

UNIVERSIDADE FEDERAL DO AMAZONAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

LUANA QUADROS DE SOUZA LEÃO

**ANÁLISE DO PERFIL PROTEÔMICO E METABOLÔMICO DE
TECIDOS COM LESÕES PRECURSORAS DO CÂNCER DE COLO DE
ÚTERO EM MULHERES DO AMAZONAS**

MANAUS

2019

LUANA QUADROS DE SOUZA LEÃO

**ANÁLISE DO PERFIL PROTEÔMICO E METABOLÔMICO DE
TECIDOS COM LESÕES PRECURSORAS DO CÂNCER DE COLO DE
ÚTERO EM MULHERES DO AMAZONAS**

Dissertação de mestrado submetida ao Programa de Pós-Graduação em Química, do Instituto de Ciências Exatas da Universidade Federal do Amazonas como requisito para obtenção do título de Mestre em Química na área de concentração Produtos Naturais e Biomoléculas

Orientadora: Dr^a Priscila Ferreira de Aquino

MANAUS

2019

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

L687a Leão, Luana Quadros de Souza
Análise do perfil proteômico e metabolômico de tecidos com
lesões precursoras do câncer de colo de útero em mulheres do
Amazonas / Luana Quadros de Souza Leão . 2019
162 f.: il. color; 31 cm.

Orientadora: Priscila Ferreira de Aquino
Dissertação (Mestrado em Química - Orgânica) - Universidade
Federal do Amazonas.

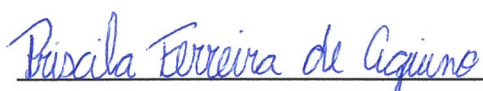
1. Neoplasias intraepiteliais cervicais. 2. Espectrometria de
massas. 3. Proteômica. 4. Metabolômica. I. Aquino, Priscila Ferreira
de. II. Universidade Federal do Amazonas III. Título

Análise do perfil proteômico e metabolômico de tecidos com
lesões precursoras do câncer de colo de útero em mulheres do
Amazonas

LUANA QUADROS DE SOUZA LEÃO

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Química,
do Instituto de Ciências Exatas da Universidade Federal do Amazonas como
requisito parcial para a obtenção do título de Mestre em Química.

Aprovada em 29 de novembro de 2019.



Profa. Dra. PRISCILA FERREIRA DE AQUINO
Presidente/Orientador - FioCruz/AM



Profa. Dra. CRISTINA MARIA BORBOREMA DOS SANTOS
Membro Externo - UFAM



Prof. Dr. RAFAEL GARRETT DA COSTA
Membro Externo - IQ/UFRJ

Dedico a todas as mulheres que lutam contra o c ncer de colo
de  tero e suas les es precursoras.

AGRADECIMENTOS

Agradeço, primeiramente a Deus, pela energia, coragem, saúde e garra para a realização dessa dissertação. Obrigada por ser meu controle, meu porto seguro e por me proporcionar momentos de grande aprendizagem tanto emocional quanto profissional durante essa etapa da minha vida. Obrigada por ter colocado pessoas maravilhosas em meu caminho, me dado forças para não desistir, oportunidades para errar, aprender e persistir para a realização desse sonho.

À minha família, especialmente às rainhas da minha vida, minha mãe Maria Alcilene Leonel Quadros e minha avó Josefa Leonel Quadros. Agradeço por terem me criado com tanto amor e carinho, por sempre me apoiarem nos momentos mais difíceis e decisivos da minha vida, eu amo vocês. Agradeço também ao meu pai, Ruy de S. Leão Filho por todo o apoio.

Ao meu avô Renato Oliveira Quadros (*in memoriam*) por ter sido o melhor avô do mundo, mesmo que infelizmente não tenhamos convivido por muito tempo. O senhor foi a minha motivação para a realização desse trabalho.

À minha orientadora Dra. Priscila Ferreira de Aquino por acreditar na minha capacidade para a execução desse estudo. Obrigada pela orientação, pelas noites sem dormir, pelo companheirismo, dedicação e por me proporcionar experiências enriquecedoras no âmbito científico. Obrigada por todas as palavras de incentivo, pelas críticas e pelos conselhos durante esse período. Sou grata por deixar essa singela frase marcada em minha vida: “Vai ser com emoção hein”. Literalmente, você estava certa.

Às pacientes da FCECON, que aceitaram participar voluntariamente desse estudo.

Aos médicos Jonas Balan de Padua, Zeliene Araújo de Souza, Gilson José Correa e Anderson Ferreira Gonçalves pela colaboração nas coletas das amostras, pelo cuidado, paciência e compreensão durante esse período. Agradeço pelos apelidos carinhosos: Pesquisadora mirim, Papa-cone e principalmente Caça-cone, no qual me tornou conhecida no centro cirúrgico; isso sem dúvidas ficará guardado em minhas lembranças. Agradeço também a todos os residentes e enfermeiros que contribuíram para a execução dessa etapa.

Ao meu grupo de pesquisa: Sofia Angiole, Juliana Leal, Kemily Nunes, Genésio

Pontes, Valderjane Aprígio, Jéssica Kelly Silva, Cláudia Araújo, Renato Hosanah e Marissa Anjos, por todo o auxílio, apoio, gargalhadas, reuniões, comilanças e principalmente por toda a contribuição durante o processo de desenvolvimento dessa dissertação. Sou muito grata por ser membro dessa equipe.

Ao Prof. Dr. Afonso Duarte Leão de Souza pela oportunidade e auxílio no início das análises de metabolômica. Agradeço em especial à Dra Elzalina Ribeiro pela motivação, ensinamentos, risadas, paciência e companheirismo durante essa etapa.

Ao Prof Dr. Rafael Garrett da Costa pelo apoio, contribuição, conhecimentos transmitidos e por me proporcionar uma grande experiência na área de metabolômica. Agradeço toda a sua equipe do Laboratório de Metabolômica (LABMETA): Evelyn Penaloza, Paula de Araújo, Ana Carolina Rosa, Carol Silva, Hygor de Souza, Vivian Noronha e Clarice Torres, pelo acolhimento, troca de conhecimentos, café, carinho e por toda a atenção e cuidado recebido. Sou grata especialmente à Dra. Marina Amaral por toda atenção, ensinamentos, ajuda, companheirismo e auxílio nas análises de metabolômica.

Ao doutorando Felipe Mansoldo e sua orientadora Prof^a Dra Alane Beatriz Vermelho, do Laboratório BIOINOVAR-UFRJ, pelo auxílio nas análises de metabolômica.

Ao Prof Dr. Fábio César Sousa Nogueira, pela contribuição, conhecimentos transmitidos e pela oportunidade de aprendizagem no ramo da abordagem proteômica. Agradeço toda a sua equipe do Laboratório de Proteômica (LABPROT), especialmente ao Joseph Evaristo, Gabriel Reis, Yasmin Araújo e Adrieli Alcaires, pelo apoio, parceria e auxílio durante as análises de proteômica.

Ao amigo MSc. Wagner Picanço Moreira pela paciência, ensinamentos, cuidado e apoio em todos os momentos durante o período desse estudo. Sem dúvidas, você foi essencial para essa conquista.

Aos amigos Josiana Moreira, Laiane Silva e Joed Augusto, por todo apoio e companheirismo adquiridos durante o período desse mestrado. Sou grata pela existência de cada um, pelas distrações, risadas e choros que fizeram parte desse momento.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Código de Financiamento 001, pela bolsa de estudos concedida.

Ao Programa de Pós-Graduação em Química da Universidade Federal do Amazonas (UFAM) e ao ILMD/FIOCRUZ-Amazonas, pela estrutura e participação

intensa nesse estudo.

Ao PPSUS, FAPEAM e o Governo do Estado do Amazonas, pelo financiamento do projeto.

À FCECON, pela estrutura e auxílio nas coletas de amostras.

Por fim, a todos que contribuíram de forma direta ou indiretamente para a realização desse trabalho. Muito obrigada.

*“Não é necessário ver toda a estrada. Apenas dê
o primeiro passo”.*

Martin Luther King Jr.

