravelmente o uso de testes moleculares no diagnóstico da DW.²²

O mecanismo da DW afeta diretamente o transporte de Cu⁺ pelo ATP7B transportador, conforme anteriormente apresentado para o metabolismo do metal (pag. 2). O Cu⁺, que se encontra no citosol, não é transportado pela ATP7B em decorrência da sua mutação genética e interferindo, portanto, no processo seguinte de ligação do Cu com a Cp e, conseqüentemente, na eliminação/redução dos níveis cúpricos celulares.³⁰ Como resultado, o Cu⁺ livre passa a reagir com o peróxido de hidrogênio presente no corpo, através de reações Fenton, formando radicais hidroxila (Equação 1).¹⁹ Os radicais hidroxila causam danos aos tecidos ocasionando à lesão dos hepatócitos e resultando na liberação de íons de Cu²⁺ tóxico na corrente sanguínea, levando também ao acúmulo de Cu em outros tecidos (Figura 1).



Equação 1- Reações de Fenton dos íons cobre (A, B e global)



Adaptado ³⁰

Figura 1- Distribuição de $\text{Cu}^+/\text{Cu}^{+2}$ no hepatócito, com acúmulo do metal e radicais hidroxila

As primeiras manifestações clínicas da DW estão relacionadas principalmente a problemas hepáticos, os quais incluem um espectro bastante variado, dentre eles: hepatomegalia e esplenomegalia assintomática; elevação persistente dos níveis de alanina aminotransferase (ALT) e aspartato aminotransferase (AST); esteatose hepática, quadros de hepatite autoimune (HAI), que podem progredir para quadros mais avançados com cirrose hepática e insuficiência hepática aguda.^{31,32} A doença hepática aguda, geralmente caracterizada por icterícia com coloração amarelada da pele e dos olhos (Figura 2A-B), além de fadiga, perda de apetite e dor abdominal. Em estágios mais avançados, a DW pode progredir para hepatite crônica, com ou sem cirrose hepática (Figura 2C), levando a complicações graves como ascite, que se trata do acúmulo de líquido na cavidade abdominal, sangramento gastrointestinal e insuficiência hepática. Dentre os pacientes diagnosticados com a DW, 50% desses apresentam sintomas hepáticos.^{31,32}

Além do fígado, outros órgãos são afetados pelo acúmulo de Cu, sendo o cérebro o segundo órgão mais afetado.^{10,18} Os sintomas neurológicos são amplos, podendo incluir tremores, movimentos involuntários, rigidez muscular e dificuldade de coordenação. Problemas de fala, como dificuldade em articular palavras, também são observados. Alterações comportamentais e psicológicas são comuns nos sintomas, como mudanças de humor, ansiedade, depressão, irritabilidade e dificuldade de concentração. Em estágios avançados, a DW pode levar a complicações neurológicas graves, como demência, convulsões e distúrbios psiquiátricos.^{31,32} Vale ressaltar que estes sintomas podem aparecer até muitos anos depois do diagnóstico da DW, evidenciando que o acompanhamento constante e tratamento adequado da doença são muito importantes para evitar agravamento.

Os olhos também podem ser afetados pela DW, com a formação de depósitos de cobre na córnea, chamados de anéis de Kayser-Fleischer (K-F) (Figura 2D). Os anéis de K-F podem acometer cerca de 95% dos pacientes com problemas neurológicos e até 50% dos pacientes com sintomas hepáticos causados pela DW e de 20-30% dos casos assintomáticos.³³ Observa-se que estes anéis possuem coloração esverdeada ou acastanhada ao redor da íris, sendo um sinal importante e evidente para o diagnóstico da DW.

A DW pode também levar a outras complicações, como problemas renais, incluindo a perda de função renal, a formação de cálculos renais e alterações na excreção de Cu na urina.³⁴ Problemas ósseos, como osteoporose e osteoartrite, também podem ocorrer.³³ O comprometimento cardíaco, incluindo cardiomiopatia, arritmias cardíacas e ainda cardiomiopatia diabética, são complicações provenientes do metabolismo deficiente do Cu^{+2} , ocasionados pelo acúmulo do metal no miocárdio, com comprometimento da função mitocondrial e stress oxidativo.¹⁶

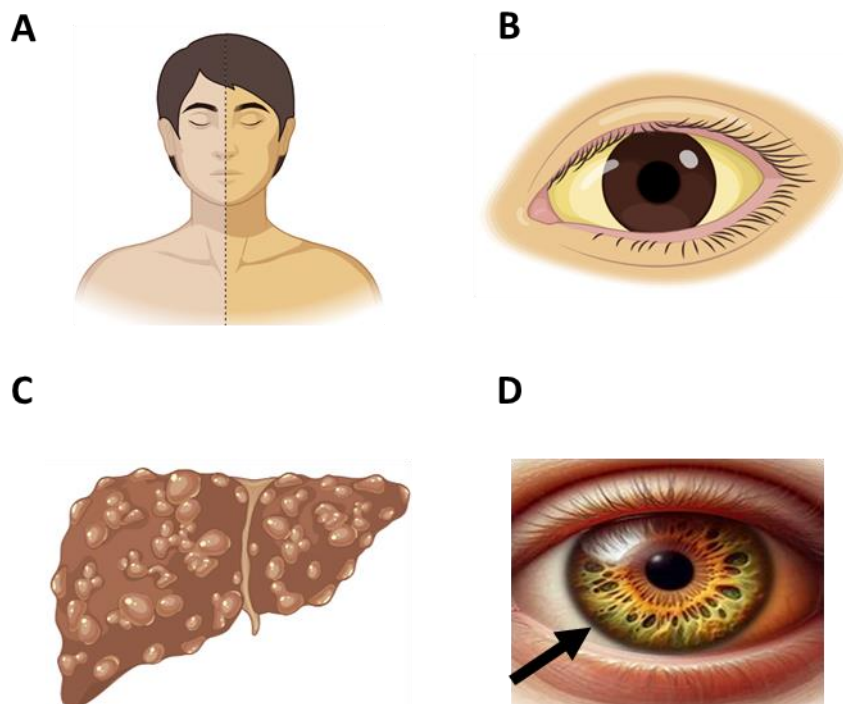


Figura 2- Características Clínicas da DW - (A) coloração amarelada da pele; (B) coloração amarelada dos olhos; (C) fígado com cirrose; (D) anéis de Kayser-Fleischer. A, B e C: Biorender D: *Image Creator from Designer. Powered by DAA E 3*

Em crianças e adolescentes diagnosticados com a DW, há alterações neurológicas que podem ocasionar mudanças de comportamento, psicoses e até distúrbios de fala, que se sobrepõem aos sintomas de outras doenças, evidenciando um desafio em se conduzir um correto diagnóstico da patologia.^{27,35}

1.3. Diagnóstico da Doença de Wilson

O diagnóstico da DW no estágio inicial é realizado através de exames de nível plasmático de Cp, níveis de Cu na urina, níveis de Cu sérico e hepático e a presença de anéis de K-F. O emprego da quantidade de Cu livre, não ligado à Cp, tem sido proposto como um biomarcador de diagnóstico, no entanto, há inúmeras controvérsias com relação ao seu emprego.³⁶ Paralelamente, há relatos quanto à dificuldade do diagnóstico na prática clínica, no qual se mostra ser atípico o paciente apresentar os sintomas mais clássicos como doença hepática, tremor ou distonia e a presença de anéis de K-F, sendo o último não específico para DW, podendo ocorrer também em doenças colestáticas, como por exemplo, a colestase biliar primária (CBP).³⁷ Desta forma, outros aspectos clínicos e bioquímicos do paciente devem ser

examinados, incluindo estudo genético, biópsia hepática e determinações de mutações no organismo do paciente.

Para suprir a falta de especificidade na determinação de Cu livre, foi proposto o emprego do nível de Cu-trocável (CuExc; do inglês: *exchangeable copper*) e do nível relativo de Cu-trocável (REC; do inglês: *exchangeable copper*). A determinação de CuExc envolve a incubação do soro do paciente em solução de EDTA por 1 hora e posterior ultrafiltração da solução e então o REC é calculado com emprego da equação (2) $REC(\%) = (CuExc/Cu\ total) \times 100$.³⁶

Na busca por um sistema de diagnóstico mais assertivo, sociedades científicas publicaram diretrizes com diferentes critérios de diagnóstico. A Associação Americana para o Estudo de Doenças Hepáticas (AASLD; do inglês: *American Association for the Study of Liver Diseases*)³² propôs um algoritmo clínico/bioquímico para o diagnóstico da DW e a Associação Européia para Estudo do Fígado (EASL; do inglês: *European Association for the Study of the Liver*)³⁸ juntamente com a Sociedade Europeia de Pediatria Gastroenterologia, Hepatologia e Nutrição (ESPGHAN; do inglês: *European Society for Paediatric Gastroenterology, Hepatology and Nutrition*)³⁹ defendem o uso de um sistema de pontuação de *Leipzig*.

O algoritmo proposto pela AASLD propõe uma abordagem algorítmica clínica e bioquímica, na qual em uma triagem inicial considera a presença oftalmológica dos anéis de K-F, níveis de Cp sérica e de Cu urinário/24h. Esse mesmo algoritmo é também dependente da presença de sintomas neuropsiquiátricos e/ou hepáticos nos pacientes. Quando há resultados inconsistentes, recomenda-se o emprego de biópsia hepática ou do teste genético.³⁶

Já a estratégia recomendada pela EASL e ESPGHAN é baseada em um sistema de pontuação de *Leipzig*, que considera exames clínicos, bioquímicos e genéticos. Nesse caso, o emprego de testes genéticos, que são menos invasivos, mas com altos custos, são preferidos ao emprego da biópsia hepática, como no AASLD. Ainda assim, o sistema de *Leipzig* apresenta limitações no sistema de pontuação, uma vez que ele se baseia em recomendações feitas por um painel de especialistas e não em estudos populacionais.⁴⁰

O estudo familiar é um importante fator no diagnóstico da DW, na triagem de parentes de primeiro grau testes clínicos e genéticos são preferidos na AASLD e ESPGHAN, enquanto a EASL prefere apenas testes genéticos. No quesito tratamento da DW, há um consenso, embora a ESPGHAN favoreça o uso de zinco (Zn) ao invés

de outros quelantes na fase de manutenção e com pacientes assintomáticos. Em todas as diretrizes, o transplante de fígado está indicado em casos de insuficiência hepática aguda decorrente da DW, mas não se recomenda esse procedimento em casos em que as doenças neuropsiquiátricas são predominantes. Existem recomendações limitadas sobre circunstâncias especiais, incluindo gravidez, cirurgia e risco de malignidade na DW.⁴⁰

As manifestações clínicas da DW não são específicas, podendo ser confundidas com outras patologias, como hepatite, cirrose hepática, doença de *Parkinson* ou, até mesmo *Alzheimer*. Além da falta de especificidade também dos parâmetros bioquímicos que são influenciados por outras condições médicas. Por exemplo, níveis anormais de Cu e Cp podem ser observados em outras doenças hepáticas crônicas de causas diversas ou distúrbios metabólicos como diabetes tipo 2, obesidade, hiperlipedemia e problemas cardiovasculares,^{41,42} portanto um diagnóstico preciso é fundamental para iniciar o tratamento adequado e melhorar a qualidade de vida dos pacientes.

Dentre as principais alternativas clínicas empregadas no diagnóstico da DW, vale destacar:^{31,32,36}

- **Testes laboratoriais**

São realizados testes sanguíneos para avaliar os níveis de Cu, Cp, aspartato aminotransferase (AST) e ácido úrico. Pacientes com DW geralmente apresentam níveis baixos de Cp e ácido úrico e níveis elevados de Cu e AST no sangue. Além disso, a análise da urina pode revelar a presença de Cu em excesso com a excreção de Cu na urina. O parâmetro da quantidade de Cu livre (não ligado à Cp), não tem sido considerado devido à grandes controvérsias sobre sua utilidade.³⁶ Os principais parâmetros e valores de referência estão na Tabela 1.

Tabela 1- Sumário dos principais parâmetros e valores de referência no diagnóstico da DW

Teste Diagnóstico	Valores de referência	Valores diagnóstico DW
Cp sérica	20-40 mg/dL	< 20 mg/dL
Ácido úrico	2,7 – 8,5 mg/dL	diminuído
Cu total	700-1.520 µg/L	variável
Cu urinário 24 hr	< 40 µg/24hr	> 100 µg/24hr
Cu hepático	15-55 µg/g peso seco	> 250 µg/g peso seco
RMN	ausente	imagem T2 mostrando leões talâmicas no meio do cérebro
TC	revela ou exclui outras doenças neurológicas	-
Testes genéticos	negativo	positivo

Cp: ceruloplasmina; RMN: ressonância magnética nuclear; TC: tomografia computadorizada

- **Anéis de Kayser-Fleischer**

Os anéis de Kayser-Fleischer (K-F) são anéis dourados ou esverdeados ao redor da córnea. Esses anéis se desenvolvem com o acúmulo de Cu. Em alguns casos da DW, os anéis de K-F são o primeiro sinal da patologia.

- **Testes genéticos**

Os testes genéticos podem ser realizados para identificar mutações no gene ATP7B, responsável pela DW. Esses testes podem confirmar o diagnóstico em casos ambíguos ou em pacientes assintomáticos com histórico familiar da doença.

- **Biópsia hepática**

Em alguns casos, pode ser realizada uma biópsia hepática para analisar a quantidade de cobre presente no fígado. Esse procedimento envolve a retirada de uma pequena amostra de tecido hepático para análise em laboratório.

- **Testes de imagem do cérebro**

Testes de imagem por ressonância magnética nuclear (RMN) são empregadas para estabelecer um diagnóstico de pacientes com sintomas neurológicos ou psiquiátricos inexplicados. Pacientes com alterações expressivas, como de cavidade de tecido, se beneficiam pouco com o tratamento. A RMN não é considerada útil como prognóstico ou para monitoramento da progressão da DW.

Além da RMN, tomografia computadorizada (CT) também pode ser empregada, no entanto, essa técnica de imagem não irá fornecer dados sobre a DW e sim irá revelar ou excluir a presença de outras patologias relacionadas ao sistema nervoso central.

1.4. Tratamento da Doença de Wilson

Existem algumas opções de tratamento para a DW, incluindo dieta restritiva de alimentos ricos em Cu e o uso de terapias medicamentosas específicas. Alguns dos medicamentos mais comumente utilizados são a α -penicilamina e o dicloridrato de trietilnotetramina, comercialmente chamado de trientina⁴³ (Figura 3). Ambos são agentes quelantes que se ligam ao Cu no organismo, facilitando sua excreção e reduzindo os níveis do metal no fígado e outros tecidos. Além disso, o tetratiomolibdato de amônio⁴⁴ (Figura 3) também é utilizado para reduzir a absorção de Cu no intestino, controlando a quantidade de entrada deste metal no organismo. O Zn, na forma de acetato de zinco, é usado no tratamento ou na manutenção da DW, principalmente quando há efeitos colaterais ou ineficácia em pacientes inicialmente tratados com as moléculas quelantes com Cu. O mecanismo de ação do acetato de Zn inclui competição com o Cu na absorção intestinal e promove ainda a elevação da síntese de metalotioneína, proteína com elevada afinidade por Cu. Como consequência, a administração de acetato de Zn aumenta consideravelmente a excreção de Cu via intestino.^{45,46} Em casos de DW em estágio crônico, há um comprometimento significativo do fígado com o acúmulo de Cu e o transplante de fígado passa a ser considerado uma opção de tratamento.⁴⁶

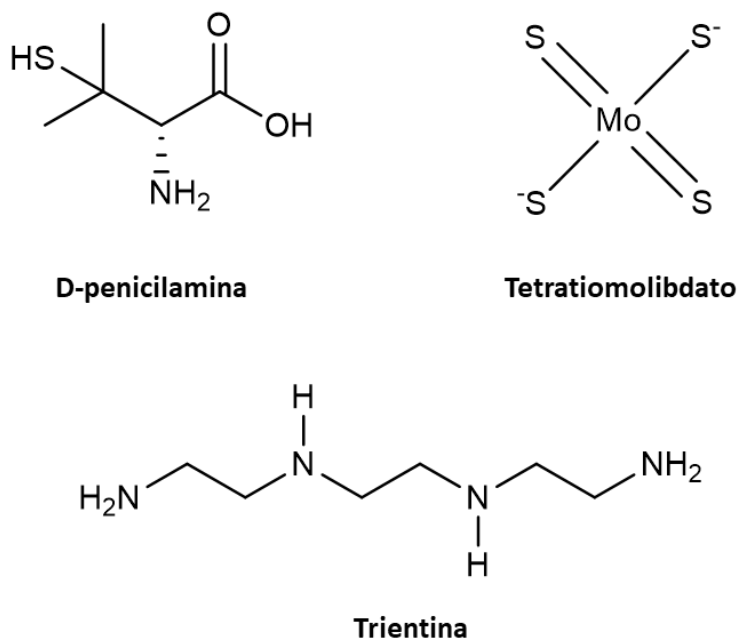


Figura 3- Estruturas químicas dos fármacos empregados no tratamento da DW.

É possível ter uma vida normal e com qualidade convivendo com a DW, porém é fundamental que o diagnóstico seja realizado precocemente e que o tratamento adequado e contínuo seja seguido para evitar complicações graves, logo sem um diagnóstico e tratamento adequados, a DW pode ter consequências fatais.

1.5. Metabolômica

Nas últimas décadas, as técnicas analíticas passaram por várias revoluções, que foram impulsionadas pelo desenvolvimento das ciências ômicas, associadas à análise de dados pelas ferramentas de bioinformática.⁴⁷

O fortalecimento do uso das ciências ômicas para o entendimento de organismos de forma ampla tem crescido exponencialmente, impactando mudanças e novos desafios em diversas áreas de pesquisa científica. A metabolômica tornou-se uma aliada nesse impulsionamento, sendo uma área de estudo ômica promissora em estudos clínicos e de biomarcadores, em estudos toxicológicos, nutricionais, de produtos naturais, microbiológicos, parasitologia, ambientais, dentre outros.^{48,49} A Figura 4 ilustra a evolução do número de artigos publicados com o termo “*metabolomic*” na plataforma de pesquisa “*Pubmed*” nos últimos 10 anos.

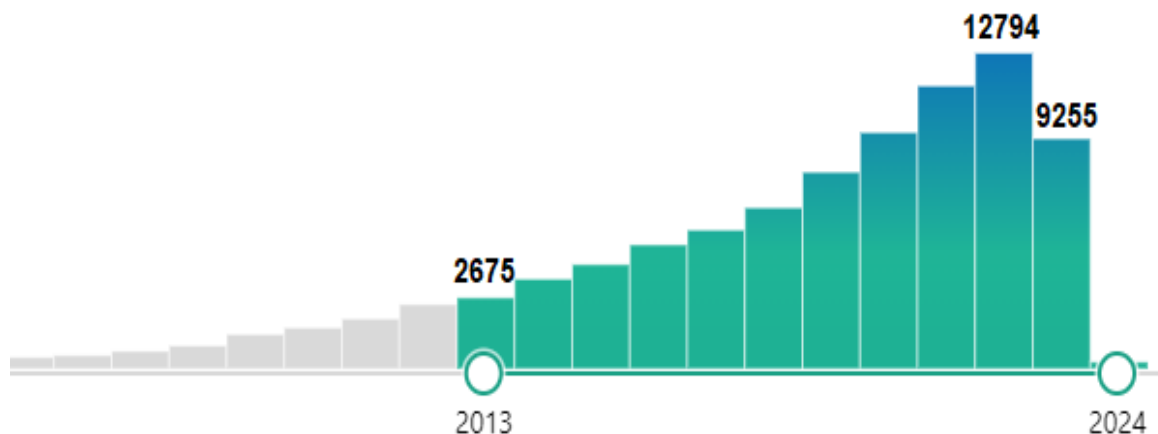
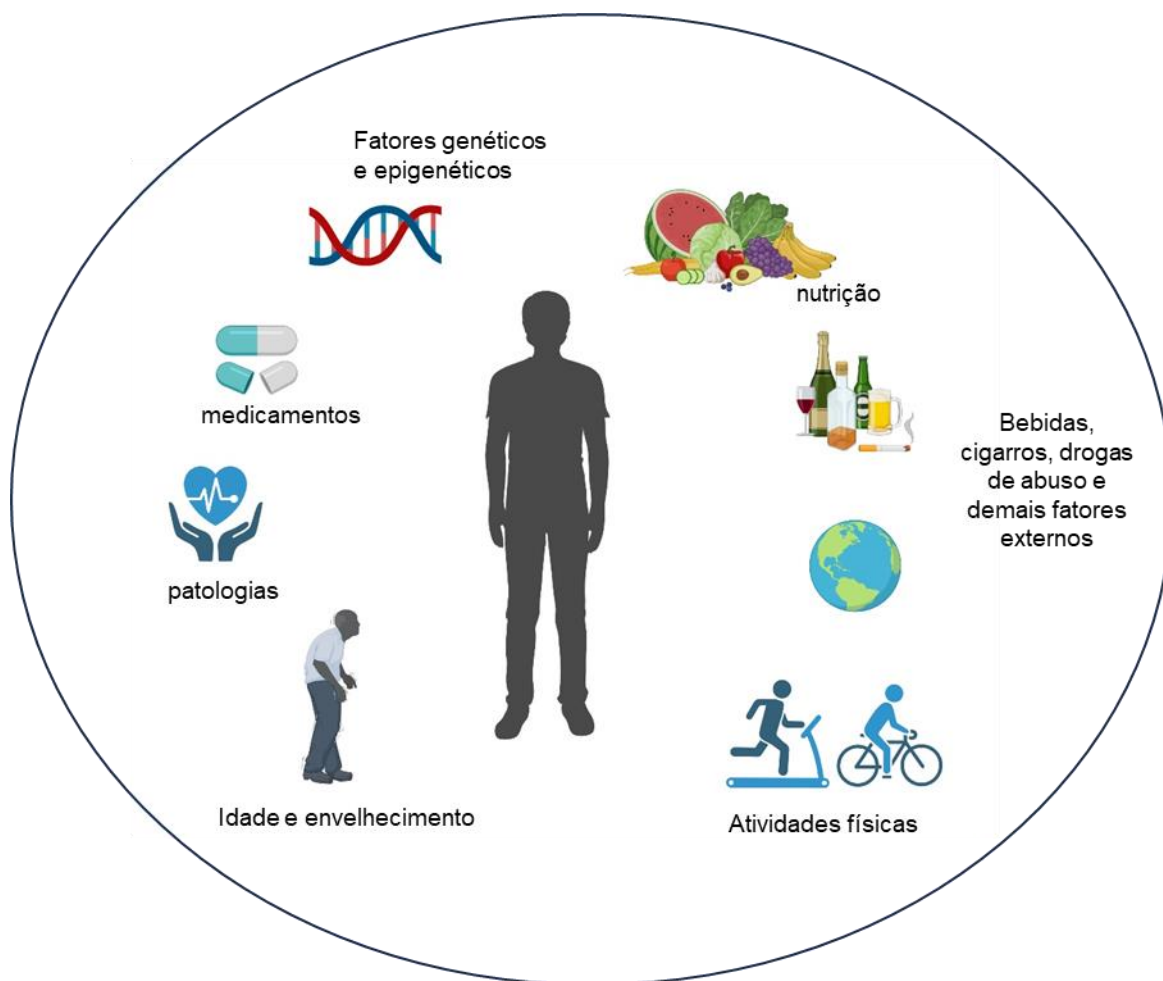


Figura 4- Evolução do número de artigos publicados na plataforma “*pubmed*” com o termo “*metabolomic*”. Pesquisa realizada em 12/09/2023

As ciências ômicas oferecem a possibilidade de medir de forma abrangente a expressão de genes (transcriptômica), a abundância de proteínas (proteômica) ou os níveis de metabólitos (metabolômica) em um sistema biológico. As alterações de níveis de metabólitos estão associadas a presença de patologias, assim como a exposição a agentes químicos e tóxicos ou fármacos. Desta forma, a metabolômica reflete a atividade bioquímica e o estado das células/tecidos, representando os fenótipos dos sistemas biológicos. Ainda, permite avaliar as alterações genômicas, muitas vezes associadas a fatores externos, como contaminantes ambientais, uso de fármacos, saúde física, estilo de vida, idade, prática de atividades físicas, consumo de álcool ou cigarro e hábitos de alimentação (Figura 5).^{50,51}



Fonte: Elaborada pela autora.

Figura 5- Representação de fatores que podem influenciar o fenótipo de um indivíduo.

O termo metaboloma foi introduzido pela primeira vez na literatura em 1998 por Stephen Oliver⁵², objetivando identificar conjuntos de compostos de baixas massas moleculares que o organismo sintetiza.⁵³ A partir dos diversos avanços genômicos que a medicina obteve, o emprego da metabolômica ampliou-se como uma forma de entender os mecanismos do funcionamento celular a nível molecular, levando à compreensão das mudanças fenotípicas.^{54,55}

Desta forma, a metabolômica permite a análise qualitativa e quantitativa de forma abrangente e detalhada de metabólitos presentes em amostras biológicas, células e tecidos, possibilitando a identificação de alterações bioquímicas em um organismo. Ou seja, determina moléculas de baixa massa molecular (< 1500 Da) e monitora, em nível molecular, uma informação integrativa da função celular, definindo assim o fenótipo de um indivíduo em resposta a alterações químicas, ambientais ou genéticas. A identificação de metabólitos é um processo complementar aos estudos transcriptômicos e proteômicos para determinar a função de um gene.^{56,57}

Metabólitos são substratos, produtos intermediários e produtos finais provenientes das biotransformações do metaboloma. Representam diferentes classes químicas com ampla faixa de concentração e diversidade físico-química, sendo de baixa a alta polaridade, moléculas voláteis e não voláteis, aminoácidos, ácidos orgânicos, lipídeos, açúcares, vitaminas e minerais.⁵⁸

Como uma ciência emergente, as análises metabolômicas têm sido empregadas em diversas áreas relacionadas a medicina. Do ponto de vista clínico, o estudo das alterações metabólicas que ocorrem em resposta a diferentes processos fisiológicos ajuda a estabelecer mecanismos relacionados a patologias.^{59,60} Na medicina personalizada, abordagens farmacometabonômicas podem servir para prever a ação de medicamentos em um indivíduo.⁶¹ Além disso, o microbioma intestinal influencia a eficácia de medicamentos, o que pode ser deliberadamente modificado para otimizar os benefícios e minimizar efeitos adversos de um determinado tratamento.⁶² Ainda, esta abordagem ajuda a compreender como os fármacos atuam durante o tratamento dos pacientes ou como eles influenciam a evolução do tratamento.⁶³

Dentro deste contexto, a descoberta de novos biomarcadores de prognóstico, diagnóstico ou de monitoramento clínico de patologias tem recebido grande atenção.⁶⁴ Um biomarcador é um metabólito avaliado como um indicador de processos biológicos normais, patogênicos ou de respostas farmacológicas a um agente terapêutico. Assim, os biomarcadores podem permitir uma detecção mais confiável e precoce da doença, aumentar a seleção de pacientes com maior probabilidade de responder à terapia medicamentosa e permitem monitoramento em tempo real. Ainda, podem monitorar a eficácia do tratamento. No desenvolvimento de novos fármacos, os biomarcadores são vistos como facilitadores *go/no go* na tomada de decisões, acelerando a promoção de compostos ativos ou rejeitando precocemente aqueles compostos destinados a falhar.⁶⁵

Na condução de análises metabolômicas realiza-se um estudo comparativo entre um designado grupo controle e um grupo estudo. O grupo controle é composto por indivíduos saudáveis (ausente da patologia) ou amostras celulares sem alterações, enquanto o grupo estudo é constituído por indivíduos com a patologia ou amostras celulares com alterações genéticas, *knockdown* de genes/proteínas ou uso de moléculas terapêuticas. As estratégias metabolômicas são comumente

classificadas em duas categorias⁶⁶⁻⁶⁸: **1-** metabolômica global (do inglês: *untarget metabolomic*) e **2-** metabolômica alvo (do inglês: *target metabolomic*).

A metabolômica global, também designada como não direcionada, é uma análise abrangente e utilizada na geração de hipóteses. É realizada para coletar e analisar todos os metabólitos detectáveis em uma amostra. Como uma abordagem de descoberta, a análise global e não direcionada também pode revelar novos compostos e incógnitas químicas, potencialmente levando a novas descobertas sobre o mecanismo ou tratamento de uma patologia. Como resultado, vários níveis de informação podem ser alcançados. Em contraste, a metabolômica alvo ou direcionada permite a quantificação absoluta de um subconjunto definido de metabólitos, previamente caracterizados e anotados bioquimicamente, por exemplo, metabólitos específicos de uma via bioquímica (Figura 6).

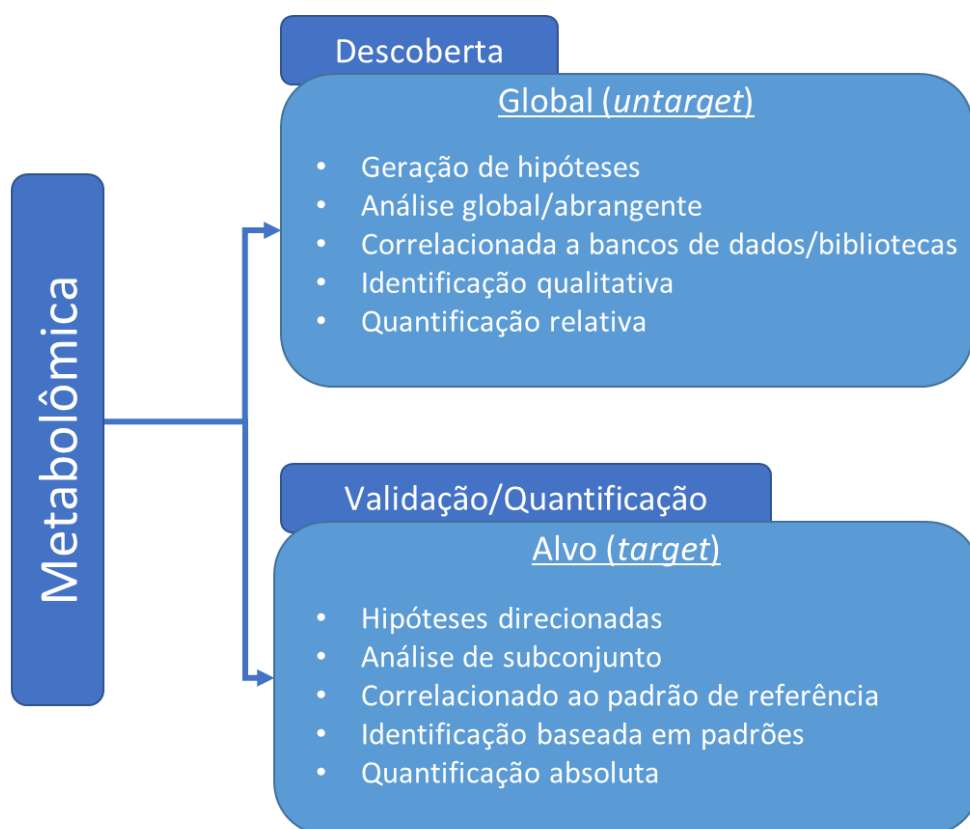
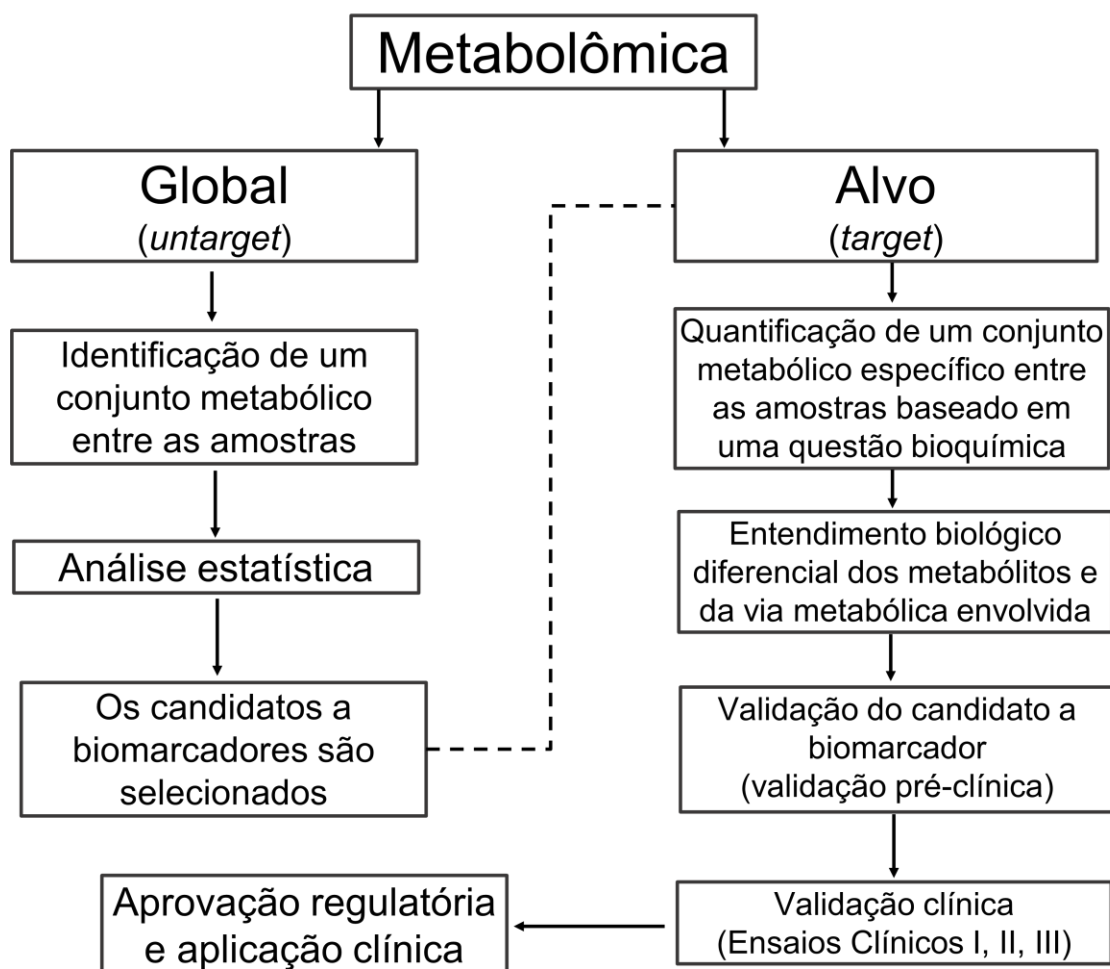


Figura 6- Classificação das abordagens metabolômicas

É importante destacar que as abordagens metabolômicas global e alvo não são mutuamente exclusivas; pelo contrário, são complementares. Ao integrá-las, é possível descobrir potenciais biomarcadores que aprofundam a compreensão das

vias bioquímicas relacionadas ao alvo do estudo. A Figura 7 ilustra os conceitos, evidenciando que a descoberta de biomarcadores por meio da análise metabolômica segue várias etapas. Inicialmente, identificam-se conjuntos de metabólitos entre as amostras em estudo, utilizando inicialmente a abordagem global ou não direcionada. Em seguida, com o emprego da abordagem alvo conduz a quantificação e validação de moléculas candidatas a biomarcadores, culminando na validação clínica. Este processo integrado enriquece a compreensão das vias metabólicas subjacentes e facilita a descoberta de biomarcadores clínicos relevantes.⁶⁹



Adaptado ⁶⁹

Figura 7- Estratégias metabolômicas na descoberta e validação de biomarcadores

1.5.1. Vantagens e Desvantagens da Metabolômica Global e da Metabolômica Alvo

Da mesma forma que em outros métodos de análise em Química Analítica, as análises metabolômicas também possuem vantagens e desvantagens a serem previamente consideradas. A metabolômica global está principalmente voltada para a descoberta e geração de hipóteses, com isso, elimina-se a necessidade de uma compreensão exaustiva de todos os metabólitos identificados, mas sim a obtenção de um perfil metabólico (Figura 6). Fatores adicionais incluem o fato de que na metabolômica global o preparo das amostras biológicas são procedimentos simples, em que majoritariamente emprega-se precipitação de proteínas com solventes orgânicos, o que permite detectar simultaneamente um elevado número de moléculas. Adicionalmente, a metabolômica global não necessariamente requer o uso de padrões internos isotopicamente marcados. No entanto, o potencial da técnica permite desvendar metabólitos já conhecidos e desconhecidos, proporcionando à descoberta de alterações bioquímicas ainda não identificadas. Na análise metabolômica global, a aquisição de dados é conduzida no modo *full MS scan*, ou seja, sem fragmentação na detecção por espectrometria de massas (MS). Com isso, é possível reinvestigar as amostras pela busca de novos constituintes químicos, sem a necessidade de reanálise.⁷⁰

Por outro lado, as desvantagens da metabolômica global estão relacionadas à exigência de um processamento intensivo dos dados e à aplicação de análises estatísticas complexas, necessárias para garantir a confiabilidade dos resultados. A ausência de padrões de referência é outro desafio no que concerne à identificação de *features* moleculares desconhecidos; padrões de fragmentação não reportados ou pouco informativos; dificuldade em interpretar as taxas de falsas descobertas (FDR; do inglês: *false discovery rate*); a reduzida precisão e exatidão nas análises metabolômicas globais, devido à análise relativa dos metabólitos; a necessidade de recursos bioestatísticos adicionais para selecionar o modelo estatístico apropriado e, não menos importante, desvios na detecção de metabólitos de baixa e elevada abundância na mesma amostra biológica.⁷¹

No caso da metabolômica alvo, as suas vantagens estão primeiramente associadas com a identificação e caracterização de metabólitos já previamente conhecidos. Os métodos alvos permitem uma ampla compreensão dos processos

metabólicos, da cinética enzimática e das vias moleculares estabelecidas para obter uma compreensão elucidativa dos mecanismos bioquímicos. Adicionalmente, pode-se fazer uso de padrões analíticos marcados isotopicamente, reduzindo a probabilidade de falsos positivos e de artefatos analíticos. A quantificação absoluta dos metabólitos fornece melhor precisão e exatidão dos dados em comparação com a análise relativa na metabolômica global. Para o preparo de amostras, a metabolômica alvo pode ser otimizada para os compostos de interesse, assegurando extrações mais eficientes e precisas, mesmo que os níveis de concentração dos constituintes da amostra sejam diferentes entre si.^{72,73}

Em contraste, as análises metabolômicas alvo também apresentam desvantagens, dentre elas pode-se citar a necessidade do conhecimento prévio do conjunto de metabólitos a ser quantificado, o que também resulta em uma compreensão pouco abrangente dos dados biológicos. Outra limitação reside na dificuldade do uso de padrões internos marcados isotopicamente para cada molécula de interesse, quando esses não estão disponíveis comercialmente ou devido aos elevados custos de aquisição. Na etapa de aquisição de dados, as análises metabolômica alvo empregam o monitoramento de reações selecionadas (SRM), o qual isola o íon precursor e monitora os íons produtos gerados.⁷³ Apesar dessa abordagem propiciar elevada especificidade, seletividade e sensibilidade na detecção por MS, não é possível reinvestigar os dados após a aquisição. Assim, caso haja metabólitos que possam ser de interesse, mas que não tenham sido previamente contemplados nas análises, torna-se necessário a reinjeção das amostras. De forma geral, considerando-se as vantagens e desvantagens das análises metabolômica global e alvo, é primordial uma avaliação cuidadosa dos objetivos esperados do estudo antes de conduzir as análises.

Além dos aspectos previamente apresentados, as análises metabolômicas globais apresentam outros desafios científicos. Uma das questões mais desafiadoras de um estudo metabolômico é a confirmação da identidade de um *feature* molecular.⁷⁴ Essa etapa é um passo essencial para compreender as mudanças biológicas que ocorrem no sistema e continua a ser um grande limitador nas investigações metabolômicas globais. Adicionalmente, a falta de protocolos que permitam a harmonização nos métodos de amostragem, preparo de amostras com extração eficiente de moléculas com diferentes propriedade físico-químicas e em diferentes níveis de concentração, e a não uniformização nas etapas de pré- e pós-

processamento de dados entre laboratórios, dificulta não somente a reprodutibilidade de dados inter- e intralaboratórios, mas também a transposição da metabolômica dos laboratórios acadêmicos para as indústrias farmacêuticas, levando ao avanço da medicina de precisão.^{75,76}

1.6. Técnicas de Análise Empregadas em Metabolômica

O metaboloma é quimicamente diverso e requer métodos diferentes para a separação e extração dos constituintes das amostras, assim como para a aquisição e detecção de dados, com o objetivo de capturar a maioria dos compostos em uma mistura complexa.

Os metabólitos presentes em uma amostra biológica apresentam diversas funções químicas, variando de compostos altamente polares a compostos apolares. Adicionalmente, diferentes metabólitos apresentam diferentes níveis de concentração nas amostras, o que aumenta o desafio analítico. Assim, as diferentes propriedades físico-químicas dos metabólitos presentes nas amostras representam um desafio adicional no desenvolvimento de métodos, uma vez que não é possível empregar um único método analítico capaz de abranger todo o metaboloma. Desta forma, é importante o emprego de técnicas analíticas e de modos de separação ortogonais.⁷⁷

As principais plataformas analíticas utilizadas em análises metabolômica são a ressonância magnética nuclear (RMN)^{78,79}, a cromatografia gasosa acoplada à espectrometria de massas (GC-MS)^{80,81} e a cromatografia líquida de alta eficiência acoplada à espectrometria de massa de alta resolução (LC-HRMS)⁵⁷. Da mesma forma, as técnicas de RMN, GC-MS e de LC-HRMS possuem vantagens e limitações e nenhuma delas consegue de forma individualizada detectar todos os metabólitos presentes no metaboloma.

A espectroscopia de RMN baseia-se nas propriedades magnéticas de determinados núcleos atômicos, o que permite determinar a estrutura e dinâmica molecular de amostras. Apenas alguns átomos absorvem radiação, induzindo uma transição dos *spins* nucleares, sendo esses ¹H, ¹³C, ¹⁴N, ¹⁵N, ¹⁷O, ¹⁹F, ²³Na e ³¹P. A espectroscopia de RMN é uma técnica não destrutiva e não invasiva e a RMN de prótons (¹H) é a mais frequentemente aplicada em metabolômica global.⁷⁹

A análise metabolômica global por RMN permite a determinação da concentração absoluta de cada metabólito, a qual pode ser medida a partir de um

espectro de RMN. Isso é possível porque o sinal produzido por um único núcleo ^1H não depende do ambiente magnético, o que permite que uma única calibração de intensidade seja aplicada a todos os demais metabólitos.

A técnica de GC-MS é uma técnica cromatográfica aplicada ao estudo de metabólitos voláteis que possuem baixo ponto de ebulição e que estarão presentes na fase gasosa na faixa de temperatura de 50-350°C. Os metabólitos podem ser voláteis na sua forma biologicamente nativa ou podem passar por uma derivatização química, para que se tornem voláteis e, conseqüentemente, detectáveis. Esta abordagem aumenta consideravelmente o número de metabólitos que podem ser detectados por GC-MS em uma amostra biológica e é comumente aplicada em metabolômica global e metabolômica alvo.⁸¹

No trabalho em questão, a análise metabolômica global baseou-se na técnica de LC-HRMS e, portanto, essa será discutida com maiores detalhes.

1.6.1. Cromatografia Líquida Acoplada à Espectrometria de Massas

A técnica de cromatografia líquida acoplada à espectrometria de massas (LC-MS) se tornou uma tecnologia líder nas análises de pequenas e grandes moléculas polares e apolares.⁸²⁻⁸⁵ A grande popularidade da técnica de LC-MS decorre de sua versatilidade, permitindo a análise de uma ampla variedade de moléculas, com diferentes propriedades físico-químicas.^{86,87} No uso de LC-MS, não estão envolvidas altas temperaturas como nas análises de GC-MS e, portanto, não há a necessidade de as moléculas serem estáveis termicamente, descartando a necessidade de derivatização química da amostra.

Antes da etapa de detecção por MS, ocorre um processo de separação das moléculas na coluna cromatográfica, sendo essa etapa de grande importância para a análise de amostras biológicas complexas. A cromatografia líquida (LC) permite a separação de moléculas com ampla faixa de polaridade usando métodos de eluição isocrático ou eluição gradiente, sendo esse último o preferido nas análises metabolômicas global, devido à diversidade de constituintes na amostra com distintos fatores de retenção cromatográfico (k).⁸⁸

Com o objetivo de alcançar uma ampla cobertura do metaboloma, nas análises metabolômicas emprega-se preferencialmente o modo de separação no

modo reverso de eluição (RPLC, do inglês: *reversed phase liquid chromatography*), com o emprego de colunas cromatográficas com fases estacionárias C18 ou C8, adequadas para analitos com características físico-químicas de baixa e média polaridade, como por exemplo, fosfolipídeos e ácidos graxos. No entanto, analitos polares e altamente hidrofílicos não são adequadamente retidos ($k > 2$) nas colunas C18 e para isso emprega-se a cromatografia de interação hidrofílica (HILIC). A cromatografia baseada no modo HILIC de separação é a escolha na análise de compostos com características físico-químicas de alta e média polaridade, como por exemplo, açúcares e ácidos orgânicos.^{88,89} As separações cromatográficas no modo RPLC e HILIC são ortogonais, ou sejam, possuem mecanismos de interação distintos entre a fase estacionária e os analitos.^{90,91}

A detecção dos compostos de interesse com emprego de espectrometria de massas (MS) é bastante versátil e de elevada especificidade quando se trata de análise de misturas complexas, como é o caso das análises metabolômicas.

O equipamento de MS é composto por um sistema de introdução de amostra, uma fonte de ionização, um analisador de massas e um detector de íons, que juntos propiciam resultados com elevada precisão, exatidão e sensibilidade (Figura 8).⁹²

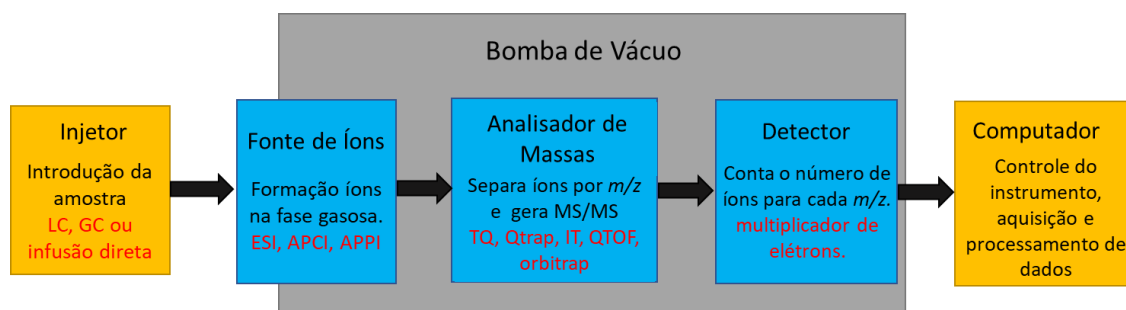


Figura 8- Representação de componentes de um Espectrômetro de Massas.

Quando acoplado a um sistema de análise por LC, as amostras são introduzidas no MS na forma líquida e são vaporizadas e ionizadas na fonte de íons. As fontes de ionização por *electrospray* (ESI) e a ionização química à pressão atmosférica (APCI) são duas fontes de ionização comumente usadas em sistemas de LC-MS. No entanto, existem outras fontes comercialmente disponíveis como ionização química por fotoionização (APPI), ionização por dessorção a laser assistida por matriz (MALDI), ionização por dessorção via *electrospray* (DESI) e análise direta em tempo real (DART), dentre outras inovações tecnológicas.^{93,94}

Embora tanto o modo de ionização por ESI quanto por APCI possam resultar em analitos ionizados, o processo de ionização difere entre ambas as fontes de ionização. Na ionização por ESI, a ionização do analito ocorre na fase líquida, enquanto a ionização por APCI, ocorre na fase gasosa e, para que isso aconteça, valores elevados de temperatura e gases de dessolvatação são empregados. ESI é a fonte de ionização preferencialmente utilizada nas análises metabolômicas, pois essa é indicada para moléculas de baixa e alta massa molecular, de média e alta polaridade. Já a fonte APCI é recomendada para moléculas de baixa e média massa molecular, e de média e baixa polaridade. Ambas as fontes de ionização propiciam ionização de analitos no modo positivo e negativo de ionização.⁹⁵

Após a etapa de ionização, a técnica de MS detecta os íons provenientes dos analitos constituintes da amostra de acordo com a razão massa-carga (m/z). Com isso, é possível obter informações sobre a massa molecular e a estrutura química, permitindo a identificação e quantificação dos compostos de interesse.

As análises metabolômicas globais fazem uso de analisadores de massa de alta resolução (HRMS) que incluem analisadores de massas híbrido do tipo quadrupolo acoplados a tempo de voo (QqTOF) e analisadores do tipo *Orbitrap*. Existem três aspectos principais a serem considerados com o uso de HRMS: alto poder de resolução de massas, precisão de massa exata (erro < 2 ppm) e a obtenção de íons extraídos com uma estreita janela de extração em massa (*Orbitrap* < 2 ppm e QqTOF < 5ppm). Durante a aquisição de dados, os analisadores de massas podem ser operados no modo *full MS scan* (ou modo de varredura completa) para um amplo intervalo de m/z 50-2000. Paralelamente ao modo *full MS scan*, pode ser associada as aquisições no modo fragmentação MS/MS, gerando espectros de íons produtos, os quais auxiliam a identificação estrutural e, em alguns casos, podem ainda maximizar a detecção de alguns metabólitos por diminuição de íons interferentes.⁹⁶⁻

98

1.7. Processamento de dados

1.7.1. Análises Multivariadas e Univariadas

Após a análise instrumental, o processamento de dados é uma das etapas essenciais das análises metabolômicas, afinal uma quantidade muito expressiva de dados é gerada revelando dificuldades na análise manual e requerem processamento de acordo com cada técnica que foi utilizada para a sua aquisição.⁹⁶

No caso das análises por LC-HRMS, é necessário criar uma matriz de dados que será usada no processamento dos dados. A matriz de dados é obtida pelo processo de extração, detecção e deconvolução dos *features* moleculares, gerando um conjunto de dados que é agrupado e alinhado, relacionando o valor de *m/z* a um tempo de retenção. Durante essa etapa, os dados já podem ser normalizados para que a variância e o escalonamento desejados sejam alcançados.⁹⁹

Os *softwares* empregados para extrair os *features* moleculares, gerar a matriz de dados e processamento dos dados são específicos de cada fabricante dos sistemas de LC-HRMS, sendo esses os softwares *Profile Analysis*[®] (Brucker), *Progenesis QI* (Waters), *CompoundDiscovery*[™] (Thermo Scientific), *MetaboScape*[®] (Brucker) e *MassHunter Profinder* (Agilent). Da mesma forma, vários desenvolvedores têm trabalhado no desenvolvimento de softwares de livre acesso que podem ser empregados nas etapas de processamento dos dados com análises estatísticas, sendo os mais comumente usados o XCMS¹⁰⁰, MZmine¹⁰¹, MetaboAnalyst^{102,103} e MS-dial¹⁰⁴.

Vale ressaltar que o processamento dos dados metabolômicos necessitam de análises estatísticas multivariada e univariada, para uma melhor visualização de dados relevantes, conseqüentemente, uma melhor interpretação dos *features* moleculares que sejam diferenciados nos grupos estudados.¹⁰⁵

A **análise multivariada** é importante nos estudos metabolômicos por ser capaz de inferir, a partir de uma base de dados complexos, mais de uma variável simultaneamente em múltiplas dimensões, considerando-se os efeitos de todas as variáveis nas respostas de interesse. No caso da análise metabolômica global, as múltiplas variáveis são classes diferentes a partir de um conjunto de dados, os quais serão agrupados quanto à similaridade das amostras. Por exemplo, grupo estudo e grupo controle.^{106,107}

Na análise estatística multivariada emprega-se dois métodos para reconhecimento de padrões, a análise de componentes principais (PCA; do inglês: *partial component analysis*) e a análise discriminante por mínimos quadrados parciais (PLS-DA; do inglês: *partial least squares discriminat analysis*). A análise por PCA é um método de análise não supervisionado que tem como objetivo visualizar a presença de tendências de separação das amostras. Observa-se assim que as variáveis sofrem uma redução na dimensionalidade dos dados, sendo reduzida de três para duas dimensões. A PCA emprega combinações lineares e não ortogonais de

variáveis, preservando a variância dos dados originais tanto quanto possível. A PCA é melhor visualizada através de gráficos de *score plotting* com a plotagem das amostras em relação a dois ou três componentes principais (PC1, PC2, PC3), facilitando a visualização dos resultados obtidos.^{107,108}

É interessante destacar ainda que a PCA realiza um agrupamento das amostras por padrões, demonstrando as informações de cada variável em relação a um determinado grupo de variáveis, mesmo sem nenhuma informação da amostra analisada, portanto é muito utilizada nos estudos metabolômicos globais.

A análise estatística com emprego do método PLS-DA, uma análise multivariada supervisionada, é também amplamente utilizada, pois combina a redução da dimensionalidade com a análise discriminante em um algoritmo. A diferença entre PCA e PLS-DA é que o método por PCA modela a variância máxima (PC1 + PC2) enquanto o PLS-DA modela a covariância máxima entre (PC1 e PC2). Com isso, ocorre uma predição de qual variável (metabólito) está mais relacionada com a resposta.^{105,107} No caso em questão, quais variáveis são discriminantes para a DW.

A PLS-DA busca realizar uma separação dos grupos a partir de uma combinação linear das variáveis, resultando em um agrupamento mais claro, quando comparado com o modelo obtido com a PCA. Adicionalmente à PLS-DA, tem-se também a análise discriminante por mínimos quadrados parciais ortogonais (OPLS-DA).^{106,109}

A análise multivariada OPLS-DA possui o mesmo poder preditivo da PLS-DA, mas permite uma melhor separação das classes no espaço, por incorporação de um sinal ortogonal de correção. Assim há a separação das variáveis preditivas das variáveis não correlacionadas com a resposta obtida, logo a visualização dos dados através de gráficos de dispersão é facilitada.¹⁰⁷ Nas análises PLS-DA ou OPLS-DA, obtém-se valores de R^2 e Q^2 , que são parâmetros relacionados ao coeficiente de regressão e a capacidade de predição do modelo, respectivamente.¹¹⁰ Importante ressaltar que os modelos PLS-DA ou OPLS-DA devem sempre ser validados, ou seja, o uso de validação cruzada indicará a capacidade preditiva do modelo e a sua adequação. Nessa etapa, as curvas característica de operação do receptor (ROC; do inglês: *receiver operating characteristic*) também são empregadas como ferramentas estatísticas de diagnóstico. A curva ROC é um método simples e muito útil que possibilita avaliar a variação da sensibilidade (taxa de verdadeiro positivo) e da especificidade (taxa de

falso positivo) de um teste de classificação. A área completa sob uma determinada curva ROC, ou área sob a curva ROC (AUC; do inglês: *area under the curve*), representa a probabilidade de que a predição estará correta quando se observa uma variável de teste.¹¹¹

Portanto, o intuito do processamento de dados por análises multivariadas com emprego de PCA, PLS-DA ou OPLS-DA é o de realizar um aperfeiçoamento dos resultados e, conseqüentemente, uma melhor interpretação dos dados encontrados.

A **análise univariada** consiste no teste de hipótese estatística para uma diferença de médias em uma variável específica entre dois grupos de amostras. A análise de dados metabolômicos por métodos estatísticos univariados é essencial, pois uma variável pode ser não significativa no modelo multivariado, mas significativa no contexto univariado. A razão para isso decorre do fato que os modelos multivariados podem, muitas vezes mascarar variáveis significativas à medida que todos os metabólitos e as relações entre os metabólitos são avaliados simultaneamente. Outra consequência é a dificuldade de métodos multivariados em estimar a presença de covariáveis.¹¹² Com isso, as análises univariadas assumem um papel de grande relevância na validação de potenciais biomarcadores.

A normalidade da distribuição dos dados deve ser inicialmente avaliada para a seleção do teste estatístico posterior. De forma geral, *Shapiro-Wilks* é geralmente mais empregado para detectar ausência de normalidade em pequenos conjuntos de amostras ($n < 50$); enquanto *Kolmogorove-Smirnov* (K-S) e *Lilliefors* (modificação de K-S) são mais eficazes quando o número de amostras for maior ($n > 50$). Se a normalidade da distribuição é confirmada, a discriminação entre as variáveis significativas empregará os testes paramétricos de *t-Student* ($n = 2$ grupos) ou a análise de variância (ANOVA, $n > 2$ grupos). Caso contrário, métodos não paramétricos devem ser usados, como o Mann-Whitney U ($n = 2$ grupos) ou o Kruskal-Wallis ($n > 2$ grupos).^{105,113,114}

A transformação dos dados (raiz quadrada, transformações logarítmicas, hiperbólicas, Box-Cox)¹¹⁵ é usual a fim de obter uma distribuição normal dos dados para serem empregados métodos paramétricos, para se controlar o efeito de covariáveis. As covariáveis são variáveis que não estão entre as principais variáveis de pesquisa, mas podem afetar a variável dependente e sua relação com a variável independente, por exemplo, gênero e idade entre indivíduos estudados. Nesses

casos, a análise de covariância (ANCOVA; do inglês: *analysis of covariance*) deverá ser empregada. A ANCOVA é um método estatístico que combina ANOVA e a análise de regressão para ajustar o efeito linear à covariável.¹¹⁶

Nos estudos metabolômicos globais, cada amostra biológica resulta em um elevado número de *features* moleculares, sendo que cada *feature* precisa ser investigado por métodos univariados para evitar a ocorrência de resultados falsos positivos. Para minimizar as taxas de falsos positivos, são empregados métodos estatísticos, como de *Sidak*, *Bonferroni*, *Bonferroni-Holm* e de *Benjamini-Hochberg*, para corrigir os valores de significância p . A correção de *Benjamini-Hochberg* é também conhecida como taxa de falsa descoberta (FDR) e é um dos métodos recomendados para controlar a proporção de falsos positivos em todos os resultados significativos.^{105,117}

2. Objetivos

Os objetivos desse trabalho foram de obter uma melhor compreensão dos mecanismos bioquímicos associados à Doença de Wilson e explorar novos métodos que auxiliem o diagnóstico para que a doença seja precocemente e corretamente detectada, prevenindo complicações clínicas. Adicionalmente, identificar potenciais candidatos a biomarcadores que permitam, além de auxiliar no diagnóstico, monitorar a progressão da doença e a eficácia medicamentosa e contribuir para o desenvolvimento de terapias medicamentosas alternativas às atuais.

Para alcançar esses objetivos, uma análise abrangente e comparativa de amostras de plasma e urina de pacientes diagnosticados com a doença de Wilson e de um grupo de indivíduos saudáveis foi conduzida através da abordagem metabólica global com emprego da cromatografia líquida acoplada à espectrometria de massas de alta resolução (LC-HRMS).

3. Materiais e Métodos

3.1. Generalidades

Os experimentos foram realizados no *SEPARARE* - Núcleo de Pesquisa em Cromatografia do Departamento de Química da Universidade Federal de São Carlos (UFSCar). Os reagentes e solventes utilizados no preparo de amostras e análises por LC-HRMS foram todos de alto grau de pureza, sendo ácido fórmico (grau MS, Sigma-Aldrich, Alemanha), acetonitrila (ACN) (grau HPLC, Merck, Estados Unidos), metanol (MEOH) (grau HPLC, Merck, Estados Unidos) e etanol (grau HPLC, Merck, Estados Unidos). Todas as soluções aquosas utilizaram água ultrapura proveniente de um sistema Milli-Q (Millipore, Brasil).

Para o preparo de amostras, utilizou-se micropipetas Gilson (Gilson, Estados Unidos) e/ou Eppendorf ((Eppendorf, Alemanha), vórtex MS3 basic (IKA, Alemanha) e centrífuga refrigerada 5415R (Eppendorf®, Alemanha).

3.2. Casuística e Amostragem

Este estudo foi conduzido de acordo com a Declaração de *Helsinki* e aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Brasil (protocolo: 3.728.710, aprovado em 27 de novembro de 2019 e CAAE: 25138319.0.0000.0068). Todos os participantes receberam esclarecimentos sobre a pesquisa e concordaram em participar, assinando um Termo de Consentimento Livre e Esclarecido (TCLE). As amostras biológicas de plasma e urina dos pacientes com Doença de Wilson (WD) foram coletadas pela enfermeira Celina Teixeira Prado, no Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Clínica de Gastroenterologia. Os pacientes foram selecionados e monitorados pelo Prof. Dr. Eduardo Luiz Rachid Cançado. O grupo controle foram pessoas saudáveis e as amostras biológicas foram coletadas na Universidade Federal de São Carlos.

Por se tratar de patologia rara, não se aplicaram cálculos estatísticos para determinar o número de indivíduos (n), buscando-se a inclusão de um número de mínimo de n=20 participantes, conforme critérios de inclusão e exclusão. Foram coletadas amostras de plasma de indivíduos portadores da DW (n=22) e de indivíduos saudáveis (n=35), pertencentes ao grupo controle. Para a coleta de urina, foram coletadas amostras de indivíduos portadores da DW (n=23) e de indivíduos saudáveis (n= 37), pertencentes ao grupo controle. A faixa etária média dos participantes da

pesquisa está descrita na Tabela 3 dos Resultados e Discussão (pág. 41). Os participantes foram divididos em dois grupos: indivíduos com a doença de Wilson (DW) e indivíduos saudáveis (GC), sem diagnóstico para a doença de Wilson.

3.3. Critérios de Inclusão e Exclusão

3.3.1. Critérios de Inclusão

Foram incluídos os pacientes adultos (idade > 18 anos) cujo diagnóstico de DW foi estabelecido mediante achados clínicos, laboratoriais e de exames complementares, de acordo com os seguintes critérios.³²

- Alterações hepáticas clínicas, bioquímicas, anatomopatológicas ou ultrassonográficas;
- Doença neurológica com distonia, extra-piramidal, cerebelar ou convulsões, com ou sem manifestações psiquiátricas associadas;
- Presença na córnea de anéis de Kayser-Fleischer;
- Níveis de Cp sérica < 20 mg/dL;
- Cuprúria > 100 µg/24 h, ou após estímulo com uso de penicilamina (1,0 g) por via oral > 1600 µg/24 h em crianças ou > 1400 µg/24 h em adultos;
- Cobre hepático > 250 µg/tecido seco;
- Biópsia hepática com achados como glicogênio nucleolar, esteatose hepática, fibrose, hepatite crônica ou cirrose, reatividade para coloração de rodanina;
- Ressonância magnética de encéfalo com evidências de deposição de cobre em gânglios da base.

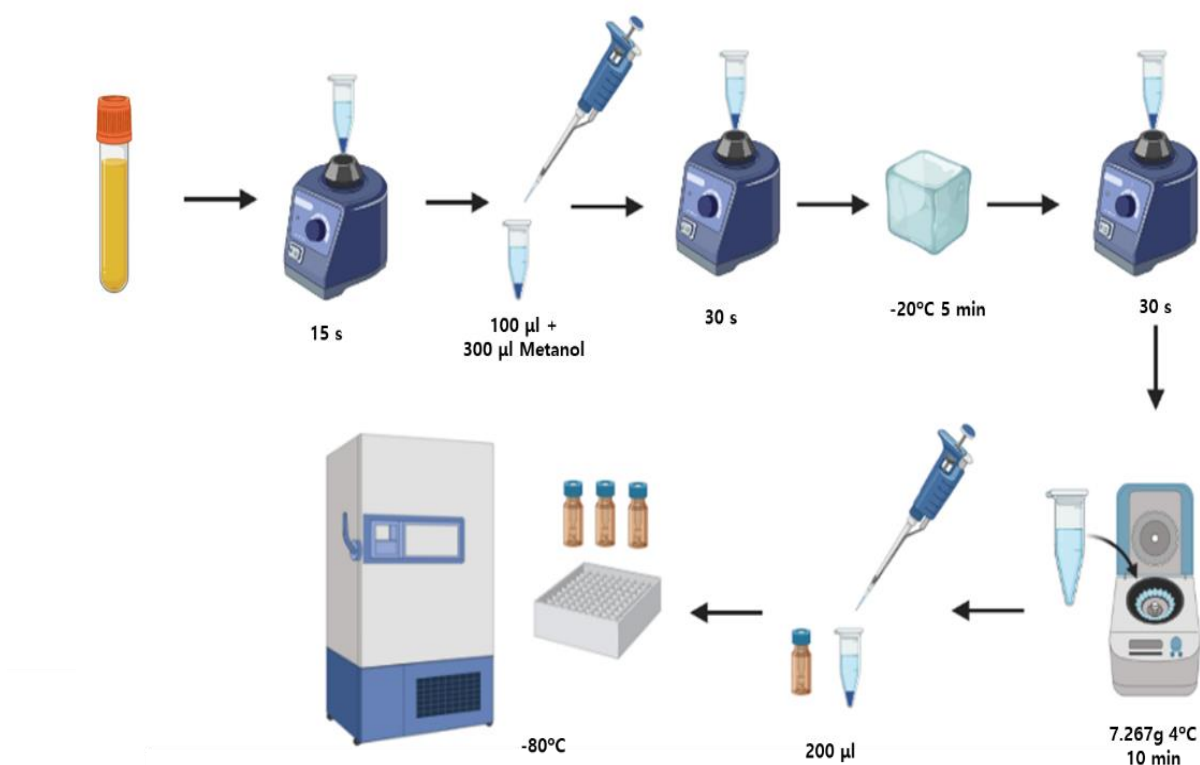
3.3.2. Critérios de Exclusão

Foram usados como critérios de exclusão pacientes que não tinham diagnóstico para DW confirmado, menores de 18 anos de idade, sem uso de álcool, drogas ilícitas e medicações hepatotóxicas e que tenham recusado fornecer amostras para fins de pesquisa acadêmica.

3.4. Preparo de amostras para análise metabolômica

3.4.1. Plasma

As amostras de plasma foram degeladas em banho de gelo, homogeneizadas em vórtex por 15 s e posteriormente extraídas via precipitação de proteínas. Uma alíquota de 100 μL de plasma de cada amostra foi transferida para tubos de centrifugação (1,5 mL) e nelas adicionada uma alíquota de 300 μL de metanol gelado (1:3 v/v plasma:MeOH). Na sequência, os tubos de centrifugação foram agitados durante 30 s em vórtex e, então, a mistura foi mantida a -20°C por 5 min. Após esse período, os tubos de centrifugação foram agitados novamente por 30 s e as amostras foram centrifugadas a 7.267 g a 4°C por 10 min. A seguir, alíquotas de 200 μL do sobrenadante foram transferidas para *vials* de injeção e armazenados a -80°C até a análise por LC-HRMS. A amostra de controle de qualidade (QC) foi preparada após a etapa de precipitação das proteínas, na qual uma alíquota de 15 μL do sobrenadante de cada amostra individual, dos pacientes e indivíduos saudáveis ($n=67$), foi transferida para um tubo de centrifugação de 5 mL, sendo a mistura final homogeneizada em vórtex por 10 s. A amostra de controle de qualidade da amostra em branco (QB) foi composta por 100 μL de água, substituindo o volume de plasma e extraída usando o mesmo cronograma de preparo. Um esquema experimental do procedimento empregado na extração dos metabólitos do plasma está descrito na Figura 9.



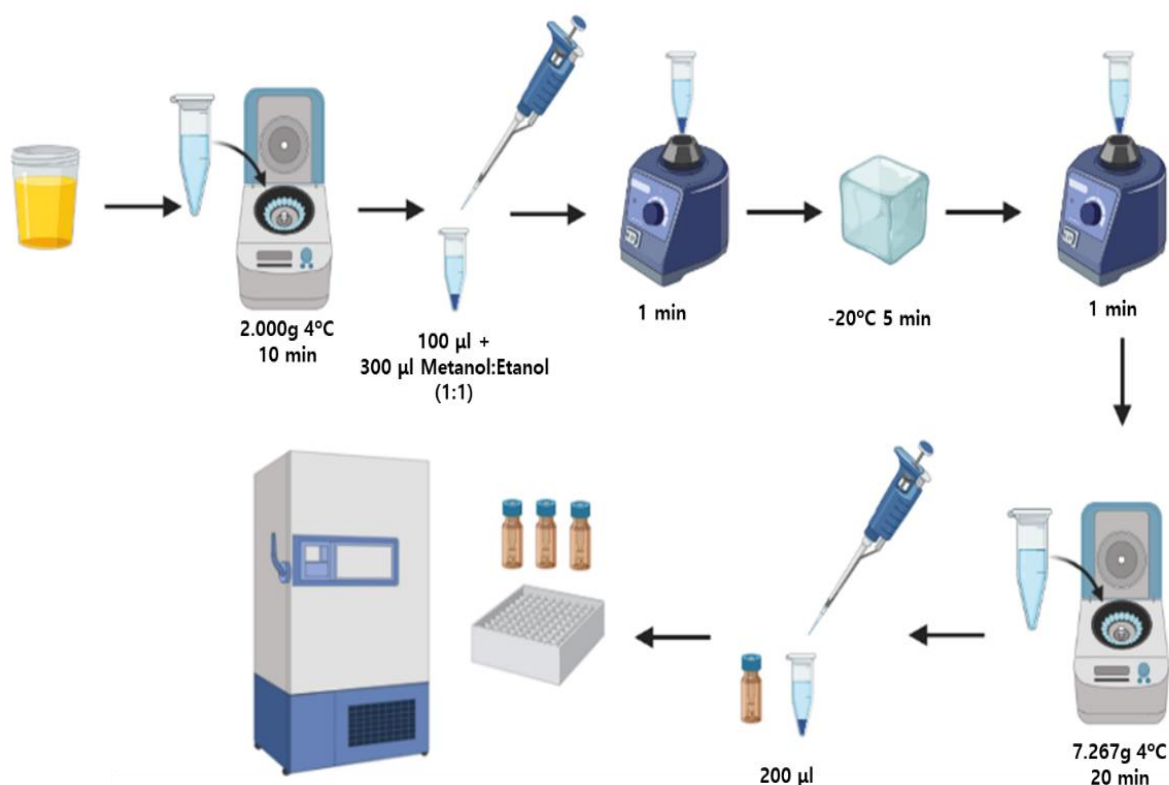
Fonte: Elaborada pela autora.

Figura 9- Esquema simplificado do preparo das amostras de plasma.

3.4.2. Urina

As amostras de urina foram degeladas em banho de gelo e centrifugadas a 2.000 g a 4°C por 10 min em tubos de centrifugação (2,0 mL). Em seguida, foi retirada uma alíquota de 100 µL de cada amostra e transferida para novo tubo de centrifugação (1,5 mL), ao qual foram adicionados 300 µL da mistura etanol:metanol (1:1 v/v) gelada, na mesma proporção de 1:3 (v/v) (urina:solvente orgânico). Na sequência, os tubos de centrifugação foram agitados durante 1 min e em seguida a mistura foi mantida a -20°C por 5 min. Após esse período, as amostras foram agitadas novamente por 1 min e centrifugadas a 7.267 g a 4°C por 20 min. A seguir, alíquotas de 200 µL do sobrenadante foram transferidas para *vials* de injeção e armazenados a -80°C até a análise por LC-HRMS. A amostra de controle de qualidade (QC) foi preparada após a etapa de extração da urina, na qual uma alíquota de 15 µL do sobrenadante de cada amostra individual, dos pacientes e indivíduos saudáveis (n=69), foi transferida para um tubo de centrifugação de 5 mL, sendo a mistura final homogeneizada em vórtex por 10 s. A amostra branco (QB) foi preparada por 100 µL

de água, substituindo o volume de urina. Um esquema experimental do procedimento empregado na extração dos metabólitos da urina está descrito na Figura 10.



Fonte: Elaborada pela autora.

Figura 10- Esquema simplificado de extração das amostras de urina.

3.5. Análise Instrumental por LC-HRMS

As análises metabolômicas foram realizadas com o emprego de um cromatógrafo líquido de ultra eficiência (UHPLC) Agilent 1290 Infinity II (Agilent Technologies, Santa Clara, CA, Estados Unidos) composto por uma bomba binária (G7120A – *High speed Pump*), autoinjeter (G7129B) e um compartimento de coluna (G7116B) acoplado a um espectrômetro de massa de alta resolução com analisador híbrido do tipo quadrupolo - tempo de voo (QqTOF), modelo Impact HD (Bruker Daltonics, Bremen, Alemanha), com uma fonte de ionização por eletrospray operando no modo positivo (ESI+) ou no modo negativo (ESI-). A aquisição e processamento dos dados foi realizada utilizando o software Data Analysis® 4.2 (Bruker Daltonics GmbH, Bremen, Alemanha) e o Profile Analysis® 2.1 (Bruker Daltonics GmbH, Bremen, Alemanha).

3.5.1. Separação Cromatográfica no Modo Reverso de Separação

A separação cromatográfica de metabólitos no plasma e na urina no modo de eluição em fase reversa (RPLC) foi realizada usando uma coluna Zorbax Eclipse XDB C18 (3.0 x100mm i.d; 3,5 μ m) (Agilent Technologies), mantida a 40°C. As fases móveis consistiram em água (solvente A) e acetonitrila (solvente B), ambos com 0,1% de ácido fórmico. O tempo total de análise foi de 30 min, a uma vazão de 0,4 mL/min, volume de injeção de 3 μ L e acondicionamento das amostras a 15°C. O gradiente de eluição usado seguiu a seguinte ordem: $t_0=0$ min, 1% B; 0-3,0 min, 1-2% B; 3-10 min, 2-30% B; 10-15 min, 30-50% B; 15-18 min, 50-80% B; 18-20 min, 80-90% B; 20-22 min, 90-95% B; 22-26 min, 95-99% B; 26,01-28 min, 99% B, para limpeza da coluna, e 28,01-30 min, 1% B, para reequilíbrio da coluna cromatográfica. Durante a limpeza da coluna, a válvula de 6-caminhos desviou a vazão da fase móvel para o descarte, enquanto propiciou a infusão da solução de calibração (formiato de sódio 1mmol/L em ACN) na fonte de ionização. Essa etapa permite a aquisição de íons de referência para a calibração de massas a cada análise.

3.5.2. Separação Cromatográfica por Interação Hidrofílica

3.5.2.1. Amostras de Plasma

A separação cromatográfica de metabólitos no plasma no modo de eluição por interação hidrofílica (HILIC) foi realizada usando uma coluna Zorbax Hilic Plus (2.1x100mm i.d; 3,5 μ m) (Agilent Technologies), mantida a 30°C. As fases móveis consistiam em água (solvente A) e acetonitrila (solvente B), ambos com 0,1% de ácido fórmico. O tempo total de análise foi de 21 min, a uma vazão de 0,4 mL/min, volume de injeção de 3 μ L e acondicionamento das amostras a 15°C. O seguinte gradiente de eluição foi usado: $t_0=0$ min, 99% B; 0-3,0 min, 99-98% B; 3-10 min, 98-80% B; 10-16 min, 80-50% B; 16,50-18 min, 1% B, para limpeza da coluna, e 18,10-21 min, 99% B, para reequilíbrio da coluna. Durante a limpeza da coluna, a válvula de 6-caminhos desviou a vazão da fase móvel para o descarte, enquanto propiciou a infusão da solução de calibração (formiato de sódio 1mmol/L em ACN) na fonte de ionização. Essa etapa permite a aquisição de íons de referência para a calibração de massas a cada análise.

3.5.2.2. Amostras de Urina

A separação cromatográfica de metabólitos na urina no modo de eluição por HILIC foi realizada usando uma coluna Infinitylab Poroshell 120 Hilic (2.1 x100mm i.d; 1,9 μ m) (Agilent Technologies), mantida a 30°C. As fases móveis consistiam em água (solvente A) com de formiato de amônio (0,1 mol/L; pH=3) e acetonitrila (solvente B). O tempo total de análise foi de 23 min, a uma vazão de 0,4 mL/min, volume de injeção de 1 μ L e acondicionamento das amostras a 15°C. O seguinte gradiente de eluição foi usado: $t_0=0$ min, 99% B; 0-3,0 min, 99-98% B; 3-10 min, 98-80% B; 10-18 min, 80-50% B; 18,50-20 min, 1% B, para limpeza da coluna, e 20,10-23 min, 99% B, para reequilíbrio da coluna. Durante a limpeza da coluna, a válvula de 6-caminhos desviou a vazão da fase móvel para o descarte, enquanto propiciou a infusão da solução de calibração (formiato de sódio 1mmol/L em ACN) na fonte de ionização. Essa etapa permite a aquisição de íons de referência para a calibração de massas a cada análise.

3.6. Parâmetros para a Aquisição de Dados no Sistema LC-HRMS

Os dados foram adquiridos utilizando um espectrômetro de massas de alta resolução (HRMS) contendo um analisador de massas híbrido do tipo quadrupolo-tempo de voo (QqTOF), modelo Impact HD QTOF™ (Bruker Daltonics, Bremen, Alemanha) equipado com uma interface *elektrospray* (ESI) operando em modo de ionização positivo ou negativo. O modo de aquisição centróide foi usado para coletar e armazenar os dados. Primeiramente, os espectros de massas foram obtidos no modo de ionização positivo (ESI+) e depois no modo de ionização negativo (ESI-).

Os parâmetros ideais para as análises no modo RPLC foram: voltagem do capilar de +4,0 e -3,0 kV para modo de ionização positivo e negativo, respectivamente; potencial do cone de transferência (*end plate offset*) 500 V; gás nebulizador (N₂) 4 bar; temperatura do gás de secagem de 180 °C (N₂); vazão do gás de secagem (N₂) 8 L/min; energia da célula de colisão 5 eV e intervalo de varredura completa MS (*full MS scan*) de m/z 50-1300.

Para as análises HILIC, os parâmetros foram voltagem do capilar de +4,5 kV e -3,0 kV para o modo positivo e negativo, respectivamente; potencial do cone de transferência (*end platte offset*) 450 V; gás nebulizador 4 bar; temperatura do gás de

secagem (N₂) 180 °C; vazão do gás de secagem (N₂) 8 L/min; e intervalo de varredura completa MS (*full MS scan*) de *m/z* 50-1300.

O modo MS/MS de aquisição dependente de dados (DDA) utilizou parâmetros dinâmicos com radiofrequência de colisão (RF) variando de 200,0 a 550,0 Vpp; tempo de transferência de íons de 50,0 a 90,0 µs; com temporização de 50,0% em cada; RF 1 e RF 2 do funil foram de 250,0 e 150,0 Vpp, respectivamente. A RF do hexapolo foi de 50,0 Vpp e a energia do quadrupolo foi de 5,0 eV com armazenamento de íons pré-pulso de 6,0 µs. A energia do quadrupolo e a energia da célula de colisão foram ambas mantidas fixas em 5eV (empregadas somente para deslocamento dos íons). Os parâmetros utilizados para acionar a fragmentação MS/MS foram de velocidade de aquisição de 2,0 Hz para íons de baixa intensidade (10.000 contagem/por soma de 1.000 íons) e 4,0 Hz para íons de alta intensidade (100.000 contagens/por soma de 1.000 íons), utilizando um tempo total de ciclo de 3 s. A intensidade absoluta foi de 1.491 contagens (302 contagens/por soma de 1.000 íons), com exclusão ativa de 1 espectro e liberação após 0,90 min. A aquisição no modo MS (*full MS scan*) foi ajustada a uma velocidade de aquisição de 2,0 Hz. A energia de colisão utilizada para fragmentação de íons foi programada para variar de 250,0 a 100,0% do valor fixo de 20 eV inicialmente definido, com a seguinte janela de isolamento: *m/z* 100, 200 e 300: largura do pico de 4; para *m/z* 700 e 1000: largura do pico de 6. A calibração interna do espectrômetro de massas foi realizada com a infusão de formiato de sódio em acetonitrila (1 mmol/L), utilizando um modelo de calibração de alta precisão (do inglês: High Precision Calibration- HPC). A solução de calibração foi injetada no final de cada análise e todos os espectros foram recalibrados antes do processamento dos dados. Os dados foram adquiridos utilizando o *Qtof Control Software v3.4*, e os dados brutos foram processados com o *Brucker Data Analysis v4.3*, ambos da Brucker (Bruker Daltonics, Bremen, Alemanha).

3.7. Pré-Processamento dos Dados

Utilizou-se o *software Bruker Compass ProfileAnalysis v2.1* para detectar os picos, extrair as características moleculares e realizar o alinhamento por tempo dos dados adquiridos por LC-HRMS, gerando uma matriz de dados contendo *m/z versus* tempo de retenção (*t_R*) *versus* intensidade (*count*). Essa etapa do pré-processamento de dados ocorreu no intervalo de análise de 0 a 28 minutos para a eluição cromatográfica no modo reverso para as amostras de plasma e urina e de 0 a

18 minutos para a eluição cromatográfica no modo HILIC para as amostras de plasma e de 0 a 20 minutos para as amostras de urina, não sendo considerados os tempos de análise durante as etapas de limpeza e recondicionamento da coluna. O intervalo de massa empregado foi m/z 50 a 1300 e o parâmetro para detecção de dados foi de 70%, ou seja, somente *features* moleculares detectados em pelo menos 70% das amostras foram extraídos. A geração da matriz de variáveis (*bucket table*) empregou os seguintes parâmetros adicionais: *threshold* de correlação de eficiência de 0.2, aquisição mínima de 10 espectros (*scans*) e uma suavização (*smoothing*) de 1. Os adutos considerados para ESI(+) foram: $[M+H]^+$; $[M+Na]^+$; $[M+K]^+$; $[M-H_2O+H]^+$; $[M-CO_2+H]^+$; $[M-CH_3OH+H]^+$; $[M+CH_3OH+H]^+$; $[M+CH_3CN+H]^+$; $[2M+Na]^+$ e $[2M+K]^+$ e para ESI(-) foram: $[M-H]^-$; $[M-H_2O-H]^-$; $[M+Na-H_2]^-$; $[M+K-H_2]^-$; $[M+Cl]^-$; $[M+HCOOH-H]^-$; $[M+CH_3COOH-H]^-$; $[2M-H]^-$; $[2M+HCOOH-H]^-$.

Nessa etapa, obteve-se uma matriz de dados no formato *.txt*, que relaciona um *feature* molecular (m/z) com um respectivo tempo de retenção (t_R) (par m/z - t_R) em função das intensidades dos sinais. Nessa matriz em formato *.txt* todos os pontos foram substituídos por vírgula e o pré-processamento continuou no Excel (Microsoft Office 2016).

Na sequência, empregando-se como referência as amostras de controle de qualidade das amostras branco (QB), realizou-se a subtração do ruído instrumental e de possíveis contaminantes não pertencentes às amostras biológicas de todas as amostras em processamento. O próximo passo envolveu a análise da qualidade dos dados, empregando-se somente as amostras QCs como referência. Nessa etapa foram excluídos os *features* moleculares de todas as amostras que apresentaram média da intensidade dos picos inferiores à média da intensidade do sinal do QB, através da relação ($\bar{X}_{QC} \leq 1.05\bar{X}_{QB}$). A repetibilidade das amostras de QCs foram avaliadas através coeficiente de variação (CV), sendo excluídos os *features* moleculares com $CV \geq 30\%$ dos QCs e dos grupos em estudo. Com relação a lacunas vazias (do inglês: *missing values*), os *features* moleculares que apresentavam *missing values* > 80% foram excluídos.

A continuidade de processamento de dados ocorreu na plataforma Noreva 2.0 (<http://idrblab.cn/noreva/>),¹¹⁸ a qual foi empregada para realizar a imputação de dados utilizando-se o método de k-vizinhos mais próximos (KNN; do inglês: *k-nearest neighbors*).

Na remoção de erros sistemáticos, variabilidade inter- e intra-amostra e na remoção da variabilidade instrumental dos grupos de amostras e QCs, foi utilizado o método de normalização por suavização de gráfico de dispersão ponderada localmente (LOWESS/Spline; do inglês: *locally weighted scatterplot smoothing*) versão 1.1 (<http://prime.psc.riken.jp/compms/others/main.html>). Nas amostras de urina, a diluição das amostras foi normalizada com o emprego da normalização por sinal MS total (MSTUS; do inglês: *MS total useful signal*).¹¹⁹ A partir dessa etapa, a matriz de dados pré-processada foi utilizada para conduzir as análises estatísticas nos softwares *MetaboAnalyst 5.0* (<https://www.metaboanalyst.ca/>) e IBM SPSS Statistics versão 25.0.

3.7.1. Análises Multivariadas e Univariadas

Para realizar as análises estatísticas multivariadas entre os grupos estudo (DW) e grupo controle (GC), os dados nos modos de ionização positivo e negativo foram combinados em uma única matriz de dados. Em seguida, foram aplicadas a Análise de Componentes Principais (PCA) e a Análise Discriminante de Mínimos Quadrados Parciais (PLS-DA/OPLS-DA) para análises exploratórias e discriminatórias, respectivamente. A PCA, PLS-DA/OPLS-DA foram então dimensionados usando auto-escalonamento (*autoscaling*) no módulo de análises estatísticas da plataforma *MetaboAnalyst 5.0* (<https://www.metaboanalyst.ca/>). Adicionalmente, conduziu-se a realização da validação cruzada do modelo avaliando os valores de coeficiente de determinação (R^2) e o coeficiente de predição (Q^2), através do teste de permutação. A utilização dessas ferramentas permitiu a obtenção da Importância da Variável na Projeção ($VIP \geq 1$), destacando as variáveis de maior relevância no estudo.

Utilizando os *features* moleculares com $VIPs \geq 1$, foi conduzida a análise univariada. Inicialmente, a normalidade dos dados foi avaliada com o teste de Shapiro-Wilk. Os dados transformados foram avaliados pelo teste de *t-Student* ($p < 0,05$) e pelo teste Qui-Quadrado ($p < 0,05$) para avaliar a influência da idade e do gênero dos indivíduos dos dois grupos, respectivamente. Na sequência, os dados foram transformados com a técnica de Box-Cox com o software Develve (<https://develve.net>). Após a transformação dos dados, as análises estatísticas prosseguiram com a análise de variância (ANCOVA). A ANCOVA avalia a relação entre uma variável dependente e uma variável independente, enquanto controla os

efeitos de outras variáveis que não são de interesse principal na análise. Na ANCOVA foi incluindo o indicador do status da doença DW como o principal preditor de interesse e outras covariáveis (idade) para controlar possíveis efeitos de confusão na associação entre abundância de metabólitos e DW. Os p -valores do teste ANCOVA foram ajustados com uma Taxa de Falsa Descoberta (FDR) de 10% para minimizar erros do tipo I. O erro do tipo I rejeita a hipótese nula (H_0) quando ela é verdadeira. Todas essas análises foram realizadas utilizando o software IBM *SPSS Statistics* versão 25.0.

Valores de *fold-change* (FC) foram expressos como regulação positiva (*up regulation*) quando [valores médios FC = DW/GC] $\geq 1,5$ ou regulação negativa (*down regulation*) quando [valores médios FC = -1/ (valores médios DW/GC)]. Assim, foram considerados potenciais metabólitos discriminantes somente aqueles que apresentaram um intervalo de $0,7 \geq FC \geq 1,5$ e $p \leq 0,05$.

O grau de relacionamento entre duas variáveis (metabólitos) foi avaliado com o emprego de teste de coeficiente de correlação de Pearson (r), sendo essa uma medida adimensional e que pode assumir valores de +1 a -1. Foram avaliados somente os metabólitos que apresentaram VIP scores ≥ 1 e $0,7 \geq FC \geq 1,5$. Dessa forma, quando existe correlação direta com $r = +1$ ou $r = -1$, existe uma correlação perfeitamente positiva/negativa, em que as variáveis aumentam e/ou diminuem na mesma direção. Se a correlação for indireta, negativa e/ou positiva, $-0,5 \geq r \geq +0,5$, as variáveis variam inversamente; já valores próximos a zero, sugerem ausência de correlação linear entre as variáveis. Nesse trabalho foram consideradas correlações importantes àquelas com valores mínimos no intervalo de $-0,5 \geq r \geq +0,5$, com base na seguinte classificação (Tabela 2).

Tabela 2- Classificação do coeficiente de correlação de Pearson (r)¹²⁰

Valor de r (+ ou -)	Correlação
0 (zero)	Nula
0 - 0,30	Biologicamente desprezíveis
0,31 – 0,50	Fraca
0,51 – 0,70	Moderada
0,71 – 0,90	Forte
0,91 – 0,99	Muito forte
1	Perfeita

3.8. Anotação Putativa dos Metabólitos

A anotação putativa dos *features* moleculares estatisticamente significativos foi conduzida utilizando o software *Progenesis*[™] QI v2.4 (Waters Corporation[™] © Nonlinear Dynamics, Newcastle, Reino Unido). Os dados brutos (.d) foram processados utilizando o descompactador *CompassXtract* 3.1 (Bruker Daltonics). A anotação dos *features* moleculares foi baseada em critérios como a massa exata do íon molecular com uma exatidão ≤ 5 ppm, perfil de fragmentação (MS/MS) com um erro de massa ≤ 10 ppm, similaridade isotópica e conhecimento biológico dos metabólitos. Essas características relacionam-se com o nível 2 de confiança na anotação ref.¹²¹ Adicionalmente, foram considerados os seguintes adutos: $[M+H]^+$, $[M+K]^+$, $[M+NH_4]^+$ e $[M+H-H_2O]^+$ no modo positivo de ionização e no modo negativo foram considerados $[M-H]^-$, $[M+Cl]^-$ e $[M-H_2O-H]^-$. Para auxiliar na anotação foram utilizadas bibliotecas externas como *LipidMaps* (<http://www.lipidmaps.org/>), *Human Metabolome Database* (<http://www.hmdb.ca/metabolites>) e as bibliotecas disponíveis no *MoNA - MassBank of North America* (<https://mona.fiehnlab.ucdavis.edu/>). Essas bibliotecas fornecem informações sobre a estrutura química, propriedades físico-químicas e a importância biológica dos metabólitos, auxiliando na anotação das moléculas presentes nas amostras analisadas.

Após a anotação putativa dos metabólitos foram geradas curvas características de operação do receptor (ROC; do inglês: *receiver operating characteristic*), exibindo taxas de sensibilidade e especificidade, com um intervalo de confiança de 95%.

3.9. Identificação das Rotas Bioquímicas

As vias bioquímicas com importância biológica foram obtidas a partir dos metabólitos que apresentaram alterações estatísticas significativas entre as amostras dos indivíduos com DW e GC e foram identificadas utilizando a ferramenta *Pathway Analysis* do programa *MetaboAnalyst* 5.0, com base no banco de dados da *Kyoto Encyclopedia of Genes and Genomes* (KEGG) *Homo sapiens*.

4. Resultados e Discussão

4.1. Dados Demográficos

O número de indivíduos participantes foi dividido em dois grupos: pacientes diagnosticados com doença de Wilson, designado como grupo estudo (DW) e um grupo controle composto por indivíduos saudáveis, sem diagnóstico para a doença (GC). Foram obtidas 57 amostras de plasma, sendo 22 amostras provenientes do grupo DW e 35 amostras do GC e 60 amostras de urina, sendo 23 amostras do grupo DW e 37 amostras do GC. Informações adicionais sobre gênero e faixa etária estão descritas na Tabela 3.

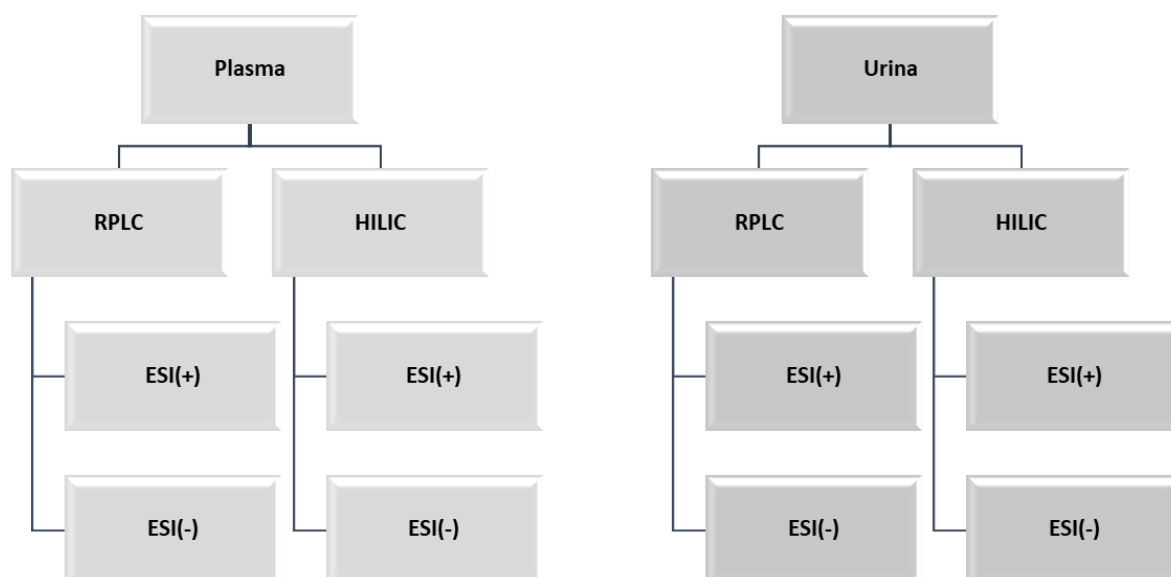
Tabela 3- Informações demográficas das amostras de plasma e urina que integraram o grupo estudo (DW) e o grupo controle (GC).

	Matriz biológica			
	Plasma		Urina	
Categoria	DW	GC	DW	GC
Número de indivíduos (n)	22	35	23	37
Gênero feminino (%)	9	20	9	20
Gênero masculino (%)	13	15	14	17
Idade (Média ± DP)	35 ± 13	32 ± 10	38 ± 12	31 ± 9

4.2. Pré-Processamento dos Dados de Metabolômica

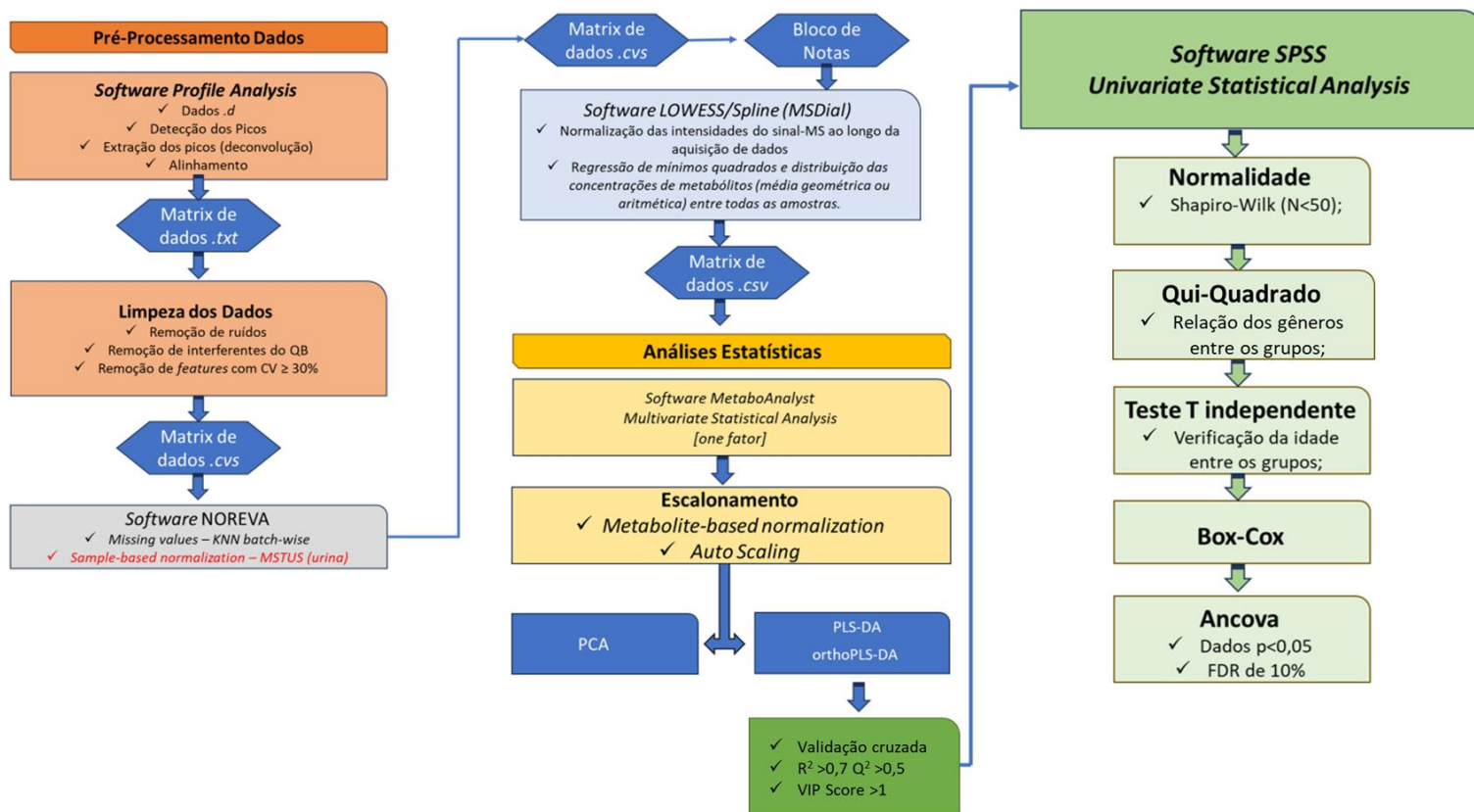
Global

Nesse estudo foram empregadas duas matrizes biológicas (plasma e urina) e cada uma delas foi analisada por LC-HRMS empregando os modos de separação RPLC e HILIC e os modos de ionização ESI(+) e ESI(-), conforme Fluxograma 1. Ao estudar o perfil metabolômico de plasma e da urina, objetivou-se identificar processos bioquímicos que poderiam não ser devidamente detectados se apenas um biofluido fosse avaliado, o que poderia acontecer devido à diferença de níveis de concentração e devido ao efeito de supressão de ionização em uma ou outra matriz biológica. Com isso, o uso de duas diferentes matrizes biológicas tem o potencial de propiciar uma abordagem metabolômica mais abrangente.



Fluxograma 1- Modos de separação e ionização empregados para cada matriz biológica nas análises metabolômicas por LC-HRMS.

Os cromatogramas de LC-HRMS referentes a cada categoria de análise estão disponíveis no decorrer da discussão. Inicialmente, as matrizes de dados dos modos ESI(+) e ESI(-) dentro de cada categoria de separação cromatográfica foram combinadas em uma única matriz, após as etapas de pré-processamento de dados. Desta forma, os modos de eluição RPLC ESI(+) e RPLC ESI(-) foram agrupados e analisados conjuntamente no software MetaboAnalyst para cada uma das matrizes, assim como as análises no modo de eluição HILIC ESI(+) e HILIC ESI(-). A partir das matrizes combinadas, foram conduzidas as etapas de processamento dos dados, as análises estatísticas multivariada e univariada e a identificação dos metabólitos para cada matriz biológica estudada, com consequente análise das rotas bioquímicas impactadas, conforme Fluxograma 2. Desta forma, nos tópicos seguintes, serão feitas abordagens gerais do processamento de dados, os quais foram igualmente aplicados para cada categoria de amostras analisadas por LC-HRMS.



Fluxograma 2- Etapas de Processamento dos Dados Metabolômicos.

4.2.1. Detecção, Deconvolução e Alinhamentos de Picos

Previamente à análise estatística, os dados de LC-HRMS foram extraídos com o emprego do *software Profile Analysis*. Essa etapa permitiu converter os dados brutos (instrumental) em dados extraídos, o qual gerou matrizes de características de dados contendo os *features* moleculares detectados em cada amostra. O *feature* molecular é tipicamente um pico ou sinal (m/z) que representa um composto químico (metabólito) e as matrizes características contêm as intensidades dos íons (*count*) presentes na amostra *versus* o par m/z -tempo de retenção (t_R), descrevendo uma impressão digital metabolômica.¹²² Nessa etapa, os dados brutos de LC-HRMS consistem de matrizes em 3D relacionando m/z vs. t_R vs. intensidade dos íons (*count*), os quais são convertidos em uma matriz de dados 2D, relacionando uma par m/z - t_R . do *feature* molecular vs. a intensidade de sinal. Com isso, realiza-se a detecção de picos e deconvolução (do inglês: *peak picking/detection and deconvolution*). Durante a geração da matriz de dados, os *features moleculares* detectados e extraídos também passam por alinhamento para propiciar o

agrupamento de picos detectados na amostra, agrupando m/z a um intervalo t_R , e propiciando a resposta em termos de altura do pico ou da área do pico.¹²³

As matrizes de dados consistiram de linhas contendo os pares $(m/z)-t_R$ e relacionando-as com as colunas contendo a intensidade do pico (*count*) em formato *.txt* para cada modo de separação e ionização utilizado, conforme Fluxograma 2.

4.2.2. Inspeção dos Controles de Qualidade das Amostras Branco (QB)

Após a detecção/extração e alinhamento dos picos realizados no *software Profile Analysis*, foi conduzida a inspeção dos controles de qualidade das amostras branco (QB) e realizada a subtração de ruído instrumental, sinais de interferentes e/ou contaminantes presentes nas amostras branco (QB).

4.2.3. Avaliação da Qualidade dos Dados

Após a etapa de pré-processamento ("limpeza") dos dados, seguiu-se com a avaliação da qualidade dos dados. Em metabolômica global, essa etapa é conduzida através da análise das amostras de controle de qualidade (QCs), conforme proposto por Karaman¹²³ e empregando diferentes etapas de avaliação: **1)** Presença dos *features* moleculares das amostras em estudo nas amostras dos QCs; **2)** Correção dos desvios relacionados à intensidade dos picos; **3)** repetibilidade das análises dos QCs, injetados em intervalos regulares; **4)** linearidade dos dados. A linearidade dos dados não foi avaliada nesse trabalho.

1) Presença nos QCs: Nessa etapa verificou-se a presença dos *features* moleculares em pelo menos 80% do total de amostras QCs analisadas, para assegurar a boa qualidade dos dados. *Features* moleculares com $CV \geq 30\%$ foram excluídos das amostras de QC e das amostras em estudo (DW e GC).

2) Correção dos desvios relacionados à intensidade dos picos: As amostras QCs foram usadas para corrigir erros sistemáticos relacionados à variabilidade instrumental, o qual acarreta perda de sensibilidade no espectrômetro de massas (MS), decorrentes de desvios relacionados à ordem de injeção e análise.¹²⁴ Nessa etapa aplicou-se uma abordagem denominada *LOWESS/Spline*. A técnica *LOWESS/Spline* trata-se de uma técnica robusta baseada no uso das amostras QCs para promover uma suavização da dispersão dos dados que é estimada apenas amostras de controle de qualidade localmente por correção de sinal (QC- RLSC; do

inglês: *locally weighted scatter-plot smoothing*). Nessa abordagem, o desvio do sinal de cada *feature* moleculares (metabólito) nas amostras do estudo (DW e GC) é corrigido pela mudança no sinal do mesmo *feature* molecular, presente nas amostras de QCs. A correção é de variabilidade temporal do sinal de um mesmo *feature* molecular das amostras em estudos e QCs.

3) Repetibilidade das análises dos QCs: as amostras de QCs devem apresentar um desvio padrão relativo (RSD) aceitável em todas as réplicas de análise durante a execução experimental, verificado através do coeficiente de variação (CV). Os *features* moleculares que mostraram $CV \geq 30\%$ foram excluídos das amostras QCs.

4) Linearidade: Uma série de amostras de QCs com sucessivas diluições podem ser preparadas e posteriormente plotadas com relação às intensidades correspondentes de cada *feature* molecular, gerando valores de coeficiente de regressão (R^2). *Features* moleculares com $R^2 \leq 0$ devem ser removidos, já valores de $0,5 \leq R^2 \leq 0,7$ são recomendados. Nesse trabalho, o estudo da linearidade não foi empregado.