RESUMO

O câncer é uma das doenças que causa mais mortes no mundo. Dentre os cânceres mais diagnosticados em mulheres no Brasil, especialmente no Amazonas, encontra-se o câncer de colo de útero. Esse geralmente é antecedido por lesões precursoras conhecidas como neoplasias intraepiteliais cervicais (NICs) de grau 1, 2 e 3, sendo os estágios 2 e 3 considerados de alto grau. A fim de prevenir a progressão dessas lesões precursoras a células cancerígenas, é necessário um melhor entendimento sobre o processo de evolução das NICs bem como buscar novos métodos capazes de identificá-las e diagnosticá-las previamente. Atualmente, a utilização das abordagens proteômica e metabolômica vem crescendo na área da pesquisa oncológica, permitindo a investigação de possíveis alterações e/ou interações de proteínas e metabólitos em decorrência de uma patologia. Nesse sentido, o presente estudo teve o objetivo de realizar análises do perfil proteico e metabólico em tecidos com NICs de alto grau, obtidos de mulheres voluntárias na Fundação Centro de Controle de Oncologia do Estado do Amazonas (FCECON). Para isso, amostras de tecidos com lesões de alto grau e suas respectivas margens adjacentes foram coletadas e submetidas a análise proteômica e metabolômica. Em proteômica, utilizou-se a abordagem proteômica *shotgun* para a inferência do perfil proteico, sendo a quantificação feita por *label-free* através dos cromatogramas de íons extraídos (XICs). Em metabolômica, foram empregadas análises estatísticas multivariadas como PCA e PLS-DA bem como a abordagem *Random Forest* a fim de verificar os metabólitos diferenciais presentes nessas amostras. Como resultado, verificou-se na análise proteômica que cada tipo de tecido possui diferenças entre si, apresentando proteínas unicamente expressas em cada condição e ainda proteínas superexpressas correlacionadas ao câncer cervical. Alguns exemplos foram: cortactina, mucina-5B, calnexina e cofilina-1, onde a maior abundância destas podem estar relacionadas à persistência ou progressão molecular dessas lesões precursoras. Já na análise metabolômica, verificou-se alterações nos perfis de metabólitos entre os tecidos de NIC 2 e NIC 3 e suas respectivas margens adjacentes, sendo condizente com os dados obtidos em proteômica. Além disso, foi possível observar a presença de alguns metabólitos que colaboram para a distinção desses tecidos como a histamina, presente principalmente nas margens adjacentes às lesões precursoras do câncer de colo de útero (CCU). Portanto, os dados encontrados nesse trabalho demonstram características interessantes, onde esperamos que possam contribuir para o entendimento mais abrangente em nível molecular sobre os possíveis mecanismos de progressão em tais células estudadas.

Palavras-chave: Neoplasias intraepiteliais cervicais, espectrometria de massas, proteômica, metabolômica.

ABSTRACT

Cancer is one of the diseases with high mortality in the world. Among the most diagnosed cancers in women in Brazil, especially in the Amazonas, cervical cancer stands out. This type of cancer is usually preceded by precursor lesions known as grade 1, 2 and 3 cervical intraepithelial neoplasia (CIN), where the stages 2 and 3 are considered high grade. In order to prevent the progression of these precursor lesions to cancer cells, a better understanding of the process of CIN evolution is needed, as well as search new methods capable of identifying and diagnosing them. Currently, the use of proteomic and metabolomic approaches has been growing in the area of cancer research, allowing the investigation of possible alterations or/and interactions of proteins and metabolites due to a pathology. In this sense, the present study aims to analyze the proteomic and metabolomic profile of tissues with high-grade CIN, obtained from volunteer women at the Oncology Control Foundation Center of the Amazonas state (FCECON). For this, samples of tissues with high-grade lesions and their respective adjacent margins were collected and submitted to proteomic and metabolomic analysis. In proteomics, the shotgun proteomic approach was used to infer the protein profile with label-free quantitation by extracted ion chromatograms (XICs). In metabolomics, multivariate statistical analyzes such as PCA and PLS-DA were employed as well as the Random Forest approach to verify the differential metabolites present in these samples. As a result, it was verified in the proteomic analysis that each type of tissue has differences, presenting proteins abundant in each condition and overexpressed proteins correlated to cervical cancer. Some examples were cortactin, mucin-5B, calnexin and cofilin-1, which may be related to the persistence or molecular progression of these precursor lesions. In the metabolomic analysis, it was observed changes in the metabolite profiles between the tissues of CIN 2 and CIN 3 and their respective adjacent margins, being consistent with the data obtained in proteomics. In addition, it was noted the presence of some metabolites that can contribute to the distinction of these tissues such as histamine, present mainly in the margins adjacent to the precursor lesions of the CCU. Therefore, the data found in this work demonstrate interesting features which we hope to contribute to a broader molecular understanding of the possible mechanisms of progression in such cells studied.

Keywords: Cervical intraepithelial neoplasms, mass spectrometry, proteomics, metabolomics.

LISTA DE FIGURAS

Figura 1 - Representação da evolução das lesões precursoras até o câncer de colo de útero.	29
Figura 2 - Diferença entre a cérvix do colo uterino de mulheres com o colo uterino normal, com a NIC de baixo grau, com a NIC de alto grau e com câncer cervical.	30
Figura 3 – Fluxograma das ciências ômicas mais conhecidas e pesquisadas atualmente, sendo associadas ao grupo de biomoléculas específicas de cada abordagem.	32
Figura 4 – Representação de métodos de análise de dados utilizados na metabolômica: PCA e PLS-DA.	34
Figura 5 - Esquema dos principais componentes de um espectrômetro de massas.	35
Figura 6 - Representação esquemática do funcionamento da fonte de ionização por ESI.	36
Figura 7 - Representação esquemática do analisador quadrupolo.	37
Figura 8 – Representação do analisador <i>Orbitrap</i> , onde a detecção da corrente elétrica gerada pelo movimento dos íons é convertida em um espectro de massas através da transformada de Fourier.	38
Figura 9 – Representação geral das três principais abordagens proteômicas.	42
Figura 10 – Colo do útero de uma paciente definindo a área com a cerne da lesão e a margem de ressecção.	47
Figura 11 - Representação de um cromatograma de íons extraídos (XIC).	51
Figura 12 – Eletroforese unidimensional das amostras selecionadas para o estudo proteômico. Legenda: Amostras 7 e 19 – lesão NIC 3; Amostras 8 e 20 – margens NIC 3. Amostras 21 e 17 – lesão NIC 2; Amostras 22 e 18 – margens NIC 2.	58
Figura 13 – Eletroforese unidimensional das amostras selecionadas de NIC 2 para o estudo proteômico. Legenda: Amostras 3, 13 e 11 – lesão NIC 2; Amostras 4, 14 e 12 – margem NIC 2.	59
Figura 14 - Eletroforese unidimensional das amostras selecionadas de NIC 3 para o estudo proteômico. Legenda: Amostras 1, 9 e 5 – lesão de NIC 3; Amostras 2, 10 e 6 – margens de NIC 3.	60
Figura 15 – Rede de interações da proteína KRT17 através do programa STRING 11.0. As linhas na cor cinza mostram confiança de evidência na interação, sendo que quanto mais intenso é a cor e a linha, mais alta é a confiança (<i>score</i> > 0,9).	68
Figura 16 - Rede de interações da proteína RPLP1 através do programa STRING 11.0. As linhas na cor cinza mostram confiança de evidência na interação, sendo que quanto mais intenso é a cor e a linha, mais alta é a confiança (<i>score</i> > 0,9).	69
Figura 17 – Diagrama da distribuição diferencial de proteínas abundantes nas lesões de NIC 2 (lesão2) e NIC 3 (lesão3) associadas a proliferação celular.	70
Figura 18 – Rede de interações da proteína colágeno de cadeia alfa-1 (COL6A1) através do programa STRING 11.0. As linhas na cor cinza mostram confiança de evidência na interação, sendo que quanto mais intenso é a cor e a linha, mais alta é a confiança (<i>score</i> > 0,9).	74
Figura 19 – Rede de interações da proteína ribossômica 60S L3 (RPL3) e L31 (RPL31) através do programa STRING 11.0. As linhas na cor cinza mostram confiança de	