De forma geral, a Tabela 4 ilustra o número de *features* moleculares detectados e extraídos durante a etapa de geração das matrizes de dados e após as etapas de pré-processamento dos dados. Para as amostras de plasma, no modo RPLC ESI(+) e RPLC ESI(-) foram obtidos 288 e 173 *features* moleculares nas matrizes de dados, respectivamente. Após as etapas de pré-processamento, os *features* moleculares que atenderam os parâmetros de inclusão pré-estabelecidos foram de 134 para o modo RPLC ESI(+) e de 109 para o modo RPLC ESI(-). Já no modo HILIC de eluição para as amostras de plasma, foram detectados e extraídos 348 e 602 *features* moleculares para HILIC ESI(+) e HILIC ESI(-), respectivamente. Destes, 146 *features* moleculares para HILIC ESI(+) e 343 *features* moleculares no modo HILIC ESI(-) atingiram os limites de inclusão pré-estabelecidos. Para as amostras de urina, no modo RPLC ESI(+) e RPLC ESI(-) foram obtidos 292 e 202 *features* moleculares na matriz de dados, respectivamente. Após as etapas de pré-processamento, os *features* moleculares que atenderam os parâmetros de inclusão nos dados foram de 113 para o modo RPLC ESI(+) e de 102 para o modo RPLC ESI(-). Já no modo HILIC de eluição para as amostras de urina, foram detectados e extraídos 209 e 436 *features* moleculares para HILIC ESI(+) e HILIC ESI(-), respectivamente. Destes, 112 *features* moleculares para HILIC ESI(+) e 173 *features*

moleculares no modo HILIC ESI(-) atingiram os limites de inclusão estabelecidos nas etapas de pré-processamento dos dados.

Tabela 4- Descrição do número de *features* moleculares em cada modo de eluição e de ionização empregado nas análises por LC-HRMS, detectados e extraídos e após o pré-processamento dos dados.

Matriz Biológica	Modo de Eluição	Features Moleculares detectados e extraídos (n)	Features Moleculares após pré-processamento (n)
Plasma	RPLC ESI(+)	288	134
	RPLC ESI(-)	173	109
	HILIC ESI(+)	348	146
	HILIC ESI(-)	602	343
Urina	RPLC ESI(+)	292	113
	RPLC ESI(-)	202	102
	HILIC ESI(+)	209	112
	HILIC ESI(-)	436	173

4.2.4. Preenchimento de Lacunas (*missing values*)

A razão da presença de lacunas de dados pode ser decorrente de: 1) picos malformados devido a presença de poucos *scans* (<< 10) ou em decorrência da instabilidade do íon, fazendo com que o sinal seja confundido com ruído instrumental; 2) picos de baixa intensidade que estão abaixo do limite de detecção do detector e 3) valores negativos provenientes de artefatos espectrais ou ruído, os quais são automaticamente convertidos em 'zero'. Assim, a presença de lacunas implica na perda de dados durante o processo de detecção/extração dos picos.

Para contornar a presença de lacunas de dados foi empregada a “regra dos 80%”.¹²⁵ Ou seja, um *feature* molecular deve ser detectado em pelo menos 80% de todas as amostras para ele ser mantido. Em termos práticos, se tivermos um estudo com 100 amostras entre os grupos estudo, controle e QCs e um *feature* molecular é detectado em 80 amostras do total, ele é mantido para prosseguir para as análises posteriores. Se esse mesmo *feature* molecular somente foi detectado em 70 amostras, ele deverá ser excluído.

Nesse trabalho, a “regra dos 80” foi empregada, com o emprego posterior de métodos de imputação de dados, como o método de K-vizinho mais próximo (KNN - sample-wise). O KNN foi selecionado baseando-se na sua praticidade e facilidade de uso e pela robustez apresentada em estudos metabolômicos.¹²⁶ Na imputação de K-vizinho mais próximo (KNN) os valores ausentes foram substituídos pela média dos valores do *feature* molecular específico (não ausente) mais próximos ($k = 10$) em termos de distância euclidiana das respostas. Portanto, um valor único é imputado para cada valor ausente.¹²⁷

4.2.5. Normalização dos Dados

A normalização de dados é uma etapa importante na etapa de pré-processamento em análises metabolômicas.^{124,128,129} A normalização visa garantir a remoção de erros sistemáticos, variabilidade inter- e intra-amostras e variabilidades instrumentais. Em estudos de metabolômica clínica, a urina é um biofluido muito interessante pela conveniência na coleta não invasiva e em maiores quantidades. Por exemplo, amostras de urina podem ser coletadas de forma repetida ao longo de um período, oferecendo um perfil metabolômico mais abrangente do que as amostras de plasma, que são coletadas em um determinado momento. Por outro lado, a urina apresenta dificuldades adicionais relacionadas às diferentes concentrações dos metabólitos devido às variações do nível de hidratação do indivíduo e essas variações devem ser corrigidas para que somente a variabilidade biológica seja visualizada.

Para a normalização de amostras de urina, o emprego da creatinina é um parâmetro clínico comumente empregado, juntamente com osmolalidade, massa seca de urina, volume de urina e gravidade específica. Todos esses métodos apresentam vantagens e desvantagens. No caso da creatinina, esse biomarcador relaciona-se com a filtração glomerular em condições fisiológicas normais.^{130,131} No entanto, em estudos metabolômicos envolvendo a urina, geralmente têm-se como alvos patologias ou condições fisiológicas anormais e, portanto, esse biomarcador pode ser alterado. No caso da DW, há evidência do comprometimento dos rins devido ao acúmulo de Cu no organismo,^{33,34} portanto o emprego de creatinina como parâmetro de normalização das amostras não foi considerado nesse estudo.

Em artigo de revisão de Zhang e Watson¹³², estão descritos estudos metabolômicos com urina nos quais se indica o uso majoritário de normalização por

sinal útil total MS (MSTUS; do inglês: *MS total useful signal*), creatinina, osmolalidade e aproximadamente 50% dos estudos não reportam como a urina foi normalizada.

Dieterle *et al*¹³³ propuseram um método de normalização por quociente probabilístico (PQN; do inglês: *probabilistic quotient normalization*) para ser empregado na remoção da variabilidade por diluição de amostras biológicas complexas. Este método dimensiona os dados com base na diluição mais proeminente. Cada pico presente em uma amostra é dividido pelo correspondente pico na amostra referência gerando um quociente. O quociente com maior frequência nas amostras é selecionado para conduzir a normalização.¹¹⁰ Os autores aplicaram o método proposto na análise metabolômica por ressonância magnética nuclear de hidrogênio (RMN ¹H), na qual foram avaliados mais de 4.000 espectros de amostras, sendo demonstrado que a normalização PQN foi mais exata e mais robusta do que outros métodos de normalização amplamente difundidos, como exemplo, a normalização pela integral ou normalização pela soma constante.

Estudos mais recentes exploraram métodos de normalização compostos por suavização de gráfico de dispersão estimada localmente (*LOESS*, do inglês: *locally estimated scatterplot smoothing*) associado a MSTUS ou *LOESS* associado a PQN.¹¹⁹ Uma comparação entre creatinina, densidade e PQN foi também conduzida pelo grupo de Johnson *et al.*,¹³⁴ que concluíram que gravidade e PQN foram os dois métodos que melhor ajustaram os dados.

Nesse estudo, para as amostras de plasma e urina foi empregada a normalização das intensidades do sinal-MS ao longo da aquisição de dados com a técnica de *LOWESS/Spline*, classificada como normalização baseada na amostra, conforme previamente descrito para as amostras QCs (item 4.2.3, pag. 58). Adicionalmente, para as amostras de urina, foi ainda empregada a normalização MSTUS para a correção da diluição entre amostras. A normalização PQN também foi avaliada, no entanto, não mostrou resultados satisfatórios para as amostras de urina analisadas pelo modo de eluição RPLC. Já a normalização MSTUS, indicou resultados satisfatórios para as amostras de plasma e de urina analisadas tanto no modo RPLC quanto no modo HILIC. A escolha da melhor estratégia de normalização foi confirmada por meio de análises multivariadas, que permitiram comparar de forma robusta o desempenho das abordagens PQN e MSTUS.

4.2.6. Centralização, Escalonamento e Transformação dos Dados

A variabilidade nos dados provenientes de uma análise metabolômica é caracterizada como a soma da variação induzida (biológico - proveniente da patologia) e da variação total não induzida.¹³⁵ A variação total não induzida é toda a variação originada de fatores biológicos não induzidos, como a falta de repetibilidade na extração das amostras biológicas, erros durante a coleta e amostragem (decomposição, temperatura, erros volumétricos, reações enzimáticas), diferenças na ordem de magnitude e de capacidade de ionização das moléculas, instabilidade química durante análise que promovem flutuações nos níveis detectados. A variação total não induzida promove resultados heteroscedásticos.¹³⁵ Com isso o emprego de métodos de centralização e de escalonamento é uma prática comum em metabolômica.

A centralização converte todas as flutuações nas concentrações em torno do zero, ao invés da média da concentração do metabólito. Assim, a centralização compensa as diferenças entre metabólitos muito abundantes e metabólitos pouco abundantes, deixando apenas a variação entre as amostras como relevante. A etapa de centralização de dados é empregada em combinação com todos os métodos de escalonamento.^{128,135}

Os métodos de escalonamento são abordagens baseadas no dimensionamento de metabólitos e esses dividem cada variável por um fator de escala diferente para cada variável. O escalonamento objetiva ajustar às diferenças nas escalas ou dimensionamento dos dados (*fold-change*) entre os diferentes metabólitos convertendo os dados em diferenças apresentadas como concentração relativa. Desta forma, com o escalonamento, há uma padronização da escala dos dados, fazendo com que todas as variáveis contribuam de maneira igualitária, ou seja, na mesma escala.^{110,135}

Há diferentes estratégias estatísticas que podem ser usadas no escalonamento de dados metabolômicos global, incluindo: autoescalonamento (do inglês: *auto scaling*), escalonamento por pareto (do inglês: *pareto scaling*), escalonamento por estabilidade de variável (do inglês: *vast scaling*), dimensionamento de intervalo (do inglês: *range scaling*) e dimensionamento de nível (do inglês: *level scaling*).^{123,136} Nesse trabalho, as amostras de plasma e urina dos grupos DW e GC foram escalonadas com o emprego de autoescalonamento. O método de

autoescalamento, também conhecido como escalamento de variação unitária, baseia-se no emprego do desvio padrão como fator de escala, proporcionando pesos iguais a todos os sinais da amostra. No método de autoescalamento, o desvio padrão de cada *feature molecular* é igual a 1.¹²⁸

A transformação de dados também é uma etapa importante em metabolômica global. O objetivo da transformação de dados é corrigir a heterocedasticidade, fazer das distribuições distorcidas mais simétricas e converter relações multiplicativas em relações aditivas.¹²⁹ Embora a transformação do tipo logarítmica (*log*) é a mais comum, aqui empregou-se a transformação por Box-Cox.¹³⁷

4.3. Análise Estatísticas

Os dados obtidos para as análises dos grupos DW e GC, conforme Fluxogramas 1 e 2, foram inicialmente processadas por análise multivariada não supervisionada com emprego da análise de componentes principais (PCA), incluindo os grupos DW e GC e as amostras dos QCs para análise da qualidade dos dados, observando a detecção de tendência e/ou padrões de agrupamentos. A análise por PCA foi também realizada com a ausência dos QCs com o intuito de observar a dimensionalidade intrínseca do conjunto de dados.

Em seguida, com o objetivo de obter uma correlação entre as variáveis químicas e um conjunto de dados (modelo), usado para a previsão e classificação, empregou-se análises supervisionadas através da análise discriminante por mínimos quadrados parciais (PLS-DA) ou da análise discriminante por mínimos quadrados parciais ortogonais (OPLS-DA) entre os grupos DW e GC. Em ambas as análises supervisionadas empregadas, procedeu-se também com a validação do modelo através de testes de permutação com a validação cruzada (n= 2000).

A validação cruzada é uma abordagem que tem como finalidade lidar com a variabilidade nos dados, proporcionando uma análise da exatidão do modelo preditivo em classificar os dados, permitindo verificar se as diferenças entre a classificação das amostras no modelo são corretas. Neste teste, a classificação das amostras é intencionalmente forçada ao erro, sendo que, uma das amostras é selecionada para compor o conjunto de validação e as demais amostras remanescentes são usadas para o conjunto de treinamento.¹²⁸

Os autores aplicaram o método proposto na análise metabolômica por ressonância magnética nuclear de hidrogênio (RMN ¹H), na qual foram avaliados mais

de 4.000 espectros de amostras, sendo demonstrado que a normalização PQN foi mais exata valores de $R^2 \geq 0,7$, $Q^2 \geq 0,5$ e p -permutação $< 0,05$.

Após a validação cruzada da OPLS-DA, identificou-se os valores de $VIP \geq 1$, gerando os *features* moleculares caracterizados como as mais importantes variáveis para a discriminação dos grupos DW e GC, dentro do modelo projetado no OPLS-DA.

Para cada categoria de análise por LC-HRMS (Fluxograma 1), os *features* moleculares provenientes com valores $VIPs \geq 1$, foram submetidos a análises univariadas.^{112,117} Com isso, inicialmente realizou-se testes de normalidade dos dados empregando-se testes de Shapiro-Wilk (nível de confiança, $p < 0,05$). O teste de normalidade permite selecionar entre a aplicação de métodos estatísticos paramétricos para um conjunto de dados que segue uma distribuição normal, como o teste de *t-Student* ou ANOVA, ou métodos estatísticos não paramétricos para conjunto de dados que segue uma distribuição não-normal., como o Mann-Whitney U ou Kruskal-Wallis.

Nos testes de Shapiro-Wilk, se $p < 0,05$, a hipótese nula (H_0) é que os dados seguem uma distribuição normal. Já se $p > 0,05$, a H_0 pode ser rejeitada e considera-se que os dados possuem uma distribuição não normal.

Depois do teste de Shapiro-Wilk, foram realizados teste de *t-Student*, com nível de confiança $p < 0,05$, para amostras independentes, para avaliar o impacto da idade no grupo de dados, juntamente com um teste Qui-quadrado de Pearson (Q-Q) para avaliar a dependência entre variáveis categóricas (gênero). Esses dois testes foram empregados para comparar as características físicas, como idade e gênero, entre os indivíduos dos grupos em estudo e verificar a existência de covariáveis.

Nos testes de *t-Student* a hipótese nula (H_0) é que as variáveis são dependentes. Assim, valores $p < 0,05$ evidenciam que a H_0 não pode ser rejeitada e há evidências de que as variáveis são dependentes e causam influências nas variáveis em estudo. Já se $p > 0,05$, a H_0 pode ser rejeitada e considera-se que as variáveis são independentes e que não devem ser consideradas como covariáveis.

Nos testes de Qui-Quadrado a hipótese nula (H_0) é que as variáveis são dependentes. Assim, valores $p < 0,05$ evidencia que a H_0 não pode ser rejeitada e há evidências de que as variáveis são dependentes e causam influências nas variáveis em estudo. Já se $p > 0,05$, a H_0 pode ser rejeitada e considera-se que as variáveis são independentes e que não devem ser consideradas como covariáveis.

Para as séries de *features* moleculares com distribuição não normal, essas foram submetidas a uma transformação Box-Cox para estabilizar a variabilidade e melhorar a aproximação da distribuição dos dados à normalidade.

Após a transformação do Box-Cox e com a confirmação das covariáveis impactantes, as análises estatísticas prosseguiram com a análise de variância (ANCOVA). A ANCOVA combina as técnicas de análise de variância (ANOVA) e regressão linear simples com a incorporação de uma variável quantitativa (covariável) junto com a variável nominal em um único modelo. O uso principal da ANCOVA é aumentar a precisão de experimentos aleatorizados.¹³⁸ Os *p*-valores do teste ANCOVA foram ajustados com uma Taxa de Falsa Descoberta (FDR) de 10% para minimizar erros do tipo I.

Ao final das análises estatísticas foram considerados para a etapa de identificação de potenciais candidatos a biomarcadores os *features* moleculares que apresentaram $VIP \geq 1,0$, $p \leq 0,05$ e com *p*-valores do teste ANCOVA menores do que os valores FDR de 10% (0,01).

4.4. Amostras Biológicas de Plasma

4.4.1. Plasma - Modo Reverso de Eluição

No modo RPLC de eluição, o cromatograma de LC-HRMS apresentado na Figura 11 é o resultado das análises das amostras de plasma do controle de qualidade (QC) no modo de ionização ESI(+) (Figura 11A) e no modo de ionização ESI(-) (Figura 11B).

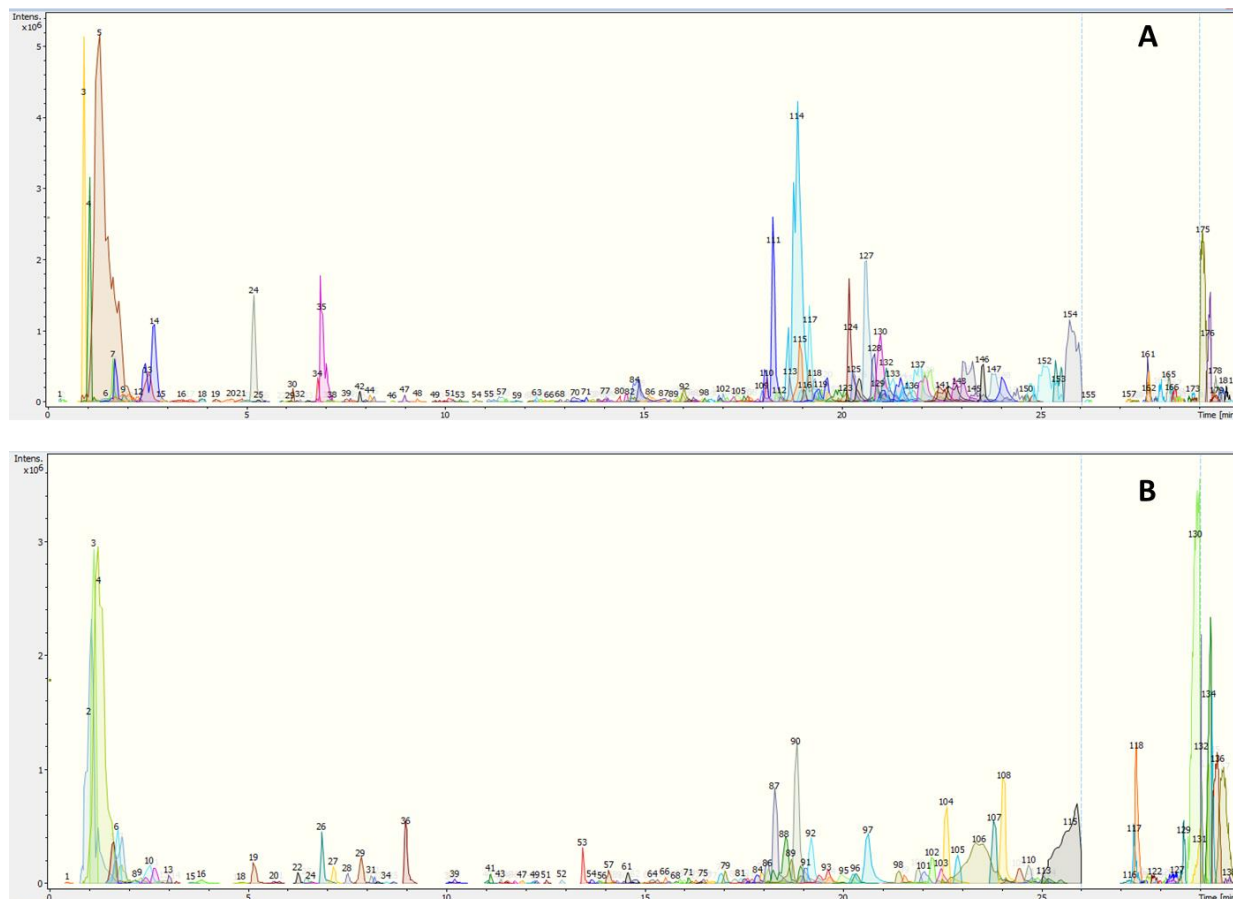


Figura 11- Cromatogramas de picos extraídos de LC-HRMS representativos das análises da amostra de controle de qualidade (QC) de plasma no **(A)** modo RPLC ESI(+) e **(B)** RPLC ESI(-).

A Figura 12 ilustra a análise multivariada não supervisionada com emprego da análise de componentes principais (PCA). É possível observar que os grupos GC e DW foram parcialmente discriminadas com 18% da variância total dos dados, sendo 10,3% em PC1 e 7,7% em PC2. Observa-se que as amostras QCs encontram-se centralizadas e agrupadas, indicando que não houve variabilidade instrumental durante a sequência de aquisição de dados por LC-HRMS. A análise por PCA foi também realizada com a ausência dos QC's com o intuito de observar a

dimensionalidade dos dados. Da mesma forma, não houve uma discriminação distinta entre DW e GC, com 10,7% em PC1 e 8,8% em PC2, totalizando 19,5% da variância total.

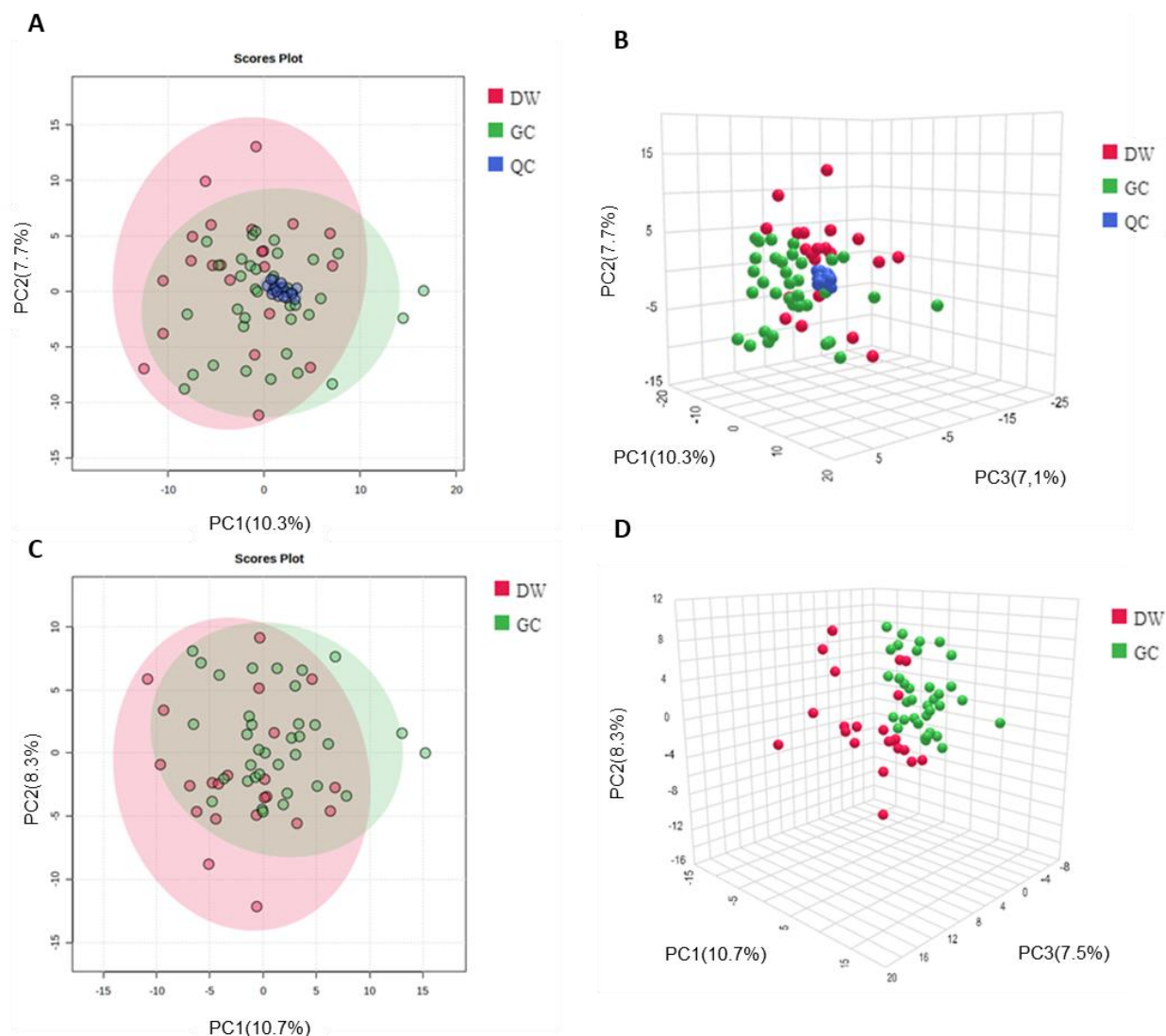


Figura 12- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de *scores plots* 2D e 3D baseadas nos modelos PCA, no modo RPLC ESI(+) combinado com RPLC ESI(-).

(A) *Score plots* 2D das amostras de plasma dos grupos DW, GC e controle de qualidade (QC); **(B)** *Score plots* 3D das amostras de plasma dos grupos DW, GC e controle de qualidade (QC); **(C)** *Score plots* das amostras de plasma dos grupos DW e GC; **(D)** *Score plots* das amostras de plasma dos grupos DW e GC. Vermelho: grupo estudo (DW); Verde: grupo controle (GC) e Azul: controles de qualidade (QC). *Software: MetaboAnalyst.*

Em seguida, com o objetivo de melhorar os resultados de previsão e classificação, gerou-se um gráfico de análise discriminante por mínimos quadrados parciais ortogonais (OPLS-DA) entre os grupos DW e GC (Figura 13). Realizou-se a validação do modelo OPLS-DA através do teste de permutação com a validação cruzada. O teste de permutação no modelo de OPLS-DA, com 2000 permutações, resultou em valores de $R^2=0,958$, $Q^2=0,785$ e p -permutação $<5e^{-4}$, sendo esses valores considerados aceitáveis para mostrar que há confiabilidade no modelo OPLS-DA obtido. A Figura 13 ilustra o gráfico de scores do modelo OPLS-DA, teste de permutação (validação cruzada) e importância da variável na projeção (VIP).

Após a validação cruzada da OPLS-DA, foi possível identificar os valores de $VIP \geq 1$, a partir dos quais obtiveram-se 84 *features* moleculares caracterizados como os mais importantes para a discriminação dos grupos DW e GC, dentro do modelo projetado no OPLS-DA (Figura 13C). Os 84 *features* moleculares provenientes dos $VIP \geq 1$, foram submetidos às análises univariadas. Iniciou-se com o teste de normalidade dos dados empregando-se testes de Shapiro-Wilk (nível de confiança, $p < 0,05$).

Para as amostras de plasma no modo RPLC ESI(+)/ESI(-), o teste de Shapiro-Wilk resultou em um valor de $p < 0,05$, indicando uma distribuição não normal dos dados. Na sequência, foi realizado o teste de *t-Student*, $p < 0,05$, para avaliar o impacto da idade no grupo de dados, juntamente com um teste Qui-quadrado de Pearson (Q-Q) para avaliar a dependência da variável gênero dos indivíduos. Esses dois testes foram empregados para comparar as características físicas, como idade e gênero, entre os indivíduos dos grupos em estudo e verificar a existência de covariáveis (idade e/ou gênero).

Após o teste de *t-Student* e o teste Qui-quadrado de Pearson (Q-Q), a série de 84 *features* moleculares foi submetida a uma transformação Box-Cox. Para as amostras de plasma no modo RPLC ESI(+)/ESI(-). O teste de *t-Student* resultou em um valor de $p = 0,021$ (Anexo 1, Tabela A1), indicando a idade como significativa e, portanto, uma covariável. Para as amostras de plasma no modo RPLC ESI(+)/ESI(-), o teste Q-Q indicou um valor de $p = 0,233$ (Anexo 1, Tabela A2), indicando que o gênero dos indivíduos selecionados não é uma covariável.

Com a confirmação da idade dos indivíduos como uma covariável, as análises estatísticas prosseguiram com a análise de covariância (ANCOVA). Durante

esse processo, os p -valores obtidos foram examinados quanto à significância estatística e posterior classificação de potenciais biomarcadores.

Da análise univariada, foram selecionados 81 *features moleculares* estatisticamente significativos com $VIP \geq 1,0$, $p < 0,05$ e com p -valores do teste ANCOVA menores do que os valores FDR de 10% (Anexo 1, Tabela A3), os quais seguiram para a etapa de anotação putativa.

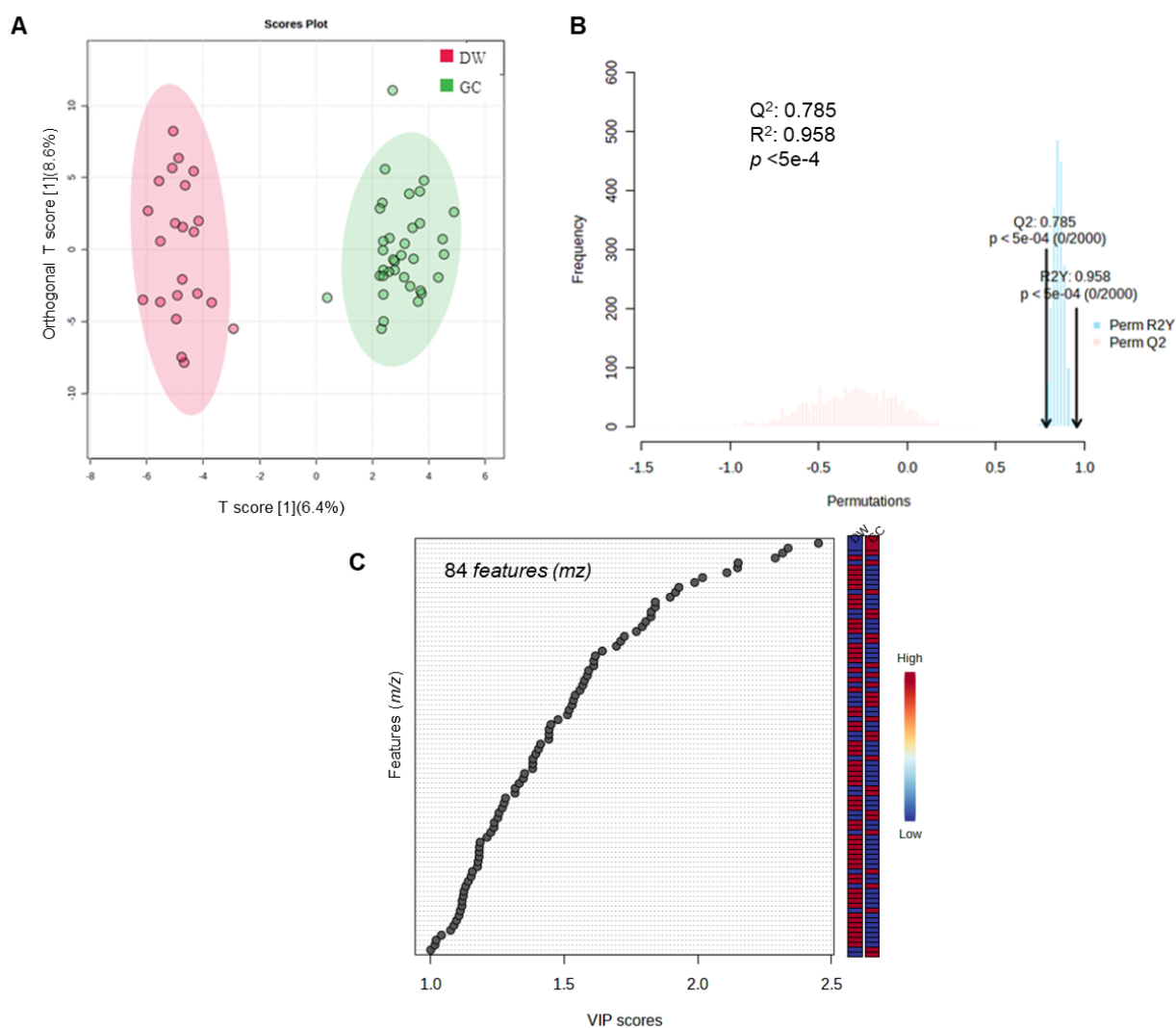


Figura 13- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de *score plots* 2D do modelo OPLS-DA no modo RPLC ESI(+) combinado com RPLC ESI(-), teste de permutação do modelo OPLS-DA e importância da variável na projeção (VIP).

(A) *Score plots* 2D das amostras de plasma dos grupos DW e GC; **(B)** Teste de permutação do modelo OPLS-DA; **(C)** OPLS-DA de acordo com VIP scores no eixo x das amostras de plasma ($VIP \geq 1$). As caixas coloridas indicam se a concentração do metabólito está aumentada (vermelho) ou diminuída (azul). *Software: MetaboAnalyst.*

4.4.2. Plasma - Modo HILIC de Eluição

No modo HILIC de eluição, o cromatograma de LC-HRMS representado na Figura 14 é o resultado das análises das amostras de plasma do controle de qualidade (QC) no modo de ionização ESI(+) (Figura 14A) e no modo de ionização ESI(-) (Figura 14B).

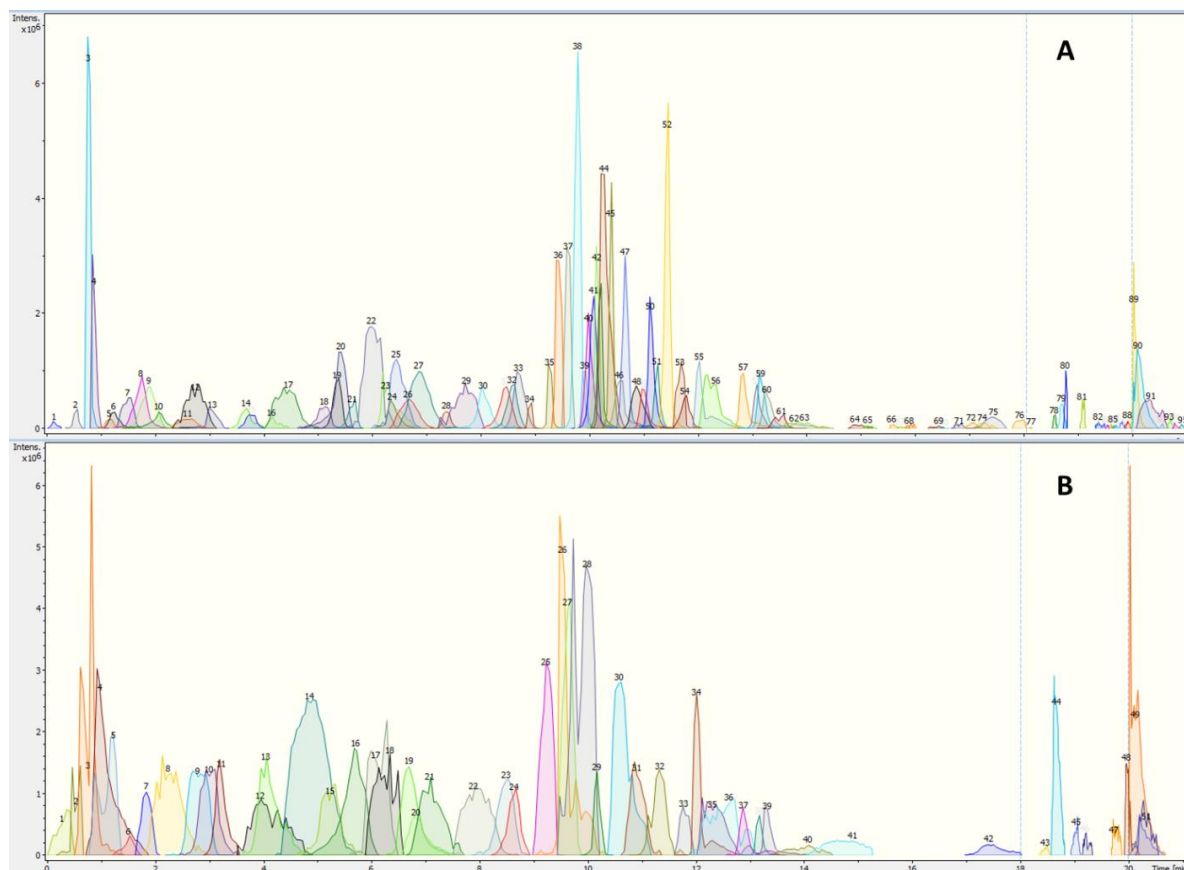


Figura 14- Cromatogramas de picos extraídos de LC-HRMS representativos das análises da amostra de controle de qualidade (QC) de plasma no **(A)** modo HILIC ESI(+) e **(B)** HILIC ESI(-).

A Figura 15 ilustra a análise multivariada não supervisionada com emprego da análise de componentes principais (PCA). É possível observar que os grupos GC e DW foram parcialmente discriminados com 20,6% da variância total dos dados, sendo 14,1% em PC1 e 6,5% em PC2. Observa-se que as amostras QCs encontram-se centralizadas e agrupadas, indicando que não houve variabilidade instrumental durante a sequência de aquisição de dados por LC-HRMS. A análise por PCA foi também realizada com a ausência dos QC's com o intuito de observar a dimensionalidade dos dados. Da mesma forma, não houve uma discriminação distinta

entre DW e GC, com 15,0% em PC1 e 6,9% em PC2, totalizando 21,9% da variância total.

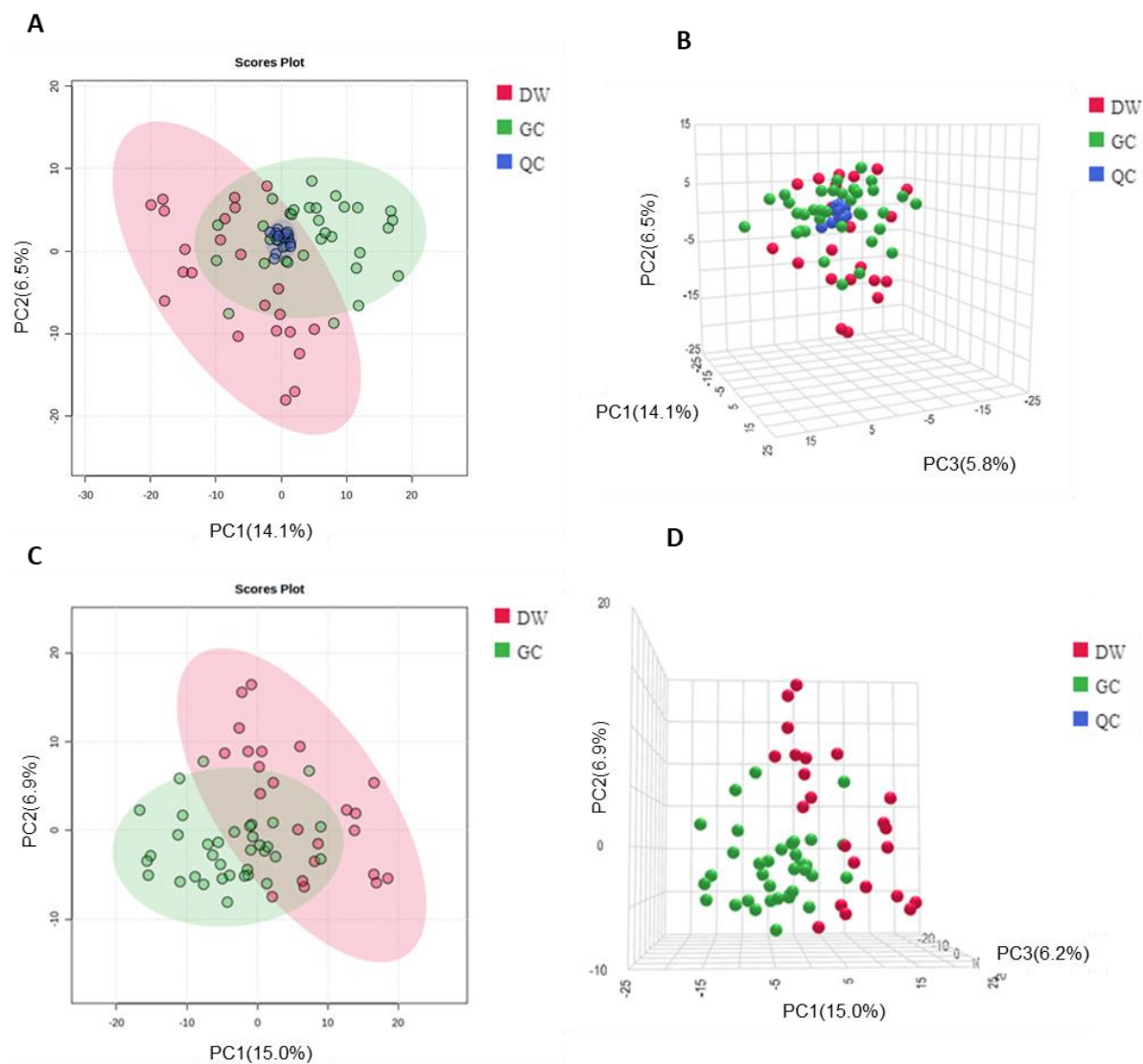


Figura 15- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de *scores plots* 2D e 3D baseadas nos modelos PCA, no modo HILIC ESI(+) combinado com HILIC ESI(-).

(A) *Score plots* 2D das amostras de plasma dos grupos DW, GC e controle de qualidade (QC); **(B)** *Score plots* 3D das amostras de plasma dos grupos DW, GC e controle de qualidade (QC); **(C)** *Score plots* das amostras de plasma dos grupos DW e GC; **(D)** *Score plots* das amostras de plasma dos grupos DW e GC. Vermelho: grupo estudo (DW); Verde: grupo controle (GC) e Azul: controles de qualidade (QC). *Software: MetaboAnalyst.*

Em seguida, a Figura 16 ilustra os resultados de previsão e classificação, sem as amostras QCs, com modelo OPLS-DA entre os grupos DW e GC. A validação

do modelo foi conduzida pelo teste de permutação com a validação cruzada com 2000 permutações, resultando em valores de $R^2=0,977$, $Q^2=0,845$ e p -permutação $<5e^{-4}$, sendo esses valores considerados aceitáveis para mostrar que há confiabilidade no modelo OPLS-DA obtido. A Figura 16 ilustra o gráfico de *scores* do modelo OPLS-DA, o teste de permutação (validação cruzada) e a importância da variável na projeção (VIP).

Após a validação cruzada da OPLS-DA, foi possível identificar os valores de $VIP \geq 1$, a partir dos quais obtiveram-se 165 *features* moleculares foram caracterizados como os mais importantes para a discriminação dos grupos DW e GC, dentro do modelo projetado no OPLS-DA (Figura 16C). Os 165 *features* moleculares provenientes dos $VIP \geq 1$, foram submetidos às análises univariadas, da mesma forma como conduzido para as análises RPLC ESI(+)/ESI(-).

Os resultados do teste *t-Student* e o teste Qui-quadrado de Pearson (Q-Q) para a série de amostras de plasma no modo HILIC ESI(+)/ESI(-), indicaram um valor de $p = 0,021$ e $p = 0,233$, respectivamente (Anexos Tabela A1 e Tabela A2), indicando a idade como significativa e, portanto, uma covariável no conjunto de dados.

A análise de covariância (ANCOVA) forneceu para os dados HILIC ESI(+)/ESI(-), 162 *features* moleculares estatisticamente significativos com $VIP \geq 1,0$, $p < 0,05$ e com p -valores do teste ANCOVA menores do que os valores FDR de 10% (Anexo 1, Tabela A4) e que foram empregados na anotação putativa.

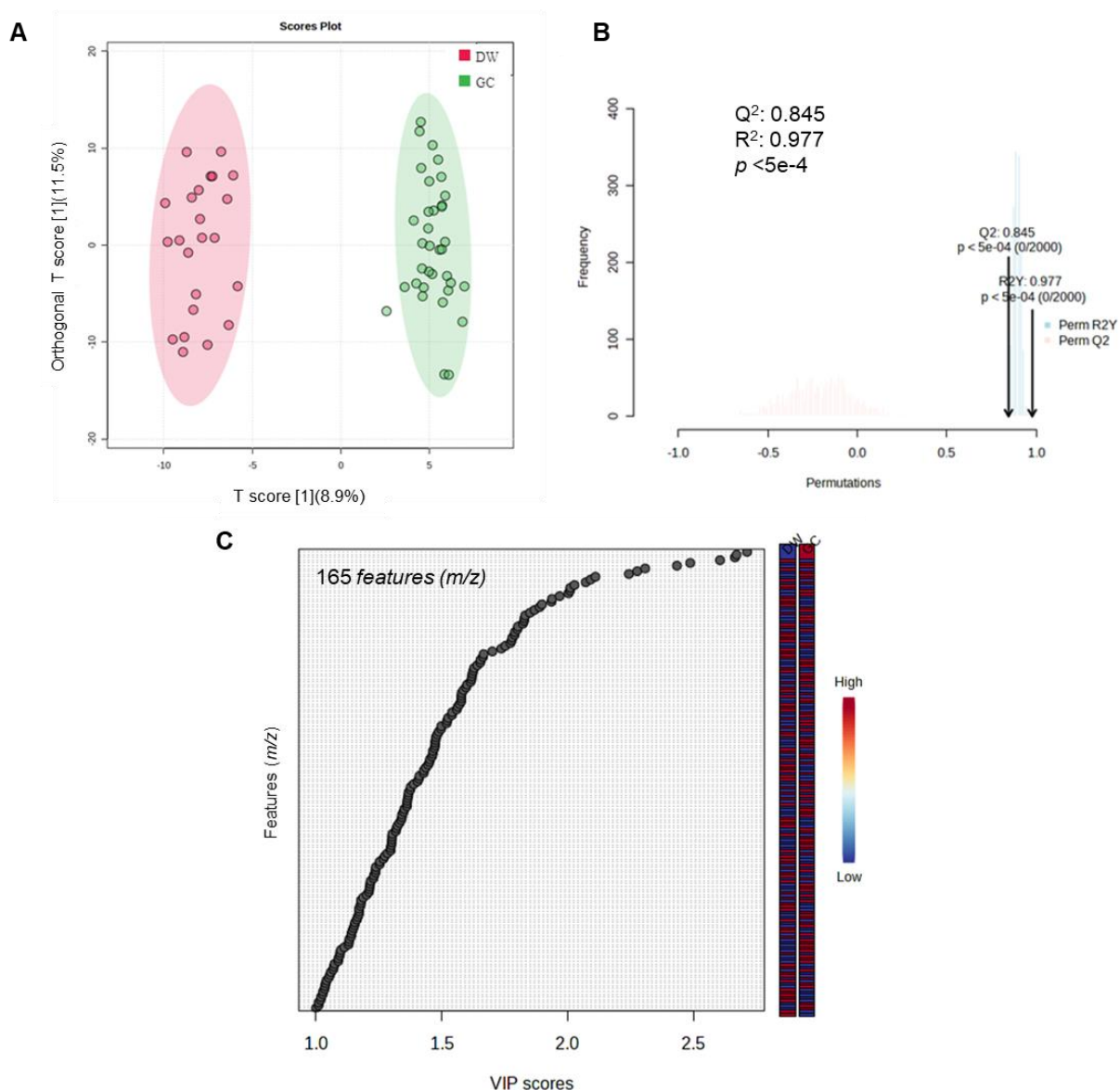


Figura 16- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de score plots 2D do modelo OPLS-DA no modo HILIC ESI(+) combinado com HILIC ESI(-), teste de permutação do modelo OPLS-DA e importância da variável na projeção (VIP).

(A) Score plots 2D das amostras de plasma dos grupos DW e GC; **(B)** Teste de permutação do modelo OPLS-DA; **(C)** OPLS-DA de acordo com VIP scores no eixo x das amostras de plasma (VIP ≥ 1). As caixas coloridas indicam se a concentração do metabólito está aumentada (vermelho) ou diminuída (azul). *Software: MetaboAnalyst.*

4.5. Amostras Biológicas de Urina

4.5.1. Urina - Modo Reverso de Eluição

No modo RPLC de eluição, o cromatograma de LC-HRMS apresentado na Figura 17 é o resultado das análises das amostras de urina do controle de qualidade (QC) no modo de ionização ESI(+) (Figura 17A) e no modo de ionização ESI(-) (Figura 17B).

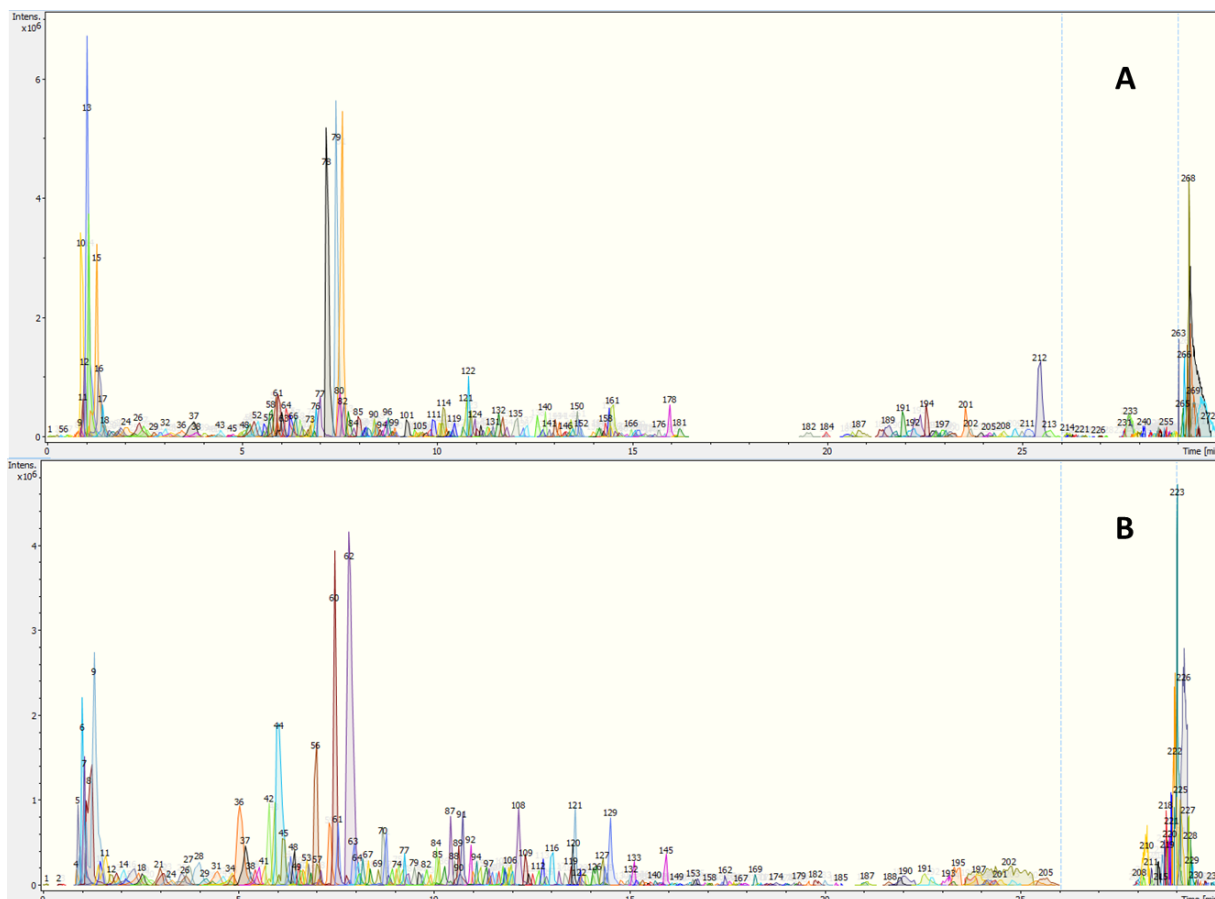


Figura 17- Cromatogramas de picos extraídos de LC-HRMS ilustrativos das análises da amostra de controle de qualidade (QC) de plasma no **(A)** modo RPLC ESI(+) e **(B)** RPLC ESI(-).