evidência na interação, sendo que quanto mais intenso é a cor e a linha, mais alta é a confiança (<i>score</i> > 0,9).....	76
Figura 20 – Diagrama de distribuição diferencial de proteínas abundantes nas margens adjacentes de NIC 2 (margem2) e NIC 3 (margem3) associadas a processos de regulação biológica.	77
Figura 21 – Rede de interações da proteína TMPO através do programa STRING 11.0. As linhas na cor cinza mostram confiança de evidência na interação, sendo que quanto mais intenso é a cor e a linha, mais alta é a confiança (<i>score</i> > 0,9).	80
Figura 22 – Rede de interações da proteína HMGB1 através do programa STRING 11.0. As linhas na cor cinza mostram confiança de evidência na interação, sendo que quanto mais intenso é a cor e a linha, mais alta é a confiança (<i>score</i> > 0,9).	82
Figura 23 – Rede de interações da proteína CDC42 através do programa STRING 11.0. As linhas na cor cinza mostram confiança de evidência na interação, sendo que quanto mais intenso é a cor e a linha, mais alta é a confiança (<i>score</i> > 0,9).	84
Figura 24 – Rede de interações da proteína FTL através do programa STRING 11.0. As linhas na cor cinza mostram confiança de evidência na interação, sendo que quanto mais intenso é a cor e a linha, mais alta é a confiança (<i>score</i> > 0,9).	86
Figura 25 - Representação esquemática de duas abordagens empregadas em metabolômica, a <i>targeted</i> e a <i>untargeted</i>	92
Figura 26 – Metabólitos diferenciais encontrados ao comparar pacientes com câncer de colo de útero, neoplasias intraepiteliais cervicais e pacientes normais.	95

LISTA DE TABELAS

Tabela 1 - Distribuição proporcional dos dez tipos de câncer mais incidentes estimados para o biênio 2018 – 2019, em mulheres na Região Norte, exceto os de pele não melanoma.	27
Tabela 2 – Perfil socioeconômico das mulheres envolvidas no estudo proteômico. ..	53
Tabela 3 – Perfil clínico das mulheres envolvidas no estudo proteômico (Parte 1). ...	54
Tabela 4 – Perfil clínico das mulheres envolvidas no estudo proteômico (Parte 2). ...	54
Tabela 5 - Lista de algumas das proteínas diferencialmente abundantes quando comparados lesões de NIC 2 e lesões de NIC 3*.	71
Tabela 6 - Lista de algumas das proteínas diferencialmente abundantes quando comparados margem adjacente de NIC 2 com de NIC 3*.	77
Tabela 7 – Lista de algumas proteínas diferencialmente abundantes quando comparados cerne da lesão de NIC 2 e margem adjacente de NIC 2*.	87
Tabela 8 – Lista de algumas proteínas diferencialmente abundantes quando comparados cerne da lesão de NIC 3 e margem adjacente de NIC 3*.	88
Tabela 9 – Ressuspensão das amostras da otimização para o estudo metabolômico.	97
Tabela 10 – Ressuspensão e volume de injeção das amostras para o estudo metabolômico.	98
Tabela 11 – Perfil socioeconômico das mulheres envolvidas no estudo metabolômico.	102
Tabela 12 – Perfil clínico das mulheres envolvidas no estudo metabolômico (Parte 1).	102
Tabela 13 – Perfil clínico das mulheres envolvidas no estudo metabolômico (Parte 2).	103

LISTA DE QUADRO

Quadro 1 – Características gerais das mulheres envolvidas no estudo proteômico. .	52
--	----

LISTA DE GRÁFICOS

Gráfico 1 - Cromatogramas de uma réplica técnica de amostras do cerne da lesão de NIC 2 (preto) e NIC 3 (vermelho) durante uma análise de 180 minutos.....	56
Gráfico 2 - Cromatogramas de uma réplica técnica de amostras da margem adjacente de NIC 2 (preto) e NIC 3 (vermelho) durante uma análise de 180 minutos.....	56
Gráfico 3 - Resultado da análise no <i>Gene Ontology</i> referente ao cerne das lesões de NIC 2 e NIC 3, mostrando as funções das proteínas identificadas em porcentagem nos tecidos analisados.	62
Gráfico 4 - Gráfico do resultado da análise no <i>Gene Ontology</i> referente as margens adjacentes de NIC 2 e NIC 3, mostrando as funções das proteínas identificadas em porcentagem nos tecidos analisados.	63
Gráfico 5 - Diagrama de Venn referente às proteínas expressas nas lesões de NIC 2 e NIC 3. Legenda: Proteínas expressas unicamente nas lesões de NIC 2 (roxo), proteínas expressas unicamente em NIC 3 (vermelho) e proteínas compartilhadas em ambas (rosa).	66
Gráfico 6 - Diagrama de Venn referente às proteínas expressas nas margens adjacentes de NIC 2 e NIC 3. Legenda: Proteínas exclusivamente abundantes nas margens de NIC 2 (azul), exclusivas em margens de NIC 3 (amarelo) e compartilhadas em ambas as condições (azul-amarelo).....	73
Gráfico 7 - Diagrama de Venn referente às proteínas expressas nas lesões de NIC 2 e suas margens adjacentes. Legenda: Proteínas expressas exclusivamente nas lesões de NIC 2 (rosa), proteínas expressas apenas na margem (lilás) e proteínas compartilhadas em ambas (roxo).	79
Gráfico 8 - Diagrama de Venn referente às proteínas expressas nas lesões de NIC 3 e suas margens adjacentes. Legenda: Proteínas expressas exclusivamente nas lesões de NIC 3 (rosa), proteínas expressas apenas na margem (azul turquesa) e proteínas compartilhadas em ambas (azul).	83
Gráfico 9 - Distribuição das mulheres selecionadas para o estudo metabolômico de acordo com o tipo de lesão.	101
Gráfico 10 - Cromatogramas das amostras relacionados a metodologia 1 (M1), 2 (M2) e 3 (M3) – teste 1.	104
Gráfico 11 - Cromatogramas das amostras relacionados a metodologia 1 (M1), 2 (M2) e 3 (M3) – teste 2.	105
Gráfico 12 – Cromatogramas das amostras do estudo metabolômico.....	106
Gráfico 13 - PCA das amostras NIC 2 – margem (verde) e lesão (vermelho), e NIC 3 – margem (marrom) e lesão (laranja).....	107
Gráfico 14 - PLS-DA das amostras NIC 2 - margem (azul), NIC 3 - lesão (cinza), NIC 2 – lesão (laranja) e NIC 3 – margem (laranja), apresentando semelhanças entre as amostras de lesão de NIC 2 e a margem adjacente de NIC 3..	108
Gráfico 15 - PLS-DA em 3D das amostras NIC 2 - margem (azul), NIC 3 - lesão (verde), NIC 3 – margem (laranja) e NIC 2 – lesão (laranja), apresentando semelhanças entre as amostras de lesão de NIC 2 e a margem adjacente de NIC 3.	108
Gráfico 16 - PLS-DA referente as variáveis de impacto presentes nas amostras de T e M.	109

Gráfico 17 - PCA das amostras de lesões NIC 2 e NIC 3 (representados por T) e margens adjacentes à lesão NIC 2 e NIC 3 (representados por M).....	110
Gráfico 18 - Possíveis outliers baseados na distância ortogonal referentes as amostras de T e M.	110
Gráfico 19 - PLS-DA das amostras das lesões NIC 2 e NIC 3 (T – em laranja) e margens adjacente NIC 2 e NIC 3 (M – em azul), após remoção dos outliers.....	111
Gráfico 20 - PLS-DA em 3D das amostras de lesões NIC2/NIC3 (T – em laranja) e margens adjacentes de NIC 2/NIC 3 (M – em azul).	112
Gráfico 21 - <i>Loadings</i> referentes as variáveis de maior relação com a componente 1 (em A) e a componente 2 (em B).	112
Gráfico 22 - Classificação das amostras por <i>Random Forest</i> através das variáveis mais importantes para o modelo.....	113
Gráfico 23 - Diferença significativa das variáveis do modelo <i>Random Forest</i> com o grupo T (lesões NIC 2 e NIC 3) sendo superior ao grupo M (margens NIC 2 e NIC 3) ($p < 0,05$) (Parte 1).....	114
Gráfico 24 - Diferença significativa das variáveis do modelo <i>Random Forest</i> com o grupo T (lesões NIC 2 e NIC 3) sendo superior ao grupo M (margens NIC 2 e NIC 3) ($p < 0,05$) (Parte 2).....	115
Gráfico 25 - Diferença significativa da variável do modelo <i>Random Forest</i> com o grupo M (margens NIC 2 e NIC 3) sendo superior ao grupo T (lesões NIC 2 e NIC 3) ($p < 0,05$).	115
Gráfico 26 - Espectros de MS2 da histamina. Em preto, o espectro MS2 experimental com m/z de 112.08694. Em vermelho, o espectro de MS2 do banco de dados MONA utilizando o <i>software</i> livre MS-DIAL.....	116
Gráfico 27 - Espectro de massas em MS2 do íon de 112,05 m/z ($[M+H]^+$).	117
Gráfico 28 - Diferença significativa da variável X11.869x112.08693 no modelo <i>Random Forest</i> com o grupo M (margens NIC 2 e NIC 3) sendo superior ao grupo T (lesões NIC 2 e NIC 3) ($p < 0,05$).	117