A Figura 18 ilustra a análise multivariada não supervisionada com emprego da análise de componentes principais (PCA). É possível observar que os grupos GC e DW foram parcialmente discriminadas com 24,2% da variância total dos dados, sendo 17,7% em PC1 e 6,5% em PC2. Observa-se que as amostras QCs encontram-se centralizadas e agrupadas, indicando que não houve variabilidade instrumental durante a sequência de aquisição de dados por LC-HRMS. A análise por

PCA foi também realizada com a ausência dos QC's com o intuito de observar a dimensionalidade dos dados. Da mesma forma, não houve uma discriminação distinta entre DW e GC, com 19,1% em PC1 e 6,7% em PC2, totalizando 25,8% da variância total.

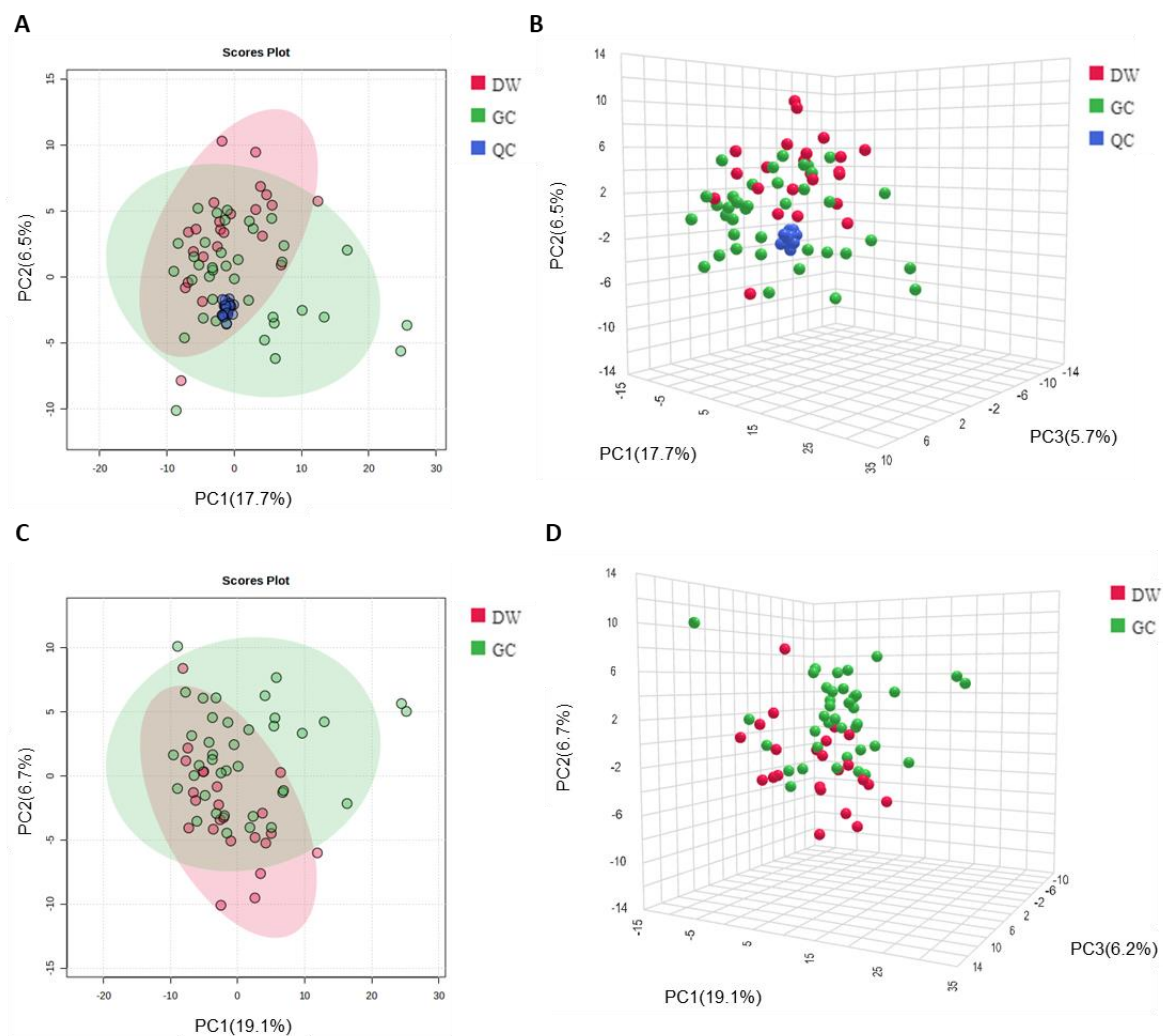


Figura 18-Análises multivariadas de metabolômica global por LC-HRMS com gráficos de *scores plots* 2D e 3D baseadas nos modelos PCA, no modo RPLC ESI(+) combinado com RPLC ESI(-).

(A) *Score plots* 2D das amostras de urina dos grupos DW, GC e controle de qualidade (QC); **(B)** *Score plots* 3D das amostras de urina dos grupos DW, GC e controle de qualidade (QC); **(C)** *Score plots* das amostras de urina dos grupos DW e GC; **(D)** *Score plots* das amostras de urina dos grupos DW e GC. Vermelho: grupo estudo (DW); Verde: grupo controle (GC) e Azul: controles de qualidade (QC).

Software: MetaboAnalyst.

A Figura 19 ilustra os resultados de previsão e classificação por OPLS-DA entre os grupos DW e GC e os resultados da validação do modelo OPLS-DA através do teste de permutação com a validação cruzada. O teste de permutação no modelo de OPLS-DA, com 2000 permutações, resultou em valores de $R^2=0,936$, $Q^2=0,717$ e p -permutação $<5e^{-4}$, sendo esses valores considerados aceitáveis para mostrar que há confiabilidade no modelo OPLS-DA obtido. A Figura 19 ilustra o gráfico de scores do modelo OPLS-DA, teste de permutação (validação cruzada) e importância da variável na projeção (VIP).

Após a validação cruzada da OPLS-DA, foi possível identificar os valores de $VIP \geq 1$, a partir dos quais obtiveram-se 64 *features* moleculares foram caracterizados como os mais importantes para a discriminação dos grupos DW e GC, dentro do modelo projetado no OPLS-DA (Figura 19C). Os 64 *features* moleculares provenientes dos $VIP \geq 1$, foram submetidos às análises univariadas. Da mesma forma como descritos para as análises estatísticas das amostras de plasma, foi conduzido o teste de normalidade para os dados da urina no modo RPLC ESI(+)/ESI(-) (teste de Shapiro-Wilk, nível de confiança, $p < 0,05$), transformação dos dados por Box-Cox, teste de *t-Student* ($p < 0,05$) e teste Qui-quadrado de Pearson (Q-Q) para avaliar a dependência de covariáveis.

Para as amostras de urina, o teste de *t-Student* resultou em um valor de $p = 0,011$ (Anexo 2, Tabela A5), indicando a idade como significativa e, portanto, uma covariável e o teste Q-Q indicou um valor de $p = 0,261$ (Anexo 2, Tabela A6), indicando que a idade dos indivíduos não é uma covariável. Na sequência, empregou-se a análise de covariância (ANCOVA), que resultou em 34 *features* moleculares estatisticamente significativos com $VIP \geq 1,0$, $p < 0,05$ e com p -valores do teste ANCOVA menores do que os valores FDR de 10% (Anexo 2, Tabela A7), os quais seguiram para a etapa de anotação putativa.

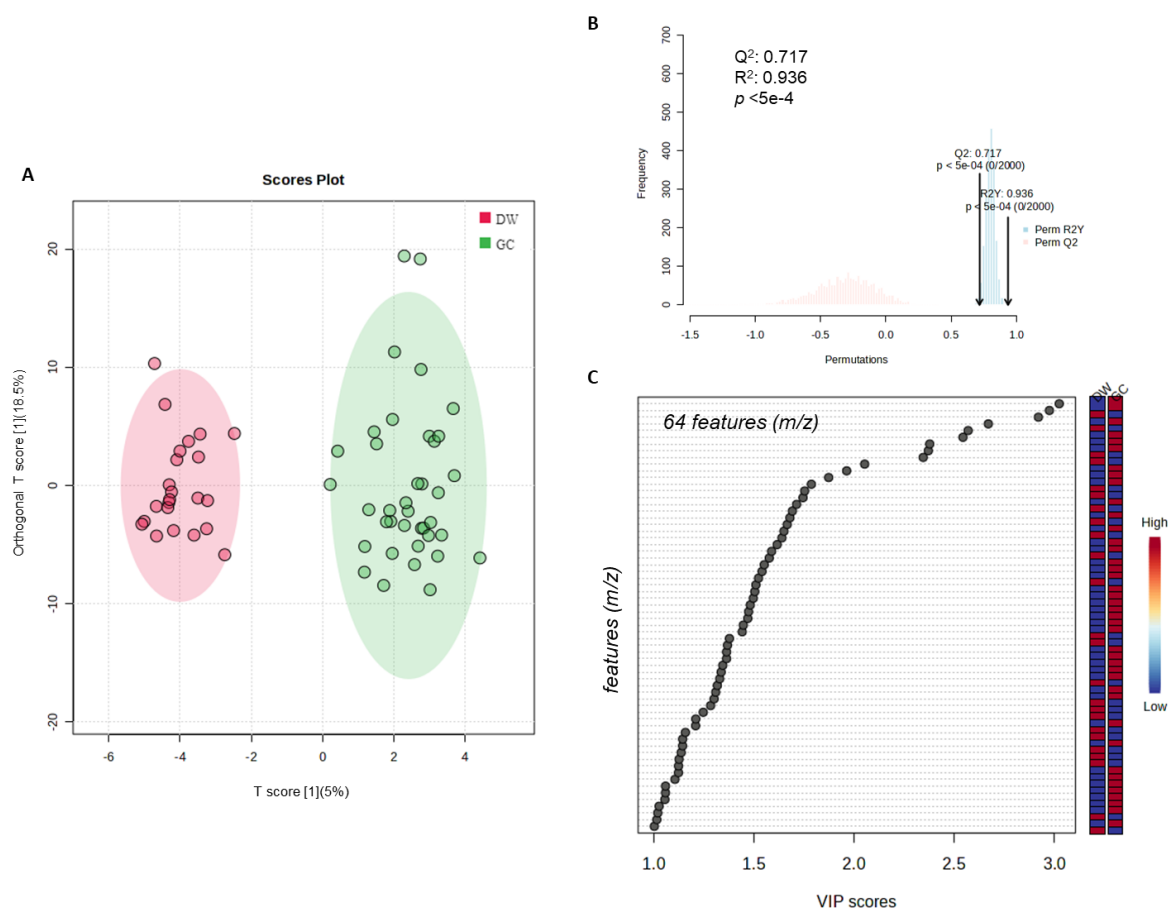


Figura 19- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de *score plots* 2D do modelo OPLS-DA no modo RPLC ESI(+) combinado com RPLC ESI(-), teste de permutação do modelo OPLS-DA e importância da variável na projeção (VIP).

(A) *Score plots* 2D das amostras de urina dos grupos DW e GC; **(B)** Teste de permutação do modelo OPLS-DA; **(C)** OPLS-DA de acordo com VIP scores no eixo x das amostras de urina (VIP ≥ 1). As caixas coloridas indicam se a concentração do metabólito está aumentada (vermelho) ou diminuída (azul). *Software: MetaboAnalyst.*

4.5.2. Urina - Modo HILIC de Eluição

No modo HILIC de eluição, o cromatograma de LC-HRMS apresentado na Figura 20 é o resultado das análises das amostras de urina do controle de qualidade (QC) no modo de ionização ESI(+) (Figura 20A) e no modo de ionização ESI(-) (Figura 20B).

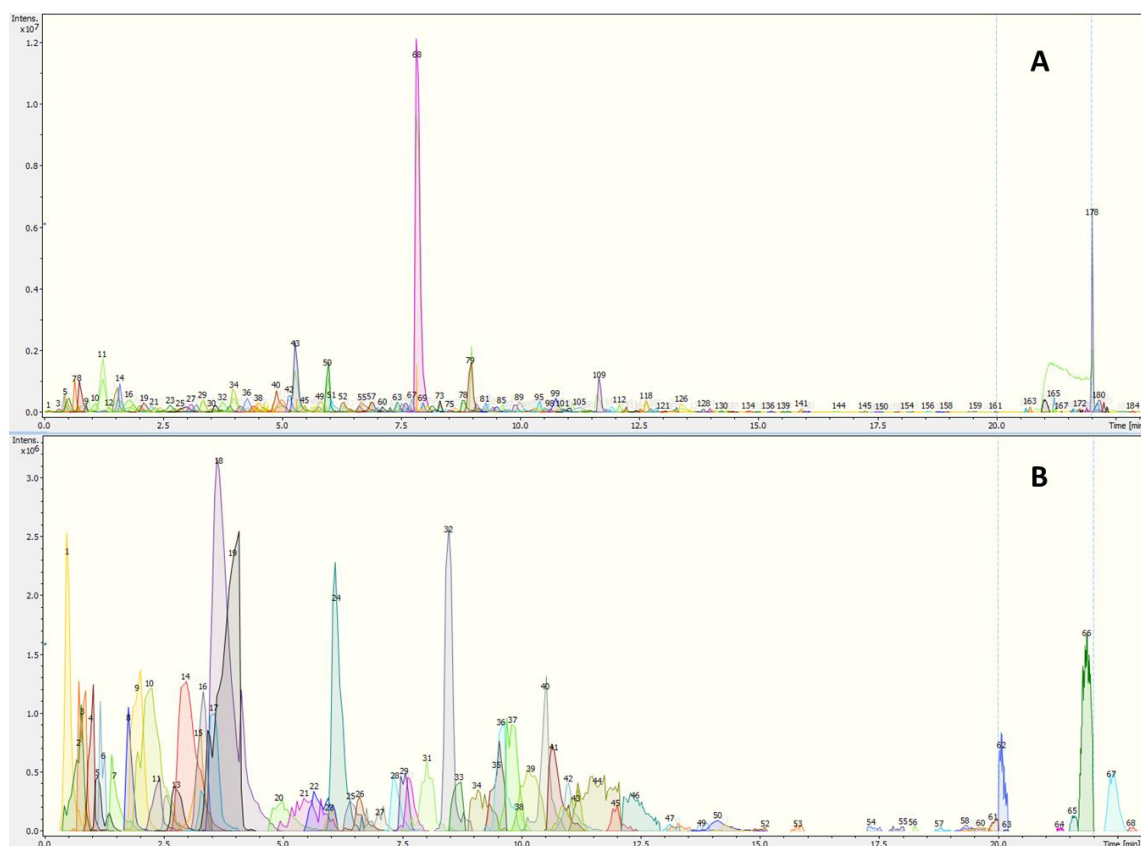


Figura 20- Cromatogramas de picos extraídos de LC-HRMS ilustrativos das análises da amostra de controle de qualidade (QC) de plasma no **(A)** modo HILIC ESI(+) e **(B)** HILIC ESI(-).

A Figura 21 ilustra a análise multivariada não supervisionada com emprego da análise de componentes principais (PCA). É possível observar que os grupos GC e DW foram parcialmente discriminadas com 31,5% da variância total dos dados, sendo 25,1% em PC1 e 6,4% em PC2. Observa-se que as amostras QCs encontram-se centralizadas e agrupadas, indicando que não houve variabilidade instrumental durante a sequência de aquisição de dados por LC-HRMS. A análise por PCA foi também realizada com a ausência dos QC's com o intuito de observar a dimensionalidade dos dados. Da mesma forma, não houve uma discriminação distinta entre DW e GC, com 25,3% em PC1 e 7,1% em PC2, totalizando 32,4% da variância total.

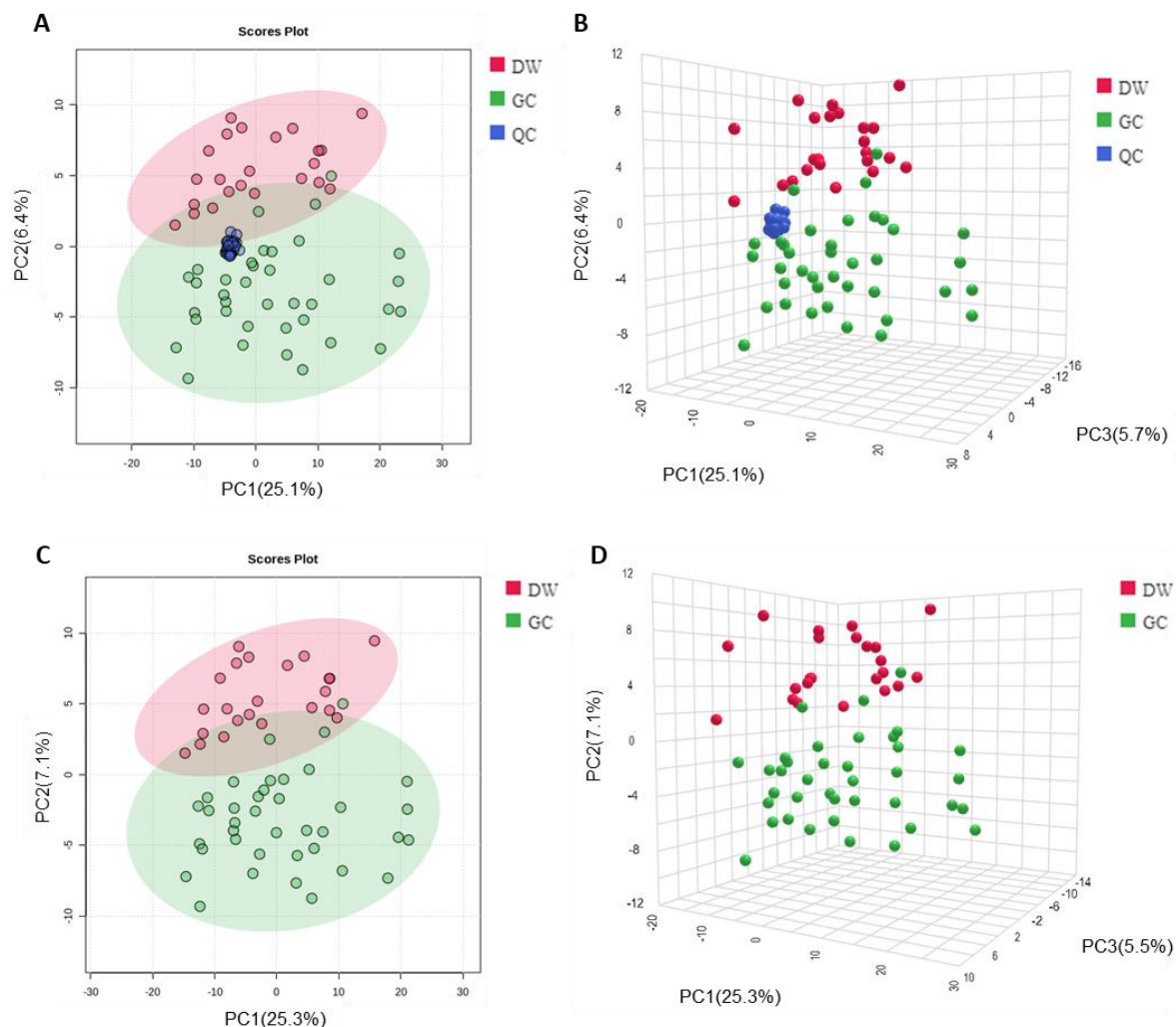


Figura 21- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de *scores plots* 2D e 3D baseadas nos modelos PCA, no modo HILIC ESI(+) combinado com HILIC ESI(-).

(A) *Score plots* 2D das amostras de urina dos grupos DW, GC e controle de qualidade (QC); **(B)** *Score plots* 3D das amostras de urina dos grupos DW, GC e controle de qualidade (QC); **(C)** *Score plots* das amostras de urina dos grupos DW e GC; **(D)** *Score plots* das amostras de urina dos grupos DW e GC. Vermelho: grupo estudo (DW); Verde: grupo controle (GC) e Azul: controles de qualidade (QC).

Software: MetaboAnalyst.

Em seguida, com o objetivo de melhorar os resultados de previsão e classificação, gerou-se um gráfico de análise discriminante por mínimos quadrados parciais (PLS-DA) entre os grupos DW e GC (Figura 22). Realizou-se a validação do modelo PLS-DA através do teste de permutação com a validação cruzada. O teste de

permutação no modelo de PLS-DA, com 2000 permutações, resultou em valores de $R^2=0,819$, $Q^2=0,664$ e p -permutação= $0,019$, sendo esses valores considerados aceitáveis para mostrar que há confiabilidade no modelo PLS-DA obtido. A Figura 22 ilustra o gráfico de *scores* do modelo PLS-DA, teste de permutação (validação cruzada) e importância da variável na projeção (VIP).

Após a validação cruzada da PLS-DA, foi possível identificar os valores de $VIP \geq 1$, a partir dos quais obtiveram-se 82 *features* moleculares caracterizados como os mais importantes para a discriminação dos grupos DW e GC, dentro do modelo projetado no PLS-DA (Figura 22C). Os 82 *features* moleculares provenientes dos $VIP \geq 1$, foram submetidos às análises univariadas, seguindo-se a mesma sequência estatística usada para as amostras de urina RPLC ESI(+)/ESI(-).

A análise de covariância (ANCOVA) resultou em 64 *features* moleculares estatisticamente significativos com $VIP \geq 1,0$, $p \leq 0,05$ e com p -valores do teste ANCOVA menores do que os valores FDR de 10% (Anexo 2, Tabela A8), os quais seguiram para a etapa de anotação putativa.

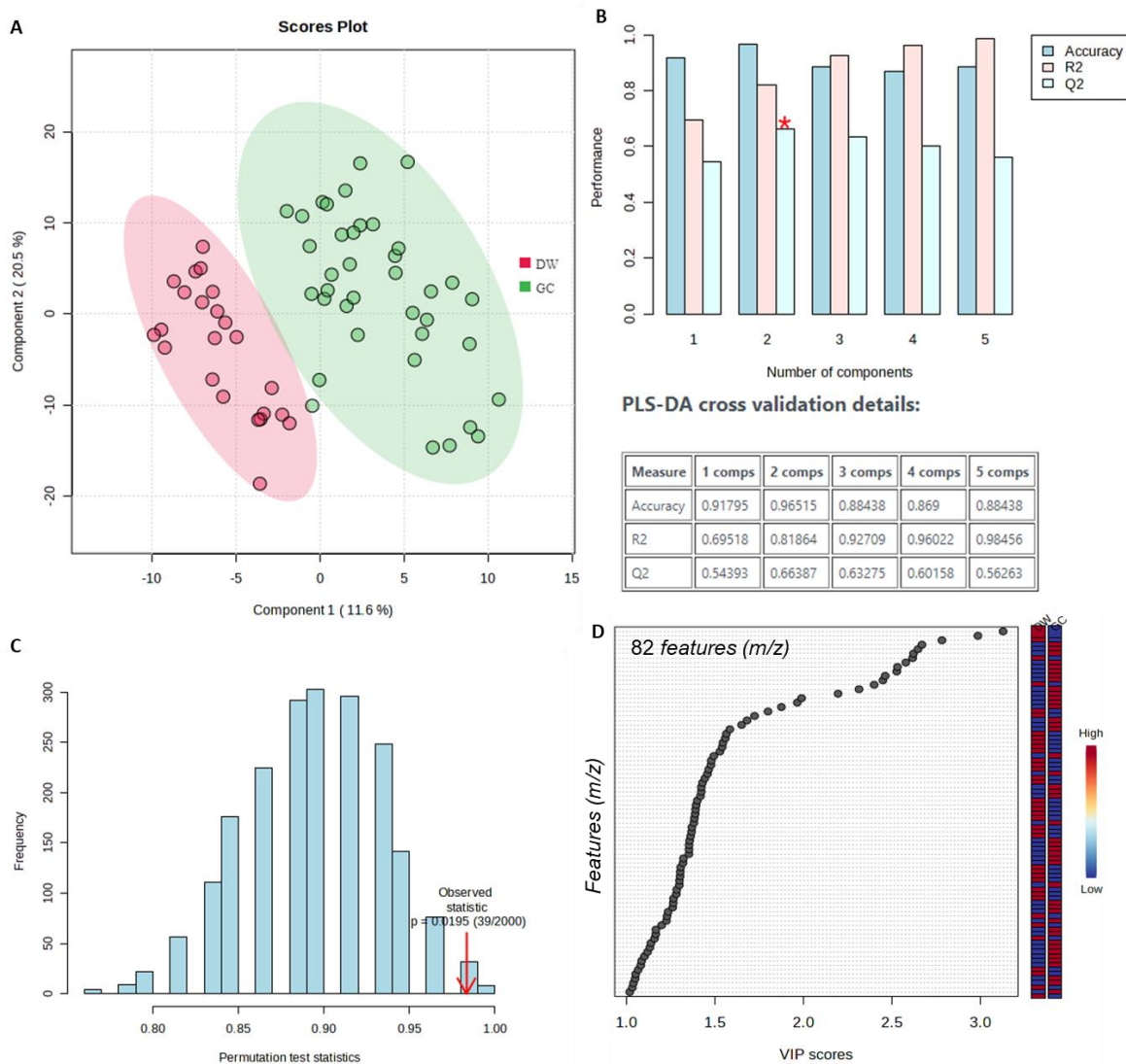


Figura 22- Análises multivariadas de metabolômica global por LC-HRMS com gráficos de *score plots* 2D do modelo PLS-DA no modo HILIC ESI(+) combinado com HILIC ESI(-), teste de permutação do modelo PLS-DA e importância da variável na projeção (VIP). **(A)** *Score plots* 2D das amostras de urina dos grupos DW e GC; **(B)** Teste de permutação do modelo PLS-DA; **(C)** PLS-DA de acordo com VIP scores no eixo x das amostras de urina (VIP \geq 1). As caixas coloridas indicam se a concentração do metabólito está aumentada (vermelho) ou diminuída (azul). *Software: MetaboAnalyst.*

4.6. Anotação Putativa do Metabólitos e Níveis de Identificação

Esse trabalho, por seguir uma abordagem de metabolômica global, é inerentemente exploratório. Na seção atual será discutida a seleção de metabólitos que foram classificados como relevantes para DW vs. GC. Sendo essa uma das etapas finais nos estudos de metabolômica global.

A identificação dos *features* moleculares provenientes dos experimentos de metabolômica global é, sem dúvida, a etapa mais desafiadora nas abordagens ômicas. Apesar dos consideráveis avanços nas tecnologias analíticas e nas plataformas de análise e identificação de dados ao longo dos anos, essa etapa continua sendo um grande obstáculo. Por outro lado, essa fase é fundamental, pois permitirá identificar os *features* moleculares e, posteriormente, permitir a análise das alterações nas vias bioquímicas do objeto de estudo.^{139,140}

Apenas uma pequena porcentagem, geralmente entre 2% e 10%, dos *features* moleculares detectados têm a possibilidade de serem identificados na metabolômica global. Os demais *features* moleculares permanecem sem identificação, sendo classificados como “desconhecidos”, representando um desafio considerável para as associações bioquímicas.¹⁴¹ Existem recomendações específicas para relatar o nível de confiança dos resultados obtidos com relação à identificação dos *features* moleculares na metabolômica, o qual foi proposto pelo grupo de *Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI)*.¹²¹ Os níveis de identificação de metabólitos estão categorizados de nível 1 a 4 e são uma forma de categorizar o grau de confiança na identificação de um metabólito já reportado ou um metabólito inédito. Na classificação, o nível 1 é o mais elevado e o nível 4 o menos elevado.

- **Nível 1:** A identificação de nível 1 é obtida ao comparar valores de massa exata (m/z), contribuição isotópica, tempo de retenção (t_R), e padrões de fragmentação (MS^2) do metabólito de interesse com um padrão comercial disponível e/ou pelo isolamento do composto seguido de sua caracterização estrutural através de técnicas como RMN (geralmente para substâncias não disponíveis comercialmente ou para metabólitos inéditos).

- **Nível 2:** A anotação de nível 2 ocorre quando se possui dois ou mais tipos de dados experimentais, por exemplo dados de MS^1 e MS^2 e contribuição

isotópica, que são comparados com dados de uma biblioteca espectral ou da literatura. No entanto, deve-se ter cautela ao comparar dados espectrais adquiridos em diferentes instrumentos, uma vez que parâmetros experimentais como o poder de resolução de massas, energia de colisão e a energia de ionização podem não ser compatíveis com os dados experimentais. Essas diferenças podem ser ainda mais significativas quando se utiliza diferentes aditivos nas fases móveis. Desta forma, embora não se obtenha a confirmação com um padrão autêntico, a identificação é apoiada por perfis de fragmentação e massa exata.

- **Nível 3:** O nível 3 é quando apenas um dos dados experimentais (MS^1 ou MS^2) é compatível com o banco de dados. Dessa forma, existem evidências para uma provável estrutura, porém não são suficientes para a confirmação exata, como é o caso de isômeros de posição. Nestes casos, a anotação por uso de ferramentas *in silico* são geralmente classificadas como nível 3. Neste nível, a anotação é suportada pela concordância entre a massa observada e a massa teórica do metabólito.

- **Nível 4:** O nível 4 de anotação é inferido quando a fórmula molecular pode ser associada à informação espectral, como por exemplo, átomos específicos e/ou isótopos. No entanto, não há evidências o suficiente para determinar a estrutura química. Apesar de ser o menor nível de confiança na anotação de metabólitos, essas informações podem ser utilizadas para rastreamento em estudos futuros. Assim, os *features* moleculares não são identificados.

Esses níveis de identificação ajudam a comunicar a confiança na identificação dos metabólitos em um estudo de metabolômica global.¹⁴² Neste estudo, após as análises estatísticas do plasma e da urina dos grupos DW e GC, nos modos RPLC e HILIC, foi possível propor anotações putativas para uma série de metabólitos que se encontram todos reportados na Tabela 5. O Anexo 3 contém exemplos de metabólitos identificados, mostrando os espectros obtidos experimentalmente através de análises LC-HRMS e comparando-os com os espectros teóricos disponíveis na biblioteca do MoNA. Essa comparação é importante para validar e confirmar a identificação dos metabólitos detectados.

Tabela 5 - Metabólitos discriminantes identificados com $VIP \geq 1,0$, $p \leq 0,05$ e Taxa de Falsa Descoberta (FDR) $< 0,10$.

Metabólito	HMDB	tr (min)	m/z	Aduto	Fórmula	VIP	p-Valor	FDR	Matriz/aquisição	Nível
Ácido glucárico	HMDB0000663	1.09	191.0202	[M-H ₂ O-H] ⁻	C ₆ H ₁₀ O ₈	1.42888	0.01701	0.08424	Plasma HILIC -	2
Sulfafenazol	HMDB0015667	1.19	349.0386	[M+Cl] ⁻	C ₁₅ H ₁₄ N ₄ O ₂ S	1.03658	0.02704	0.0897	Plasma HILIC -	3
Ácido aspártico	HMDB0000191	1.35	132.0303	[M-H] ⁻	C ₄ H ₇ NO ₄	1.57752	1.9E-05	0.01697	Plasma HILIC -	3
Ácido glutárico	HMDB0000661	0.30	177.0404	[M+FA-H] ⁻	C ₅ H ₈ O ₄	1.33893	0.00471	0.06848	Plasma HILIC -	3
Hidroxitubirilglicina	HMDB0094716	1.55	160.0615	[M-H] ⁻	C ₆ H ₁₁ NO ₄	1.82677	0.03397	0.09212	Plasma HILIC -	2
Sulfato de androsterona	HMDB0002759	0.85	369.1730	[M-H] ⁻	C ₁₉ H ₃₀ O ₅ S	1.06012	0.03851	0.09455	Plasma HILIC -	3
Sulfato de 3b,16a-dihidroxiandrostenona	HMDB0000386	2.67	383.1530	[M-H] ⁻	C ₁₉ H ₂₈ O ₆ S	1.37709	0.01212	0.08121	Plasma HILIC -	2
Fenilacetilglutamina	HMDB0006344	1.91	263.1036	[M-H] ⁻	C ₁₃ H ₁₆ N ₂ O ₄	1.17123	0.08473	0.09818	Plasma HILIC -	2
Alanil-hidroxi prolina	HMDB0028688	10.09	247.0933	[M+FA-H] ⁻	C ₈ H ₁₄ N ₂ O ₄	2.71218	5.3E-14	0.00061	Plasma HILIC -	2
PC O-(16:1/18:2)	LMGP01020257	10.29	786.5647	[M+FA-H] ⁻	C ₄₂ H ₈₀ NO ₇ P	1.01133	0.00961	0.07818	Plasma HILIC -	2
PE (22:2)/PGI2	HMDB0283954	10.24	894.5469	[M+FA-H] ⁻	C ₄₇ H ₈₀ NO ₁₀ P	1.07474	0.01607	0.08242	Plasma HILIC -	2
Aspartil-hidroxi prolina	HMDB0011160	10.59	291.0839	[M+FA-H] ⁻	C ₉ H ₁₄ N ₂ O ₆	1.47579	0.00356	0.06545	Plasma HILIC -	2
PC (18:0/12:0)	LMGP01010736	10.64	750.5279	[M+FA-H] ⁻	C ₃₈ H ₇₆ NO ₈ P	1.01127	0.03848	0.09394	Plasma HILIC -	2
Urolitina A 3-sulfato	HMDB0240567	10.59	352.9969	[M+FA-H] ⁻	C ₁₃ H ₈ O ₇ S	1.80277	3.4E-05	0.01879	Plasma HILIC -	2
SM (42:3);O2	HMDB0240615	11.09	855.6583	[M+FA-H] ⁻	C ₄₇ H ₉₁ N ₂ O ₆ P	1.36219	0.00212	0.05455	Plasma HILIC -	2
SM (36:2);O2	LMSP03010050	11.14	773.5809	[M+FA-H] ⁻	C ₄₁ H ₈₁ N ₂ O ₆ P	1.21299	0.00144	0.04909	Plasma HILIC -	2
Ácido monocarboxílico de malationa	HMDB0060626	11.49	347.0029	[M+FA-H] ⁻	C ₈ H ₁₅ O ₆ PS ₂	1.62197	0.00046	0.03758	Plasma HILIC -	2
Citrulina	HMDB0000904	11.34	174.0885	[M-H] ⁻	C ₆ H ₁₃ N ₃ O ₃	1.09003	0.04487	0.09576	Plasma HILIC -	2
LPC (16:1)	HMDB0010383	11.68	538.3148	[M+FA-H] ⁻	C ₂₄ H ₄₈ NO ₇ P	1.29988	0.00755	0.07515	Plasma HILIC -	2
SM (38:1);O2	LMSP03010061	11.83	803.6271	[M+FA-H] ⁻	C ₄₃ H ₈₇ N ₂ O ₆ P	1.0953	0.00071	0.04121	Plasma HILIC -	2
Lisina	HMDB0000182	13.16	145.0982	[M-H] ⁻	C ₆ H ₁₄ N ₂ O ₂	1.05139	0.00681	0.07273	Plasma HILIC -	2
N-(gamma-glutamil)etanolamina	HMDB0039222	13.51	189.0881	[M-H] ⁻	C ₇ H ₁₄ N ₂ O ₄	2.6643	8.7E-13	0.00121	Plasma HILIC -	2
Serilserina	HMDB0029048	2.92	173.0567	[M-H ₂ O-H] ⁻	C ₆ H ₁₂ N ₂ O ₅	2.0913	2.6E-06	0.00727	Plasma HILIC -	2
Sorbose	HMDB0001266	3.17	179.0563	[M-H] ⁻	C ₆ H ₁₂ O ₆	1.96807	5.8E-05	0.02121	Plasma HILIC -	2

Tabela 5- Metabólitos discriminantes identificados com $VIP \geq 1,0$, $p \leq 0,05$ e Taxa de Falsa Descoberta (FDR) $< 0,10$ (continuação).

Metabólito	HMDB	tr (min)	m/z	Aduto	Fórmula	VIP	p-Valor	FDR	Matriz/aquisição	Nível
Metionil-Tirosina	HMDB0028985	3.94	347.0839	[M+Cl] ⁻	C ₁₄ H ₂₀ N ₂ O ₄ S	1.34214	0.03799	0.09333	Plasma HILIC -	2
Ácido Eritrônico	HMDB0000613	5.50	135.0304	[M-H] ⁻	C ₄ H ₈ O ₅	1.20212	0.00061	0.04	Plasma HILIC -	2
Ácido N-Acetilneuramínico	HMDB0000230	5.80	308.0982	[M-H] ⁻	C ₁₁ H ₁₉ NO ₉	1.77293	7.3E-05	0.02303	Plasma HILIC -	3
Taurina	HMDB0000251	7.53	124.0073	[M-H] ⁻	C ₂ H ₇ NO ₃ S	2.07138	8.9E-06	0.01152	Plasma HILIC -	2
PE (24:0/20:4)	HMDB0009729	7.83	886.6095	[M-Cl] ⁻	C ₄₉ H ₉₀ NO ₈ P	1.66394	0.00033	0.03212	Plasma HILIC -	2
PE (20:4/18:0)	LMGP02010943	8.08	748.5271	[M-H ₂ O-H] ⁻	C ₄₃ H ₇₈ NO ₈ P	1.29739	0.00124	0.04848	Plasma HILIC -	2
Glicil-Hidroxiprolina	HMDB0011173	9.84	233.0778	[M+F _A -H] ⁻	C ₇ H ₁₂ N ₂ O ₄	1.30301	0.02281	0.08727	Plasma HILIC -	2
DG(18:1(9Z)-O(12,13)/0:0/i-15:0)	HMDB0299052	0.73	595.4917	[M+H] ⁺	C ₃₆ H ₆₆ O ₆	1.13023	0.02592	0.08909	Plasma HILIC +	2
5-Acetilamino-6-amino-3-metiluracil	HMDB0004400	1.89	181.0720	[M+H-H ₂ O] ⁺	C ₇ H ₁₀ N ₄ O ₃	1.14543	0.05006	0.09636	Plasma HILIC +	2
2-Penten-1-ol	HMDB0031604	10.09	104.1070	[M+NH ₄] ⁺	C ₅ H ₁₀ O	2.00973	3E-05	0.01818	Plasma HILIC +	2
PC (32:0)	HMDB0000564	10.93	734.5688	[M+H] ⁺	C ₄₀ H ₈₀ NO ₈ P	1.64136	0.00042	0.03636	Plasma HILIC +	2
SM (40:1);O2	HMDB0012092	10.93	809.6517	[M+Na] ⁺	C ₄₅ H ₉₁ N ₂ O ₆ P	1.58635	0.00537	0.07091	Plasma HILIC +	2
SM (34:1);O2	HMDB0013464	11.13	703.5752	[M+H] ⁺	C ₃₉ H ₇₉ N ₂ O ₆ P	1.40429	0.00091	0.04485	Plasma HILIC +	2
Ácido 3-Piridilacético	HMDB0001538	11.17	138.0547	[M+H] ⁺	C ₇ H ₇ NO ₂	1.15362	0.04357	0.09515	Plasma HILIC +	2
LPC (18:1)	HMDB0010385	11.51	522.3545	[M+H] ⁺	C ₂₆ H ₅₂ NO ₇ P	1.21391	0.00208	0.05394	Plasma HILIC +	2
LPC (16:1)	HMDB0010383	11.60	494.3238	[M+H] ⁺	C ₂₄ H ₄₈ NO ₇ P	1.36505	0.00245	0.05758	Plasma HILIC +	2
LPC (14:0)	HMDB0010379	11.70	468.3082	[M+H] ⁺	C ₂₂ H ₄₆ NO ₇ P	1.17951	0.01095	0.07879	Plasma HILIC +	2
Dissulfeto de Cisteinilglicina	HMDB0000709	12.21	298.0522	[M+H] ⁺	C ₈ H ₁₅ N ₃ O ₅ S ₂	1.25408	0.00207	0.05333	Plasma HILIC +	2
Ácido {[3,4,5-trihidroxi-6-(2,4,6-trihidroxifenil)oxan-2-il]metoxi}sulfônico	HMDB0132678	13.14	333.0280	[M+H-2H ₂ O] ⁺	C ₁₂ H ₁₆ O ₁₁ S	1.78973	1.2E-05	0.01394	Plasma HILIC +	3
Ácido diaminopimélico	HMDB0001370	13.49	191.1022	[M+H] ⁺	C ₇ H ₁₄ N ₂ O ₄	2.43369	1.4E-09	0.00364	Plasma HILIC +	2
1-Metilhistidina	HMDB0000001	13.57	170.0921	[M+H] ⁺	C ₇ H ₁₁ N ₃ O ₂	1.2379	0.0032	0.06182	Plasma HILIC +	2
Valina	HMDB0000883	2.04	118.0861	[M+H] ⁺	C ₅ H ₁₁ NO ₂	1.82837	1.3E-05	0.01455	Plasma HILIC +	2

Tabela 5- Metabólitos discriminantes identificados com $VIP \geq 1,0$, $p \leq 0,05$ e Taxa de Falsa Descoberta (FDR) $< 0,10$ (continuação).

Metabólito	HMDB	tr (min)	m/z	Aduto	Fórmula	VIP	p-Valor	FDR	Matriz/aquisição	Nível
Ácido 4-O- α -D-Galactopiranosil-D-galacturônico	HMDB0039721	4.14	371.0805	[M+H] ⁺	C ₁₂ H ₁₈ O ₁₃	1.01903	0.03464	0.09273	Plasma HILIC +	2
10-Pentacoseno	LMFA11000068	5.41	368.4246	[M+NH ₄] ⁺	C ₂₅ H ₅₀	1.29728	0.00351	0.06485	Plasma HILIC +	3
Hipoxantina	HMDB0000157	5.36	137.0457	[M+H] ⁺	C ₅ H ₄ N ₄ O	1.04296	0.00473	0.06909	Plasma HILIC +	2
Tirosina	HMDB0000158	5.51	146.0596	[M+H-2H ₂ O] ⁺	C ₉ H ₁₁ NO ₃	1.37528	0.00635	0.07152	Plasma HILIC +	2
1-O-(2-hidroxi-hexadecil)-sn-glicerol	LMGL01020063	6.98	371.2553	[M+K] ⁺	C ₁₉ H ₄₀ O ₄	1.59683	0.0025	0.05939	Plasma HILIC +	2
5'-N-Metilcarboxamidoadenosina	HMDB0246883	6.98	277.1044	[M+H-H ₂ O] ⁺	C ₁₁ H ₁₄ N ₆ O ₄	1.85057	8.3E-05	0.02424	Plasma HILIC +	2
Brassinolida	HMDB0034081	7.08	503.3338	[M+Na] ⁺	C ₂₈ H ₄₈ O ₆	1.17722	0.05443	0.09758	Plasma HILIC +	2
Creatinina	HMDB0000562	8.60	114.0661	[M+H] ⁺	C ₄ H ₇ N ₃ O	1.57779	0.00273	0.06061	Plasma HILIC +	2
Acetilcarnitina	HMDB0000201	9.45	204.1230	[M+H] ⁺	C ₉ H ₁₇ NO ₄	1.17989	0.02358	0.08788	Plasma HILIC +	2
Ácido Cinâmico	HMDB0000567	9.11	166.0859	[M+NH ₄] ⁺	C ₉ H ₈ O ₂	1.61465	0.00036	0.03515	Plasma HILIC +	2
Piperidina	HMDB0034301	9.26	86.0963	[M+H] ⁺	C ₅ H ₁₁ N	1.57116	0.00017	0.02667	Plasma HILIC +	2
Ácido Glutâmico	HMDB0000148	9.55	148.0601	[M+H] ⁺	C ₅ H ₉ NO ₄	1.44618	0.00171	0.05152	Plasma HILIC +	3
Ácido (2-hidroxi-2-{9-hidroxi-7-oxo-2H,3H,7H-furo[3,2-g]cromen-2-il}propoxi)sulfônico	HMDB0132923	1.17	393.0053	[M+Cl] ⁻	C ₁₄ H ₁₄ O ₉ S	1.1816	0.016	0.062	Plasma RPLC -	2
Ácido Edético	HMDB0015109	1.12	291.0839	[M-H] ⁻	C ₁₀ H ₁₆ N ₂ O ₈	1.5413	0.000	0.011	Plasma RPLC -	2
Ácido Glutâmico	HMDB0000148	1.90	128.0352	[M+H-H ₂ O] ⁺	C ₅ H ₉ NO ₄	1.6104	0.008	0.049	Plasma RPLC -	2
Ácido azelaico	HMDB0000784	11.07	187.0974	[M-H] ⁻	C ₉ H ₁₆ O ₄	1.17544	0.018	0.065	Plasma RPLC -	2
Sulfato de 3 β ,16 α -dihidroxiandrostenona	HMDB0000386	11.37	383.1530	[M-H] ⁻	C ₁₉ H ₂₈ O ₆ S	2.10885	0.000	0.007	Plasma RPLC -	2
Sulfato de 3 β ,16 α -dihidroxiandrostenona	HMDB0001032	13.47	367.1587	[M-H] ⁻	C ₁₉ H ₂₈ O ₅ S	1.61095	0.014	0.057	Plasma RPLC -	2
LPC (16:1)	HMDB0010383	17.83	538.3147	[M+FA-H] ⁻	C ₂₄ H ₄₈ NO ₇ P	1.69548	0.006	0.045	Plasma RPLC -	2
LPE (22:6)	HMDB0011526	18.17	524.2778	[M-H] ⁻	C ₂₇ H ₄₄ NO ₇ P	1.31673	0.010	0.052	Plasma RPLC -	2
LPE (18:2)	HMDB0011507	18.22	476.2781	[M-H] ⁻	C ₂₃ H ₄₄ NO ₇ P	1.18193	0.028	0.077	Plasma RPLC -	2
LPC(P-16:0/0:0)	HMDB0010407	19.36	524.3356	[M+FA-H] ⁻	C ₂₄ H ₅₀ NO ₆ P	1.4434	0.017	0.064	Plasma RPLC -	2

Tabela 5- Metabólitos discriminantes identificados com $VIP \geq 1,0$, $p \leq 0,05$ e Taxa de Falsa Descoberta (FDR) $< 0,10$ (continuação).

Metabólito	HMDB	tr (min)	m/z	Aduto	Fórmula	VIP	p-Valor	FDR	Matriz/aquisição	Nível
2-benzil-7-hidroxiocetanal	HMDB0134134	19.51	233.1546	[M-H] ⁻	C ₁₅ H ₂₂ O ₂	1.58657	0.003	0.035	Plasma RPLC -	2
SM(d18:1/16:1(9Z))	HMDB0240613	21.11	745.5494	[M+FA-H] ⁻	C ₃₉ H ₇₇ N ₂ O ₆ P	1.18549	0.031	0.082	Plasma RPLC -	2
Ácido δ-linolênico	HMDB0003073	21.60	277.2171	[M-H] ⁻	C ₁₈ H ₃₀ O ₂	1.82466	0.012	0.054	Plasma RPLC -	2
Ácido 3E-hexadecenoico	LMFA01030266	22.24	253.2171	[M-H] ⁻	C ₁₆ H ₃₀ O ₂	1.51874	0.029	0.079	Plasma RPLC -	2
Ácido araquidônico	HMDB0001043	22.29	303.2328	[M-H] ⁻	C ₂₀ H ₃₂ O ₂	1.10647	0.038	0.088	Plasma RPLC -	2
Ácido 2-hexadec-7-enilico-8,11-dienodioico	HMDB0255383	22.61	559.4725	[M-H] ⁻	C ₃₆ H ₆₄ O ₄	1.38263	0.027	0.076	Plasma RPLC -	2
Ácido linoleico	HMDB0000673	22.61	279.2332	[M-H] ⁻	C ₁₈ H ₃₂ O ₂	1.44234	0.016	0.061	Plasma RPLC -	2
Ácido palmítico	HMDB0000220	23.80	255.2331	[M-H] ⁻	C ₁₆ H ₃₂ O ₂	1.41145	0.020	0.069	Plasma RPLC -	2
Ácido horméico	LMFA01140034	24.39	307.2638	[M-H] ⁻	C ₂₀ H ₃₆ O ₂	1.02078	0.050	0.090	Plasma RPLC -	2
Ácido hexacosanodioico	LMFA01170040	24.81	425.3631	[M-H] ⁻	C ₂₆ H ₅₀ O ₄	1.01628	0.036	0.087	Plasma RPLC -	2
Ácido 2,5-dihidroxibenzenossulfônico	HMDB0245498	5.68	188.9862	[M-H] ⁻	C ₆ H ₆ O ₅ S	1.21122	0.050	0.089	Plasma RPLC -	2
Ácido adípico	HMDB0000448	6.60	145.0505	[M-H] ⁻	C ₆ H ₁₀ O ₄	1.39367	0.007	0.048	Plasma RPLC -	2
Fenilacetilglutamina	HMDB0006344	8.20	263.1035	[M-H] ⁻	C ₁₃ H ₁₆ N ₂ O ₄	1.61631	0.008	0.05	Plasma RPLC -	2
(5R,6R)-3-(2-acetamidoetilsulfanil)-6-etil-7-oxo-1-azabicyclo[3.2.0]hept-2-eno-2-carboxílico	HMDB0256886	1.11	337.0615	[M+K] ⁺	C ₁₃ H ₁₈ N ₂ O ₄ S	2.14867	2.36E-05	0.006	Plasma RPLC +	2
2-Amino-2-metil-1-propanol	HMDB0244974	1.61	72.0807	[M+H-H ₂ O] ⁺	C ₄ H ₁₁ NO	1.44313	0.015	0.058	Plasma RPLC +	2
LPC (16:1)	HMDB0010383	17.83	494.3241	[M+H] ⁺	C ₂₄ H ₄₈ NO ₇ P	1.45046	0.032	0.085	Plasma RPLC +	2
LPC (20:4)	HMDB0010396	18.28	544.3399	[M+H] ⁺	C ₂₈ H ₅₀ NO ₇ P	1.8257	0.000	0.017	Plasma RPLC +	2
MG (18:0)	HMDB0011131	18.48	400.3421	[M+ACN+H] ⁺	C ₂₁ H ₄₂ O ₄	1.12235	0.030	0.080	Plasma RPLC +	2
12-Hidroxi-12-octadecanoilcarnitina	HMDB0013154	18.63	426.3576	[M+H-H ₂ O] ⁺	C ₂₅ H ₄₉ NO ₅	1.12235	1.32E-05	0.005	Plasma RPLC +	2
LPC (16:0)	HMDB0010382	18.82	496.3408	[M+H] ⁺	C ₂₄ H ₅₀ NO ₇ P	1.09706	0.031	0.081	Plasma RPLC +	2
LPC (18:0)	HMDB0010384	20.60	524.3717	[M+H] ⁺	C ₂₆ H ₅₄ NO ₇ P	1.34708	0.006	0.043	Plasma RPLC +	2
Ácido benzoquinonaacético	HMDB0002334	20.20	149.0230	[M+H-H ₂ O] ⁺	C ₈ H ₆ O ₄	1.91662	0.000	0.020	Plasma RPLC +	2
SM (34:2)	HMDB0240613	20.85	701.5590	[M+H] ⁺	C ₃₉ H ₇₇ N ₂ O ₆ P	1.12499	0.019	0.068	Plasma RPLC +	2

Tabela 5- Metabólitos discriminantes identificados com $VIP \geq 1,0$, $p \leq 0,05$ e Taxa de Falsa Descoberta (FDR) $< 0,10$ (continuação).