LISTA DE SÍMBOLOS E ACRÔNIMOS

1-DE	-	Eletroforese unidimensional
2D-DIGE	-	Eletroforese em gel de diferença bidimensional (do inglês <i>two-dimensional Difference Gel Electrophoresis</i>)
2-DE	-	Eletroforese bidimensional
Aa	-	Aminoácido
ACN	-	Acetonitrila
ACTR2	-	Proteína 2 relacionada à actina
ACTR3	-	Proteína 3 relacionada à actina
AGC	-	Controle de ganho automático (do inglês <i>Automatic Gain Control</i>)
ATP	-	Adenosina trifosfato
BSA	-	Albumina sérica bovina
CAF	-	Cirurgia de alta frequência
CAPG	-	Proteína de captação de macrófagos
CCU	-	Câncer de colo de útero
CDC42	-	Homólogo da proteína 42 de controle da divisão celular
CEC	-	Carcinoma espinocelular cervical
CECS	-	Carcinoma de células escamosas cervicais
CE-MS	-	Eletroforese Capilar acoplada a Espectrometria de Massas (do inglês <i>Capillary electrophoresis–mass spectrometry</i>)
CEP	-	Comitê de Ética em Pesquisa
CL-EMAR	-	Cromatografia líquida acoplada à espectrometria de massas de alta resolução
CML	-	Citologia em meio líquido
COL6A1	-	Colágeno de cadeia alfa-1
COL6A3	-	Colágeno de cadeia alfa-3
DC	-	Voltagem de corrente direta (do inglês <i>direct-current</i>)
DDA	-	Aquisição dependente de dados (do inglês <i>Data Dependent Acquisition</i>)
DMBT1	-	Excluído no tumor cerebral maligno 1 (do inglês <i>Deleted in malignant brain tumors 1 protein</i>)
DNA	-	Ácido Desoxirribonucleico
DP	-	Desvio padrão
DST	-	Doenças sexualmente transmissíveis
DTT	-	Ditiotreitol
ESI	-	Ionização por Electrospray (do inglês <i>Electrospray ionization</i>)
FCECON	-	Fundação Centro de Controle de Oncologia do Estado do Amazonas
FCGBP	-	Proteína de ligação a IgGFc (do inglês <i>IgGFc-binding protein</i>)
FDR	-	Taxa de Falsos Positivos (do inglês <i>False Discovery Rate</i>)
FIGO	-	Federação Internacional de Ginecologia e Obstetrícia (do inglês <i>International Federation of Gynecology and Obstetrics</i>)
FTH1	-	Cadeia pesada da ferritina
FT-ICR	-	Ressonância ciclotrônica de íons com transformada de Fourier (do inglês <i>Fourier-transform ion cyclotron resonance</i>)
FTL	-	Cadeia leve da ferritina (do inglês <i>Ferritin light chain</i>)
GC-MS	-	Cromatografia gasosa acoplada a espectrometria de massas
HCD	-	Dissociação por colisão de alta energia (do inglês <i>High Energy Collision Dissociation</i>)

HIV	-	Vírus da Imunodeficiência Humana
HMGB1	-	Proteína B1 do grupo de alta mobilidade
HPV	-	Papilomavírus Humano (do inglês <i>Human Papiloma Virus</i>)
HSA	-	Albumina sérica humana
HSIL	-	Lesões intraepiteliais escamosas de alto grau
IAA	-	Iodoacetamida
INCA	-	Instituto Nacional de Câncer José Alencar Gomes da Silva
IUCC	-	União internacional para controle do câncer (do inglês <i>International Union Cancer</i>)
KDa	-	Quilodalton
KRT1	-	Queratina 1 do citoesqueleto, tipo II
KRT17	-	Queratina 17 do citoesqueleto, tipo I
Kv	-	Quilovolt
LC	-	Cromatografia líquida
LC-MS	-	Cromatografia líquida acoplada a espectrometria de massas
LCN2	-	Lipocalina associada a gelatinase neutrófila
LMNB1	-	Lâmina B1
LSIL	-	Lesões intraepiteliais escamosas de baixo grau
m/z	-	Razão massa/carga
MALDI	-	Ionização/dessorção a laser auxiliada por matriz (do inglês <i>Matrix Assisted Laser Desorption Ionization</i>)
MARCKS	-	Substrato c-cinase rico em alanina miristoilada (do inglês <i>Myristoylated alanine-rich C-kinase substrate</i>)
MIC	-	Carcinoma microinvasivo
mM	-	Milimolar
MS	-	Espectrometria de massas (do inglês <i>Mass Spectrometry</i>)
MS/MS	-	Espectrometria de massas sequencial (do inglês <i>Tandem Mass Spectrometry</i>)
MS1	-	Perfil do espectro de massas
MS2	-	Varredura de íons produtos
MTBE	-	Éter metil-terc-butílico
NCE	-	Energia de colisão normalizada
NH₄HCO₃	-	Bicarbonato de amônio
NIC 1	-	Neoplasia intraepitelial cervical de grau I
NIC 2	-	Neoplasia intraepitelial cervical de grau II
NIC 3	-	Neoplasia intraepitelial cervical de grau III
NICs	-	Neoplasias intraepiteliais cervicais
nLC-MS/MS	-	Nanocromatografia líquida acoplada ao espectrômetro de massas sequencial
OMS	-	Organização Mundial da Saúde
OT	-	<i>Orbitrap</i>
p/v	-	razão peso/volume
p53	-	Proteína supressora tumoral
PC	-	Componente Principal
PCA	-	Análise de Componentes Principais (do inglês <i>Principal Component Analysis</i>)
PCOLCE	-	Proteína potenciadora de procolágeno C-endopeptidase 1 (do inglês <i>Procollagen C-endopeptidase enhancer 1</i>)
PLS-DA	-	Análise Discriminante por Mínimos Quadrados Parciais (do inglês <i>Partial Least Squared Discriminant Analysis</i>)

Q	-	Quadrupolo
QC	-	Controle de Qualidade (do inglês <i>Quality Control</i>)
RF	-	Fase reversa (do inglês <i>Reverse Phase</i>)
RF*	-	<i>Random Forest</i>
RMN ¹H	-	Ressonância magnética nuclear de Hidrogênio
RNA	-	Ácido ribonucleico
RPL10	-	Proteína ribossômica 60S L10
RPL3	-	Proteína ribossômica 60 S L3
RPL31	-	Proteína ribossômica 60 S L31
RPL9	-	Proteína ribossômica 60S L9
RPLP1	-	Proteína ribossômica ácida 60S P1 (do inglês <i>60S acidic ribosomal protein P1</i>)
SDS	-	Dodecil sulfato de sódio
SDS-PAGE	-	Eletroforese em gel de poliacrilamida (do inglês <i>polyacrylamide gel electrophoresis</i>)
SELDI	-	Ionização/dessorção por laser em superfície melhorada (do inglês <i>Surface-enhanced laser desorption/ionization</i>)
SERPINA7	-	Globulina de ligação à tiroxina (do inglês <i>Thyroxine-binding globulin</i>)
Sisgen	-	Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado
STRING	-	Ferramenta de pesquisa para a recuperação de genes/proteínas em interação (do inglês <i>Search Tool for the Retrieval of Interacting Genes/Proteins</i>)
TBS	-	Sistema Bethesda (do inglês <i>The Bethesda System</i>)
TCLE	-	Termo de Consentimento Livre e Esclarecido
TEMED	-	Tetrametiletilenodiamina
TFA	-	Ácido trifluoroacético
TMPO	-	Polipeptídeo 2 associado à lâmina
TNM	-	Sistema de Classificação dos Tumores Malignos
TOF	-	Tempo de voo (do inglês <i>Time of flight</i>)
TPD52	-	Proteína tumoral D52
Tris-HCl	-	Tris (hidroximetil) aminometano - Ácido Clorídrico (do inglês <i>Tris (hydroxymethyl) aminomethane - Hydrochloric acid</i>)
UPLC-QTOF-MS	-	<i>Ultra-high performance liquid chromatography-quadrupole time-of-flight mass spectrometry</i>
v/v	-	razão volume/volume
XIC	-	Cromatograma de íon extraído (do inglês <i>Extracted-Ion Chromatogram</i>)