Metabólito	HMDB	tr (min)	m/z	Aduto	Fórmula	VIP	p-Valor	FDR	Matriz/aquisição	Nível
Ftalato de dipentilo	HMDB0251441	21.36	307.1902	[M+H] ⁺	C ₁₈ H ₂₆ O ₄	1.15171	0.015	0.060	Plasma RPLC +	2
Galabiosilceramida (18:1/16:0);O2	HMDB0004833	23.84	862.6250	[M+H] ⁺	C ₄₆ H ₈₇ NO ₁₃	1.17761	0.000	0.013	Plasma RPLC +	2
Propionilcarnitina	HMDB0000824	3.39	218.1384	[M+H] ⁺	C ₁₀ H ₁₉ NO ₄	1.04054	0.031	0.083	Plasma RPLC +	2
4-Metoxibenzeno propanol 1-(2-sulfoglicosídeo)	HMDB0034757	0.59	389.0906	[M-H ₂ O-H] ⁻	C ₁₆ H ₂₄ O ₁₀ S	1.7997	0.00937	0.05122	Urina HLIC -	2
Alantoína	HMDB0000462	1.97	157.0366	[M-H] ⁻	C ₄ H ₆ N ₄ O ₃	1.3698	0.00515	0.04634	Urina HLIC -	2
Ácido silícico	HMDB0031436	10.95	140.9859	[M+FA-H] ⁻	H ₄ O ₄ Si	1.5279	0.00018	0.03293	Urina HLIC -	2
8-Azahipoxantina	HMDB0247423	2.07	182.0321	[M+FA-H] ⁻	C ₄ H ₃ N ₅ O	1.199	0.05513	0.07317	Urina HLIC -	2
Ácido pimélico	HMDB0000857	5.48	159.0662	[M-H] ⁻	C ₇ H ₁₂ O ₄	1.1193	0.04988	0.07195	Urina HLIC -	2
Neomentol-glucuronídeo	HMDB0060012	6.68	331.1757	[M-H] ⁻	C ₁₆ H ₂₈ O ₇	1.3838	0.01346	0.05854	Urina HLIC -	2
Ácido 4-hidroxi-2-butenóico gama-lactona	HMDB0032330	7.29	129.0193	[M+FA-H] ⁻	C ₄ H ₄ O ₂	1.6509	0.01215	0.0561	Urina HLIC -	3
Eritrosa	HMDB0002649	7.59	119.0348	[M-H] ⁻	C ₄ H ₈ O ₄	1.9653	0.00354	0.04512	Urina HLIC -	3
Rhamnalpinogenina	HMDB0031533	8.44	363.0287	[M-Cl] ⁻	C ₁₇ H ₁₂ O ₇	2.6214	3E-05	0.02805	Urina HLIC -	2
Ácido 3-hidroxissubérico	HMDB0000325	9.25	171.0661	[M-H ₂ O-H] ⁻	C ₈ H ₁₄ O ₅	1.2996	0.00804	0.04878	Urina HLIC -	2
Ácido quinurênico	HMDB0000715	0.92	190.0494	[M+H] ⁺	C ₁₀ H ₇ NO ₃	2.449	0.00021	0.03415	Urina HLIC +	2
Ácido 3-piridilacético	HMDB0001538	11.62	138.0547	[M+H] ⁺	C ₇ H ₇ NO ₂	1.5454	0.04121	0.06585	Urina HLIC +	2
Acetilcarnitina	HMDB0000201	12.23	204.1227	[M+H] ⁺	C ₉ H ₁₇ NO ₄	1.3905	0.041	0.06463	Urina HLIC +	2
Ácido imidazolopropiônico	HMDB0002271	12.52	141.0655	[M+H] ⁺	C ₆ H ₈ N ₂ O ₂	1.4182	0.02218	0.0622	Urina HLIC +	2
N1-Acetilpermidina	HMDB0001276	14.36	188.1753	[M+H] ⁺	C ₉ H ₂₁ N ₃ O	1.3045	0.04378	0.06829	Urina HLIC +	2
5-Acetamidovalerato	HMDB0012175	3.75	160.0963	[M+H] ⁺	C ₇ H ₁₃ NO ₃	1.3941	0.01148	0.05366	Urina HLIC +	2
Betaína	HMDB0000043	3.54	100.0754	[M+H-H ₂ O] ⁺	C ₅ H ₁₁ NO ₂	1.4582	0.00338	0.0439	Urina HLIC +	2
9-(1,3-Dioxolan-2-il)purina-2,6-diamina	HMDB0257744	3.95	222.0871	[M+Na] ⁺	C ₈ H ₁₀ N ₆ O ₂	1.3885	0.05698	0.07439	Urina HLIC +	2
7-Metilhipoxantina	HMDB0003162	4.67	151.0610	[M+H] ⁺	C ₆ H ₆ N ₄ O	1.0667	0.06067	0.07561	Urina HLIC +	2
Arabitol	HMDB0000568	4.57	153.0751	[M+H] ⁺	C ₅ H ₁₂ O ₅	1.541	0.00234	0.04268	Urina HLIC +	2
1-Tetradeceno	HMDB0061834	7.61	214.2524	[M+NH ₄] ⁺	C ₁₄ H ₂₈	1.3681	0.00165	0.04024	Urina HLIC +	2

Tabela 5- Metabólitos discriminantes identificados com $VIP \geq 1,0$, $p \leq 0,05$ e Taxa de Falsa Descoberta (FDR) $< 0,10$ (continuação).

Metabólito	HMDB	tr (min)	m/z	Aduto	Fórmula	VIP	p-Valor	FDR	Matriz/aquisição	Nível
Hexilamina	HMDB0032323	8.82	102.1275	[M+H] ⁺	C ₆ H ₁₅ N	2.6696	4.1E-13	0.0061	Urina HLIC +	2
1-Estearoilglicerol	HMDB0244009	23.54	393.2767	[M+Cl] ⁻	C ₂₁ H ₄₂ O ₄	1.96	0.00845	0.04063	Urina RPLC -	2
1-Hidroxi-2-(N-acetilcisteinil)-3-buteno	HMDB0244631	5.42	278.0698	[M+FA-H] ⁻	C ₉ H ₁₅ NO ₄ S	1.38	0.03085	0.04688	Urina RPLC -	2
Ácido 7-hidroxi-5-heptinoico	LMFA01050262	5.70	187.0610	[M+FA-H] ⁻	C ₇ H ₁₀ O ₃	1.14	0.02014	0.04219	Urina RPLC -	2
Indol-3-acético-ácido-O-glucuronídeo	HMDB0060001	8.76	350.0877	[M-H] ⁻	C ₁₆ H ₁₇ NO ₉	1.14	0.04412	0.05313	Urina RPLC -	2
Pregnanediol 3-O-glucuronídeo	HMDB0010318	15.16	496.3026	[M-H] ⁻	C ₂₇ H ₄₄ O ₉	1.33	0.02596	0.04531	Urina RPLC -	2
S-adenosil-L-metioninamina	HMDB0304476	22.77	355.1573	[M-H] ⁻	C ₁₄ H ₂₄ N ₆ O ₃ S ²⁺	2.67	2.1E-12	0.00781	Urina RPLC -	2
Ácido 3,3'-tiobispropanoico	HMDB0031162	3.10	177.0224	[M-H] ⁻	C ₆ H ₁₀ O ₄ S	1.79	0.00513	0.03594	Urina RPLC -	2
Sulfato de pirocatecol	HMDB0059724	4.00	188.9863	[M-H] ⁻	C ₆ H ₆ O ₅ S	1.65	0.00399	0.03125	Urina RPLC -	2
Glutamina	HMDB0000641	7.44	145.0617	[M-H] ⁻	C ₅ H ₁₀ N ₂ O ₃	1.33	0.00518	0.0375	Urina RPLC -	2
Polietileno oxidado	HMDB0032472	11.12	243.1234	[M-H] ⁻	C ₁₂ H ₂₀ O ₅	1.71	0.00464	0.03281	Urina RPLC -	2
Hepta-2,5-dienoilcarnitina	HMDB0241691	20.00	250.1444	[M-H ₂ O-H] ⁻	C ₁₄ H ₂₃ NO ₄	1	1.8E-05	0.02344	Urina RPLC -	2
Ácido imidazolopropiônico	HMDB0002271	1.06	141.0656	[M+H] ⁺	C ₆ H ₈ N ₂ O ₂	1.44	0.00027	0.025	Urina RPLC +	2
3-Palmitoil-sn-glicerol	HMDB0245964	23.76	331.2844	[M+H] ⁺	C ₁₉ H ₃₈ O ₄	1.47	1.4E-08	0.01875	Urina RPLC +	2
Ácido 4-hidroxiciclohexilcarboxílico	HMDB0001988	8.69	109.0646	[M+H-2H ₂ O] ⁺	C ₇ H ₁₂ O ₃	1.52	0.00031	0.02813	Urina RPLC +	2
N-Fructoilesteirilalanina + C2H5S	PR310819	1.69	459.1112	[M+FA-H] ⁻	C ₁₄ H ₂₆ N ₂ O ₈ S ₂	2.28	0.000	0.004	Plasma RPLC -	2

DG(18:1(9Z)-O(12,13)/0:0/i-15:0): Ácido (2S)-2-Hidroxi-3-(((9Z)-11-(3-pentiloxiran-2-il)undec-9-enoil]oxi)propil 13-metiltetradecanoíco; PC: fosfatidilcolina; PE; fosfatidiletanolamina; SM: esfingomielina; LPC: lisofosfatidilcolina; LPE: lisofosfatidiletanolamina; MG: Monoacilglicerol.

4.7. Curva Característica de Operação do Receptor (ROC)

A curva ROC é uma ferramenta relevante para avaliar a capacidade de um modelo em distinguir entre os resultados positivos e negativos, ajudando a determinar a eficácia de um diagnóstico. A curva ROC apresenta uma resposta para sensibilidade no eixo y e a especificidade no eixo x para um sistema de classificação binária, como positivo ou negativo. A sensibilidade é a capacidade de identificar o resultado positivo quando a condição está presente e a especificidade é a capacidade de identificar os resultados negativos quando a condição está ausente. A Área sob a Curva ROC (AUC) é uma métrica que avalia o desempenho de um modelo ou teste diagnóstico. Valores de AUC variam de 0,5 a 1, refletindo diferentes níveis de capacidade de discriminação. Quanto mais próxima de 1 a AUC estiver, melhor será a capacidade do modelo em distinguir entre os grupos em análise.^{143,144}

4.7.1. Curva ROC das Análises das Amostras de Plasma e Urina

Após a condução das análises univariadas e anotação putativa dos metabólitos, o método PLS-DA foi empregado para obter a área sob a curva (AUC) com base nas curvas características de operação do receptor (ROC) para as amostras de plasma e urina. As curvas ROC foram geradas usando desempenho de validação cruzada de Monte Carlo (MCCV, do inglês: *Monte Carlo cross validation*). A curva ROC permitiu a medida de desempenho do teste, ou seja, como indicativo da eficácia do diagnóstico entre indivíduos com DW e do grupo GC.

Para a análise das amostras de plasma, os dados de 90 metabólitos foram utilizados e testados com diferentes números de variáveis (3, 5, 10, 20, 45, 90), como ilustrado na Figura 23A. Para a análise das amostras de urina, foram utilizados 37 metabólitos e testados diferentes números de variáveis (2,3, 5, 10, 20, 37), como ilustrado na Figura 23B.

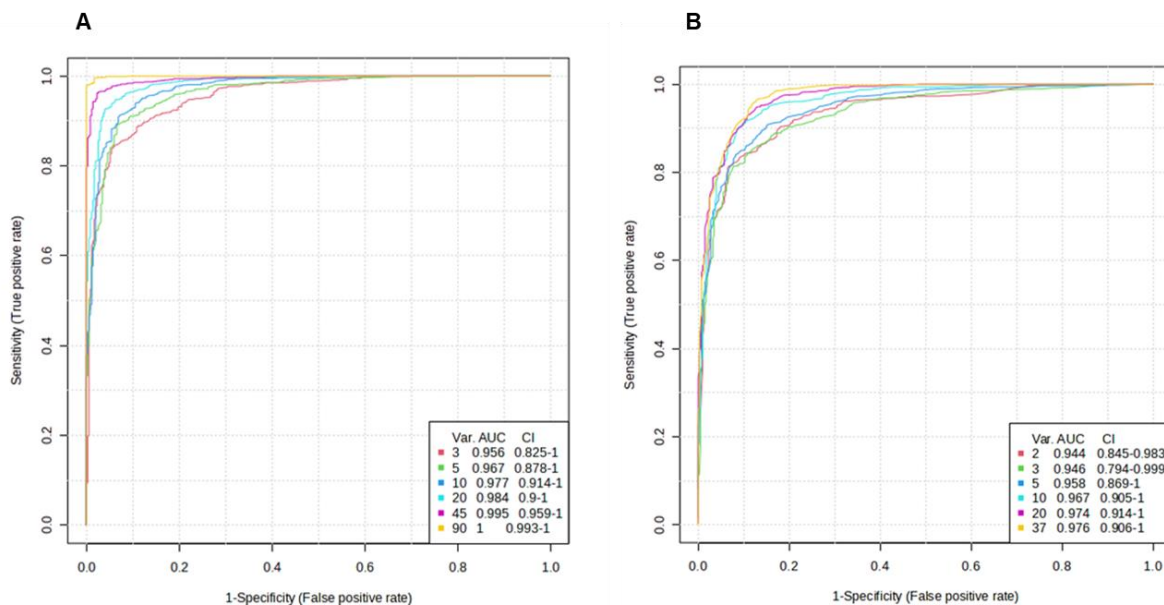


Figura 23- Curvas ROC pelo método estatístico multivariado no ponto de corte ideal definido pela distância mínima para o canto superior esquerdo. **(A)** Modelo otimizado gerou uma AUC de 0,995 (IC 95%: 0,956–1) a partir de amostras de plasma. **(B)** Modelo otimizado gerou uma AUC de 0,976 (IC 95%: 0,944–0,976) a partir de amostras de urina. *Software: MetaboAnalyst.*

A curva ROC das amostras de plasma (Figura 23A) apresentou pontuação AUC variando de 0,956 e 1 (IC 95%: 0,825-1), o que indica um excelente modelo com elevada sensibilidade, ou seja, capacidade preditiva do modelo. A curva ROC das amostras de plasma demonstrou que a melhor classificação foi obtida com 9 metabólitos, representada pela curva amarela. No entanto, 98,4% de capacidade preditiva foi obtida com 3 metabólitos (Figura 24A), sendo esse o modelo escolhido para fins de previsão. A curva ROC das amostras de urina (Figura 23B) apresentou pontuação AUC variando de 0,944 a 0,976 (IC 95%: 0,794-1), o que indica um bom modelo. Para as amostras de urina, a melhor classificação foi obtida com 37 metabólitos, representada pela curva amarela, com 91,3% na capacidade preditiva. No entanto, o modelo com 2 metabólitos apresentou pontuação AUC de 0,946 e 86,1% na capacidade preditiva (Figura 25A), sendo o escolhido para fins de previsão.

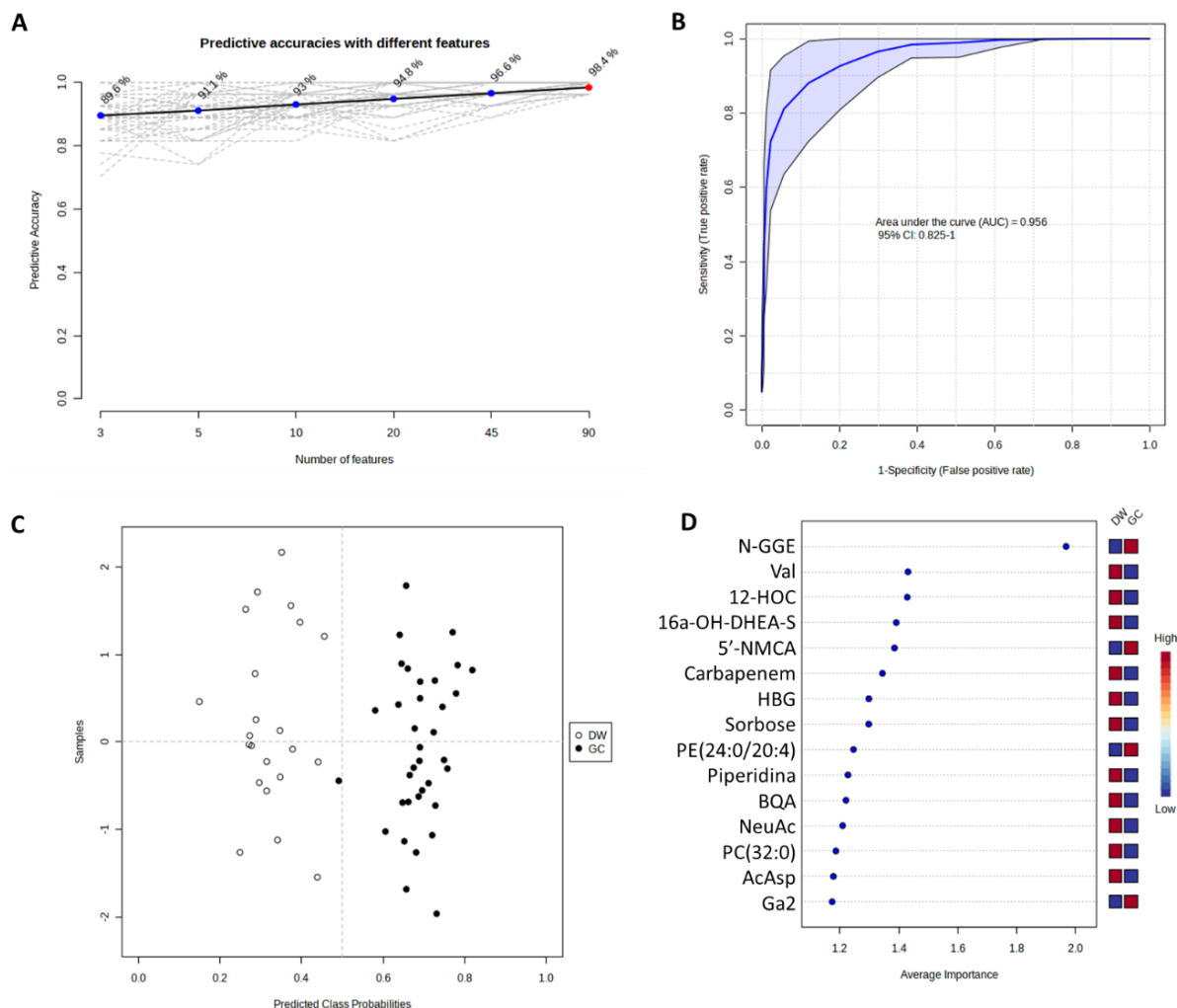


Figura 24- Resultados da Curva ROC método classificatório PLS-DA para a análise das amostras de plasma. **A)** exatidão preditiva, **B)** Curva ROC, modelo 3 metabólitos **C)** predição de classes e **D)** gráfico de VIP scores. *Software: MetaboAnalyst.*

N-GGE: N-(δ -glutamil)etanolamina; Val: Valina; 12-HOC: 12-Hidroxi-12-octadecanoilcarnitina; 16 α -OH-DHEA-S: Sulfato de 3 β ,16 α -dihidroxiandrosteno-3,20-dione; 5'-NMCA: 5'-N-Metilcarboxamidoadenosina; Carbapenem: Ácido (5R,6R)-3-(2-acetamidoetilsulfanil)-6-etil-7-oxo-1-azabicyclo[3.2.0]hept-2-eno-2-carboxílico; HBG: Hidroxibutirilglicina; PE: fosfatidiletanolamina; BQA: Ácido benzoquinonaacético; NeuAc: Ácido N-acetilneuramínico; PC: fosfatidilcolina; AcAsp: Ácido aspártico; Ga2: Galabiosilceramida (18:1/16:0);O2.

A Figura 24B mostra a curva ROC para as amostras de plasma com 3 metabólitos (IC 95%: 0,825-1) e pontuação AUC de 0,956, sendo considerado uma ótima medida para a discriminação entre os grupos DW e GC. O gráfico de predição de classes (Figura 24C) reforça a consideração da capacidade preditiva do modelo,

na qual claramente há uma discriminação dos grupos DW e GC. Outro aspecto importante analisado foi o gráfico de VIP scores, representando a média da importância de cada variável na distinção entre os grupos (Figura 24D). Este gráfico destacou 15 metabólitos, sendo esses: N-GGE, Val, 12-HOC, 16 α -OH-DHEA-S, 5'-NMCA, Carbapenem, HBG, PE(24:0/20:4), BQA, NeuAc, PC (32:0), AcAsp e Ga2. Em comparação aos outros metabólitos, o N-GGE foi o metabólito com maior importância, seguido pela Val, sendo esses componentes das vias metabólicas do metabolismo de aminoácidos, apresentando uma regulação negativa no grupo DW. Esses metabólitos evidenciaram níveis alterados de diferentes classes de compostos, constituídos por aminoácidos, ácidos graxos, esteróides, purinas, derivados de β -lactamas, carboidratos, piperidina, fosfolipídeos e esfingolipídeos (Tabela 6).

A Figura 25B mostra a curva ROC para as amostras de urina com 2 metabólitos (IC 95%: 0,845-0,983) e pontuação AUC de 0,944 o que é considerado uma ótima medida para a discriminação do grupo DW e GC. A capacidade preditiva também é demonstrada pelo gráfico de predição de classes na Figura 25C. O VIP scores dos principais metabólitos anotados nas amostras de urina destacou 15 metabólitos, sendo eles: 1-HexGlic, KU, Ery, NM-Gluc δ -lactona, 1-SG, 13-Met-TTAcon, 5-AceVal, Preg-Glu, 4-MeBenz e ImP. Esses metabólitos evidenciaram níveis alterados de diferentes classes de compostos, constituídos por quinolina, carboidrato, furanona, aminoácidos, glicosídeos terpênicos, álcoois graxos, ácidos graxos, esteróides, imidazol e hidrocarbonetos alifáticos insaturados (Tabela 6). A Tabela 6 descreve os metabólitos discriminantes provenientes dos VIP scores ≥ 1 obtidos dos modelos PLS-DA provenientes das análises das amostras de plasma e urina, incluindo a taxinomia química com a superclasse e subclasse dos metabólitos.

Os gráficos circulares descritos na Figura 26 representam as superclasses (Figura 26A) e as subclasses (Figura 26B) dos metabólitos putativamente anotados que alcançaram valores de *fold-change* no intervalo de $0,7 \geq FC \geq 1,5$. Dentre as superclasses, majoritariamente observou-se lipídeos (41%), ácidos orgânicos (26%), compostos orgânicos heterocíclicos (15%), compostos orgânicos contendo oxigênio (10%), hidrocarbonetos alifáticos insaturados (5%) e nucleotídeos (3%). Nas categorias de subclasses, tem-se majoritariamente ácidos graxos (25%), seguido de aminoácidos (17%) e esteróides sulfatados (11%). E, em menores porcentagens, temos purinas e carboidratos (8% cada), glicosídeos

terpênicos e imidazol (5% cada) e, com 3% cada, a classe de álcool, esfingolípídeos, ácidos dicarboxílicos, hidroácido, cetoácidos, amida e quinolinas.

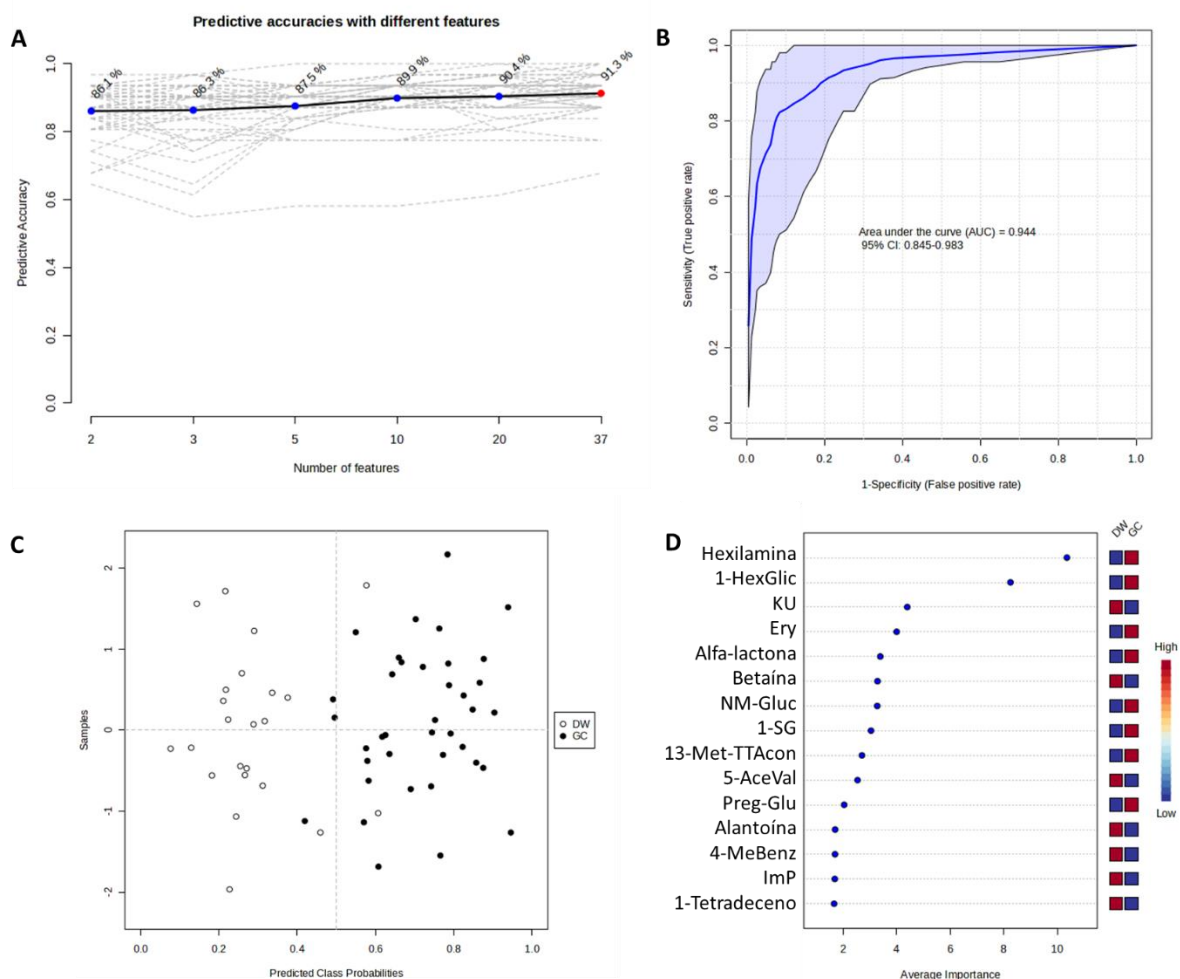


Figura 25- Resultados da Curva ROC método classificatório PLS-DA para a análise das amostras de urina. **A)** exatidão preditiva; **B)** Curva ROC, modelo 2 metabólitos; **C)** predição de classes; **D)** gráfico de VIP scores. *Software: MetaboAnalyst.*

1-HexGlic: 1-hexadecanoato de glicerol; KU: Ácido quinurênico, Ery: Eritrose; NM-Gluc: Neomentol-glucuronídeo; δ -lactona: δ -lactona do ácido 4-hidroxi-2-butenóico; 1-SG: 1-Estearoilglicerol; 13-Met-TTAcon: 13-Metil-1-tritriaconteno; 5-AceVal: 5-acetamidavalerato; Preg-Glu: Pregnanodiol 3-O-glucuronídeo; 4-MeBenz: 4-Metoxibenzenopropanol 1-(2-sulfoglicosídeo); ImP: Ácido imidazolpropiônico.

Tabela 6- Metabólitos discriminantes dos VIPs scores obtidos dos modelos PLS-DA provenientes das análises das amostras de plasma e urina

No.	Metabólito	Abreviação	Taxonomia Química Superclasse - Subclasse
Vip scores ≥ 1 - Análise das Amostras de Plasma			
1	N-(δ-glutamil)etanolamina	N-GGE	Ácidos orgânicos - aminoácido
2	Valina	Val	Ácidos orgânicos - aminoácido
3	12-Hidroxi-12-octadecanoilcarnitina	12-HOC	Ácidos orgânicos – ácido graxo
4	Sulfato de 3b,16a-dihidroxiandrosteno	16α-OH-DHEA-S	Lipídeos – esteróide
5	5'-N-Metilcarboxamidoadenosina	5'-NMCA	Nucleosídeos - purina
6	Ácido (5R,6R)-3-(2-acetamidoetil-sulfanil)-6-etil-7-oxo-1-azabicyclo[3.2.0]hept-2-eno-2-carboxílico	Carbapenem	Compostos organoheterocíclicos – β-lactama
7	Hidroxibutirilglicina	HBG	Ácidos orgânicos - aminoácido
8	Sorbose	Sorbose	Lipídeos – carboidrato
9	Fosfatidiletanolamina (24:0/20:4)	PE(24:0/20:4)	Lipídeos – glicerofosfoetanolaminas
10	Piperidina	Piperidina	Compostos orgânicos - piperidina
11	Ácido benzoquinonaacético	BQA	Compostos orgânicos - carbonílico
12	Ácido N-acetilneuramínico	NeuAc	Compostos orgânicos - carboidrato
13	Fosfatidilcolina (32:0)	PC(32:0)	Lipídeos – glicerofosfocolina
14	Ácido aspártico	AcAsp	Ácido orgânicos - aminoácido
15	Galabiosilceramida (18:1/16:0);O2	Ga2	Lipídeos – glicosídeo
Vip scores ≥ 1 - Análise das Amostras de Urina			
1	Hexilamina*	Hexilamina	Compostos orgânicos – amina
2	1-Hexadecanoato de glicerol**	1-HexGlic	Lipídeos – glicerolípido
3	Ácido cinurênico	KU	Compostos orgânicos – quinolina
4	Eritrose	Ery	Compostos orgânicos – carboidrato
5	δ-Lactona do ácido 4-hidroxi-2-butenóico	δ-lactona	Compostos organoheterocíclico – furanona
6	Betaína	Betaína	Compostos orgânicos - aminoácido
7	Neomentol-glucuronídeo	NM-Gluc	Lipídeos – glicosídeo terpênico
8	1-Estearoilglicerol**	1-SG	Lipídeos – álcool graxo
9	13-Metil-1-tritriaconteno	13-Met-TTAcon	Lipídeos – álcool graxo
10	5-Acetamidavalerato	5-AceVal	Compostos orgânicos – ácido graxo
11	Pregnanodiol 3-O-glucuronídeo	Preg-Glu	Compostos orgânicos – esteróide
12	Alantoína	All	Compostos organoheterocíclico – imidazol
13	4-Metoxibenzenopropanol 1-(2-sulfoglicosídeo)	4-MeBenz	Compostos orgânicos – acil graxos
14	Ácido imidazolpropionico	ImP	Compostos organoheterocíclico – imidazol
15	1-Tetradeceno**	1-Tetradeceno	Compostos orgânicos – hidrocarboneto insaturado alifático

*aditivo alimentos; **exossoma humano;

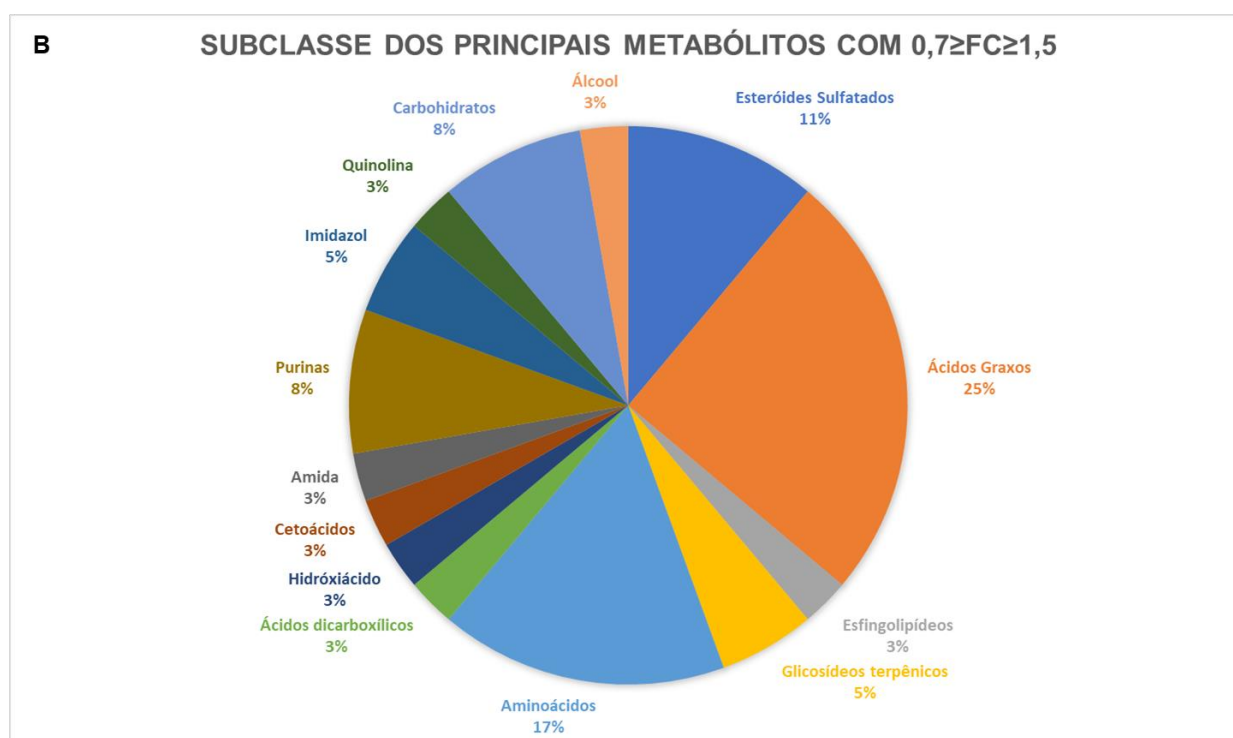
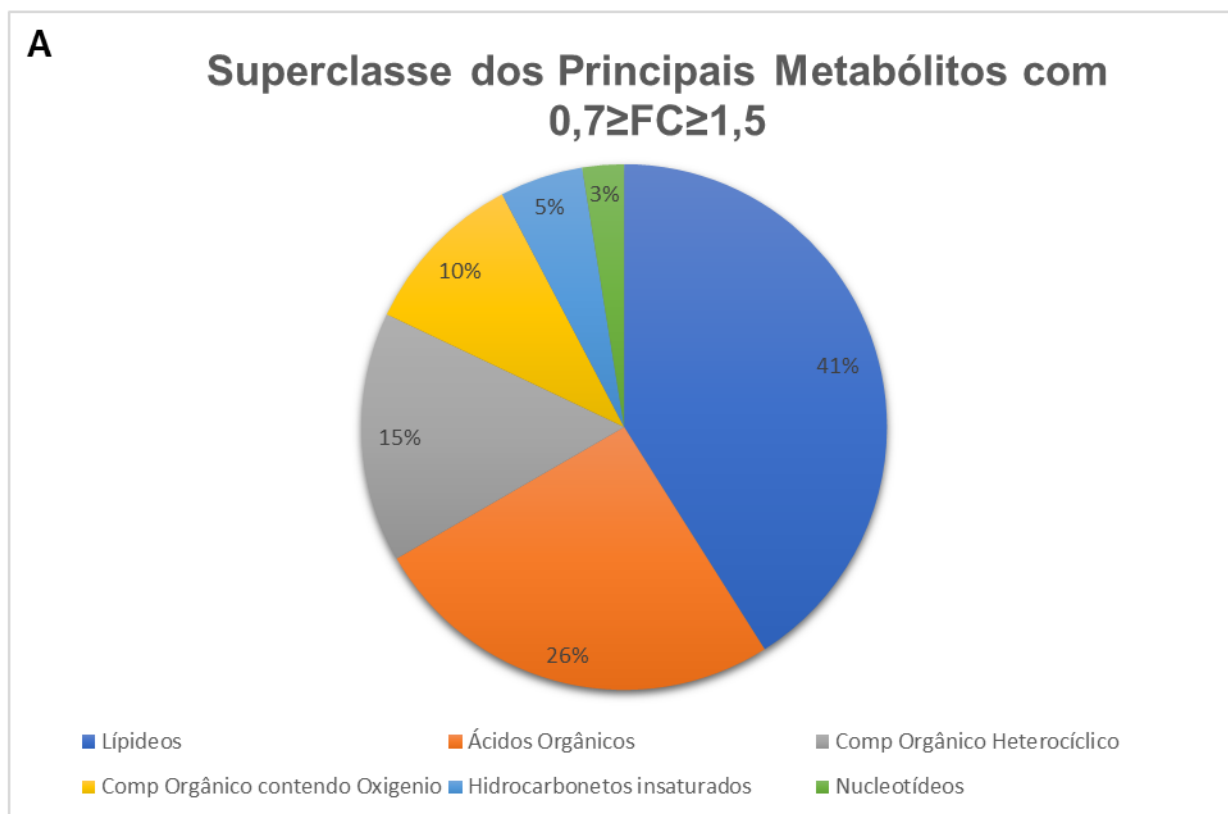


Figura 26- Gráficos circulares descrevendo as superclasses (**A**) e as subclasses (**B**) dos metabólitos putativamente anotados que alcançaram valores de *fold-change* no intervalo de $0,7 \geq FC \geq 1,5$.

4.8. Alterações Vias Bioquímicas

Análises funcionais das vias metabólicas foram conduzidas com o objetivo de obter informações sobre a importância dos metabólitos nos processos bioquímicos e suas associações com a DW. O nível significativo para a via metabólica foi determinado como $p \leq 0,05$. De forma geral, considera-se uma via metabólica é significativa quando seu impacto é maior que 0,1 e seu valor $-\log(p\text{-value})$ ultrapassa 1,21.^{145,146}

O impacto das vias metabólicas (Figura 27) representa uma estimativa objetiva da importância de uma determinada via em relação à rede metabólica global. Para essa análise, empregou-se o método de enriquecimento por teste hipergeométrico, que é um teste estatístico utilizado na análise de tabelas de contingência. Além disso, utilizamos a análise topológica com a métrica de centralidade de intermediação relativa. Para essa avaliação, utilizou-se a biblioteca de vias metabólicas *Homo sapiens* do banco de dados KEGG. Esse processo nos permitiu identificar e destacar as vias metabólicas mais significativas e relevantes em relação aos metabólitos anotados, fornecendo uma compreensão dos processos metabólicos envolvidos na DW.

A Figura 27 ilustra as vias metabólicas alteradas, nas quais os mais elevados valores de p e de impacto refletem os aspectos mais relevantes das vias metabólicas impactadas. Considerando os valores significativos para o impacto das rotas metabólicas, quatro vias metabólicas apresentaram-se mais impactadas, sendo essas: **1)** Biossíntese de arginina (impacto=0,23 e $-\log(p)=3,30$); **2)** Metabolismo do ácido linoléico (impacto=1,0 e $-\log(p)=1,23$); **3)** Metabolismo de alanina, aspartato e glutamato (impacto=0,34 e $-\log(p)=1,36$) e **4)** Biossíntese de fenilalanina, tirosina e triptofano (impacto=0,5 e $-\log(p)=1,32$). É interessante ressaltar que dentre as vias metabólicas alteradas, a da biossíntese de arginina, o metabolismo de alanina, aspartato e glutamato e a biossíntese de fenilalanina, tirosina e triptofano, já foram reportadas em estudos anteriores relacionados à DW conduzidos por meio de análises de metabolômica global.^{64,147}

As vias metabólicas da biossíntese da arginina e o metabolismo de alanina, aspartato e glutamato foram relatadas por Qiu (2023), a pesquisa avaliou plasma de pacientes com DW e empregou fase reversa no equipamento Thermo Q Exactive™.⁶⁴ Enquanto isso, a via metabólica da biossíntese de fenilalanina, tirosina

e triptofano foi relatada por Mazi (2019), que avaliou plasma de pacientes com DW e realizou análises utilizando fase HILIC em um equipamento QTOF™.¹⁴⁷

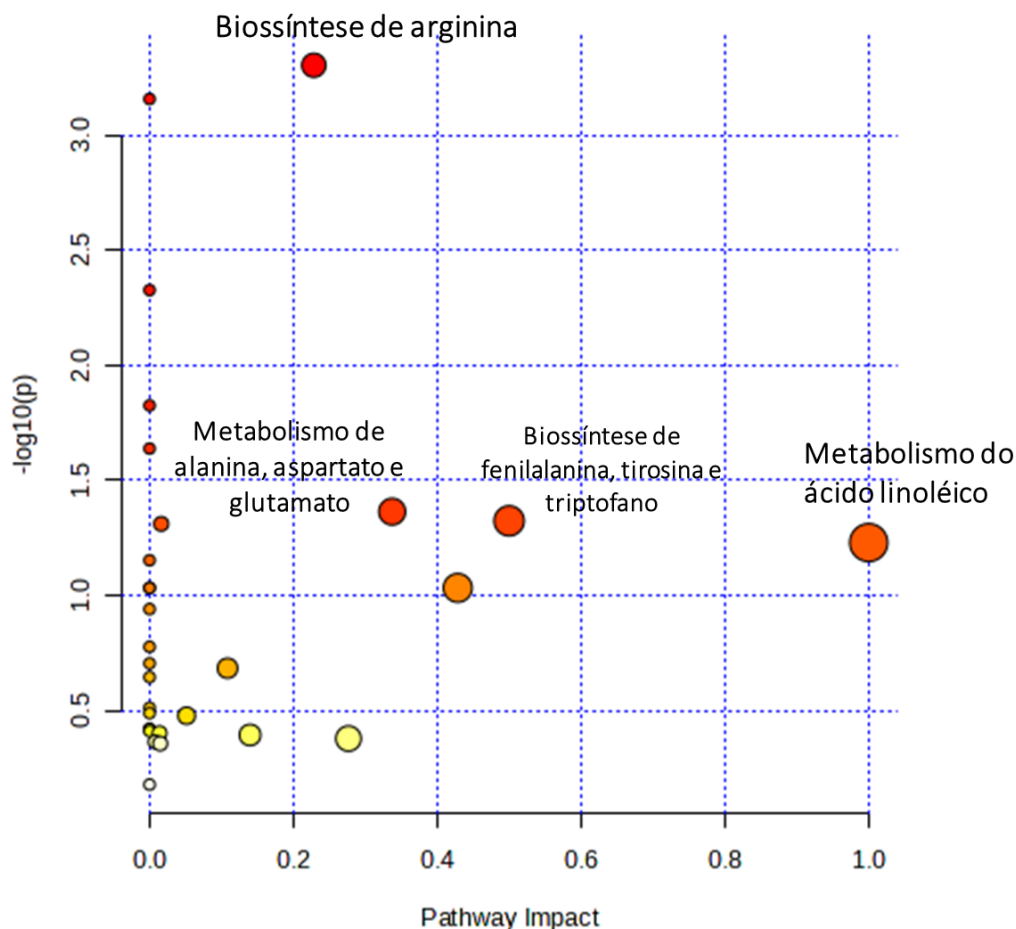


Figura 27- Mapeamento das vias metabólicas impactadas entre os grupos DW e GC em amostras de plasma e urina. O eixo x representa o impacto da via metabólica, enquanto o eixo y representa o $-\log(p\text{-value})$. A cor varia de amarelo claro a vermelho escuro, indicando o nível de significância (p) do enriquecimento e o tamanho do círculo reflete o nível da taxa de enriquecimento (impacto). *Software MetaboAnalyst*.

4.8.1. Biossíntese da Arginina

A arginina ou o ácido 2-amino-5-guanidino-pentanóico, é um aminoácido básico semi-essencial, sintetizada a partir do glutamato nas reações do ciclo da uréia. A arginina é considerada a principal transportadora de nitrogênio em humanos e animais, pois apresenta quatro átomos de nitrogênio em sua estrutura química (Figura 28). A arginina é um aminoácido importante para síntese proteica e no metabolismo do nitrogênio no ciclo da uréia.¹⁴⁸

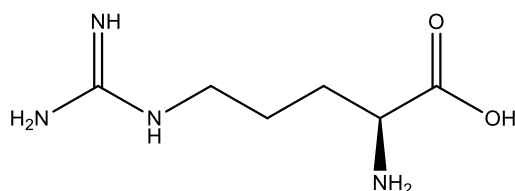


Figura 28- Estrutura química da molécula de Arginina

Os níveis de arginina no organismo são mantidos a partir da dieta (fonte exógena) e através da degradação proteica corporal (fonte endógena), porém em seres humanos adultos a maior fonte deste aminoácido provém da segunda opção, sendo degradada pelo intestino delgado.^{149,150}

A síntese da arginina ocorre via síntese da Δ 1-pirrolina-5-carboxilato (P5C) e da prolina oxidase, com a produção de diversos aminoácidos (Figura 29). Nos seres humanos a arginina, produzida pela degradação de proteínas corporais, encontra-se principalmente no intestino delgado, promovendo a liberação de citrulina no sangue, através da síntese de glutamato, glutamina e ornitina. Na sequência, a citrulina é extraída pelo rim e, conseqüentemente, transformada em arginina pela ação das enzimas argininosuccinato sintase (ASS) e argininosuccinato liase (ASL), no ciclo renal.¹⁴⁹ A arginina é precursora de nitritos, nitratos e do óxido nítrico, sendo este último responsável por diversos fenômenos fisiológicos e fisiopatológicos como a regulação da pressão arterial, transporte de elétrons na mitocôndria, além de auxiliar na síntese de colágeno. Observa-se ainda que a arginina tem o papel de converter a amônia em ureia dentro do organismo, fazendo com que seja eliminada na urina e nas fezes.¹⁵¹

No presente trabalho foram detectados metabólitos relacionados ao ciclo da arginina, sendo esses glutamina (FC= 2,11; $p= 0,00518$), ácido glutâmico ou glutamato (FC= 1,70, $p= 0,00171$), citrulina (FC= 1,18; $p= 0,04487$) fenilacetilglutamina (FC= 1,53; $p = 0,008$) e ácido aspártico ou aspartato (FC= 1,30; $p=1,9E-05$), todos com uma regulação positiva no grupo DW em comparação ao GC.

A arginina presente no corpo humano é proveniente da dieta, da síntese endógena e do metabolismo de proteínas. A síntese de arginina a partir da citrulina ocorre através do sinergismo entre as células epiteliais do intestino delgado e células tubulares proximal do rim.¹⁵² Assim, o rim desempenha um papel importante no metabolismo da arginina, que envolve a síntese e a reabsorção da arginina e a síntese de creatina.¹⁵³ Em adultos saudáveis, o nível endógeno de arginina é suficiente para que esse aminoácido não seja considerado essencial. No entanto, processos

inflamatórios ou condições que envolvam a disfunção do rim ou do intestino podem acarretar uma demanda por arginina.¹⁵⁴⁻¹⁵⁷

Arginina em excesso no organismo pode causar problemas de diarreia, dores de cabeça e dores estomacais, já sua deficiência pode interferir na produção de insulina, ocasionar problemas de fadiga muscular e hipertensão sistêmica quando a síntese de óxido nítrico (NO) se torna deficiente no indivíduo.¹⁵⁸ A arginina pode ainda ser deficiente no organismo humano devido ao catabolismo desregulado da argilase, presente no fígado e no ciclo da ureia, sendo assim podem ocorrer problemas hepáticos, de fenilcetonúria e de hipertensão pulmonar em pacientes com anemia falciforme.¹⁵⁹

Em adultos, a biossíntese da arginina inicia-se com a síntese da citrulina através da glutamina, glutamato, prolina ou ornitina nas mitocôndrias dos enterócitos, sendo então liberada do intestino delgado e captada pelos rins para a produção de arginina.^{151,160} Em revisão de Wu et al¹⁵¹, os autores mostram evidências de que a arginina tem um papel importante na reprodução, desenvolvimento fetal e pós-natal, cicatrização de feridas e integridade de tecidos, síndrome metabólicas, disfunção endotelial, vasculopatia da doença falciforme, injúria renal, fibrose hepática, doenças gastrointestinais, doenças hepáticas e função imunológica.

As principais enzimas envolvidas na síntese da citrulina são a pirrolina-5-carboxilato sintase (P5C), prolina oxidase e N-acetilglutamato (NAG) sintase. Essas três enzimas são expressas em células dos enterócitos de mamíferos, o que faz com que a citrulina seja um biomarcador de insuficiência intestinal em crianças e adultos.^{161,162}

Nas condições de análise empregadas nesse trabalho, a arginina não foi detectada nas amostras em estudo, embora glutamina (FC = 2,13; $p= 0,00518$) e glutamato (FC = 1,70; $p= 0,00171$) tenham se mostrado positivamente reguladas e citrulina (FC = 1,18; $p= 0,04487$) e aspartato (FC = 1,30; $p=1,9E-05$) tenham sido detectados reguladas positivamente, com leve abundância.

A ornitina, produzida pela 5-ornitina aminotransferase, também pode servir como precursor para a síntese de prolina (Figura 29)¹⁵², podendo possibilitar a produção de proteínas ricas em prolina, como o colágeno, o qual representa ~33% das proteínas corpóreas.¹⁶³ Na hidroxilação via modificação pós-traducional da prolina por enzimas prolil-hidroxilases (PH), ocorre a formação da hidroxiprolina (Hyp), um aminoácido não proteínogênico e que em níveis elevados está associada com a

formação de quelóides, vitiligo e processos de cicatrização.¹⁶⁴ Nas mitocôndrias, Hyp é convertido para glioxilato e glicolato e, então, glicina.¹⁶⁵

O catabolismo do colágeno ocorre via prolidase, uma metaloenzima. A prolidase ou peptidase D é uma das únicas enzimas capaz de promover a quebra de ligação C-terminal de prolina e hidroxiprolina nos imidopeptídeos.¹⁶⁶

Alterações na atividade enzimática da prolidase podem ter um papel no desenvolvimento de diversas doenças, como esquizofrenia, transtorno bipolar, depressão, ansiedade, diabetes mellitus tipo 2, hepatite crônica, osteoporose, asma e fibrose hepática.^{167,168}

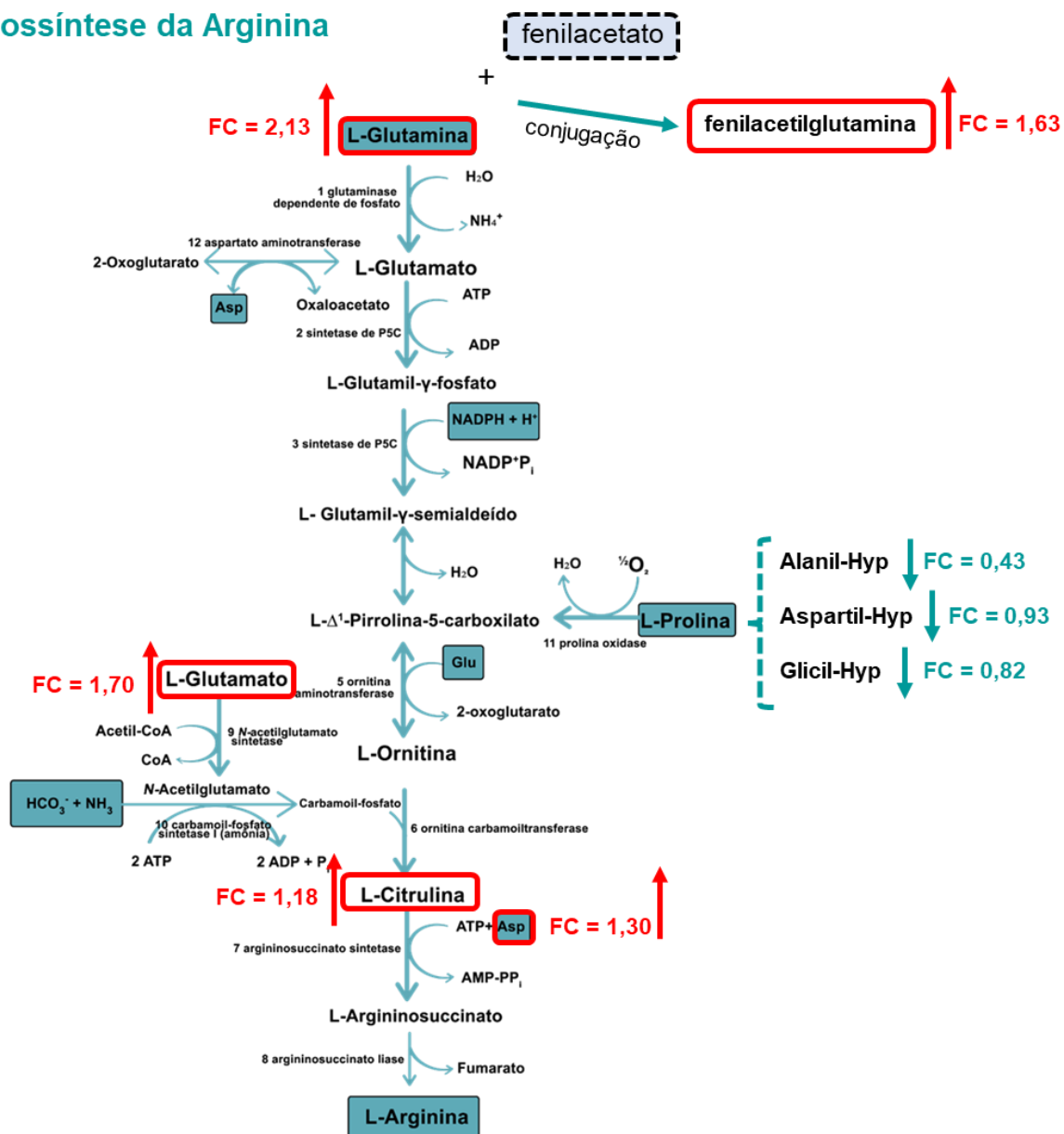
A deficiência em prolidase ou iminopeptiduria caracteriza-se como uma doença genética rara que apresenta elevados níveis de dipeptídeos formados com prolina e Hyp e que são secretados na urina.^{169,170} Em trabalho de Jandke e Spitteller¹⁷¹, os autores reportam a eliminação na urina de dipeptídeos contendo prolina ou hidroxiprolina, associado a problemas no metabolismo e reciclo do colágeno. No presente momento, não foram encontrados na literatura relatos da ocorrência ou influência da prolidase na DW. No entanto, há investigações sobre o uso da prolidase como um biomarcador para diferentes condições fisiopatológicas, uma vez que a atividade da prolidase pode refletir em fibrose hepática e osteoporose, sendo essa última relacionada ao aumento da degradação do colágeno e ainda com processos de stress oxidativos com o reciclo de colágeno.¹⁷²

Na DW, alguns distúrbios neurológicos e hepáticos relacionados com a patologia são também evidenciados em indivíduos diagnosticados com a deficiência em prolidase, o que estimula novos estudos de atividade da prolidase em amostras de pacientes diagnosticados com a DW. No contexto prolidase, destaca-se a detecção no presente trabalho dos dipeptídeos alanil-hidroxiprolina (Ala-Hyp) (FC= 0,43; $p= 5,3E-14$), aspartil-hidroxiprolina (Asp-Gly) (FC= 0,93; $p= 0,00356$) e o glicil-hidroxiprolina (Gly-Hyp) (FC= 0,82; $p= 0,02281$), os quais foram detectadas nas amostras do grupo com DW, no entanto, somente alanil-hidroxiprolina apresentou *fold-chance* < 0,7.