SUMÁRIO

INTRODUÇÃO	23
OBJETIVOS	25
Geral	25
Específicos	25
Capítulo I	26
Câncer de colo de útero, NICs e Ciências 'Ômicas'	26
1.1 Câncer de colo de útero	27
1.2 Neoplasias intraepiteliais cervicais (NICs)	29
1.3 Ciências 'Ômicas'	31
1.4 Quimiometria aplicada a Ciências 'Ômicas'	33
1.5 Espectrometria de massas	35
Capítulo II	40
Análise Proteômica	40
2.1 Proteômica	41
2.2 Proteômica aplicada ao câncer	43
2.3 MATERIAIS E MÉTODOS	45
2.3.1 Seleção de pacientes	45
2.3.2 Coleta de dados	46
2.3.3 Coleta e preservação das amostras	46
2.3.4 Estudo proteômico das NICs de alto grau	47
2.3.4.1 Pulverização e extração proteica	47
2.3.4.2 Eletroforese unidimensional	47
2.3.4.3 Coloração com Coomassie R-250	48
2.3.4.4 Digestão proteica em solução	48
2.3.4.5 Preparação das amostras para análise nLC – MS/MS	49
2.3.4.6 Aquisição de dados por nanocromatografia líquida acoplada ao espectrômetro de massas sequencial (nLC – MS/MS)	49
2.3.4.7 Análise dos dados	50
2.4 RESULTADOS E DISCUSSÃO	52
2.4.1 Características gerais dos indivíduos para o estudo proteômico	52
2.4.2 Estudo proteômico	55
2.4.3 Análise do gel unidimensional (1D)	57
2.4.4 Identificação proteica	60

2.4.5 Análise diferencial das NICs e suas respectivas margens adjacentes	66
Capítulo III	90
Análise Metabolômica	90
3.1 Metabolômica	91
3.2 Metabolômica aplicada ao câncer	93
3.3 MATERIAIS E MÉTODOS	96
3.3.1 Seleção de pacientes	96
3.3.2 Estudo metabolômico das NICs de alto grau	96
3.3.2.1 Preparo das amostras	96
3.3.2.2 Método tipo <i>Matyash</i> (Adaptação)	98
3.3.2.3 Aquisição de dados por cromatografia líquida acoplada à espectrometria de massas de alta resolução (CL – EMAR)	98
3.3.2.4 Análise de dados	99
3.4 RESULTADOS E DISCUSSÃO	101
3.4.1 Características gerais dos indivíduos para o estudo metabolômico	101
3.4.2 Otimização do preparo de amostras para análise metabolômica	104
3.4.3 Estudo metabolômico	106
Capítulo IV	119
Considerações Finais	119
REFERÊNCIAS BIBLIOGRÁFICAS	122
ANEXOS	148
ANEXO A: Parecer do Comitê de ética em pesquisa – UFAM	149
ANEXO B: Parecer do Comitê de ética em Pesquisa – FCECON	151
ANEXO C: Cadastro no Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado (Sisgen)	153
APÊNDICES	154
APÊNDICE A: Termo de Consentimento Livre e Esclarecido (TCLE)	155
APÊNDICE B: Questionário	157
APÊNDICE C: Características gerais das mulheres envolvidas no estudo metabolômico	160

PREFÁCIO

Esta dissertação foi dividida em quatro capítulos. O Capítulo 1 aborda uma revisão sobre o câncer de colo de útero, as neoplasias intraepiteliais cervicais e as ciências 'ômicas'. O Capítulo 2 e o Capítulo 3 apresentam as abordagens 'ômicas' utilizadas no decorrer desta dissertação; especificamente, no Capítulo 2 aborda a parte da análise proteômica e o Capítulo 3 apresenta a parte metabolômica. Por fim, o Capítulo 4 demonstra as considerações finais do trabalho perante todas as análises realizadas.

INTRODUÇÃO

O número de casos de câncer em todo o mundo tem crescido notavelmente a partir do século passado, sendo um dos problemas de saúde pública mais importantes da atualidade. No Brasil, o câncer é a segunda causa de mortalidade da população, superado apenas pelas doenças cardiovasculares (OMS, 2017)

Dentre os tipos de cânceres, o quarto mais comum entre as mulheres, não somente no Brasil, mas em todo o mundo é o câncer de colo de útero (CCU). Segundo o Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA), é estimado um risco de 15,43 novos casos a cada 100 mil mulheres no Brasil durante o biênio 2018-2019, sendo o primeiro em incidência na Região Norte (INCA, 2017).

O câncer de colo de útero evolui através de lesões precursoras conhecidas como lesões intraepiteliais escamosas de alto grau. Estas quando são diagnosticadas de forma precoce e tratadas adequadamente, podem impedir a progressão para o câncer (SPECK et al., 2015).

Neste contexto, uma variedade de técnicas está sendo utilizada para compreender os diversos mecanismos biológicos relacionados ao processo tumoral que são responsáveis por alterações desenvolvidas no decorrer da doença. Dentre elas, pode-se citar as chamadas ciências ômicas. Essas compreendem abordagens voltadas ao entendimento dos sistemas biológicos através da caracterização e identificação de biomoléculas de um determinado grupo, como as proteínas (proteômica) e metabólitos (metabolômica); bem como analisar as interações envolvidas entre eles a partir do auxílio de técnicas experimentais, *softwares* e bancos de dados que proporcionam estudos cada vez mais abrangentes (CANUTO et al., 2018; ESPINDOLA et al., 2010).

A análise proteômica e metabolômica tem dado novas oportunidades para descobertas significativas a respeito dos processos normais e anormais que acontecem no organismo humano, inclusive com relação a doenças como o câncer e suas lesões precursoras. Uma das ferramentas mais utilizadas para a análise dessas alterações é a espectrometria de massas (MS), a qual possibilita identificar e quantificar uma multiplicidade de moléculas simultaneamente através da relação massa/carga (m/z) (BANACH et al., 2017).

De maneira geral, análises em nível molecular são bastante interessantes para o combate ao câncer e suas lesões, visto que podem demonstrar características ou

biomoléculas específicas de uma dada condição biológica. Tal fato é especialmente atraente para a área clínica e/ou oncológica, e pode resultar em uma futura aplicabilidade para o paciente quanto ao rastreamento, diagnóstico diferencial, monitoramento da doença, entre outros. Logo, o presente trabalho utilizou a espectrometria de massas nas análises proteômicas e metabolômicas de tecidos das lesões precursoras do câncer de colo de útero, com o propósito de auxiliar na compreensão do processo proteico e metabólico presente nesses tecidos, que possam estar associados à uma possível evolução dessa doença.

Assim, este projeto se justificou pelos seguintes fatores:

- Na Região Norte existe uma grande incidência de casos de mulheres com câncer de colo de útero. Especialmente no Amazonas, a taxa bruta é de 40,97 a cada 100 mil mulheres de acordo com a última estimativa, sendo o estado com maior incidência. Para a cidade de Manaus, o panorama é ainda mais preocupante, com uma taxa bruta de 58,37 a cada 100 mil mulheres;
- Este é o primeiro estudo com enfoque em análises das lesões precursoras do câncer cervical utilizando a abordagem proteômica e metabolômica no Amazonas;
- Em Manaus, o índice de casos de lesões precursoras de alto grau vem aumentando no decorrer dos anos, segundo dados da Fundação Centro de Controle de Oncologia (FCECON);
- Estudos voltados a identificar biomoléculas diferenciais que auxiliem no diagnóstico precoce e na compreensão sobre a evolução das neoplasias intraepiteliais cervicais ao câncer de colo de útero são importantes, visto que podem contribuir para a sobrevida do paciente;
- Resultados presentes na literatura relatam o uso da espectrometria de massas juntamente com as ciências ômicas, em especial a proteômica e metabolômica, como um instrumento promissor para diagnósticos clínicos de diferentes doenças, incluindo o câncer;
- Adicionalmente, a utilização de ferramentas analíticas e a bioinformática aplicada a área médica possui carência na região norte, necessitando ser impulsionada a formação de recursos humanos e novos projetos de pesquisa nessa temática.

OBJETIVOS

Geral

- Analisar o perfil proteico e metabólico do tecido de pacientes com lesões precursoras do câncer de colo de útero.

Específicos

- Caracterizar o perfil proteico e metabólico das lesões precursoras bem como das regiões adjacentes à lesão por espectrometria de massas;
- Investigar por técnicas estatísticas as possíveis diferenças de metabólitos entre as amostras;
- Buscar proteínas e metabólitos diferenciais que sejam específicos das lesões precursoras de alto grau e/ou associados ao processo de malignização.

Capítulo I

Câncer de colo de útero, NICs
e Ciências ‘Ômicas’

1.1 Câncer de colo de útero

O câncer de colo de útero (CCU), conhecido também como câncer cervical, é uma doença que se desenvolve lentamente, tendo fases pré-malignas. Esse tipo de câncer possui caráter invasivo e é antecedido por diversas modificações no epitélio cervical original, envolvendo desde as lesões contidas no epitélio do colo uterino até a sua evolução a um carcinoma invasivo. Assim, as células pré-cancerígenas se transformam em tumores malignos acometendo a cérvix (subdividida em endocérvix e ectocérvix), dando origem ao câncer cervical (KOSS; GOMPEL, 2006).

Esse, por sua vez, é o quarto tipo de câncer mais frequente mundialmente na população feminina, sendo responsável por 7,5% de todas as mortes por câncer entre as mulheres (BRAY et al., 2018; DE MARTEL et al., 2017). Para 2018, foi estimado cerca 570.000 casos e 311.000 mortes em todo o mundo (BRAY et al., 2018). Adicionalmente, é a principal causa de morte entre mulheres que vivem em países em desenvolvimento (MENDONÇA et al., 2008; THULER, 2008), sendo o décimo lugar no *ranking* de mortalidade mundial (BRAY et al., 2018).

No Brasil, estima-se cerca de 16.370 novos casos de CCU para cada ano durante o biênio 2018-2019, sendo o risco estimado de 15,43 casos a cada 100 mil mulheres, ocupando a terceira posição. Na Região Norte, especialmente, o câncer de colo de útero é o primeiro mais incidente (25,62/100 mil), com exceção dos tumores de pele não melanoma (Tabela 1). No Amazonas, o panorama ainda é mais preocupante na qual estima-se 840 novos casos, uma taxa bruta de 40,97 a cada 100 mil mulheres. Desses, 640 novos casos apenas para a capital de Manaus, com uma taxa bruta de 58,37 a cada 100 mil mulheres (INCA, 2017).

Tabela 1 - Distribuição proporcional dos dez tipos de câncer mais incidentes estimados para o biênio 2018 – 2019, em mulheres na Região Norte, exceto os de pele não melanoma.

	Localização Primária	Casos	%
Mulheres	Colo do Útero	2.300	24,8%
	Mama Feminina	1.730	18,6%
	Cólon e Reto	660	7,1%
	Traqueia, Brônquio e Pulmão	520	5,6%
	Estômago	480	5,2%
	Leucemias	310	3,3%
	Ovário	270	2,9%
	Glândula Tireoide	270	2,9%
	Sistema Nervoso Central	270	2,9%
	Corpo do Útero	200	2,2%

Fonte: Adaptado de (INCA, 2017).

Esta alta incidência do câncer cervical está associada ao principal fator etiológico dessa doença, a infecção persistente do papilomavírus humano (HPV). Essa infecção pode possibilitar o surgimento das lesões precursoras do câncer cervical ou até mesmo o câncer, de forma incomum (SANCHES et al., 2017). Existem vários subtipos oncológicos do vírus HPV, dentre eles o 16, 18, 31, 35, 39, 45, 51, 52, 58, 59, 68 e 82 são considerados de alto risco ao desenvolvimento oncogênico, sendo o HPV-16 e o HPV-18 os mais prevalentes (CHELIMO et al., 2013).

Além dos fatores associados à infecção pelo HPV, outros aspectos relacionados à genética, imunidade e ao comportamento sexual podem corroborar para a persistência dessa infecção e assim surgir as lesões precursoras (neoplasias intraepiteliais cervicais – NICs) ou o câncer. Sendo assim, o histórico familiar, a iniciação sexual precoce, tabagismo, a multiplicidade de parceiros sexuais, comprometimento da imunidade, paridade, utilização de métodos contraceptivos hormonais, nível socioeconômico, além de coinfeções por *Chlamydia trachomatis* e HIV são citados também na literatura como fatores de risco desse tipo de câncer (ZARDO et al., 2014).

A progressão do estadiamento da doença depende especificamente da fase em que é detectado o câncer de colo de útero, o que influencia na sobrevida das pacientes. Quanto ao grau da doença, Nakagawa e colaboradores (2011) relata que a sobrevida das mulheres que apresentam casos da doença em metástase é desfavorável, com uma taxa de sobrevida de 28,9%. Já na ausência de metástase, a taxa de sobrevida se eleva, sendo 80%. Em adição, essa taxa aumenta significativamente quando a doença é detectada nos estágios I e II (93,5% e 100%, respectivamente), mostrando que se o câncer for detectado e tratado em sua fase inicial, há mais chances de sobrevivência e cura das paciente (NAKAGAWA et al., 2011).

Atualmente, como forma de prevenção à doença, existe a vacina quadrivalente contra o HPV que o Ministério da Saúde adotou e que confere proteção contra os genótipos de baixo risco (HPV 6 e 11) e de alto risco (HPV 16 e 18) (BRASIL, 2014). Já a triagem para o diagnóstico do câncer de colo de útero pode ser realizada por diversos métodos, dentre eles: citologia convencional, conhecido como teste de Papanicolaou (RIBEIRO; ANDRADE, 2016), citologia em meio líquido (CML) (SILVA et al., 2018) e detecção do DNA do HPV (MITTELDORF, 2016). Contudo, a forma de rastreamento mais utilizada no setor público é o exame de citologia convencional (Papanicolaou), que possui baixa sensibilidade (50%), com números de falso-negativos que varia de 20 a 40% (STABILE et al., 2012). De acordo com a Política Nacional de Atenção Integral à

Saúde da Mulher (2004), a prevenção do câncer cervical bem como o diagnóstico precoce e seu tratamento ainda necessita da implantação de medidas que envolvam a mobilização das mulheres, melhoramento nos sistemas de informação e investimento tecnológico. Assim, ainda é preciso buscar por novas estratégias que possam contribuir para a detecção desse tipo de câncer e principalmente das lesões precursoras.

1.2 Neoplasias intraepiteliais cervicais (NICs)

O câncer de colo uterino é precedido por lesões precursoras, conhecidas como Neoplasias Intraepiteliais Cervicais (NICs). Estas se desenvolvem na região do trato genital inferior e possui como característica a perda da arquitetura do epitélio cervical (CAIXETA, 2014). Estas neoplasias começam pela ação da infecção pelo HPV na cérvix normal. Sua persistência pode iniciar alterações morfológicas celulares que caracterizam as lesões precursoras da carcinogênese cervical, que se não forem diagnosticadas e tratadas precocemente podem evoluir ao câncer de colo de útero (Figura 1) (CAIXETA, 2014; SOCIETY, 2018). Este processo, da infecção às lesões precursoras e posteriormente à doença invasiva, leva em média de 10 – 15 anos, porém há relatos do desenvolvimento do câncer cervical em períodos mais curtos (HILDESHEIM et al., 1999; SOCIETY, 2018).