Outra observação com relação aos metabólitos presentes na via da arginina, é o da detecção do metabólito fenilacetilglutamina (FC = 1,53, $p = 0,08473$) (PAG), um produto da conjugação do fenilacetato e da glutamina no fígado e considerado um importante metabólito nitrogenado que se acumula na uremia.¹⁷³ Adicionalmente, o PAG tem sido avaliado como um biomarcador de prognóstico para

doenças cardiovasculares, cerebrovascular e neurológicas.^{174–176} A formação de fenilacetilglutamina consome glutamina e, desta forma, deve-se considerar o quanto esse consumo pode impactar a via de biossíntese de arginina via glutamina.

Biossíntese da Arginina



Adaptado ¹⁴⁹

Figura 29- Via Metabólica da Síntese da Arginina

No grupo DW, houve uma regulação positiva do ácido aspártico (FC = 1,30; $p = 1,90E-05$) e do ácido glutâmico (FC = 1,70; $p = 0,00171$) quando comparado ao grupo GC (Figura 29), metabólitos esses relacionados com rota metabólica da

arginina e também com a via metabólica do metabolismo da alanina, aspartato e glutamato que também se mostrou enriquecida no presente trabalho.

Esses dados estão representados em gráficos *boxplot* com emprego de *z-score* ou pontuação padrão (Figura 30). Os valores *z-score* descrevem a relação entre um valor e a média de um grupo de valores, ou seja, um *z-score* igual a zero, indica que não há diferença entre o valor e a média dos valores. Desta forma, valores > 0 indicam um aumento no nível de metabólito, enquanto valores < 0 indicam uma diminuição e, quando significativos, refletem mudanças biológicas entre os grupos DW e GC, como mostrado na Figura 30.

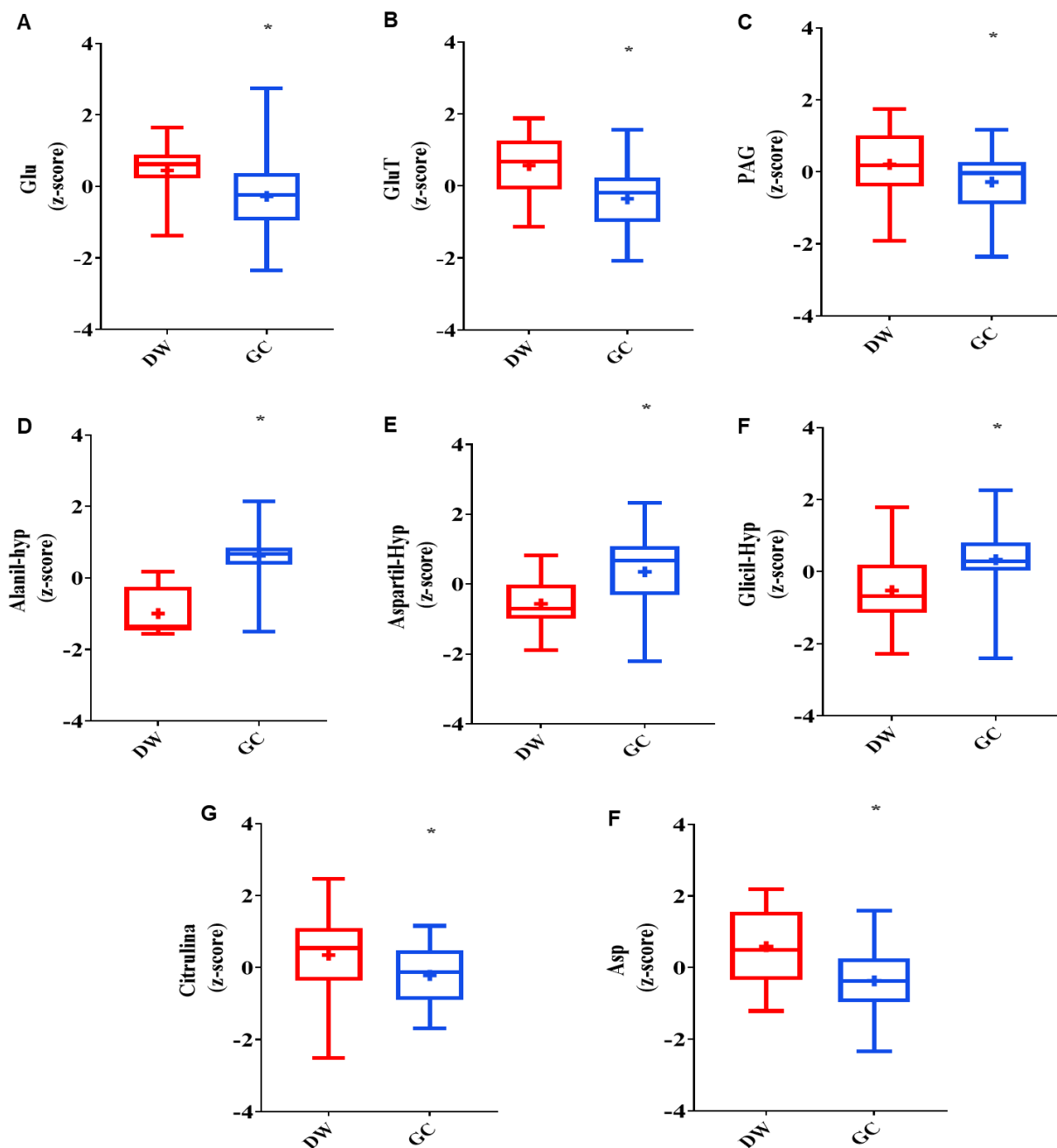
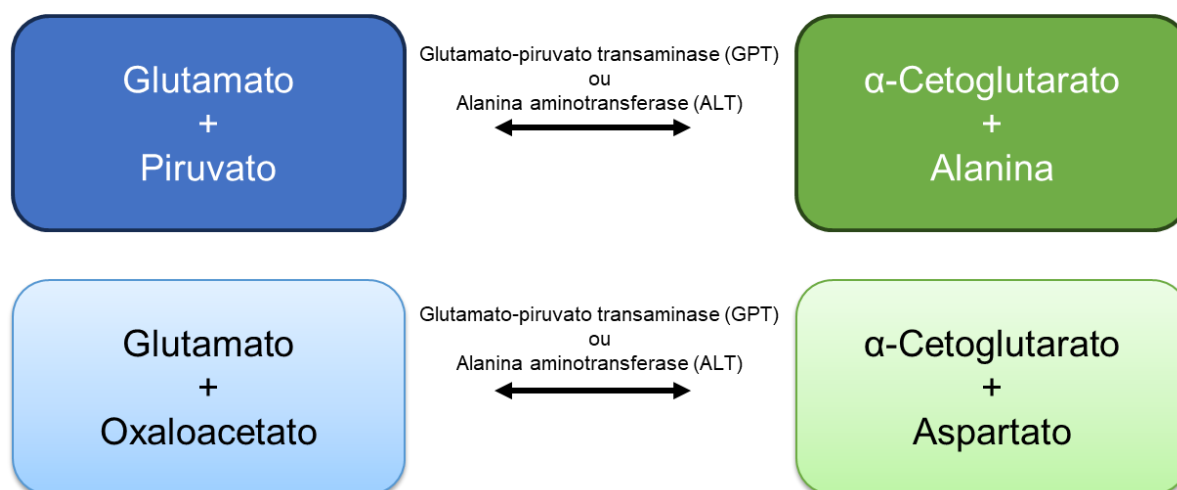


Figura 30- Boxplots (box-and-whisker) para **(A)** Glutamato (Glu); **(B)** Ácido glutâmico (GluT); **(C)** Fenilacetilglutamina (PAG); **(D)** Alanil-Hidroxirolina (Alanil-Hyp); **(E)** Aspartil-Hidroxirolina (Aspartil-Hyp); **(F)** Glicil-Hidroxirolina (Glicil-Hyp); **(G)** Citrulina e **(H)** Ácido aspártico (Asp), em amostras de indivíduos com DW vs. GC. Metabólitos discriminantes identificados com $VIP \geq 1,0$, $p \leq 0,05$ e $FDR < 0,10$.

4.8.2. Metabolismo de Alanina, Aspartato e Glutamato

A alanina (Ala) é conhecida por ser um aminoácido não essencial utilizado na síntese de proteínas e como fonte de energia celular. A Ala é sintetizada através da transaminação do piruvato, com a transferência do grupo amino para o ácido α -cetoglutárico. Da mesma forma, a transaminação do ácido α -cetoglutárico forma aspartato (Figura 31).¹⁷⁷ A Ala é o principal aminoácido gliconeogênico capaz de originar piruvato e glicose pela via inversa, sendo o piruvato um intermediário da via metabólica do ácido cítrico ou ciclo de *Krebs*. A reação de interconversão de alanina-piruvato ocorre via ação da enzima alanina transaminase (ALT), expressa no fígado, a qual possui um papel fundamental na gliconeogênese e no metabolismo de aminoácidos. A ATL é uma das enzimas mais empregadas para identificar a presença de danos hepacelulares, sendo empregada rotineiramente como biomarcador clínico.¹⁷⁸ A ATL também tem sido investigada como biomarcador de problemas cardiovasculares, doenças esteatóticas não alcoólica do fígado (NAFLD; do inglês: *non-alcoholic fatty liver disease*), doenças hepáticas alcoólicas (ALD; do inglês: *alcoholic liver disease*), hepatite C e B, hepatotoxicidade induzida por medicamentos, doenças colestáticas autoimunes e doenças hepáticas metabólicas.¹⁷⁹⁻¹⁸¹

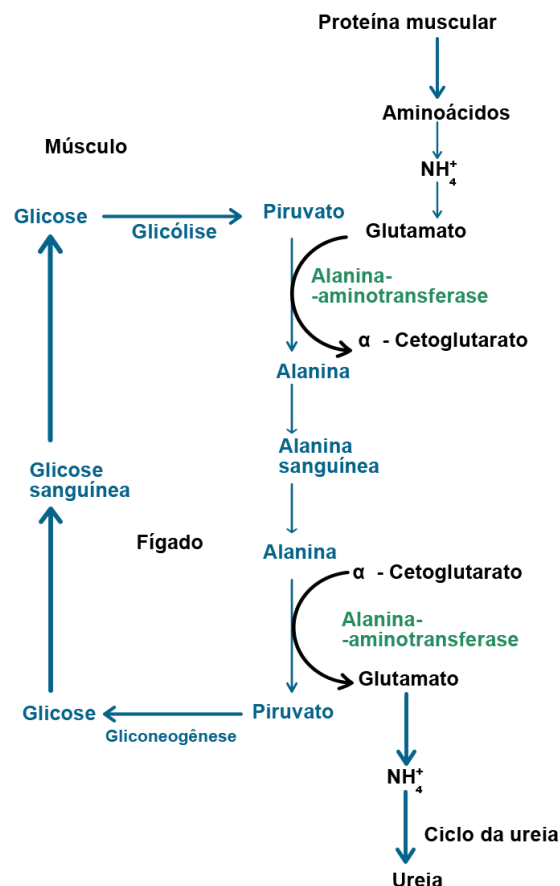


Adaptado de ¹⁷⁷

Figura 31-Reações de transaminação envolvendo o par α -cetoglutarato/glutamato como um receptor/doador de grupo amino.

O ciclo metabólico da glicose-alanina^{182,183} (Figura 32) ilustra a utilização específica do tecido no metabolismo da alanina e a cooperação entre o músculo e o

fígado, para maximizar a entrega de ATP requerida para a contração muscular. A glicose liberada pelo fígado é absorvida pelo músculo, no qual é convertida para piruvato. O glutamato transfere o grupo amino para o piruvato, produzindo alanina e α -cetogluturato. A alanina sintetizada é liberada pelo músculo e absorvida pelo fígado, no qual passa por outra transaminação, formando glutamato e piruvato. Através da gliconeogênese, o piruvato que foi regenerado, forma glicose que retorna ao músculo através do sistema circulatório e é usada como fonte de energia. O glutamato no fígado entra nas mitocôndrias e produz íons amônio através da ação da glutamato desidrogenase (GDH) que, por sua vez, participa como substrato do ciclo da uréia. O piruvato produzido pela Ala também pode ser utilizado no ciclo de Krebs para ser oxidado em outros tecidos. A Figura 32 representa o ciclo glicose-alanina.

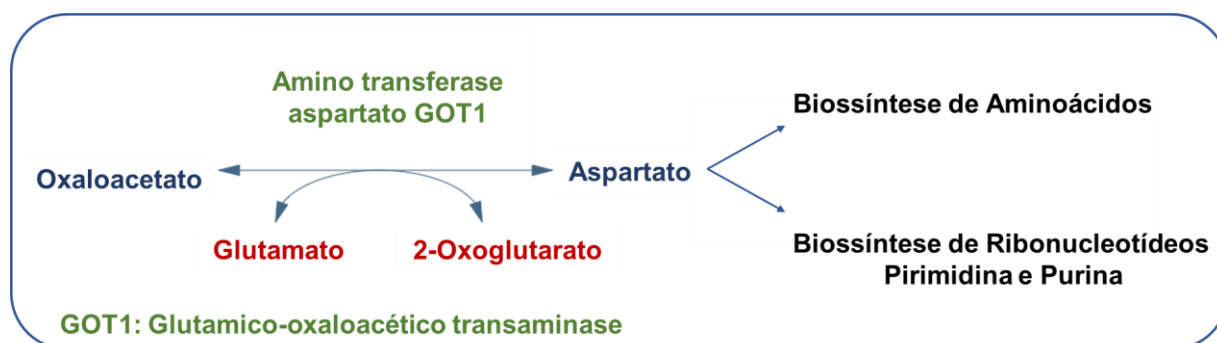


Adaptado ¹⁸³

Figura 32- Ciclo Glicose-Alanina.

Já o aspartato, um aminoácido não essencial, é sintetizado pela transaminação do ácido oxaloacético, intermediário do ciclo de *Krebs*, pela aspartato

aminotransferase ou aminoácido oxidase, que usa o grupo amino do glutamato para a síntese do aspartato. A biossíntese do aspartato está representada na Figura 33.



Adaptado ¹⁸⁴

Figura 33- Biossíntese do Aspartato

A aspartil-tRNA sintetase pode então acoplar o aspartato ao aspartil tRNA para a síntese de proteínas. A asparagina sintetase pode converter o aspartato em asparagina, enquanto o metabolismo de aspartato em arginina faz parte do ciclo da ureia. Nesse caso, o aspartato reage primeiro com a citrulina via argininosuccinato sintase para produzir argininosuccinato, que é posteriormente convertido em arginina e fumarato via argininosuccinato liase (Figura 29).¹⁸⁵

O aspartato também é um precursor de compostos de sinalização celular, como o *N*-acetil-aspartato, β -alanina, adenilsuccinato, arginino-succinato e *N*-carbamoilaspartato. O aspartato é um metabólito de importância no ciclo da ureia e está envolvido na gliconeogênese, na biossíntese de nucleotídeos e diversas outras moléculas que atuam no desenvolvimento do tecido nervoso e neurotransmissores.^{186,187} Além disso, o aspartato carrega os equivalentes de redução (NAD⁺/NADH) no transporte mitocondrial malato-aspartato, que utiliza a interconversão de aspartato e oxaloacetato. O aspartato também atua como um neurotransmissor excitatório no cérebro que ativa os receptores de *N*-metil-*D*-aspartato (NMDA).¹⁸⁸

O aspartato atua em duas vias metabólicas: 1) a síntese da asparagina e 2) a síntese dos aminoácidos essenciais lisina, treonina metionina e isoleucina.

Como anteriormente indicado, no presente trabalho, os resultados indicaram uma regulação positiva do ácido aspártico (FC = 1,30; $p = 1,9E-05$) e do ácido glutâmico (FC = 1,70; $p = 0,00171$) no grupo DW (Figura 29), alterações essas

que são responsáveis pelo impacto na rota metabólica da alanina, aspartato e glutamato.

O metabolismo do ácido aspártico é catalisado pela presença na sua via metabólica da enzima aspartato-aminotransferase (AST, do inglês: *aspartate aminotransferase*). A AST está presente majoritariamente no fígado, músculos cardíacos e esqueléticos e em outros órgãos como pâncreas e rim. Quando apresenta níveis elevados é utilizada para diagnosticar danos ao coração ou fígado.¹⁸¹ O acúmulo da AST está ainda associado a outras doenças como a hepatite B,¹⁸⁹ dengue,¹⁹⁰ Covid-19,¹⁹¹ herpes simplex,¹⁹² e o uso excessivo de medicamentos e abuso de álcool.^{193,194} Adicionalmente, Holecek (2023)¹⁸⁷ reporta uma série de doenças genéticas raras nas quais o metabolismo do aspartato é significativamente perturbado, sendo essas: ARALAR/AGC1 (transportador mitocondrial de aspartato-glutamato 1); Deficiência de AGC2 (Citrina); Deficiência de Asparagina Sintetase; Doença de Canavan; Citrulinemia Tipo 1 e Aminoacidúria Dicarboxílica.

Em pacientes diagnosticados com DW, os níveis séricos de AST apresentam ser moderadamente elevados (<5 vezes que o limite superior normal). O uso de parâmetros bioquímicos de AST como teste primário para o diagnóstico em estágio inicial da DW em pacientes assintomáticos, mas ainda com alterações bioquímicas que sejam decorrentes da intoxicação por Cu, tem sido proposto.¹⁹⁵

A alteração do ácido glutâmico em indivíduos diagnosticados com a DW é relatada por Sarode et al (2019).¹⁹⁶ Os autores relatam o aumento do glutamato em decorrência do estresse oxidativo induzido pelo Cu, sendo este responsável por inibir a ação da enzima α -cetoglutarato desidrogenase, responsável pela conversão de α -cetoglutarato a succinil-CoA. Com isso, há um aumento dos níveis de α -cetoglutarato, com consequente aumento dos níveis de ácido glutâmico. O aumento dos níveis de glutamato também foi reportado por Kalita et al (2014),¹⁹⁷ sendo os níveis mais altos observados em pacientes diagnosticados com DW e sintomas neurológicos e em pacientes sem tratamento, do que em pacientes assintomáticos e pacientes sem tratamento.

4.8.3. Biossíntese de fenilalanina, tirosina e triptofano

A fenilalanina, tirosina e triptofano são aminoácidos aromáticos (Figura 34) e estão envolvidos na síntese de proteínas e em uma variedade de síntese de metabólitos secundários em diversas outras etapas de biossíntese que se relacionem com o metabolismo de aminoácidos e na regulação hormonal via síntese de hormônios da tireóide.¹⁹⁸

A fenilalanina contém um grupo fenil, a tirosina contém um grupo 4-hidroxifenil, o que faz dela tanto um aminoácido aromático como um hidroxil aminoácido, enquanto o triptofano é um aminoácido aromático devido ao seu anel indol heterocíclico (Figura 34). Os papéis estruturais e catalíticos dos aminoácidos aromáticos nas proteínas são bem conhecidos. Por exemplo, os anéis aromáticos estabilizam as estruturas polipeptídicas através de efeitos de interação π (π), participam na catálise ácido-base como parte das tríades catalíticas e estão envolvidos na estabilização de carga e na retransmissão de elétrons no decorrer das reações de transferência de elétrons.^{199,200}

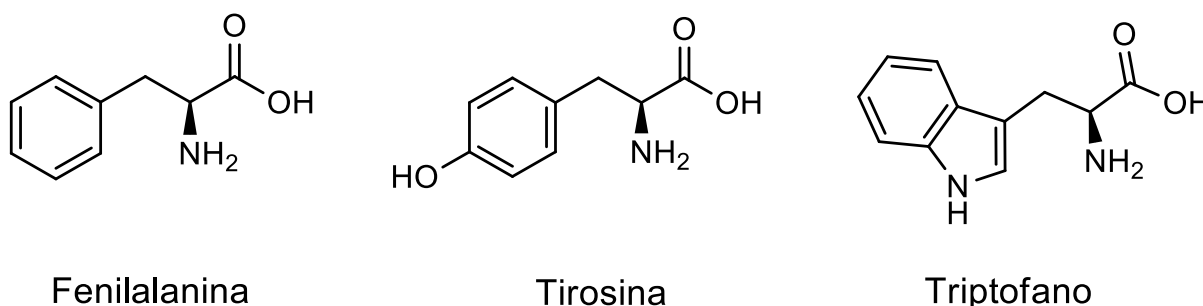
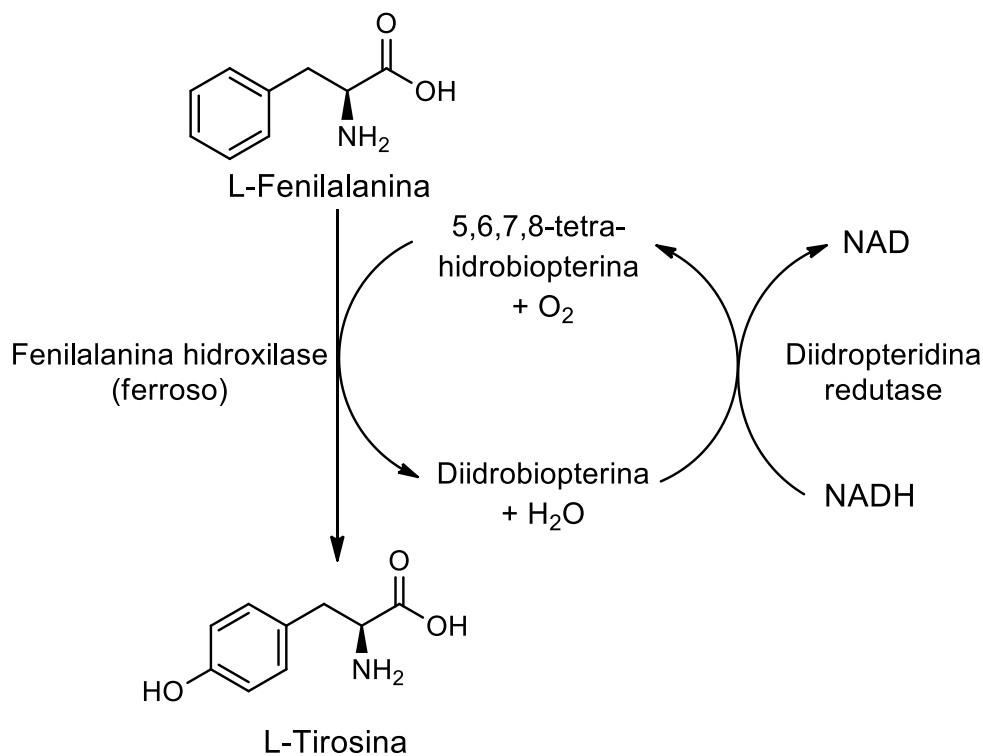


Figura 34- Estrutura química da fenilalanina, tirosina e triptofano.

A fenilalanina, um aminoácido essencial, é necessária para a produção do aminoácido não essencial tirosina (Figura 35), sendo essa conversão catalisada pela enzima fenilalanina hidroxilase (PAH; do inglês: *phenylalanine hydroxylase*), que funciona principalmente no fígado, mas demonstrou ser ativa em outros tecidos do corpo, incluindo rim, pâncreas e cérebro.²⁰¹ A deficiência na enzima PAH resulta na inabilidade de converter fenilalanina em tirosina, ocasionando hiperfenilalaninemia, também conhecida como fenilcetonúria (PKU; do inglês: *phenylketonuria*), sendo caracterizada como altos níveis de fenilalanina e baixos níveis de tirosina no sangue.²⁰² A PKU é caracterizada por comprometimento do desenvolvimento

intelectual, diversos sintomas psiquiátricos, neurológicos, eczema e defeitos na pigmentação da pele.²⁰³

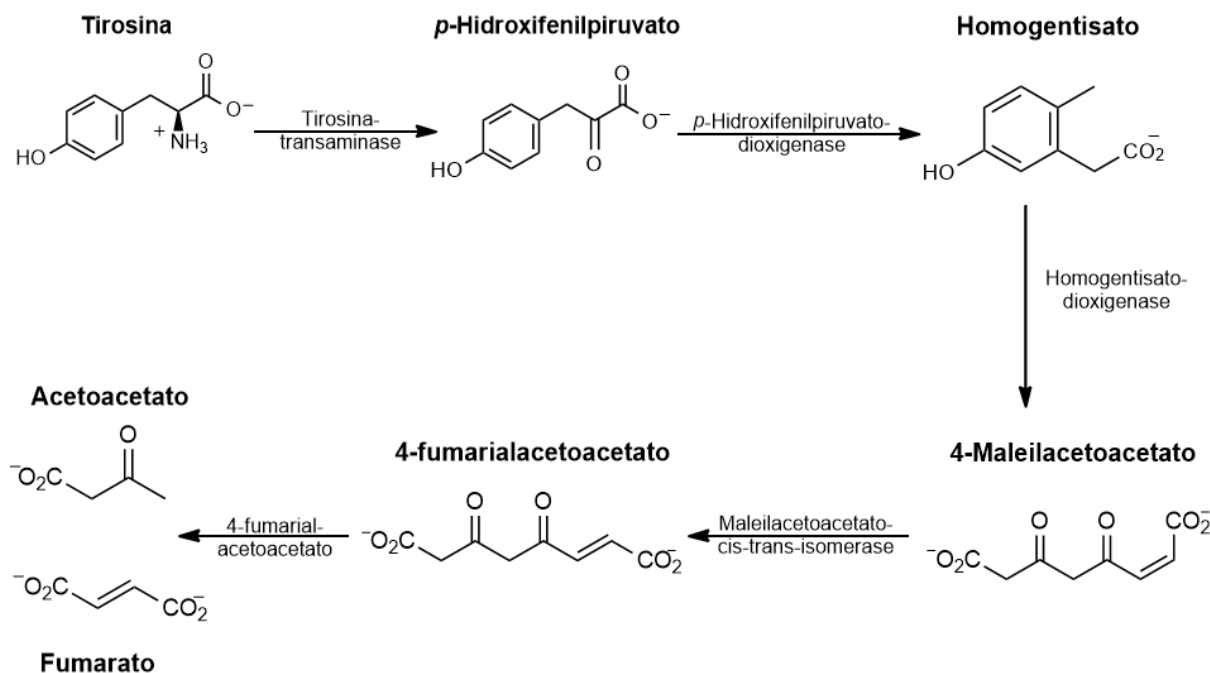
A fenilalanina é usado como fonte de energia e a sua oxidação completa requer biopterina, ascorbato, tiamina, riboflavina, niacina, vitamina B6, pantotenato, lipoato, ubiquinona, magnésio e ferro.²⁰⁴



Adaptado²⁰⁴

Figura 35- Conversão de fenilalanina em tirosina

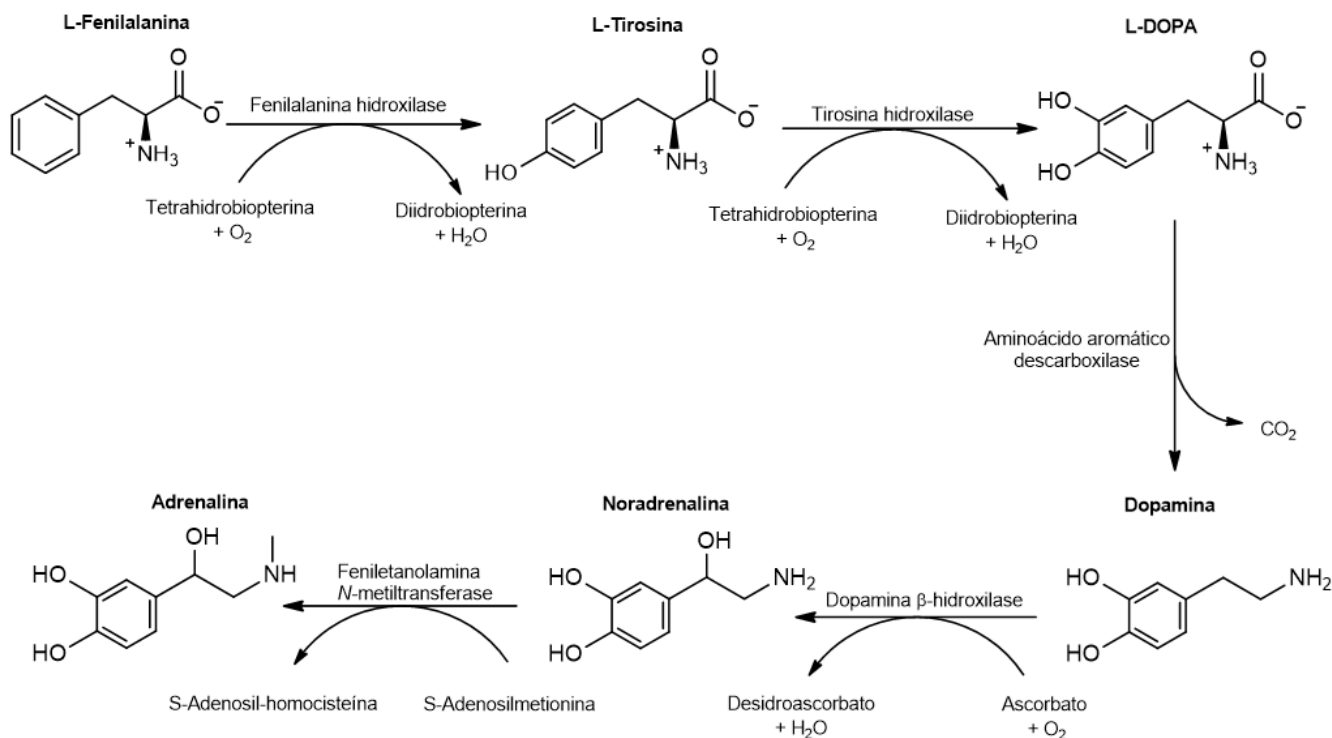
O metabolismo da fenilalanina a partir da tirosina (Figura 36) produz fumarato e acetoacetato, espécies intermediárias do ciclo de *Krebs*.



Adaptado²⁰⁴

Figura 36- Reações do catabolismo da tirosina, com produção de fumarato e acetacetato.

A tirosina é classificada como semi-essencial, já que é sintetizada a partir da fenilalanina. A tirosina é precursora de vários neurotransmissores no sistema nervoso central e periférico, incluindo L-DOPA, dopamina, norepinefrina, epinefrina, noradrenalina e adrenalina, além de hormônios como a tiroxina e a melanina.²⁰⁵ A tirosina é convertida em fenol e piruvato através da ação da enzima tirosina fenol-liase (β -tirosinase). Os compostos fenólicos são potenciais biomarcadores de câncer gastroesofágico.²⁰⁶

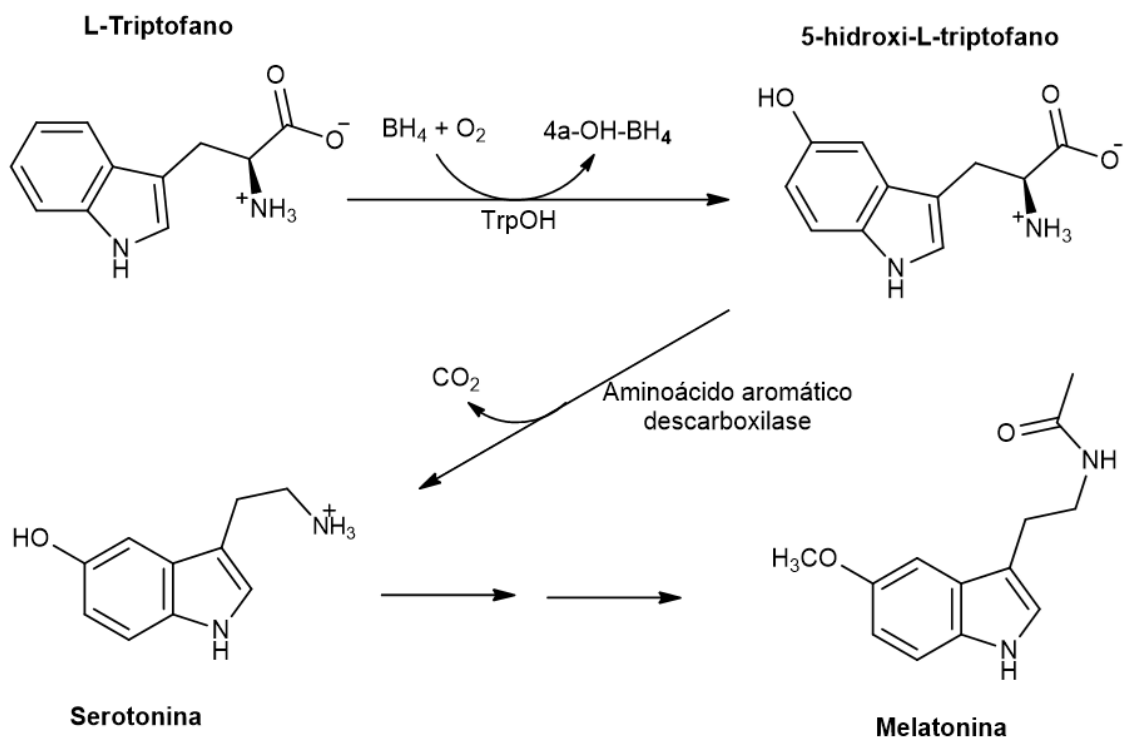


Adaptado²⁰⁵

Figura 37- Conversão de fenilalanina e tirosina em catecolaminas neurotransmissoras.

O triptofano é o único aminoácido proteico que possui um anel indol. O triptofano é também o precursor de diversas moléculas biologicamente ativas, incluindo a serotonina, um neurotransmissor serotoninérgico, e a melatonina (Figura 38).²⁰⁷ Muitos trabalhos sugerem que alterações nos níveis da serotonina desempenham papéis importantes na saúde mental de indivíduos. Os níveis alterados de atividade serotoninérgica estão associados a condições psiquiátricas envolvendo síndromes depressivas, transtorno bipolar, ansiedade, esquizofrenia, desordens compulsivas-obsessivas e déficit de atenção.^{208,209}

Em humanos, o triptofano é metabolizado por duas vias principais, a via da serotonina (Figura 38) e a via da quinurenina (Figura 39), levando à geração de compostos biologicamente ativos, como serotonina, melatonina, ácido quinurênico (KU), ácido picolínico e niacina (ácido nicotínico).²¹⁰



Adaptado ^{208,209}

Figura 38- Metabolismo do Triptofano via Biossíntese da Melatonina e Serotonina



Adaptado²¹⁰

Figura 39- Metabolismo do Triptofano via Biossíntese da Quinurenina

No presente trabalho, observou-se uma manutenção dos níveis da tirosina (FC = 0,849; $p = 0,00635$) no grupo DW quando comparado ao grupo GC e um regulação negativa do ácido quinurênico (KU) (FC = 0,69; $p = 0,00021$) (Figura 40). Até o presente momento, não foram encontrados na literatura trabalhos reportando associações do ácido quinurênico com a DW. No entanto, a via de biossíntese da quinurenina é uma rota importante e está vinculada a processos fisiológicos envolvendo respostas autoimunes, processos inflamatórios, doenças neurológicas, câncer, doenças neurodegenerativas e hormônios.²¹¹

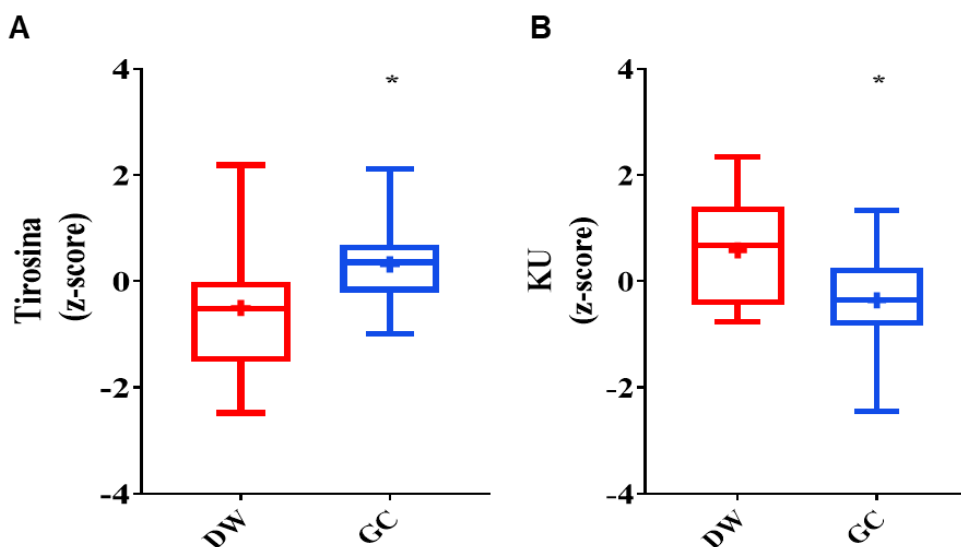


Figura 40- *Boxplots (box-and-whisker)* para **(A)** Tiroosina (Glu); **(B)** Ácido quinurênico (KU) em amostras de indivíduos com DW vs. GC. Metabólitos discriminantes identificados com $VIP \geq 1,0$, $p \leq 0,05$ e $FDR < 0,10$.

As hidroxilases de aminoácidos aromáticos tirosina hidroxilase (TH), triptofano hidroxilase 1 (TPH1) e triptofano hidroxilase 2 (TPH2) são as enzimas responsáveis por catalisar as etapas de biossíntese da tirosina e do triptofano. Desta forma, essas enzimas estão se destacando como alvos terapêuticos relacionados à processos fisiológicos de função cerebral, função cardiovascular, energia, homeostase óssea, motilidade gastrointestinal, hemostasia e função hepática.²¹²

4.8.4. Metabolismo do Ácido Linoleico

O ácido linoleico (ômega 6; 18:3n-3) é um ácido graxo poliinsaturado (PUFA; do inglês: *polyunsaturated fatty acid*) presente na alimentação dos seres humanos e relevante para os processos biológicos, auxiliando nos processos metabólicos e nas estruturas das membranas celulares. O ácido linoleico apresenta na sua estrutura química duas ligações duplas e dezoito carbonos (Figura 41), sendo considerado um ácido graxo essencial (AGE), pois não é sintetizado naturalmente pelo organismo do ser humano.²¹³

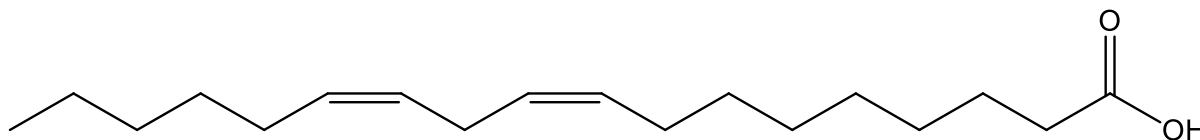
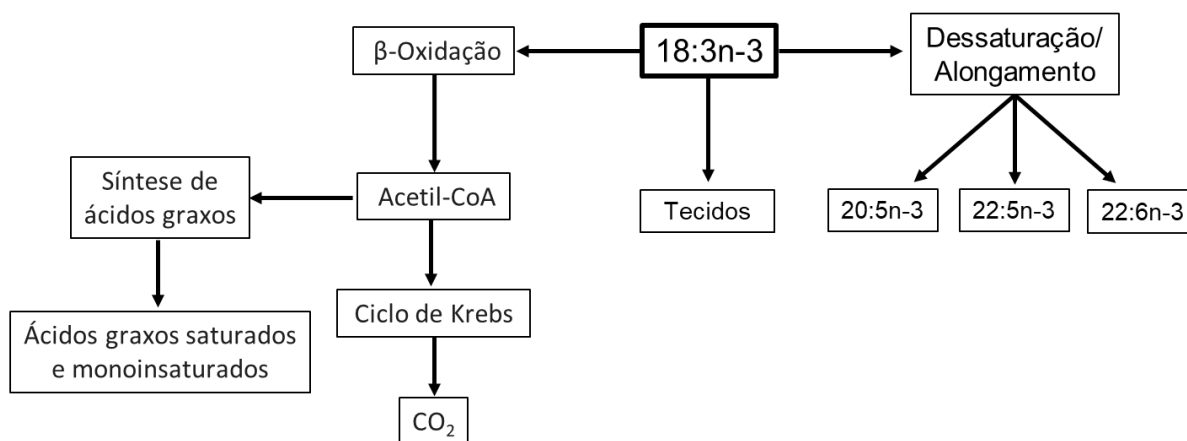


Figura 41– Estrutura química do ácido linoleico

O ácido linoleico é um precursor dos ácidos graxos poliinsaturados de cadeia longa. Assim, para que os ácidos de cadeias longas sejam formados é necessário que ocorra o metabolismo, especialmente no retículo endoplasmático do fígado, com o processo de alongamento da cadeia carbônica do ômega 6 e posterior dessaturação (Figura 42) (Figura 42). O metabolismo do ácido linoleico propicia ainda incorporação em reservatórios estruturais que sejam de transporte ou de armazenamento e reações de β -oxidação.²¹³ No metabolismo do ácido linoleico ocorre a formação de ácido araquidônico, precursor na biossíntese de prostaglandinas, tromboxanos, leucotrienos e ácidos hidroieicosatetraenólicos (HETEs), metabólitos esses atribuídos ao grupo das oxilipinas pró-inflamatórias.^{214,215} A Figura 42 ilustra o processo de metabolismo do ácido linoleico (18:3n-3) em humanos.



Adaptado ²¹³

Figura 42- Metabolismo do Ácido Linoléico (18:3n-3).

Com referência ao metabolismo de ácido linoleico, no presente trabalho foram detectados com regulação positiva os metabólitos: ácido linoleico ($F = 1,59$; $p = 0,016$); ácido δ -linoleico ($FC=1,84$; $p = 0,012$) e o ácido araquidônico ($FC = 1,77$; $p = 0,03$) no grupo DW (Figura 43).

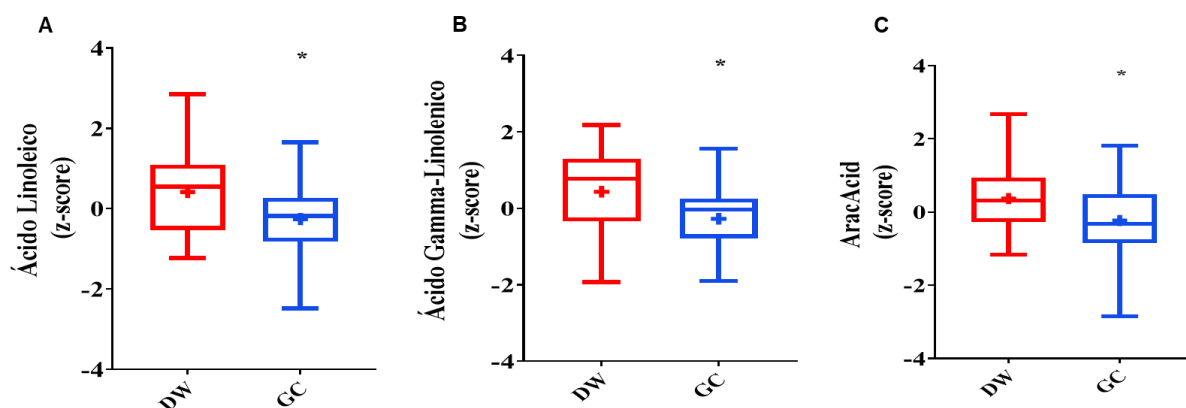


Figura 43- Boxplots (*box-and-whisker*) para **(A)** Ácido linoleico; **(B)** Ácido δ -linoleico; **(C)** Ácido araquidônico em amostras de indivíduos com DW vs. GC. Metabólitos discriminantes identificados com $VIP \geq 1,0$, $p \leq 0,05$ e $FDR < 0,10$.

Na literatura, há relatos de que o acúmulo ou a deficiência de Cu promovem modificações no metabolismo lipídico. Na DW, o Cu acumula-se principalmente no fígado devido à deficiência da ATP7B e o metabolismo lipídico é desregulado. Mazi *et al* (2020)²¹⁶ reportam que o Cu pode prejudicar as vias metabólicas através da sua ligação direta às estruturas proteicas e lipídicas ou pela geração de espécies reativas de oxigênio (ROS; do inglês: *reactive oxygen species*) que promovem danos às organelas celulares vitais para o metabolismo energético. No fígado, a sobrecarga de Cu resulta em comprometimento mitocondrial, regulação negativa do metabolismo lipídico e desenvolvimento de esteatose.

Não foram encontrados dados reportados do impacto do Cu no metabolismo específico do ácido linoleico, mas particularmente na síntese de colesterol, a qual foi regulada negativamente, indicando que a desregulação na síntese de lipídeos é um fator impactante na dislipidemia (níveis elevados de colesterol e/ou triglicerídeos) e/ou esteatose hepática (doença hepática gordurosa).²¹⁷

5. Conclusão

Esse estudo é classificado como um estudo piloto devido à quantidade limitada de amostras de pacientes com uma doença rara. Utilizando-se de uma abordagem metabolômica global nas análises das amostras de plasma e urina de pacientes com DW, foram realizadas comparações entre o grupo estudo (DW) com um grupo controle, o que possibilitou avaliar as alterações metabólicas provenientes da DW.

As análises por LC-HRMS nos modos de separação RPLC e HILIC de dois diferentes biofluidos foram empregadas, resultando em descobertas complementares e possibilitando resultados de uma análise metabolômica global, com características abrangentes.

A utilização de ferramentas estatísticas foi de extrema importância para a condução deste trabalho. As análises multivariadas, incluindo PLS-DA e OPLS-DA, permitiram avaliar a capacidade preditiva dos modelos desenvolvidos, os quais demonstraram resultados satisfatórios, dentro dos parâmetros de aceitação. Além disso, as análises univariadas foram aplicadas para examinar a proporção de falsos positivos nos dados obtidos e destacou os *features* moleculares verdadeiramente significativos para esse estudo.

Uma análise das vias metabólicas indicou que os metabólitos discriminatórios podem ter origem em diversas vias alteradas na Doença de Wilson (DW), abrangendo a biossíntese da arginina, o metabolismo do ácido linoleico, o metabolismo de alanina, aspartato e glutamato, e a biossíntese de fenilalanina, tirosina e triptofano. Algumas dessas vias já foram mencionadas na literatura, como citado nos resultados, enquanto outras representam novas descobertas desse estudo. Dentre os metabólitos identificados como impactantes nas vias metabólicas, destacando-se a glutamina, o glutamato, a citrulina, a fenilacetilglutamina, o ácido aspártico, os dipeptídeos alanil-hidroxiprolina, aspartil-hidroxiprolina e glicil-hidroxiprolina, a tirosina, o ácido quinurênico, o ácido linoleico, o ácido δ -linoleico e o ácido araquidônico.

Os resultados apresentados, embora provenientes de um estudo piloto, demonstraram ser promissores, fornecendo informações valiosas sobre uma doença que é não apenas rara, mas também carece de relatos na literatura. Isso ressalta a utilidade dos metabólitos aqui identificados no estudo e identificação de novos e

potenciais biomarcadores que possibilitem um diagnóstico rápido, preciso e futuras aplicações em medicina translacional. Os metabólitos descritos nesse trabalho destacam a importância de explorar as vias metabólicas relacionadas com aspectos bioquímicos da DW para melhorar os diagnósticos precoces da doença. No entanto, é fundamental conduzir estudos mais amplos com amostras de pacientes e validar os resultados através de análises quantitativas para determinar a eficácia desses possíveis marcadores bioquímicos para o diagnóstico clínico e também no acompanhamento terapêutico dos pacientes.

6. Limitações da pesquisa

A pandemia da Covid-19 afetou vários projetos de pesquisa, e infelizmente este não foi exceção. Isso incluiu a coleta de amostras e o tamanho amostral dos grupos em estudo (paciente e controle). O atraso na coleta de amostras, assim como a quantidade limitada de pacientes que se enquadram nos critérios de inclusão, resultou em um tamanho amostral menor do que o inicialmente planejado, o que pode ter limitado a capacidade técnica em realizar um estudo mais abrangente sobre a Doença de Wilson.

Além disso, a falta de acesso aos dados clínicos dos pacientes participantes com Doença de Wilson, como exames bioquímicos, comorbidades associadas e sintomas no momento do diagnóstico e/ou coleta das amostras, representa uma limitação significativa para a interpretação dos resultados. Essas informações são essenciais para uma análise completa e precisa dos resultados da pesquisa.

O menor tempo para a execução experimental da pesquisa, o número de amostras (n) reduzido e a falta dos dados clínicos geraram um impacto negativo na interpretação dos resultados e é importante reconhecer e destacar essas limitações. No entanto, o estudo considerou alternativas de mitigar esse impacto, com emprego de estratégias estatísticas robustas e o emprego de duas diferentes matrizes que oferecessem complementariedade e abrangência.

7. Perspectivas Futuras

Como continuidade deste projeto, é crucial conduzir um estudo de metabolômica global com uma ampliação no número de pacientes (n), incluindo grupos de pacientes que não estejam em tratamento (selecionados durante o diagnóstico) e um grupo em tratamento para garantir uma representação mais abrangente dos impactos bioquímicos do tratamento e da DW, quando comparados com um grupo de indivíduos saudáveis. Da mesma forma, a ampliação do número de indivíduos do grupo controle e a inclusão de heterogeneidade (raça, idade e gênero), tanto no grupo estudo como no grupo controle, os quais são fundamentais para avaliar a variabilidade bioquímica com a presença de covariáveis.