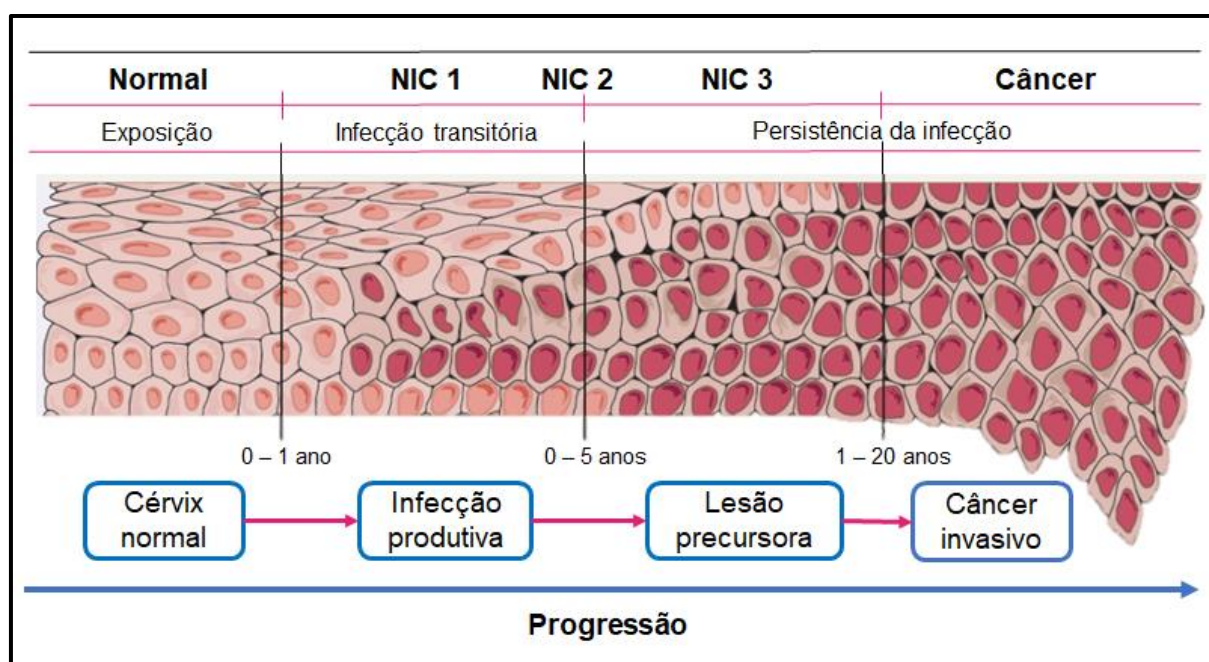


Figura 1 - Representação da evolução das lesões precursoras até o câncer de colo de útero.

Fonte: Adaptado de (CROW, 2012).

Essas lesões podem ser classificadas em graus histopatológicos através da espessura do epitélio que está sendo agredido pelas alterações. De acordo com Richart (1973), as NICs são graduadas tendo em vista a intensidade das anormalidades analisadas histologicamente. Assim, a neoplasia intraepitelial de baixo grau (NIC 1) ocorre quando acomete um terço do epitélio, no qual há perda de diferenciação e alterações celulares nas células imaturas, sendo considerada displasia leve. Já em NIC 2 é comprometido dois terços do epitélio, incluindo as camadas basal e suprabasal e NIC 3 ocupa todo o epitélio (CAIXETA, 2014; CARVALHO; QUEIROZ, 2010).

Com relação a classificação pela nomenclatura do Sistema de Bethesda (TBS – *The Bethesda System*), essas neoplasias são divididas em dois termos: lesões intraepiteliais escamosas de baixo grau (LSIL) e lesões intraepiteliais escamosas de alto grau (HSIL). As LSIL correspondem as neoplasias intraepiteliais de grau leve (NIC 1). Já as HSIL correspondem as neoplasias intraepiteliais de grau moderado (NIC 2) e de alto grau (NIC 3) (Figura 2) (JÚNIOR; LIMA; SILVA, 2018; SOUZA et al., 2004).

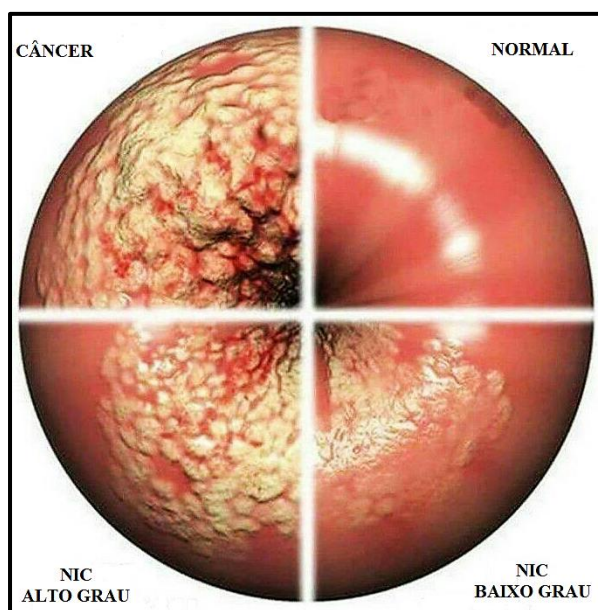


Figura 2 - Diferença entre a cérvix do colo uterino de mulheres com o colo uterino normal, com a NIC de baixo grau, com a NIC de alto grau e com câncer cervical. **Fonte:** *falasil.com*

A triagem para o diagnóstico precoce dessas lesões é realizada através de exames ginecológicos, citopatológicos e colposcópicos. Já o tratamento é feito de acordo com o tipo de lesão (INCA, 2011). No caso das NICs 2 e 3, são realizadas cirurgias sob anestesia através de um procedimento denominado de conização ou

exérese da zona de transformação. Este consiste na remoção da lesão intraepitelial cervical do colo do útero no formato de cone (KIGURE et al., 2018). Dentre os principais meios de excisão, tem-se a conização com bisturi a frio e a cirurgia de alta frequência (CAF) (LIMA et al., 2011).

A infecção persistente do papilomavírus humano (HPV) é indicado como um dos fatores responsáveis dessas neoplasias, onde aproximadamente 90% dos casos de NIC 2 e NIC 3 e dos carcinomas invasivos do colo estão envolvidos pela infecção do HPV de alto risco (BRUNI et al., 2010; MURILLO et al., 2009; ROSA et al., 2009a).

O HPV subtipos 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 e 68 são significativamente associados com NIC e a neoplasia invasiva (OLIVEIRA, 2011). A causa dessas transformações está ligada ao mecanismo de ação do vírus HPV em dois genes, o E6 e E7. O gene E6 é capaz de conectar-se à proteína p53, estimulando a degradação proteolítica de forma rápida. Já o gene E7 possui a capacidade de se unir com a proteína supressora de tumor denominada proteína retinoblastoma (pRb), tornando-a inativa. Adicionalmente, quando ocorrem alterações nas células provenientes do vírus HPV no gene E7, a proteína p16^{INK4a} é expressa e pode ser detectada através da imuno-histoquímica, sendo considerada como marcador para diagnóstico das lesões intraepiteliais escamosas de alto grau (DOORBAR et al., 2015; JUNIOR; GIRALDO; GONÇALVES, 2006; OLIVEIRA, 2011).

Um estudo interessante com relação a regressão e progressão das neoplasias intraepiteliais cervicais foi realizado por Giaccio e colaboradores (2010), que acompanhou 63 pacientes com idade média de 28,8 anos contendo a presença de LSIL pela histologia durante 18 meses. Em resumo, das 63 pacientes estudadas com LSIL na primeira consulta, 9 (14,28%) pacientes evoluíram em algum momento do acompanhamento por 18 meses para HSIL, enquanto a lesão regrediu nas demais (85,72%) (GIACCIO et al., 2010).

1.3 Ciências 'Ômicas'

Nas últimas décadas, avanços significativos na área da pesquisa biomédica têm sido estimulados através de diversas abordagens, entre elas têm-se as ciências 'ômicas'. Estas referem-se a técnicas que possuem como objetivo caracterizar e quantificar em grande escala biomoléculas de um determinado grupo, bem como suas interações através de técnicas experimentais, auxílio de bancos de dados e softwares que permitem

o fornecimento de grandes conjuntos de dados relacionados aos diversos níveis de organização biológica (BARH; BLUM; MADIGAN, 2016; STAGLJAR, 2016).