Como parte fundamental da validação dos dados, é necessário realizar análises metabolômicas alvo com o emprego de padrões analíticos dos metabólitos que se mostraram significativos na etapa de metabolômica global, a fim de obter dados de quantificação absoluta do conjunto dos metabólitos impactantes e validar os candidatos a biomarcadores. Nessa fase é importante considerar o emprego do mesmo coorte de amostras para evitar variabilidade analítica. No entanto, adiciona-se aqui o fato de realizar a separação de diferentes alíquotas das amostras para serem empregadas nas análises metabolômicas global e alvo, mantendo constante os ciclos de gelo/degelo.

Como perspectiva a longo prazo, espera-se que após a validação dos metabólitos, os selecionados possam ser empregados em validações pré-clínicas com emprego de modelos animais e que, posteriormente, possam prosseguir para estudos de medicina translacional, com estudos clínicos em humanos. Essas etapas permitem o estudo e conclusão sobre o emprego das descobertas como biomarcadores clínicos no diagnóstico precoce e preciso da Doença de Wilson, evitando a biópsia hepática e testes genéticos não acessíveis. Assim como, o emprego como biomarcadores de monitoramento seja do progresso da doença ou da eficácia do tratamento clínico.

8. Referências bibliográficas

1. "https://ods.od.nih.gov/factsheets/Copper-HealthProfessional/. Acesso 11Janeiro2024.".
2. CHAMBERS, A. *ET AL.* "An exposure-response curve for copper excess and deficiency". *J Toxicol Environ Health B Crit Rev*, **13**, 546–578, 2010.
3. TRUMBO, P., YATES, A. A., SCHLICKER, S. & POOS, M. "Dietary Reference Intakes - Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc". *J Am Diet Assoc*, **101**, 294–301, 2001.
4. WIJMENGA, C. & KLOMP, L. W. J. "Molecular regulation of copper excretion in the liver". *Proceedings of the Nutrition Society*, **63**, 31–39, 2004.
5. CHEN, J. *ET AL.* "The molecular mechanisms of copper metabolism and its roles in human diseases". *Pflugers Archiv European Journal of Physiology* vol. 472 1415–1429, Preprint at <https://doi.org/10.1007/s00424-020-02412-2> (2020).
6. GUBLER, C. J. "COPPER METABOLISM IN MAN". *J Am Med Assoc*, **161**, 530–535, 1956.
7. BOST, M. *ET AL.* "Dietary copper and human health: Current evidence and unresolved issues". *Journal of Trace Elements in Medicine and Biology* vol. 35 107–115, Preprint at <https://doi.org/10.1016/j.jtemb.2016.02.006> (2016).
8. RUIZ, L. M., LIBEDINSKY, A. & ELORZA, A. A. "Role of Copper on Mitochondrial Function and Metabolism". *Frontiers in Molecular Biosciences* vol. 8 Preprint at <https://doi.org/10.3389/fmolb.2021.711227> (2021).
9. XUE, Q. *ET AL.* "Copper metabolism in cell death and autophagy". *Autophagy* vol. 19 2175–2195, Preprint at <https://doi.org/10.1080/15548627.2023.2200554> (2023).
10. AN, Y. *ET AL.* "The Role of Copper Homeostasis in Brain Disease". *Int J Mol Sci*, **23**, 13850, 2022.
11. XIAOLONG, T. *ET AL.* "Copper in cancer: from limiting nutrient to therapeutic target". *Front Oncol*, **13**, 2023.
12. XIE, H. & KANG, Y. J. "Role of Copper in Angiogenesis and Its Medicinal Implications". *Curr Med Chem*, **16**, 1304–1314, 2009.
13. XIAO, L., YANG, C., GU, W., LIU, R. & CHEN, D. "Associations between serum copper, zinc, selenium level and sex hormones among 6–19 years old children and adolescents in NHANES 2013–2016". *Front Endocrinol (Lausanne)*, **13**, 924338, 2022.

14. GARZA, N., SWAMINATHAN, A., MAREMANDA, K., ZULKIFLI, M. & GOHIL, V. "Mitochondrial copper in human genetic disorders". *Trends in Endocrinology & Metabolism*, **34**, 2022.
15. ALTOBELLI, G., NOORDEN, S., BALATO, A. & CIMINI, V. "Copper/Zinc Superoxide Dismutase in Human Skin: Current Knowledge". *Front Med (Lausanne)*, **7**, 2020.
16. CUI, X. *ET AL.* "The Molecular Mechanisms of Defective Copper Metabolism in Diabetic Cardiomyopathy". *Oxid Med Cell Longev*, **2022**, 1–16, 2022.
17. CHEN, J., SONG, W. & ZHANG, W. "The emerging role of copper in depression". *Frontiers in Neuroscience* vol. 17 Preprint at <https://doi.org/10.3389/fnins.2023.1230404> (2023).
18. CHEN, L., MIN, J. & WANG, F. "Copper homeostasis and cuproptosis in health and disease". *Signal Transduction and Targeted Therapy* vol. 7 Preprint at <https://doi.org/10.1038/s41392-022-01229-y> (2022).
19. JOMOVA, K. *ET AL.* "Reactive oxygen species, toxicity, oxidative stress, and antioxidants: chronic diseases and aging". *Arch Toxicol*, **97**, 1–76, 2023.
20. TSVETKOV, P. *ET AL.* "Copper induces cell death by targeting lipoylated TCA cycle proteins". *Science*, **375**, 1254–1261, 2022.
21. CHANG, I. J. & HAHN, S. H. "The genetics of Wilson disease". in *Handbook of Clinical Neurology* vol. 142 19–34, (Elsevier B.V., 2017).
22. WANG, J. *ET AL.* "Identification of mutations in the ATP7B gene in 14 Wilson disease children: Case series". *Medicine (United States)*, **100**, E25463, 2021.
23. CHAUDHRY HS, A. AC. "Wilson Disease. [Updated 2023 Aug 7]". In: *StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK441990/>.*
24. KIM, J. W. *ET AL.* "Genetically confirmed wilson disease in a 9-month old boy with elevations of aminotransferases". *World J Hepatol*, **5**, 156–159, 2013.
25. CAO, C. *ET AL.* "A Rare Case of Wilson Disease in a 72-Year-Old Patient". *ACG Case Rep J*, **6**, e00024–e00024, 2019.
26. WANG, Y. *ET AL.* "Clinical and genetic characterization of pediatric patients with Wilson's disease from Yunnan province where ethnic minorities gather". *Front Genet*, **14**, 2023.
27. CAI, L. *ET AL.* "Role of gender and age in features of Wilson's disease". *Front Neurol*, **14**, 2023.
28. REIS BARBOSA, E. *ET AL.* *Wilson's Disease A case report and a historical review. Arq Neuropsiquiatr* vol. 67 (2009).

29. POLISHCHUK, R. S. "Cellular function of ATP7B (Wilson ATPase)". in *Clinical and Translational Perspectives on WILSON DISEASE* 45–56, (Elsevier, 2018). doi:10.1016/B978-0-12-810532-0.00006-9.
30. POLISHCHUK, R. S. "Cellular function of ATP7B (Wilson ATPase)". in *Clinical and Translational Perspectives on WILSON DISEASE* 45–56, (Elsevier, 2018). doi:10.1016/B978-0-12-810532-0.00006-9.
31. KASZTELAN-SZCZERBINSKA, B. & CICHOZ-LACH, H. "Wilson's disease: An update on the diagnostic workup and management". *Journal of Clinical Medicine* vol. 10 Preprint at <https://doi.org/10.3390/jcm10215097> (2021).
32. SCHILSKY, M. L. ET AL. "A Multidisciplinary Approach to the Diagnosis and Management of Wilson Disease: 2022 Practice Guidance on Wilson Disease from the American Association for the Study of Liver Diseases". *Hepatology*, 2022 doi:10.1002/hep.32801.
33. DZIEŻYC-JAWORSKA, K., LITWIN, T. & CZŁONKOWSKA, A. "Clinical manifestations of Wilson disease in organs other than the liver and brain". *Ann Transl Med*, **7**, S62–S62, 2019.
34. SCHILSKY, M. L. & MISTRY, P. K. "Chapter 42 - Wilson Disease and the Kidney". in *Genetic Diseases of the Kidney* (eds. Lifton, R. P., Somlo, S., Giebisch, G. H. & Seldin, D. W.) 709–713, (Academic Press, 2009). doi:<https://doi.org/10.1016/B978-0-12-449851-8.00042-5>.
35. SÓCIO, S. DE A. ET AL. "Doença de Wilson em crianças e adolescentes: diagnóstico e tratamento". *Revista Paulista de Pediatria*, **28**, 134–140, 2010.
36. MARTÍNEZ-MORILLO, E. & BAUÇA, J. M. "Biochemical diagnosis of Wilson's disease: An update". *Advances in Laboratory Medicine* vol. 3 103–113, Preprint at <https://doi.org/10.1515/almed-2022-0020> (2022).
37. JOÃO GALIZZI FILHO, OSVALDO FLÁVIO DE MELO COUTO & HUMBERTO OLIVA GALIZZ. "Doença de Wilson". <https://sbhepatologia.org.br/pdf/13.pdf> Acesso:25Janeiro2024.
38. FERENCI, P. ET AL. "EASL Clinical Practice Guidelines: Wilson's disease". *J Hepatol*, **56**, 671–685, 2012.
39. SOCHA, P. ET AL. "Wilson's Disease in Children: A position paper by the European Society for Paediatric Gastroenterology, Hepatology and Nutrition Committee". *J Pediatr Gastroenterol Nutr*, **66**, 1, 2017.
40. PALUMBO, C. S. & SCHILSKY, M. L. "Clinical practice guidelines in Wilson disease". *Ann Transl Med*, **7**, S65–S65, 2019.

41. QIU, Y., SU, M., XIAO, X., ZHOU, D. & XIE, L. "Metabolomic profiling of Wilson disease, an inherited disorder of copper metabolism, and diseases with similar symptoms but normal copper metabolism". 2023 doi:10.21203/rs.3.rs-2425833/v1.
42. LIU, Z., WANG, M., ZHANG, C., ZHOU, S. & JI, G. "Molecular Functions of Ceruloplasmin in Metabolic Disease Pathology". *Diabetes Metab Syndr Obes*, **Volume 15**, 695–711, 2022.
43. SCHILSKY, M. *ET AL.* "Trientine tetrahydrochloride versus penicillamine for maintenance therapy in Wilson disease (CHELATE): a randomised, open-label, non-inferiority, phase 3 trial". *Lancet Gastroenterol Hepatol*, **7**, 2022.
44. BREWER, G. "Zinc and tetrathiomolybdate for the treatment of Wilson disease". *Small Molecule Therapy for Genetic Disease*, 185–201, 2010 doi:10.1017/CBO9780511777905.014.
45. REIS, B. "Zinc therapy for Wilson's disease: study of 20 cases". *Arq Neuropsiquiatr*, **59**, 2001.
46. AVAN, A. *ET AL.* "The Role of Zinc in the Treatment of Wilson's Disease". *Int J Mol Sci*, **23**, 9316, 2022.
47. DAI, X. & SHEN, L. "Advances and Trends in Omics Technology Development". *Front Med (Lausanne)*, **9**, 2022.
48. MUTHUBHARATHI, B., THIRUMUGAM, G. & BALAMURUGAN, K. "Metabolomics-the small molecules matter large". *Mol Omics*, **17**, 2021.
49. CANUTO, G. A. B. *ET AL.* "METABOLÔMICA: DEFINIÇÕES, ESTADO-DA-ARTE E APLICAÇÕES REPRESENTATIVAS". *Quimica Nova* vol. 41 75–91, Preprint at <https://doi.org/10.21577/0100-4042.20170134> (2018).
50. PSYCHOGIOS, N. *ET AL.* "The Human Serum Metabolome". *PLoS One*, **6**, e16957, 2011.
51. HILLESHEIM, E. & BRENNAN, L. "Metabotyping and its role in nutrition research". *Nutrition Research Reviews* vol. 33 33–42, Preprint at <https://doi.org/10.1017/S0954422419000179> (2020).
52. OLIVER, S., WINSON, M., KELL, D. & BAGANZ, F. "Systematic Functional Analysis of the Yeast Genome". *TIPTECH*, **16**, 2000.
53. COLLINS, F. S. & MCKUSICK, V. A. "Implications of the Human Genome Project for Medical Science". *JAMA*, **285**, 540–544, 2001.

54. COLLINS, F. S., GREEN, E. D., GUTTMACHER, A. E., GUYER, M. S. & INSTITUTE, ON BEHALF OF THE U. S. N. H. G. R. "A vision for the future of genomics research". *Nature*, **422**, 835–847, 2003.
55. OLIVER, S., WINSON, M., KELL, D. & BAGANZ, F. "Systematic Functional Analysis of the Yeast Genome". *TIPTECH*, **16**, 2000.
56. CANUTO, G. A. B. *ET AL.* "METABOLÔMICA: DEFINIÇÕES, ESTADO-DA-ARTE E APLICAÇÕES REPRESENTATIVAS". *Quimica Nova* vol. 41 75–91, Preprint at <https://doi.org/10.21577/0100-4042.20170134> (2018).
57. SUSSULINI, A. "Metabolomics: From Fundamentals to Clinical Applications". in *Springer International Publishing* (2017).
58. TRIVEDI, D. K., HOLLYWOOD, K. A. & GOODACRE, R. "Metabolomics for the masses: The future of metabolomics in a personalized world". *New Horizons in Translational Medicine* vol. 3 294–305, Preprint at <https://doi.org/10.1016/j.nhtm.2017.06.001> (2017).
59. NEWGARD, C. B. "Metabolomics and Metabolic Diseases: Where Do We Stand?". *Cell Metab*, **25**, 43–56, 2017.
60. JOHNSON, C. H., IVANISEVIC, J. & SIUZDAK, G. "Metabolomics: Beyond biomarkers and towards mechanisms". *Nature Reviews Molecular Cell Biology* vol. 17 451–459, Preprint at <https://doi.org/10.1038/nrm.2016.25> (2016).
61. BEGER, R., SCHMIDT, M. & KADDURAH-DAOUK, R. "Current Concepts in Pharmacometabolomics, Biomarker Discovery, and Precision Medicine". *Metabolites*, **10**, 2020.
62. VERNOCCHI, P., DEL CHIERICO, F. & PUTIGNANI, L. "Gut Microbiota Profiling: Metabolomics Based Approach to Unravel Compounds Affecting Human Health". *Front Microbiol*, **7**, 2016.
63. CHAE, S., KIM, D. J. & CHO, J.-Y. "Complex influences of gut microbiome metabolism on various drug responses". *Transl Clin Pharmacol*, **28**, 7–16, 2020.
64. QIU, Y., SU, M., XIAO, X., ZHOU, D. & XIE, L. "Metabolomic profiling of Wilson disease, an inherited disorder of copper metabolism, and diseases with similar symptoms but normal copper metabolism". 2023 doi:10.21203/rs.3.rs-2425833/v1.
65. DURHAM, T. B. & BLANCO, M. J. "Target Engagement in Lead Generation". *Bioorganic and Medicinal Chemistry Letters* vol. 25 998–1008, Preprint at <https://doi.org/10.1016/j.bmcl.2014.12.076> (2015).

66. SCHRIMPE-RUTLEDGE, A., CODREANU, S., SHERROD, S. & MCLEAN, J. "Untargeted Metabolomics Strategies—Challenges and Emerging Directions". *J Am Soc Mass Spectrom*, **27**, 1–9, 2016.
67. CAJKA, T. & FIEHN, O. "Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics". *Anal Chem*, **88**, 2015.
68. SCHWAIGER-HABER, M. *ET AL.* "A Workflow to Perform Targeted Metabolomics at the Untargeted Scale on a Triple Quadrupole Mass Spectrometer". *ACS Measurement Science Au*, **1**, 35–45, 2021.
69. KHAMIS, M. M., ADAMKO, D. J. & EL-ANEED, A. "Mass spectrometric based approaches in urine metabolomics and biomarker discovery". *Mass Spectrom Rev*, **36**, 115–134, 2014.
70. GERTSMAN, I. & BARSHOP, B. "Promises and pitfalls of untargeted metabolomics". *J Inherit Metab Dis*, **41**, 2018.
71. SOUZA, A. & PATTI, G. "A Protocol for Untargeted Metabolomic Analysis: From Sample Preparation to Data Processing". in *Methods in molecular biology (Clifton, N.J.)* vol. 2276 357–382, (2021).
72. ROBERTS, L., SOUZA, A. L., GERSZTEN, R. E. & CLISH, C. "Targeted metabolomics". *Curr Protoc Mol Biol*, **30**, 1–24, 2012.
73. CAO, G. *ET AL.* "Large-scale Targeted Metabolomics Method for Metabolite Profiling of Human Samples". *Anal Chim Acta*, **1125**, 2020.
74. CHALECKIS, R., MEISTER, I., ZHANG, P. & WHELOCK, C. "Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics". *Curr Opin Biotechnol*, **55**, 44–50, 2018.
75. PINU, F., GOLDANSAZ, S. & JAINE, J. "Translational Metabolomics: Current Challenges and Future Opportunities". *Metabolites*, **9**, 2019.
76. TOLSTIKOV, V. "Metabolomics: Bridging the Gap between Pharmaceutical Development and Population Health". *Metabolites*, **6**, 20, 2016.
77. COLLINS, S. L., KOO, I., PETERS, J. M., SMITH, P. B. & PATTERSON, A. D. "Current Challenges and Recent Developments in Mass Spectrometry-Based Metabolomics". *Annual Review of Analytical Chemistry*, **14**, 467–487, 2021.
78. GOWDA, N. & RAFTERY, D. "NMR-Based Metabolomics". in *Advances in experimental medicine and biology* vol. 1280 19–37, (2021).
79. NAGANA GOWDA, G. A. & RAFTERY, D. "NMR Metabolomics Methods for Investigating Disease". *Anal Chem*, **95**, 83–99, 2023.

80. REY-STOLLE, F. *ET AL.* "Low and high resolution gas chromatography-mass spectrometry for untargeted metabolomics: A tutorial". *Anal Chim Acta*, **1210**, 339043, 2021.
81. FIEHN, O. "Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling". *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, **114**, 30.4.1-30.4.32, 2016.
82. BECCARIA, M. & CABOOTER, D. "Current Developments in LC-MS for Pharmaceutical Analysis". *Analyst*, **145**, 2020.
83. SHUKEN, S. "An Introduction to Mass Spectrometry-Based Proteomics". *J Proteome Res*, **22**, 2023.
84. PLUMB, R. *ET AL.* "Advances in high throughput LC/MS based metabolomics: A review". *TrAC Trends in Analytical Chemistry*, **160**, 116954, 2023.
85. THOMAS, S., FRENCH, D., JANNETTO, P. & CLARKE, W. "Liquid chromatography–tandem mass spectrometry for clinical diagnostics". *Nature Reviews Methods Primers*, **2**, 96, 2022.
86. LENZ, E. M. & WILSON, I. D. "Analytical Strategies in Metabonomics". *J Proteome Res*, **6**, 443–458, 2007.
87. KUEHNBAUM, N. L. & BRITZ-MCKIBBIN, P. "New Advances in Separation Science for Metabolomics: Resolving Chemical Diversity in a Post-Genomic Era". *Chem Rev*, **113**, 2437–2468, 2013.
88. GIKA, H. G., THEODORIDIS, G. A., PLUMB, R. S. & WILSON, I. D. "Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics". *J Pharm Biomed Anal*, **87**, 12–25, 2014.
89. GRECO, G. & LETZEL, T. "Main Interactions and Influences of the Chromatographic Parameters in HILIC Separations". *J Chromatogr Sci*, 2013.
90. LV, W. *ET AL.* "Alternate reversed-phase and hydrophilic interaction liquid chromatography coupled with mass spectrometry for broad coverage in metabolomics analysis". *Journal of Chromatography B*, **1152**, 122266, 2020.
91. TANG, D., ZOU, L., YIN, X. & ONG, C. "HILIC-MS for metabolomics: an attractive and complementary approach to RPLC-MS". *Mass Spectrom Rev*, **35**, 2014.
92. LEHMANN, S. *ET AL.* "Clinical mass spectrometry proteomics (cMSP) for medical laboratory: What does the future hold?". *Clinica Chimica Acta*, **467**, 51–58, 2017.
93. BHARDWAJ, C. & HANLEY, L. "Ion sources for mass spectrometric identification and imaging of molecular species". *Nat Prod Rep*, **31**, 2014.

94. ZENG, Q. *ET AL.* "Recent developments in ionization techniques for single-cell mass spectrometry". *Front Chem*, **11**, 2023.
95. HOFFMANN, E. DE & STROOBANT, V. *Mass Spectrometry: Principles and Applications*. vol. 3rd (2007).
96. PEZZATTI, J., GONZÁLEZ-RUIZ, V., BOCCARD, J., GUILLARME, D. & RUDAZ, S. "Evaluation of Different Tandem MS Acquisition Modes to Support Metabolite Annotation in Human Plasma Using Ultra High-Performance Liquid Chromatography High-Resolution Mass Spectrometry for Untargeted Metabolomics". *Metabolites*, **10**, 2020.
97. DEFOSSEZ, E., BOURQUIN, J., VON REUSS, S., RASMANN, S. & GLAUSER, G. "Eight key rules for successful data-dependent acquisition in mass spectrometry-based metabolomics". *Mass Spectrom Rev*, **42**, 2021.
98. HILAIRE, P. *ET AL.* "Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics". *Metabolites*, **10**, 158, 2020.
99. LOMMEN, A. "MetAlign: Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing". *Anal Chem*, **81**, 3079–3086, 2009.
100. SMITH, C. A., WANT, E. J., O'MAILLE, G., ABAGYAN, R. & SIUZDAK, G. "XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification". *Anal Chem*, **78**, 779–787, 2006.
101. KATAJAMAA, M., MIETTINEN, J. & OREŠIČ, M. "MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data". *Bioinformatics*, **22**, 634–636, 2006.
102. PANG, Z. *ET AL.* "Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data". *Nat Protoc*, **17**, 1–27, 2022.
103. XIA, J. (JEFF), PSYCHOGIOS, N., YOUNG, N. & WISHART, D. "MetaboAnalyst: A web server for metabolomic data analysis and interpretation". *Nucleic Acids Res*, **37**, W652-60, 2009.
104. TSUGAWA, H. *ET AL.* "MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis". *Nat Methods*, **12**, 2015.
105. CHEN, Y., LI, E.-M. & XU, L.-Y. "Guide to Metabolomics Analysis: A Bioinformatics Workflow". *Metabolites*, **12**, 2022.

106. COOK, D. W. & RUTAN, S. C. "Chemometrics for the analysis of chromatographic data in metabolomics investigations". *J Chemom*, **28**, 681–687, 2014.
107. WORLEY, B. & POWERS, R. "Multivariate Analysis in Metabolomics". *Curr Metabolomics*, **1**, 92–107, 2013.
108. FERREIRA, M. "QUIMIOMETRIA III – REVISITANDO A ANÁLISE EXPLORATÓRIA DOS DADOS MULTIVARIADOS". *Quim Nova*, **45**, 1251–1264, 2022.
109. SUGIMOTO, M., KAWAKAMI, M., ROBERT, M., SOGA, T. & TOMITA, M. *Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. Current Bioinformatics* vol. 7 (2012).
110. PILON, A. C. *ET AL.* "Plant metabolomics: Methods and challenges.". *Quim Nova*, **43**, 329–354, 2020.
111. XIA, J. (JEFF), BROADHURST, D., WILSON, M. & WISHART, D. "Translational biomarker discovery in clinical metabolomics: An introductory tutorial". *Metabolomics*, **9**, 280–299, 2013.
112. SACCENTI, E., HOEFSLOOT, H., SMILDE, A., WESTERHUIS, J. & HENDRIKS, M. "Reflections on univariate and multivariate analysis of metabolomics data". *Metabolomics*, **10**, 2013.
113. MASTRANGELO, A., FERRARINI, A., REY-STOLLE, F., GARCIA, A. & BARBAS, C. "From sample treatment to biomarker discovery: A tutorial for untargeted metabolomics based on GC-(EI)-Q-MS". *Anal Chim Acta*, **900**, 2015.
114. BOAKYE OPPONG, F. & YAO AGBEDRA, S. "Assessing Univariate and Multivariate Normality, A Guide For Non-Statisticians". **6**, 2016.
115. LEE, D. K. "Data transformation: A focus on the interpretation". *Korean J Anesthesiol*, **73**, 503–508, 2020.
116. KHAMMAR, A., YARAHMADI, M. & MADADIZADEH, F. "What Is Analysis of Covariance (ANCOVA) and How to Correctly Report its Results in Medical Research?". *Iran J Public Health*, **49**, 2020.
117. VINAIXA, M. *ET AL.* "A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data". *Metabolites*, **2**, 775–795, 2002.
118. LI, B. *ET AL.* "NOREVA: Normalization and evaluation of MS-based metabolomics data". *Nucleic Acids Res*, **45**, W162–W170, 2017.
119. GAGNEBIN, Y. *ET AL.* "Metabolomic analysis of urine samples by UHPLC-QTOF-MS: Impact of normalization strategies". *Anal Chim Acta*, **955**, 2016.

120. MIOT, H. "Análise de correlação em estudos clínicos e experimentais". *J Vasc Bras*, **17**, 2018.
121. SUMNER, L. *ET AL.* "Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI)". *Metabolomics*, **3**, 211–221, 2007.
122. HENDRIKS, M. M. W. B. *ET AL.* "Data-processing strategies for metabolomics studies". *TrAC - Trends in Analytical Chemistry* vol. 30 1685–1698, Preprint at <https://doi.org/10.1016/j.trac.2011.04.019> (2011).
123. KARAMAN, I. "Preprocessing and Pretreatment of Metabolomics Data for Statistical Analysis". in *Metabolomics: From Fundamentals to Clinical Applications* (ed. Sussulini, A.) 145–161, (Springer International Publishing, 2017). doi:10.1007/978-3-319-47656-8_6.
124. DUNN, W. B. *ET AL.* "Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry". *Nat Protoc*, **6**, 1060–1083, 2011.
125. SMILDE, A. K., VAN DER WERF, M. J., BIJLSMA, S., VAN DER WERFF-VAN DER VAT, B. J. C. & JELLEMA, R. H. "Fusion of Mass Spectrometry-Based Metabolomics Data". *Anal Chem*, **77**, 6729–6736, 2005.
126. DO, K. T. *ET AL.* "Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies". *Metabolomics*, **14**, 2018.
127. KUMAR, N., HOQUE, M. A. & SUGIMOTO, M. "Kernel weighted least square approach for imputing missing values of metabolomics data". *Sci Rep*, **11**, 2021.
128. 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*, **12**, 2016.
129. SUN, J. & XIA, Y. "Pretreating and normalizing metabolomics data for statistical analysis-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)". *Genes Dis*, **11**, 100979, 2024.
130. LIN, Y. C., BANSAL, N., VITTINGHOFF, E., GO, A. & HSU, C. "Determinants of the creatinine clearance to glomerular filtration rate ratio in patients with chronic kidney disease: a cross-sectional study (vol 15, pg 85, 2014)". *BMC Nephrol*, **15**, 2014.
131. SHAHBAZ H, G. M. "Creatinine Clearance. [Updated 2023 May 20]. ". In: *StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK544228/> .*

132. ZHANG, T. & WATSON, D. G. "A short review of applications of liquid chromatography mass spectrometry based metabolomics techniques to the analysis of human urine". *Analyst* vol. 140 2907–2915, Preprint at <https://doi.org/10.1039/c4an02294g> (2015).
133. DIETERLE, F., ROSS, A., SCHLOTTERBECK, G. & SENN, H. "Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ¹H NMR metabolomics". *Anal Chem*, **78**, 4281–4290, 2006.
134. VOLLMAR, A. *ET AL.* "Normalizing Untargeted Periconceptional Urinary Metabolomics Data: A Comparison of Approaches". *Metabolites*, **9**, 198, 2019.
135. VAN DEN BERG, R. A., HOEFSLOOT, H., WESTERHUIS, J. A., SMILDE, A. K. & VAN DERWERFL, M. J. *Centering, scaling, and transformations: improving the biological information content of metabolomics data*. *BMC Genomics* vol. 7 <http://www.biomedcentral.com/1471-2164/7/142> (2006).
136. COOK, T., MA, Y. & GAMAGEDARA, S. "Evaluation of Statistical Techniques to Normalize Mass Spectrometry-Based Urinary Metabolomics Data". *J Pharm Biomed Anal*, **177**, 112854, 2019.
137. MARIMUTHU, S., MANI, T., SUDARSANAM, T. D., GEORGE, S. & JEYASEELAN, L. "Preferring Box-Cox transformation, instead of log transformation to convert skewed distribution of outcomes to normal in medical research". *Clin Epidemiol Glob Health*, **15**, 2022.
138. ROSATO, A. *ET AL.* "From correlation to causation: analysis of metabolomics data using systems biology approaches". *Metabolomics*, **14**, 37, 2018.
139. MONGE, M. E., DODDS, J. N., BAKER, E. S., EDISON, A. S. & FERNÁNDEZ, F. M. "Challenges in Identifying the Dark Molecules of Life". *Annual Review of Analytical Chemistry*, **12**, 177–199, 2019.
140. JOHNSON, C. H. & GONZALEZ, F. J. "Challenges and opportunities of metabolomics". *J Cell Physiol*, **227**, 2975–2981, 2012.
141. DA SILVA, R. R., DORRESTEIN, P. C. & QUINN, R. A. "Illuminating the dark matter in metabolomics". *Proceedings of the National Academy of Sciences*, **112**, 12549–12550, 2015.
142. SUMNER, L. *ET AL.* "Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI)". *Metabolomics*, **3**, 211–221, 2007.
143. SWETS, J. *ET AL.* "Measuring the accuracy of diagnostic systems". *Science* (1979), 1988.

144. HOO, Z. H., CANDLISH, J. & TEARE, D. "What is an ROC curve?". *Emergency Medicine Journal*, **34**, emermed-2017, 2017.
145. LIU, X. *ET AL.* "Identification of metabolic biomarkers in patients with type 2 diabetic coronary heart diseases based on metabolomic approach:". *Sci Rep*, **6**, 30785, 2016.
146. LIU, G., LEE, D., SCHMIDT, E. & PRASAD, G. L. "Pathway Analysis of Global Metabolomic Profiles Identified Enrichment of Caffeine, Energy, and Arginine Metabolism in Smokers but Not Moist Snuff Consumers". *Bioinform Biol Insights*, **13**, 117793221988296, 2019.
147. MAZI, T. A. *ET AL.* "Dysregulated choline, methionine, and aromatic amino acid metabolism in patients with wilson disease: Exploratory metabolomic profiling and implications for hepatic and neurologic phenotypes". *Int J Mol Sci*, **20**, 2019.
148. WILMORE, D. *Arginine Metabolism: Enzymology, Nutrition, and Clinical Significance Enteral and Parenteral Arginine Supplementation to Improve Medical Outcomes in Hospitalized Patients 1. J. Nutr* vol. 134 (2004).
149. WU, G. & MORRIS, S. M. *Arginine metabolism : nitric oxide and beyond. Biochem. J* vol. 336 (1998).
150. WU, G. *ET AL.* "Arginine metabolism and nutrition in growth, health and disease". *Amino Acids*, **37**, 153–168, 2009.
151. WU, G. *ET AL.* "Arginine metabolism and nutrition in growth, health and disease". *Amino Acids*, **37**, 153–168, 2009.
152. MORRIS, S. M. *The Journal of Nutrition 6 th Amino Acid Assessment Workshop Arginine Metabolism: Boundaries of Our Knowledge 1-3. J. Nutr* vol. 137 (2007).
153. METABOLISM, R. A., BROSNAN, M. E. & BROSNAN, J. T. *Arginine Metabolism: Enzymology, Nutrition, and Clinical Significance. J. Nutr* vol. 134 (2004).
154. CAO, S. *ET AL.* "L-arginine metabolism inhibits arthritis and inflammatory bone loss". *Ann Rheum Dis*, **83**, 72–87, 2023.
155. LI, J. Y. *ET AL.* "Arginine metabolism regulates the pathogenesis of inflammatory bowel disease". *Nutr Rev*, **81**, 578–586, 2023.
156. LAU, T. *ET AL.* "Arginine, citrulline, and nitric oxide metabolism in end-stage renal disease patients". *Journal of Clinical Investigation*, **105**, 2000.
157. AU, A. *ET AL.* "Plasma arginine metabolites in health and chronic kidney disease". *Nephrol Dial Transplant*, **38**, 2023.
158. WU, G. & MORRIS, S. M. *Arginine metabolism : nitric oxide and beyond. Biochem. J* vol. 336 (1998).

159. MORRIS, S. M. "Arginases and arginine deficiency syndromes". *Current Opinion in Clinical Nutrition and Metabolic Care* vol. 15 64–70, Preprint at <https://doi.org/10.1097/MCO.0b013e32834d1a08> (2012).
160. MORRIS, S. M. "Arginine metabolism revisited". *Journal of Nutrition*, **146**, 2579S-2586S, 2016.
161. MARIC, S. *ET AL.* "Citrulline, biomarker of enterocyte functional mass and dietary supplement. Metabolism, transport, and current evidence for clinical use". *Nutrients* vol. 13 Preprint at <https://doi.org/10.3390/nu13082794> (2021).
162. AGUAYO, E., MARTÍNEZ-SÁNCHEZ, A., FERNÁNDEZ-LOBATO, B. & ALACID, F. "L-citrulline: A non-essential amino acid with important roles in human health". *Applied Sciences (Switzerland)* vol. 11 Preprint at <https://doi.org/10.3390/app11073293> (2021).
163. GELSE, K., PÖSCHL, E. & AIGNER, T. "Collagens - Structure, function, and biosynthesis". *Adv Drug Deliv Rev*, **55**, 1531–1546, 2003.
164. SRIVASTAVA, A., KHARE, P., RAGHUWANSHI, N., NAGAR, H. & SRIVASTAVA, R. "Hydroxyproline: A Potential Biochemical Marker and Its Role in the Pathogenesis of Different Diseases". *Curr Protein Pept Sci*, **17**, 596–602, 2015.
165. BELOSTOTSKY, R. & FRISHBERG, Y. "Catabolism of Hydroxyproline in Vertebrates: Physiology, Evolution, Genetic Diseases and New siRNA Approach for Treatment". *International Journal of Molecular Sciences* vol. 23 Preprint at <https://doi.org/10.3390/ijms23021005> (2022).
166. ENI-AGANGA, I., LANAGHAN, Z. M., BALASUBRAMANIAM, M., DASH, C. & PANDHARE, J. "PROLIDASE: A Review from Discovery to its Role in Health and Disease". *Frontiers in Molecular Biosciences* vol. 8 Preprint at <https://doi.org/10.3389/fmolb.2021.723003> (2021).
167. WILK, P. *ET AL.* "Structural basis for prolidase deficiency disease mechanisms". *FEBS Journal*, **285**, 3422–3441, 2018.
168. GOPALAKRISHNA, H. *ET AL.* "Chronic Liver Disease in Patients with Prolidase Deficiency: A Case Series". *Case Rep Gastroenterol*, 49–57, 2024 doi:10.1159/000536117.
169. CHARPENTIER, C., DAGBOVIE, K., LEMONNIER, A., LARREGUE, M. & JOHNSTONE, R. A. W. *Work in Progress Prolidase Deficiency with Iminodipeptiduria: Biochemical Investigations and First Results of Attempted Therapy. J. Inher. Metab. Dis* vol. 4 (1981).
170. ROSSIGNOL, F. *ET AL.* "Quantitative analysis of the natural history of prolidase deficiency: description of 17 families and systematic review of published cases".

- Genetics in Medicine* vol. 23 1604–1615, Preprint at <https://doi.org/10.1038/s41436-021-01200-2> (2021).
171. JANDKE, J. & SPITELLER, G. *DIPEPTIDE ANALYSIS IN HUMAN URINE*. *Journal of Chromatography* vol. 382 (1986).
 172. NAMIDURU, E. S. "Prolidase". *Bratislava Medical Journal* vol. 117 480–485, Preprint at https://doi.org/10.4149/BLL_2016_093 (2016).
 173. ZIMMERMAN, L., JÖRNVALL, H. & BERGSTRÖM, J. "Phenylacetylglutamine and Hippuric Acid in Uremic and Healthy Subjects". *Nephron*, **55**, 265–271, 1990.
 174. ZONG, X. *ET AL.* "Phenylacetylglutamine as a risk factor and prognostic indicator of heart failure". *ESC Heart Fail*, **9**, 2645–2653, 2022.
 175. YANG, H. H., CHEN, Y. C., HO, C. C. & HSU, B. G. "Serum Phenylacetylglutamine among Potential Risk Factors for Arterial Stiffness Measuring by Carotid–Femoral Pulse Wave Velocity in Patients with Kidney Transplantation". *Toxins (Basel)*, **16**, 2024.
 176. KRISHNAMOORTHY, N. K. *ET AL.* "Role of the Gut Bacteria-Derived Metabolite Phenylacetylglutamine in Health and Diseases". *ACS Omega* Preprint at <https://doi.org/10.1021/acsomega.3c08184> (2023).
 177. BHAGAVAN, N. V. & HA, C.-E. "Protein and Amino Acid Metabolism". in *Essentials of Medical Biochemistry* 227–268, (Elsevier, 2015). doi:10.1016/b978-0-12-416687-5.00015-4.
 178. YANG, R. Z. *ET AL.* "Alanine aminotransferase isoenzymes: Molecular cloning and quantitative analysis of tissue expression in rats and serum elevation in liver toxicity". *Hepatology*, **49**, 598–607, 2009.
 179. SCHINDHELM, R. K. *ET AL.* "Alanine aminotransferase as a marker of non-alcoholic fatty liver disease in relation to type 2 diabetes mellitus and cardiovascular disease". *Diabetes/Metabolism Research and Reviews* vol. 22 437–443, Preprint at <https://doi.org/10.1002/dmrr.666> (2006).
 180. KIM, W. R., FLAMM, S. L., DI BISCEGLIE, A. M. & BODENHEIMER, H. C. "Serum activity of alanine aminotransferase (ALT) as an indicator of health and disease". *Hepatology* vol. 47 1363–1370, Preprint at <https://doi.org/10.1002/hep.22109> (2008).
 181. NDREPEPA, G. "Aspartate aminotransferase and cardiovascular disease - A narrative review". *Journal of Laboratory and Precision Medicine* vol. 6 Preprint at <https://doi.org/10.21037/jlpm-20-93> (2021).
 182. FELIG, P. "The glucose-alanine cycle". *Metabolism*, **22**, 179–207, 1973.

183. NELSON, D. L. & COX, M. M. *Princípios de bioquímica de Lehninger*. (2014).
184. "<https://biocyc.org/pathway?orgid=HUMAN&id=ASPARTATESYN-PWY>
Acesso 19Abril2024".
185. HELENIUS, I. T., MADALA, H. R. & YEH, J. R. J. "An asp to strike out cancer? Therapeutic possibilities arising from aspartate's emerging roles in cell proliferation and survival". *Biomolecules* vol. 11 Preprint at <https://doi.org/10.3390/biom11111666> (2021).
186. LANE, A. N. & FAN, T. W. M. "Regulation of mammalian nucleotide metabolism and biosynthesis". *Nucleic Acids Research* vol. 43 2466–2485, Preprint at <https://doi.org/10.1093/nar/gkv047> (2015).
187. HOLEČEK, M. "Aspartic Acid in Health and Disease". *Nutrients* vol. 15 Preprint at <https://doi.org/10.3390/nu15184023> (2023).
188. ERRICO', F., NAPOLITANO, F., NISTICO, R., CENTONZE, D. & USIELLO, A. *D-Aspartate: An Atypical Amino Acid with Neuromodulatory Activity in Mammals. Reviews in the Neurosciences* vol. 20 (2009).
189. LIN, M. S. *ET AL*. "Alanine aminotransferase to aspartate aminotransferase ratio and hepatitis B virus on metabolic syndrome: a community-based study". *Front Endocrinol (Lausanne)*, **13**, 2022.
190. KALLURU, P. K. R. *ET AL*. "Aminotransferases in Relation to the Severity of Dengue: A Systematic Review". *Cureus*, 2023 doi:10.7759/cureus.39436.
191. ALOISIO, E., COLOMBO, G., ARRIGO, C., DOLCI, A. & PANTEGHINI, M. "Sources and clinical significance of aspartate aminotransferase increases in COVID-19". *Clinica Chimica Acta*, **522**, 88–95, 2021.
192. CHÁVEZ, S. M. *ET AL*. "Acute liver failure due to herpes simplex virus: diagnostic clues and potential role of plasmapheresis: A case report". *Medicine (United States)*, **100**, E27139, 2021.
193. KOBAYASHI, A., SUZUKI, Y. & SUGAI, S. "Specificity of transaminase activities in the prediction of drug-induced hepatotoxicity". *J Toxicol Sci*, **45**, 515–537, 2020.
194. HAN, J. H. *ET AL*. "Markedly Elevated Aspartate Aminotransferase from Non-Hepatic Causes". *J Clin Med*, **12**, 2023.
195. HAYASHI, H. *ET AL*. "Alanine aminotransferase as the first test parameter for wilson's disease". *J Clin Transl Hepatol*, **7**, 293–296, 2019.
196. SARODE, G. *ET AL*. "Metabolomics profiles of patients with Wilson disease reveal a distinct metabolic signature". *Metabolomics*, **15**, 2019.

197. KALITA, J. *ET AL.* "A study of oxidative stress, cytokines and glutamate in Wilson disease and their asymptomatic siblings". *J Neuroimmunol*, **274**, 141–148, 2014.
198. BENDER, D. "The Aromatic Amino Acids: Phenylalanine, Tyrosine and Tryptophan". in 323–376, (2012). doi:10.1002/9781118357514.ch9.
199. UYAVER, S. "Tyrosine, Phenylalanine, and Tryptophan Undergo Self-Aggregation in Similar and Different Manners". *Atmosphere (Basel)*, **13**, 2022.
200. ANJANA, R. *ET AL.* "Hypothesis Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: a study based on orientation and distance". *Bioinformation*, **8**, 1220, 2012.
201. LICHTER-KONECKI, U., HIPKE, C. M. & KONECKI, D. S. *Human Phenylalanine Hydroxylase Gene Expression in Kidney and Other Nonhepatic Tissues*. <http://www.idealibrary.com> (1999).
202. CHEN, A., PAN, Y. & CHEN, J. "Clinical, genetic, and experimental research of hyperphenylalaninemia". *Frontiers in Genetics* vol. 13 Preprint at <https://doi.org/10.3389/fgene.2022.1051153> (2023).
203. GÜTTLER, F. & LOU, H. "Phenylketonuria and Hyperphenylalaninemia". in 161–174, (1990). doi:10.1007/978-3-662-02613-7_13.
204. KOHLMEIER, M. "Phenylalanine". in 314–321, (2003). doi:10.1016/B978-012417762-8.50051-X.
205. DAUBNER, S. C., LE, T. & WANG, S. "Tyrosine hydroxylase and regulation of dopamine synthesis". *Archives of Biochemistry and Biophysics* vol. 508 1–12, Preprint at <https://doi.org/10.1016/j.abb.2010.12.017> (2011).
206. WIGGINS, T., KUMAR, S., MARKAR, S. R., ANTONOWICZ, S. & HANNA, G. B. "Tyrosine, phenylalanine, and tryptophan in gastroesophageal malignancy: A systematic review". *Cancer Epidemiology Biomarkers and Prevention* vol. 24 32–38, Preprint at <https://doi.org/10.1158/1055-9965.EPI-14-0980> (2015).
207. WANG, L., ERLANDSEN, H., HAAVIK, J., KNAPPSKOG, P. M. & STEVENS, R. C. "Three-dimensional structure of human tryptophan hydroxylase and its implications for the biosynthesis of the neurotransmitters serotonin and melatonin". *Biochemistry*, **41**, 12569–12574, 2002.
208. LIN, S. H., LEE, L. T. & YANG, Y. K. "Serotonin and mental disorders: A concise review on molecular neuroimaging evidence". *Clinical Psychopharmacology and Neuroscience*, **12**, 196–202, 2014.
209. MARAZZITI, D. "Understanding the role of serotonin in psychiatric diseases". *F1000Res*, **6**, 180, 2017.

210. COMAI, S., BERTAZZO, A., BRUGHERA, M. & CROTTI, S. "Tryptophan in health and disease". in *Advances in Clinical Chemistry* vol. 95 165–218, (Academic Press Inc., 2020).
211. SAVITZ, J. "The kynurenine pathway: a finger in every pie". *Molecular Psychiatry* vol. 25 131–147, Preprint at <https://doi.org/10.1038/s41380-019-0414-4> (2020).
212. WALØEN, K., KLEPPE, R., MARTINEZ, A. & HAAVIK, J. "Tyrosine and tryptophan hydroxylases as therapeutic targets in human disease". *Expert Opinion on Therapeutic Targets* vol. 21 167–180, Preprint at <https://doi.org/10.1080/14728222.2017.1272581> (2017).
213. BURDGE, G. C. "Metabolism of α -linolenic acid in humans". *Prostaglandins Leukot Essent Fatty Acids*, **75**, 161–168, 2006.
214. AZBUKINA, N. V. *ET AL.* "Oxylipin profiles in plasma of patients with wilson's disease". *Metabolites*, **10**, 2020.
215. MORI, T. A. & HODGSON, J. M. "Fatty acids: Health Effects of Omega-6 Polyunsaturated Fatty Acids". in *Encyclopedia of Human Nutrition* vols 2–4 209–214, (Elsevier Inc., 2012).
216. MAZI, T. A., SHIBATA, N. M. & MEDICI, V. "Lipid and energy metabolism in Wilson disease". *Liver Research* vol. 4 5–14, Preprint at <https://doi.org/10.1016/j.livres.2020.02.002> (2020).
217. HUSTER, D. *ET AL.* "High copper selectively alters lipid metabolism and cell cycle machinery in the mouse model of Wilson disease". *Journal of Biological Chemistry*, **282**, 8343–8355, 2007.

Anexo 1

Tabela A1- Resultado do teste de *t-Student* para a análise estatística referente à idade dos indivíduos dos dois grupos (DW e HC), nas amostras de plasma

Estatísticas de grupo					
	Grupo	N	Média	Erro Desvio	Erro padrão da média
Idade	DW	22	37.95	12.148	2.59
	GC	35	31.06	9.677	1.636

Teste de amostras independentes

		Teste de Levene para igualdade de variâncias	Teste-t para Igualdade de Médias
Idade	Variâncias iguais assumidas	0.138	0.021

Tabela A2- Resultado do teste de Qui-quadrado para a análise estatística referente ao gênero dos indivíduos dos dois grupos (DW e HC), nas amostras de plasma

		Grupo		Total	Qui-quadrado de Pearson
		DW	GC		
Gênero	Feminino	Contagem	9	20	0.233
		% em lable	40.9%	57.1%	
		% do Total	15.8%	35.1%	
	Masculino	Contagem	13	15	
		% em lable	59.1%	42.9%	
		% do Total	22.8%	26.3%	
Total	Contagem	22	35	57	
	% em lable	100.0%	100.0%	100.0%	
	% do Total	38.6%	61.4%	100.0%	

Tabela A3-Determinação dos *features* moleculares significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de plasma no modo RPLC ESI(+/-) de eluição (*continuação*).

Feature	Ancova p-value	Ranking	FDR 10%	Conclusão
1.21min : 329.038m/z	1.23241E-06	1	0.00119	SIG
1.22min : 583.175m/z	3.22562E-06	2	0.002381	SIG
1.69min : 459.112m/z	7.48327E-06	3	0.003571	SIG
18.65min : 426.358m/z	1.31599E-05	4	0.004762	SIG
1.15min : 337.062m/z	2.36082E-05	5	0.005952	SIG
11.24min : 383.153m/z	4.44349E-05	6	0.007143	SIG
18.78min : 265.148m/z	5.17696E-05	7	0.008333	SIG
20.19min : 279.160m/z	6.60885E-05	8	0.009524	SIG
1.54min : 291.084m/z	7.42188E-05	9	0.010714	SIG
1.65min : 162.839m/z	8.11501E-05	10	0.011905	SIG
23.83min : 862.625m/z	0.000131907	11	0.013095	SIG
0.95min : 174.991m/z	0.000146945	12	0.014286	SIG
2.58min : 136.075m/z	0.000259967	13	0.015476	SIG
18.22min : 544.340m/z	0.000288886	14	0.016667	SIG
1.60min : 347.016m/z	0.000315644	15	0.017857	SIG
20.20min : 205.086m/z	0.000337881	16	0.019048	SIG
20.57min : 149.023m/z	0.000345629	17	0.020238	SIG
19.92min : 146.938m/z	0.000462859	18	0.021429	SIG
4.88min : 209.092m/z	0.000567631	19	0.022619	SIG
1.14min : 333.053m/z	0.000643019	20	0.02381	SIG
1.61min : 113.021m/z	0.001158082	21	0.025	SIG
11.20min : 236.113m/z	0.001275526	22	0.02619	SIG
1.19min : 627.138m/z	0.00185564	23	0.027381	SIG
23.43min : 469.390m/z	0.001998515	24	0.028571	SIG
11.23min : 159.065m/z	0.002188751	25	0.029762	SIG

Tabelas A3- Determinação dos *features* moleculares significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de plasma no modo RPLC ESI(+/-) de eluição (*continuação*).

Feature	Ancova p-value	Ranking	FDR 10%	Conclusão
1.68min : 150.058m/z	0.00237948	26	0.030952	SIG
1.08min : 120.065m/z	0.002532596	27	0.032143	SIG
18.78min : 590.346m/z	0.002650177	28	0.033333	SIG
19.47min : 233.155m/z	0.00287024	29	0.034524	SIG
1.17min : 585.189m/z	0.003078263	30	0.035714	SIG
1.14min : 629.152m/z	0.003357015	31	0.036905	SIG
21.47min : 448.337m/z	0.003393948	32	0.038095	SIG
1.60min : 201.038m/z	0.003650169	33	0.039286	SIG
1.65min : 104.053m/z	0.004170133	34	0.040476	SIG
2.62min : 182.081m/z	0.005788357	35	0.041667	SIG
20.48min : 524.372m/z	0.005814912	36	0.042857	SIG
21.19min : 368.425m/z	0.006300072	37	0.044048	SIG
17.84min : 538.315m/z	0.006422067	38	0.045238	SIG
5.20min : 310.002m/z	0.006882818	39	0.046429	SIG
6.63min : 145.051m/z	0.007378214	40	0.047619	SIG
1.89min : 128.035m/z	0.008112414	41	0.04881	SIG
8.20min : 263.103m/z	0.008396419	42	0.05	SIG
1.65min : 133.031m/z	0.008987687	43	0.05119	SIG
18.15min : 524.278m/z	0.0103356	44	0.052381	SIG
21.59min : 277.217m/z	0.011659062	45	0.053571	SIG
0.91min : 162.007m/z	0.013618794	46	0.054762	SIG
17.39min : 512.299m/z	0.014109284	47	0.055952	SIG
13.78min : 367.159m/z	0.014160197	48	0.057143	SIG
1.41min : 72.081m/z	0.014716613	49	0.058333	SIG
21.37min : 307.190m/z	0.015202632	50	0.059524	SIG
22.59min : 279.233m/z	0.015845466	51	0.060714	SIG

Tabelas A3- Determinação dos *features* moleculares significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de plasma no modo RPLC ESI(+/-) de eluição (*continuação*).

Feature	Ancova p-value	Ranking	FDR 10%	Conclusão
1.18min : 393.005m/z	0.016462522	52	0.061905	SIG
17.32min : 452.203m/z	0.016480898	53	0.063095	SIG
19.33min : 524.335m/z	0.017199112	54	0.064286	SIG
11.05min : 187.097m/z	0.017500415	55	0.065476	SIG
22.16min : 438.379m/z	0.017820713	56	0.066667	SIG
20.73min : 701.559m/z	0.018817798	57	0.067857	SIG
23.81min : 255.233m/z	0.019844119	58	0.069048	SIG
24.76min : 425.363m/z	0.021320118	59	0.070238	SIG
20.29min : 1281.933m/z	0.022950179	60	0.071429	SIG
1.85min : 157.013m/z	0.022992859	61	0.072619	SIG
1.85min : 293.098m/z	0.023344797	62	0.07381	SIG
18.66min : 284.331m/z	0.02601068	63	0.075	SIG
22.58min : 559.473m/z	0.027024855	64	0.07619	SIG
18.18min : 476.278m/z	0.028174144	65	0.077381	SIG
22.21min : 253.217m/z	0.028678456	66	0.078571	SIG
18.47min : 400.342m/z	0.029741017	67	0.079762	SIG
18.77min : 496.341m/z	0.030590179	68	0.080952	SIG
21.33min : 745.550m/z	0.030817234	69	0.082143	SIG
3.39min : 218.138m/z	0.031009214	70	0.083333	SIG
17.82min : 494.324m/z	0.032035066	71	0.084524	SIG
20.41min : 463.343m/z	0.035061711	72	0.085714	SIG
24.97min : 1138.860m/z	0.035628781	73	0.086905	SIG
22.26min : 303.233m/z	0.038494	74	0.088095	SIG
5.64min : 188.986m/z	0.049608227	75	0.089286	SIG
24.35min : 307.264m/z	0.050214331	76	0.090476	SIG
13.40min : 288.232m/z	0.053148755	77	0.091667	SIG

Tabelas A3- Determinação dos *features* moleculares significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de plasma no modo RPLC ESI(+/-) de eluição.

Feature	Ancova p-value	Ranking	FDR 10%	Conclusão
20.79min : 293.175m/z	0.053454008	78	0.092857	SIG
23.83min : 355.158m/z	0.066355397	79	0.094048	SIG
18.28min : 1039.673m/z	0.069402861	80	0.095238	SIG
20.01min : 719.534m/z	0.077144478	81	0.096429	SIG
18.98min : 311.223m/z	0.097865445	82	0.097619	NS
1.24min : 343.995m/z	0.106146457	83	0.09881	NS
24.03min : 381.174m/z	0.809012835	84	0.1	NS

Legenda: **Sig** – Significativo; **NS** – Não Significativo

Tabela A4- Determinação dos *features* moleculares significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de plasma no modo HILIC ESI(+/-) de eluição (*continuação*).

Features	Ancova p-value	Ranking	FDR 10%	Conclusão
10.07min : 247.093m/z	5.25E-14	1	0.000606	SIG
13.50min : 189.088m/z	8.69E-13	2	0.001212	SIG
6.35min : 96.922m/z	1.98E-12	3	0.001818	SIG
6.59min : 170.944m/z	1.5E-11	4	0.002424	SIG
6.80min : 112.986m/z	6.91E-10	5	0.00303	SIG
13.50min : 191.102m/z	1.45E-09	6	0.003636	SIG
9.84min : 467.163m/z	3.1E-09	7	0.004242	SIG
6.32min : 98.919m/z	3.39E-08	8	0.004848	SIG
7.18min : 118.931m/z	2.23E-07	9	0.005455	SIG
10.04min : 282.119m/z	3.67E-07	10	0.006061	SIG
9.73min : 1167.355m/z	8.89E-07	11	0.006667	SIG
2.94min : 173.056m/z	2.6E-06	12	0.007273	SIG
5.69min : 174.839m/z	3.35E-06	13	0.007879	SIG
9.74min : 1021.308m/z	3.36E-06	14	0.008485	SIG
6.59min : 104.954m/z	4.14E-06	15	0.009091	SIG
8.65min : 128.960m/z	4.87E-06	16	0.009697	SIG
6.73min : 157.012m/z	6.19E-06	17	0.010303	SIG
9.65min : 1205.310m/z	7.69E-06	18	0.010909	SIG
7.55min : 124.007m/z	8.85E-06	19	0.011515	SIG
6.66min : 106.950m/z	1.09E-05	20	0.012121	SIG
7.95min : 196.947m/z	1.16E-05	21	0.012727	SIG
13.17min : 275.056m/z	1.24E-05	22	0.013333	SIG
13.20min : 333.028m/z	1.25E-05	23	0.013939	SIG
2.06min : 118.086m/z	1.34E-05	24	0.014545	SIG
13.18min : 164.928m/z	1.56E-05	25	0.015152	SIG
1.71min : 146.045m/z	1.68E-05	26	0.015758	SIG

Tabelas A4- Determinação dos *features* moleculares significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de plasma no modo HILIC ESI(+/-) de eluição (*continuação*).

Features	Ancova p-value	Ranking	FDR 10%	Conclusão
9.23min : 1205.809m/z	1.68E-05	27	0.016364	SIG
1.24min : 132.030m/z	1.93E-05	28	0.01697	SIG
14.85min : 146.939m/z	2.51E-05	29	0.017576	SIG
10.10min : 104.107m/z	3.05E-05	30	0.018182	SIG
10.78min : 353.000m/z	3.43E-05	31	0.018788	SIG
1.55min : 160.061m/z	4.07E-05	32	0.019394	SIG
7.88min : 1052.533m/z	4.13E-05	33	0.02	SIG
6.11min : 82.945m/z	5.6E-05	34	0.020606	SIG
3.57min : 179.056m/z	5.83E-05	35	0.021212	SIG
6.78min : 186.918m/z	5.97E-05	36	0.021818	SIG
6.98min : 190.911m/z	6.57E-05	37	0.022424	SIG
5.79min : 308.098m/z	7.32E-05	38	0.02303	SIG
5.74min : 172.843m/z	8E-05	39	0.023636	SIG
7.35min : 277.104m/z	8.31E-05	40	0.024242	SIG
6.11min : 80.948m/z	9.22E-05	41	0.024848	SIG
5.63min : 155.866m/z	0.000108	42	0.025455	SIG
7.43min : 459.308m/z	0.00011	43	0.026061	SIG
9.30min : 86.096m/z	0.000171	44	0.026667	SIG
10.81min : 340.000m/z	0.000176	45	0.027273	SIG
13.33min : 343.995m/z	0.000186	46	0.027879	SIG
4.92min : 134.895m/z	0.000197	47	0.028485	SIG
9.82min : 583.174m/z	0.000197	48	0.029091	SIG
4.47min : 100.934m/z	0.000199	49	0.029697	SIG
5.28min : 136.878m/z	0.000224	50	0.030303	SIG
7.90min : 968.571m/z	0.000285	51	0.030909	SIG
10.95min : 809.653m/z	0.000285	52	0.031515	SIG
7.90min : 886.609m/z	0.000326	53	0.032121	SIG

Tabelas A4- Determinação dos *features* moleculares significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de plasma no modo HILIC ESI(+/-) de eluição (*continuação*).

Features	Ancova p-value	Ranking	FDR 10%	Conclusão
10.22min : 784.585m/z	0.000327	54	0.032727	SIG
2.95min : 242.987m/z	0.000346	55	0.033333	SIG
2.94min : 244.985m/z	0.000349	56	0.033939	SIG
5.62min : 156.925m/z	0.000353	57	0.034545	SIG
9.22min : 166.086m/z	0.000361	58	0.035152	SIG
11.66min : 532.300m/z	0.000372	59	0.035758	SIG
10.93min : 734.569m/z	0.000419	60	0.036364	SIG
4.43min : 328.139m/z	0.000437	61	0.03697	SIG
11.34min : 347.003m/z	0.000457	62	0.037576	SIG
4.66min : 144.923m/z	0.000473	63	0.038182	SIG
7.89min : 800.649m/z	0.000484	64	0.038788	SIG
10.24min : 826.562m/z	0.00058	65	0.039394	SIG
5.51min : 135.031m/z	0.000607	66	0.04	SIG
2.08min : 350.326m/z	0.000645	67	0.040606	SIG
11.91min : 803.627m/z	0.000713	68	0.041212	SIG
3.94min : 875.238m/z	0.000755	69	0.041818	SIG
13.30min : 389.099m/z	0.000763	70	0.042424	SIG
4.03min : 779.190m/z	0.000778	71	0.04303	SIG
12.04min : 264.927m/z	0.000823	72	0.043636	SIG
6.31min : 94.925m/z	0.000877	73	0.044242	SIG
11.16min : 703.576m/z	0.000912	74	0.044848	SIG
7.90min : 884.611m/z	0.000917	75	0.045455	SIG
10.29min : 812.580m/z	0.000997	76	0.046061	SIG
9.16min : 90.977m/z	0.001046	77	0.046667	SIG
10.42min : 870.547m/z	0.001216	78	0.047273	SIG
5.65min : 157.863m/z	0.001227	79	0.047879	SIG
8.08min : 748.527m/z	0.001242	80	0.048485	SIG

Tabelas A4- Determinação dos *features* moleculares significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de plasma no modo HILIC ESI(+/-) de eluição (*continuação*).

Features	Ancova p-value	Ranking	FDR 10%	Conclusão
11.14min : 773.581m/z	0.001445	81	0.049091	SIG
11.66min : 598.289m/z	0.001451	82	0.049697	SIG
12.53min : 82.537m/z	0.00159	83	0.050303	SIG
1.48min : 144.030m/z	0.001625	84	0.050909	SIG
9.56min : 148.060m/z	0.00171	85	0.051515	SIG
11.66min : 989.656m/z	0.001982	86	0.052121	SIG
9.35min : 122.924m/z	0.001994	87	0.052727	SIG
12.34min : 298.052m/z	0.002071	88	0.053333	SIG
11.50min : 522.356m/z	0.002079	89	0.053939	SIG
11.08min : 855.659m/z	0.002124	90	0.054545	SIG
10.20min : 1098.658m/z	0.00218	91	0.055152	SIG
4.32min : 158.950m/z	0.002335	92	0.055758	SIG
7.25min : 321.131m/z	0.002349	93	0.056364	SIG
6.29min : 152.883m/z	0.002428	94	0.05697	SIG
11.69min : 494.324m/z	0.002447	95	0.057576	SIG
11.23min : 737.536m/z	0.002495	96	0.058182	SIG
13.15min : 388.983m/z	0.002495	97	0.058788	SIG
7.34min : 371.255m/z	0.002495	98	0.059394	SIG
5.36min : 131.961m/z	0.002659	99	0.06	SIG
8.55min : 114.066m/z	0.002727	100	0.060606	SIG
7.87min : 970.570m/z	0.002912	101	0.061212	SIG
13.53min : 170.092m/z	0.003197	102	0.061818	SIG
4.80min : 266.946m/z	0.003207	103	0.062424	SIG
10.79min : 495.068m/z	0.003208	104	0.06303	SIG
11.89min : 512.334m/z	0.003415	105	0.063636	SIG
12.42min : 71.029m/z	0.003498	106	0.064242	SIG
5.23min : 368.425m/z	0.003508	107	0.064848	SIG

Tabelas A4- Determinação dos *features* moleculares significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de plasma no modo HILIC ESI(+/-) de eluição (*continuação*).

Features	Ancova p-value	Ranking	FDR 10%	Conclusão
10.45min : 291.084m/z	0.003559	108	0.065455	SIG
6.22min : 620.546m/z	0.003702	109	0.066061	SIG
10.53min : 776.578m/z	0.003868	110	0.066667	SIG
6.23min : 793.454m/z	0.004027	111	0.067273	SIG
9.55min : 913.220m/z	0.004152	112	0.067879	SIG
1.29min : 177.040m/z	0.004705	113	0.068485	SIG
5.25min : 137.046m/z	0.004729	114	0.069091	SIG
5.09min : 353.954m/z	0.0051	115	0.069697	SIG
11.18min : 993.650m/z	0.005167	116	0.070303	SIG
10.95min : 351.013m/z	0.005373	117	0.070909	SIG
5.39min : 146.060m/z	0.006348	118	0.071515	SIG
11.07min : 923.646m/z	0.006635	119	0.072121	SIG
13.17min : 145.098m/z	0.00681	120	0.072727	SIG
4.92min : 132.901m/z	0.00708	121	0.073333	SIG
13.10min : 261.040m/z	0.007472	122	0.073939	SIG
11.14min : 841.568m/z	0.007486	123	0.074545	SIG
11.69min : 538.315m/z	0.007553	124	0.075152	SIG
9.55min : 623.130m/z	0.007891	125	0.075758	SIG
4.56min : 251.998m/z	0.008495	126	0.076364	SIG
8.79min : 349.183m/z	0.008521	127	0.07697	SIG
1.90min : 193.034m/z	0.008659	128	0.077576	SIG
10.32min : 786.565m/z	0.009614	129	0.078182	SIG
11.77min : 468.308m/z	0.01095	130	0.078788	SIG
11.69min : 844.333m/z	0.011753	131	0.079394	SIG
10.46min : 794.539m/z	0.011774	132	0.08	SIG
5.14min : 288.199m/z	0.011944	133	0.080606	SIG
1.70min : 383.153m/z	0.012115	134	0.081212	SIG

Tabelas A4- Determinação dos *features* moleculares significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de plasma no modo HILIC ESI(+/-) de eluição (*continuação*).

Features	Ancova p-value	Ranking	FDR 10%	Conclusão
7.77min : 794.614m/z	0.012863	135	0.081818	SIG
10.40min : 894.545m/z	0.016069	136	0.082424	SIG
11.58min : 568.362m/z	0.016084	137	0.08303	SIG
5.62min : 154.928m/z	0.016566	138	0.083636	SIG
1.16min : 191.020m/z	0.017012	139	0.084242	SIG
4.03min : 863.204m/z	0.018142	140	0.084848	SIG
10.68min : 1050.657m/z	0.019605	141	0.085455	SIG
1.75min : 306.058m/z	0.019756	142	0.086061	SIG
1.22min : 561.199m/z	0.020173	143	0.086667	SIG
9.91min : 547.150m/z	0.022815	144	0.087273	SIG
9.05min : 204.123m/z	0.02358	145	0.087879	SIG
10.79min : 493.072m/z	0.025432	146	0.088485	SIG
0.70min : 595.492m/z	0.025915	147	0.089091	SIG
1.21min : 349.039m/z	0.027042	148	0.089697	SIG
2.46min : 278.950m/z	0.027117	149	0.090303	SIG
5.37min : 133.960m/z	0.029636	150	0.090909	SIG
13.35min : 248.996m/z	0.030043	151	0.091515	SIG
1.55min : 367.158m/z	0.033968	152	0.092121	SIG
3.97min : 371.080m/z	0.034635	153	0.092727	SIG
4.31min : 347.082m/z	0.037986	154	0.093333	SIG
10.68min : 750.528m/z	0.038476	155	0.093939	SIG
1.66min : 369.173m/z	0.038514	156	0.094545	SIG
11.19min : 138.055m/z	0.043565	157	0.095152	SIG
11.34min : 174.089m/z	0.044868	158	0.095758	SIG
1.91min : 181.072m/z	0.050057	159	0.096364	SIG
1.26min : 559.185m/z	0.050777	160	0.09697	SIG
7.39min : 503.334m/z	0.054427	161	0.097576	SIG

Tabelas A4- Determinação dos *features* moleculares significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de plasma no modo HILIC ESI(+/-) de eluição.

Features	Ancova p-value	Ranking	FDR 10%	Conclusão
1.83min : 263.104m/z	0.084729	162	0.098182	SIG
14.64min : 159.969m/z	0.133006	163	0.098788	NS
2.21min : 195.051m/z	0.157422	164	0.099394	NS
2.80min : 209.067m/z	0.278082	165	0.1	NS

Legenda: **Sig** – Significativo; **NS** – Não Significativo

Anexo 2

Tabela A5- Resultado do teste de *t-Student* para a análise estatística referente à idade dos indivíduos dos dois grupos (DW e HC), nas amostras de urina

Estatísticas de grupo					
	Grupo	N	Média	Erro Desvio	Erro padrão da média
Idade	DW	23	38.35	11.742	2.448
	GC	37	31.05	9.554	1.571

Teste de amostras independentes			
		Teste de Levene para igualdade de variâncias	Teste-t para Igualdade de Médias
Idade	Variâncias iguais assumidas	0.187	0.011

Tabela A6- Resultado do teste de Qui-quadrado para a análise estatística referente ao gênero dos indivíduos dos dois grupos (DW e HC), nas amostras de urina

		Grupo			Qui-quadrado de Pearson	
		DW	GC	Total		
Gênero	Feminino	Contagem	9	20	29	0.261
		% em lable	39.1%	54.1%	48.3%	
		% do Total	15.0%	33.3%	48.3%	
	Masculino	Contagem	14	17	31	
		% em lable	60.9%	45.9%	51.7%	
		% do Total	23.3%	28.3%	51.7%	
Total	Contagem	23	37	60		
	% em lable	100.0%	100.0%	100.0%		
	% do Total	38.3%	61.7%	100.0%		

Tabela A7- Determinação dos *features* significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de urina no modo RPLC ESI(+/-) de eluição (*continuação*).

Feature	Ancova p-value	Ranking	FDR 10%	Conclusão
19.50min : 210.968m/z	0	1	0.0015625	SIG
21.57min : 421.352m/z	0	2	0.003125	SIG
20.10min : 193.001m/z	1.06248E-13	3	0.0046875	SIG
22.31min : 540.446m/z	5.42011E-13	4	0.00625	SIG
22.77min : 355.158m/z	2.09299E-12	5	0.0078125	SIG
24.60min : 461.262m/z	6.02318E-12	6	0.009375	SIG
23.17min : 1572.924m/z	1.10973E-11	7	0.0109375	SIG
22.78min : 433.231m/z	3.0636E-11	8	0.0125	SIG
19.97min : 190.928m/z	8.70044E-11	9	0.0140625	SIG
18.58min : 265.148m/z	1.31486E-10	10	0.015625	SIG
23.63min : 1630.967m/z	1.19241E-09	11	0.0171875	SIG
21.96min : 331.284m/z	1.38395E-08	12	0.01875	SIG
22.28min : 494.565m/z	1.9155E-08	13	0.0203125	SIG
22.75min : 1514.882m/z	3.53501E-08	14	0.021875	SIG
20.02min : 250.145m/z	1.80951E-05	15	0.0234375	SIG
1.09min : 141.065m/z	0.000265211	16	0.025	SIG
23.57min : 359.315m/z	0.000278064	17	0.0265625	SIG
8.37min : 109.064m/z	0.000308488	18	0.028125	SIG
7.46min : 84.044m/z	0.000436902	19	0.0296875	SIG
3.96min : 188.986m/z	0.003988305	20	0.03125	SIG
11.06min : 243.124m/z	0.004637411	21	0.0328125	SIG
1.44min : 100.075m/z	0.00484496	22	0.034375	SIG
3.03min : 177.023m/z	0.005127409	23	0.0359375	SIG
7.50min : 145.062m/z	0.005175454	24	0.0375	SIG
18.51min : 284.331m/z	0.006293151	25	0.0390625	SIG
23.49min : 393.277m/z	0.008445232	26	0.040625	SIG
6.15min : 187.061m/z	0.020144814	27	0.0421875	SIG

Tabelas 7A- Determinação dos *features* significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de urina no modo RPLC ESI(+/-) de eluição (*continuação*).

Feature	Ancova p-value	Ranking	FDR 10%	Conclusão
6.19min : 143.071m/z	0.02101885	28	0.04375	SIG
15.14min : 495.296m/z	0.025961903	29	0.0453125	SIG
5.39min : 278.070m/z	0.03084931	30	0.046875	SIG
19.19min : 400.378m/z	0.041235019	31	0.0484375	SIG
3.82min : 120.081m/z	0.041709671	32	0.05	SIG
18.46min : 871.264m/z	0.042246574	33	0.0515625	SIG
8.89min : 350.088m/z	0.044124324	34	0.053125	SIG
4.55min : 154.054m/z	0.060023031	35	0.0546875	NS
20.28min : 590.425m/z	0.074891025	36	0.05625	NS
8.50min : 193.035m/z	0.074943744	37	0.0578125	NS
20.40min : 368.424m/z	0.083602462	38	0.059375	NS
8.57min : 201.113m/z	0.096605075	39	0.0609375	NS
17.63min : 194.082m/z	0.097789188	40	0.0625	NS
22.74min : 384.346m/z	0.115322403	41	0.0640625	NS
14.74min : 399.127m/z	0.125278729	42	0.065625	NS
7.84min : 181.051m/z	0.188522549	43	0.0671875	NS
4.43min : 169.096m/z	0.20350523	44	0.06875	NS
3.43min : 115.040m/z	0.256378089	45	0.0703125	NS
22.73min : 338.341m/z	0.330540835	46	0.071875	NS
4.41min : 117.056m/z	0.336252427	47	0.0734375	NS
18.22min : 449.254m/z	0.362194878	48	0.075	NS
13.66min : 331.177m/z	0.375100946	49	0.0765625	NS
7.57min : 178.051m/z	0.421867351	50	0.078125	NS
2.33min : 103.040m/z	0.469215681	51	0.0796875	NS
8.06min : 116.070m/z	0.501014915	52	0.08125	NS
15.03min : 416.315m/z	0.543878024	53	0.0828125	NS
5.71min : 312.130m/z	0.544579065	54	0.084375	NS

Tabelas 7A- Determinação dos *features* significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de urina no modo RPLC ESI(+/-) de eluição.

Feature	Ancova p-value	Ranking	FDR 10%	Conclusão
5.14min : 286.103m/z	0.616782877	55	0.0859375	NS
15.30min : 228.195m/z	0.630221942	56	0.0875	NS
24.45min : 283.264m/z	0.63822559	57	0.0890625	NS
4.97min : 227.997m/z	0.678723135	58	0.090625	NS
6.16min : 159.065m/z	0.679077596	59	0.0921875	NS
1.26min : 119.035m/z	0.801881273	60	0.09375	NS
2.98min : 271.092m/z	0.825595015	61	0.0953125	NS
2.89min : 181.049m/z	0.867038969	62	0.096875	NS
5.45min : 195.076m/z	0.869714702	63	0.0984375	NS
3.07min : 227.997m/z	0.891915857	64	0.1	NS

Legenda: **Sig** – Significativo; **NS** – Não Significativo

Tabela A8- Determinação dos *features* significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de urina no modo HILIC ESI(+/-) de eluição (*continuação*).

Features	ANCOVA p-value	Ranking	FDR 10%	Conclusão
8.91min : 210.931m/z	1.11022E-16	1	0.00122	SIG
9.26min : 430.835m/z	4.21885E-15	2	0.002439	SIG
9.92min : 317.114m/z	2.17604E-14	3	0.003659	SIG
9.26min : 372.866m/z	2.4869E-14	4	0.004878	SIG
8.82min : 102.128m/z	4.10005E-13	5	0.006098	SIG
9.24min : 294.888m/z	4.60409E-13	6	0.007317	SIG
15.12min : 340.874m/z	1.86923E-11	7	0.008537	SIG
16.71min : 149.009m/z	8.0306E-11	8	0.009756	SIG
7.54min : 211.976m/z	3.84821E-10	9	0.010976	SIG
11.20min : 292.891m/z	9.51956E-10	10	0.012195	SIG
0.64min : 179.882m/z	2.05042E-08	11	0.013415	SIG
11.51min : 294.888m/z	2.11689E-08	12	0.014634	SIG
10.97min : 422.813m/z	9.65513E-08	13	0.015854	SIG
6.18min : 318.977m/z	1.16027E-07	14	0.017073	SIG
7.88min : 221.902m/z	1.2569E-07	15	0.018293	SIG
4.23min : 134.894m/z	1.96821E-07	16	0.019512	SIG
3.40min : 279.038m/z	7.47052E-07	17	0.020732	SIG
9.22min : 232.925m/z	1.53891E-06	18	0.021951	SIG
11.01min : 424.813m/z	1.79489E-06	19	0.023171	SIG
4.19min : 136.892m/z	2.32905E-06	20	0.02439	SIG
11.74min : 155.129m/z	2.03437E-05	21	0.02561	SIG
5.32min : 325.048m/z	2.98185E-05	22	0.026829	SIG
8.42min : 363.028m/z	3.01863E-05	23	0.028049	SIG
11.24min : 163.965m/z	3.39409E-05	24	0.029268	SIG
10.86min : 116.928m/z	4.50764E-05	25	0.030488	SIG
5.07min : 257.818m/z	5.14049E-05	26	0.031707	SIG
11.00min : 140.986m/z	0.000179394	27	0.032927	SIG
0.90min : 190.049m/z	0.000207472	28	0.034146	SIG

Tabelas 8A- Determinação dos *features* significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de urina no modo HILIC ESI(+/-) de eluição (*continuação*).

Features	ANCOVA p-value	Ranking	FDR 10%	Conclusão
1.00min : 161.913m/z	0.000738345	29	0.035366	SIG
12.04min : 98.956m/z	0.000783189	30	0.036585	SIG
10.12min : 360.878m/z	0.000938726	31	0.037805	SIG
3.08min : 306.057m/z	0.001423169	32	0.039024	SIG
7.62min : 214.252m/z	0.001654192	33	0.040244	SIG
3.51min : 100.933m/z	0.001736974	34	0.041463	SIG
5.63min : 153.075m/z	0.002341038	35	0.042683	SIG
3.81min : 100.075m/z	0.003382528	36	0.043902	SIG
7.64min : 119.035m/z	0.003538999	37	0.045122	SIG
1.84min : 157.034m/z	0.005149426	38	0.046341	SIG
6.22min : 236.009m/z	0.007058921	39	0.047561	SIG
9.24min : 171.066m/z	0.008040172	40	0.04878	SIG
4.92min : 290.196m/z	0.008262207	41	0.05	SIG
0.59min : 107.967m/z	0.009371408	42	0.05122	SIG
0.70min : 389.089m/z	0.009611754	43	0.052439	SIG
3.79min : 160.096m/z	0.011484178	44	0.053659	SIG
1.83min : 292.903m/z	0.011807573	45	0.054878	SIG
7.31min : 129.019m/z	0.012145336	46	0.056098	SIG
6.46min : 294.830m/z	0.01242629	47	0.057317	SIG
6.78min : 331.169m/z	0.013463728	48	0.058537	SIG
4.31min : 227.024m/z	0.018190676	49	0.059756	SIG
8.35min : 172.957m/z	0.018528104	50	0.060976	SIG
12.53min : 141.066m/z	0.022177325	51	0.062195	SIG
5.30min : 279.990m/z	0.039670019	52	0.063415	SIG
12.24min : 204.123m/z	0.040999225	53	0.064634	SIG
11.63min : 138.055m/z	0.04121364	54	0.065854	SIG
1.81min : 159.075m/z	0.041558764	55	0.067073	SIG
14.38min : 188.175m/z	0.043777406	56	0.068293	SIG

Tabelas 8A- Determinação dos *features* significativos utilizando ANCOVA com valor de $p \leq 0.05$ e FDR 10% das amostras biológicas de urina no modo HILIC ESI(+/-) de eluição.

Features	ANCOVA p-value	Ranking	FDR 10%	Conclusão
10.77min : 246.169m/z	0.049027473	57	0.069512	SIG
1.75min : 197.047m/z	0.049875434	58	0.070732	SIG
5.67min : 159.063m/z	0.049876671	59	0.071951	SIG
4.40min : 182.032m/z	0.055131067	60	0.073171	SIG
3.96min : 245.076m/z	0.056981761	61	0.07439	SIG
5.36min : 151.061m/z	0.06067468	62	0.07561	SIG
7.54min : 352.102m/z	0.066726885	63	0.076829	SIG
1.89min : 224.916m/z	0.070482482	64	0.078049	SIG
5.96min : 137.035m/z	0.083879246	65	0.079268	NS
1.74min : 198.062m/z	0.084841934	66	0.080488	NS
2.52min : 182.032m/z	0.087352054	67	0.081707	NS
4.30min : 146.081m/z	0.090547748	68	0.082927	NS
7.31min : 144.923m/z	0.098934484	69	0.084146	NS
13.37min : 175.118m/z	0.105918265	70	0.085366	NS
2.44min : 117.055m/z	0.107585024	71	0.086585	NS
1.46min : 138.055m/z	0.109820451	72	0.087805	NS
5.12min : 131.034m/z	0.115761346	73	0.089024	NS
6.27min : 310.115m/z	0.117042206	74	0.090244	NS
10.12min : 104.107m/z	0.123417078	75	0.091463	NS
2.66min : 167.055m/z	0.154213348	76	0.092683	NS
2.86min : 137.036m/z	0.157231625	77	0.093902	NS
2.95min : 217.121m/z	0.220912229	78	0.095122	NS
12.85min : 269.124m/z	0.30414232	79	0.096341	NS
10.68min : 128.035m/z	0.313948959	80	0.097561	NS
2.96min : 139.050m/z	0.33793549	81	0.09878	NS
3.13min : 600.283m/z	0.450068208	82	0.1	NS

Legenda: Sig – Significativo; NS – Não Significativo

Anexo 3

Ácido quinurênico

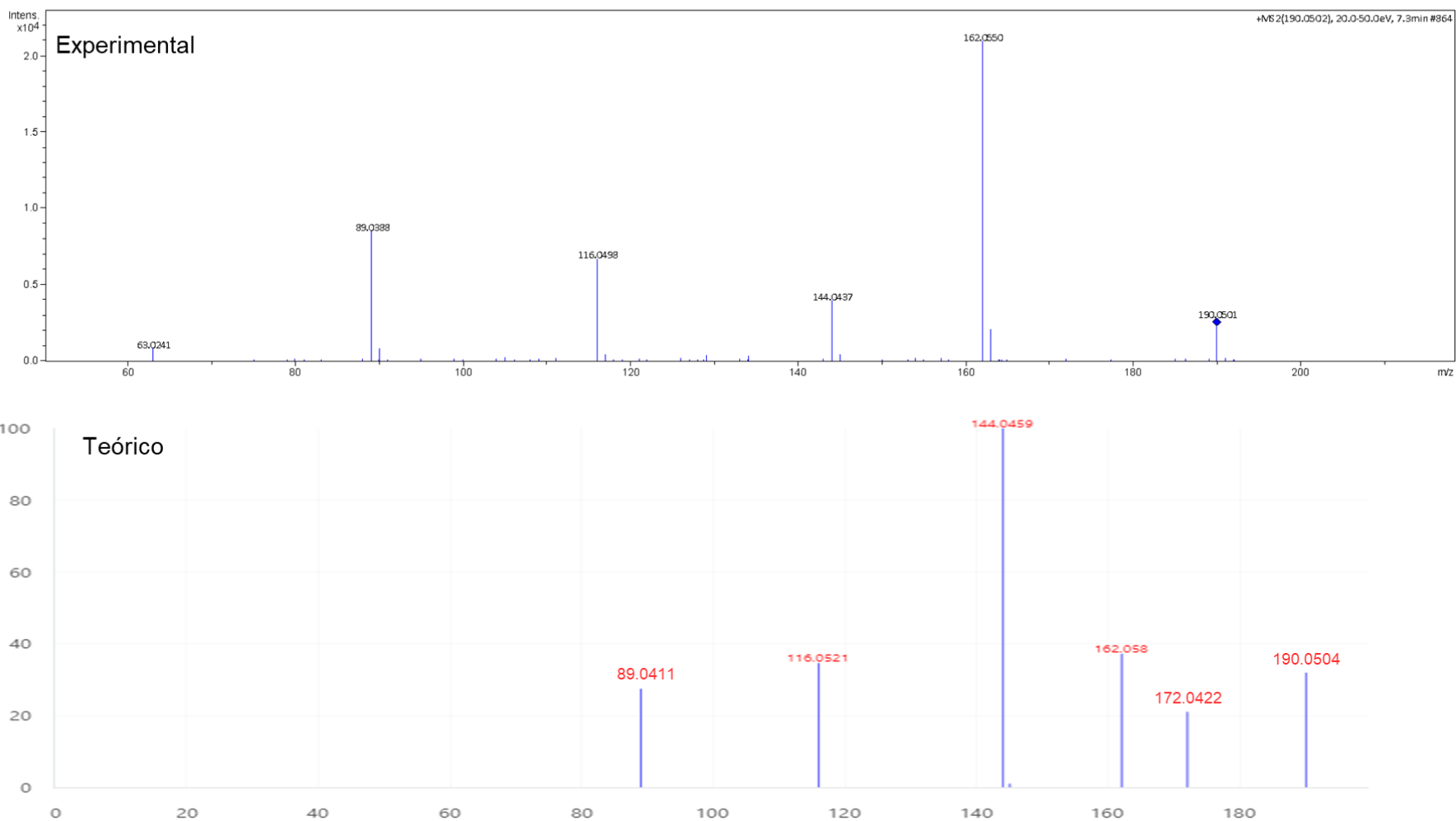


Figura 44- Identificação do ácido quinurênico por meio da comparação entre o espectro experimental obtido nas análises LC-HRMS e o espectro de massas teórico disponível na biblioteca do MoNA.

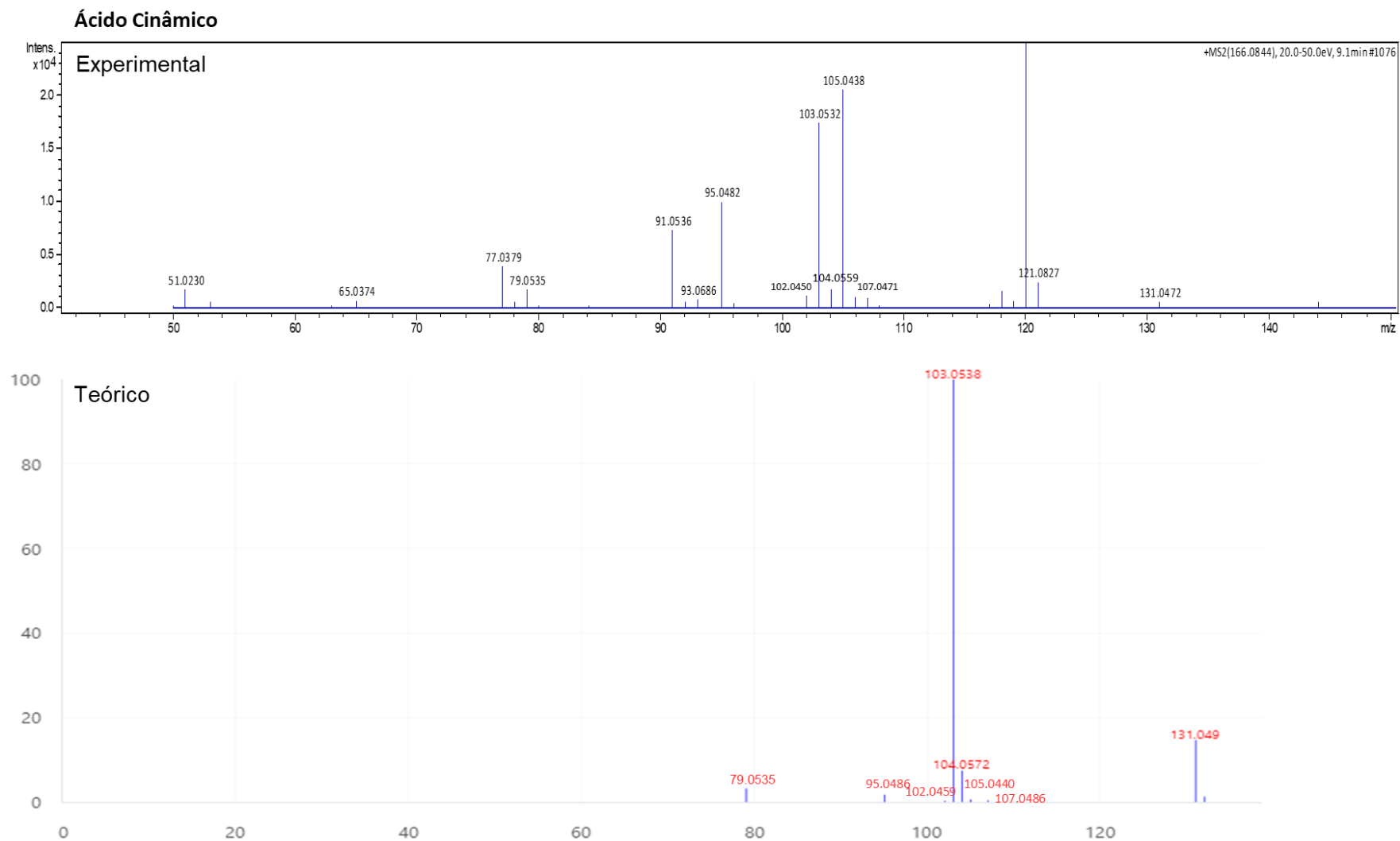


Figura 45- Identificação do ácido cinâmico por meio da comparação entre o espectro experimental obtido nas análises LC-HRMS e o espectro de massas teórico disponível na biblioteca do MoNA.

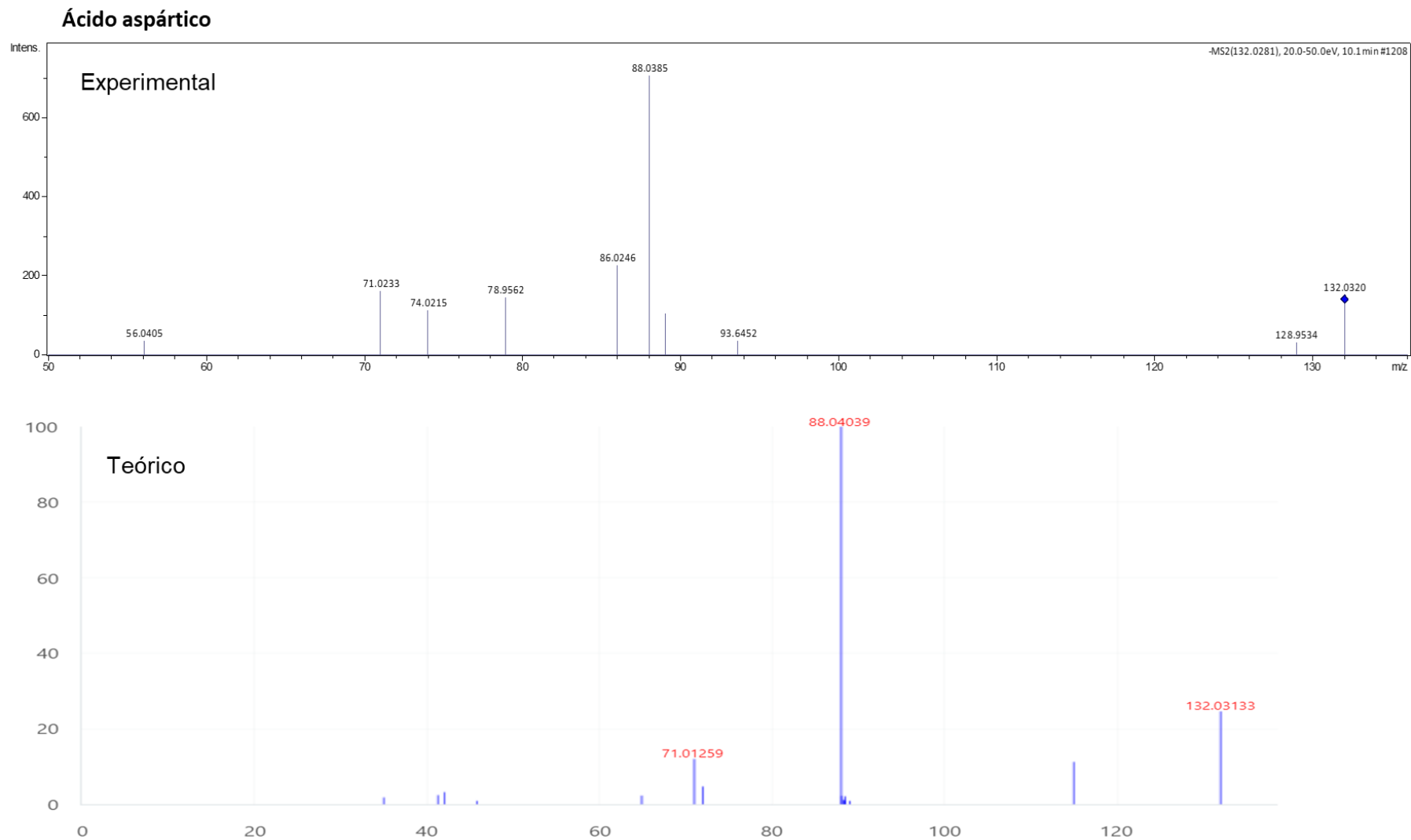


Figura 46- Identificação do ácido aspártico por meio da comparação entre o espectro experimental obtido nas análises LC-HRMS e o espectro de massas teórico disponível na biblioteca do MoNA.

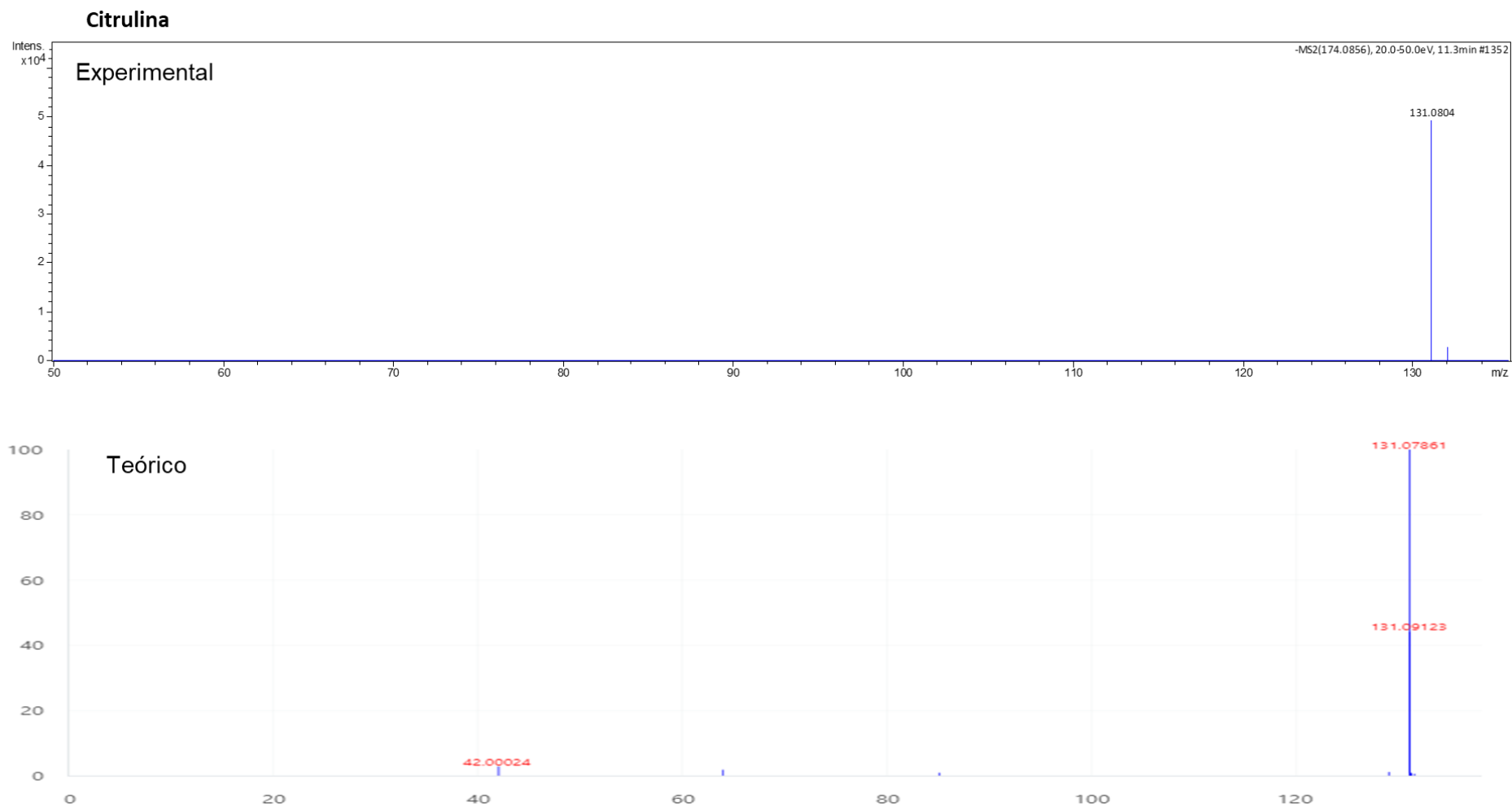


Figura 47- Identificação da citrulina por meio da comparação entre o espectro experimental obtido nas análises LC-HRMS e o espectro de massas teórico disponível na biblioteca do MoNA.

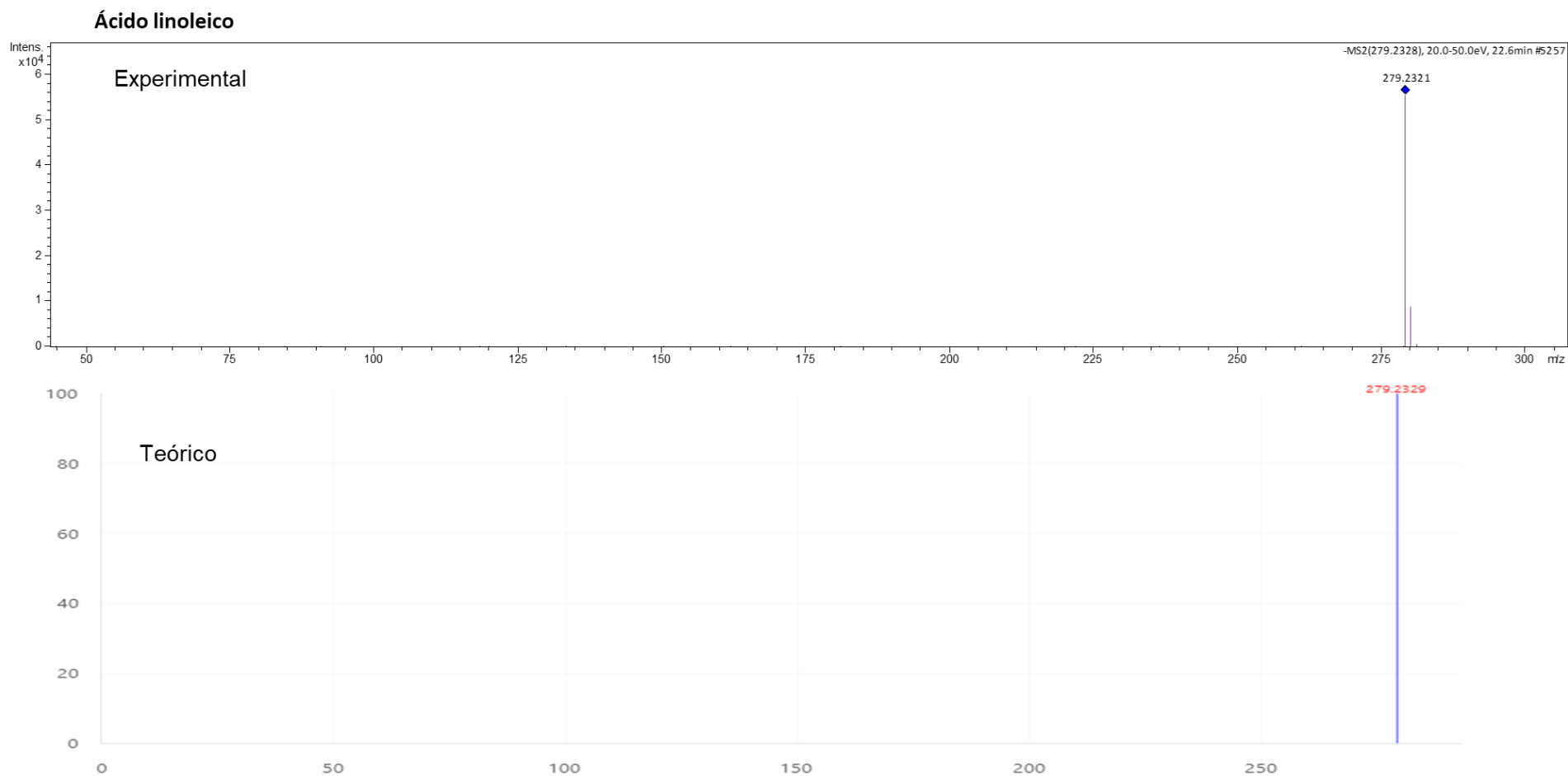


Figura 48- Identificação do ácido linoleico por meio da comparação entre o espectro experimental obtido nas análises LC-HRMS e o espectro de massas teórico disponível na biblioteca do MoNA.

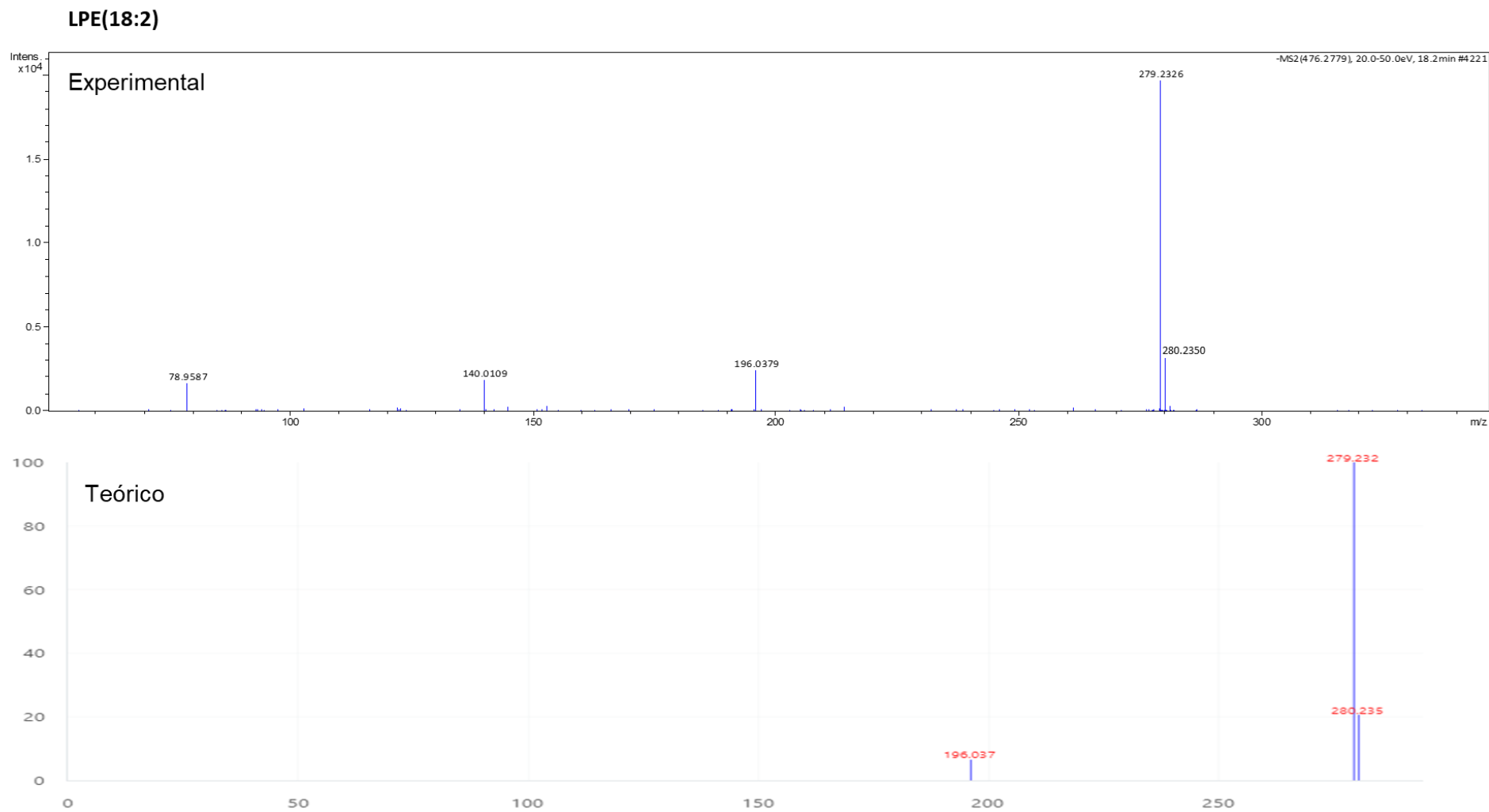


Figura 49- Identificação do LPE (18:2) por meio da comparação entre o espectro experimental obtido nas análises LC-HRMS e o espectro de massas teórico disponível na biblioteca do MoNA.