Essas tecnologias englobam quatro grupos principais de campos de estudo: a genômica, a transcriptômica, a proteômica e a metabolômica (Figura 3). A genômica é a abordagem utilizada para o estudo do genoma, onde obtêm-se dados referentes à sequência completa do material genético, ou seja, o DNA (ácido desoxirribonucleico) do indivíduo a ser estudado. A transcriptômica corresponde ao estudo do transcriptoma (RNAs) das células, na qual os RNAs possuem a capacidade de determinar quais genes são expressos e como os níveis de expressão podem se alterar no decorrer da vida do organismo, sendo diferente do DNA pois esses se mantem estático. A proteômica é o estudo do conteúdo proteico, compreendendo as funções além de elucidar a identidade destas. Já a metabolômica envolve a análise de um conjunto de moléculas (metabólitos) contidas no metabolismo celular presente nas células, fluidos, tecidos corporais, entre outros (BARH; BLUM; MADIGAN, 2016; SOUZA; RHODEN; PAMPHILE, 2014).

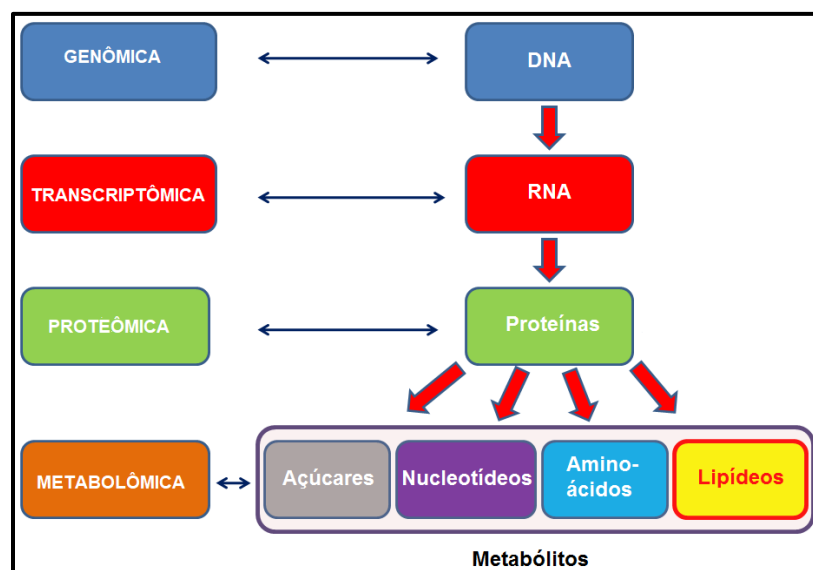


Figura 3 – Fluxograma das ciências ômicas mais conhecidas e pesquisadas atualmente, sendo associadas ao grupo de biomoléculas específicas de cada abordagem.

Fonte: Adaptado de temasbio.ufscar.br

O interesse pelas ciências ômicas vem aumentando entre a comunidade científica. Estas têm sido empregadas no estudo de microrganismos (SÁNCHEZ et al., 2013), medicamentos (MATTHEWS; HANISON; NIRMALAN, 2016), biotecnologia marinha (ABID et al., 2018), na avaliação de riscos de produtos químicos (SAUER et al.,

2017) e em doenças como o câncer, na qual o foco é no perfil e validação dessas técnicas para a detecção da doença utilizando tecido, soro, saliva, linhas celulares e/ou urina (RAI et al., 2018). Assim, neste projeto, será utilizado como campo de estudo duas abordagens das ciências ômicas para a análise de tecidos das lesões precursoras do câncer de colo de útero: a proteômica e a metabolômica.

1.4 Quimiometria aplicada a Ciências ‘Ômicas’

Com o avanço de métodos analíticos com alto desempenho e resolução, obteve-se a aquisição de uma grande quantidade de dados, especialmente nas áreas relacionadas às ciências ‘ômicas’. Consequentemente, é comum a coleta de uma ampla variedade de variáveis por amostra. Com isso, tornou-se cada vez mais importante a mineração dos dados adquiridos utilizando tratamentos estatísticos, visando extrair as informações mais relevantes e validadas para o respectivo estudo (XU; HOEFSLOOT; SMILDE, 2011) (SOUSA, 2013).

Atualmente, as análises multivariadas vêm contribuindo para esse fim, sendo bastante aplicada na área das ciências ômicas especialmente em metabolômica, com o objetivo de reduzir a dimensionalidade dos resultados obtidos (GORROCHATEGUI et al., 2016). Nas análises multivariadas, tanto a classificação quanto a distinção de entidades ou metabólitos responsáveis pela diferenciação dos grupos entre as amostras são feitas a partir da avaliação do conjunto da matriz de dados obtidos durante as análises realizadas. Assim, quando o interesse é verificar essas correlações, são empregados métodos não supervisionados como a análise de componentes principais – do inglês *Principal Component Analysis* (PCA) e métodos supervisionados como a análise discriminante por mínimos quadrados parciais – do inglês *Partial Least Squared Discriminant Analysis* (PLS-DA) (SUGIMOTO et al., 2012).

A análise de componentes principais consiste na transformação dos dados com alta dimensão para uma menor dimensão, programando-os em um subespaço de dimensões reduzidas, porém mantendo a maior variação possível dos dados originais. Esse processo repete-se até que os conjuntos de dados possam ser mostrados normalmente em duas ou três dimensões (NYAMUNDANDA; BRENNAN; GORMLEY, 2010). Com isso, a visualização das amostras é simplificada pela sua distribuição no espaço do componente principal (PC), utilizando os gráficos de pontuação. A distância Euclidiana entre as amostras individuais em *score plots* mostra o grau de variação

sistemática nos perfis de metabólitos entre as amostras. Já as *loading plots* demonstram a contribuição de metabólitos individuais relacionados a cada PC, como mostrado na Figura 4 (BOCCARD; VEUTHEY; RUDAZ, 2010).

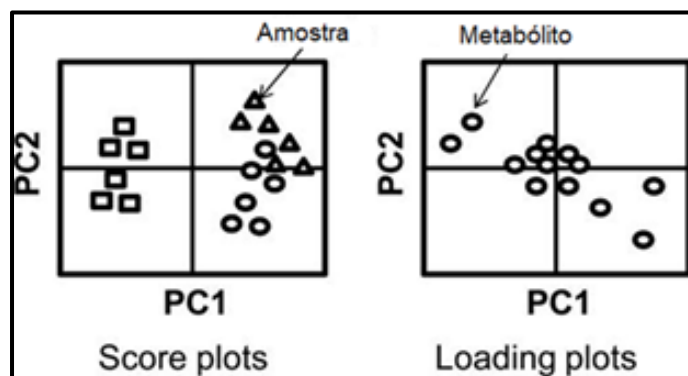


Figura 4 – Representação de métodos de análise de dados utilizados na metabolômica: PCA e PLS-DA. **Fonte:** SUGIMOTO et al., 2012.

Já a análise discriminante por mínimos quadrados parciais é um método de classificação supervisionado, onde a obtenção do agrupamento observado leva em consideração as classes previamente determinadas pela análise, ou seja, o agrupamento das amostras ocorre a partir de uma informação específica fornecida à análise (SANTOS, 2015). Similarmente às plotagens de carregamento no PCA, o PLS-DA também visa a redução de dimensionalidade; porém, o conjunto de amostras com variáveis multidimensionais são reduzidas as chamadas Variáveis Latentes (também chamadas de PCs). Essas sempre apontam na direção de maior variância sobre uma determinada informação fornecida ao sistema. Assim, no *loading plots* observa-se a covariância e a correlação entre os metabólitos juntamente com a classe estabelecida (Figura 4), isto é, o PLS-DA correlaciona a variância do espaço multidimensional (matrix X) com uma resposta informada (matriz Y) (SANTOS, 2015; SUGIMOTO et al., 2012; WOLD; SJÖSTRÖM; ERIKSSON, 2001). De forma geral, o PLS-DA têm sido bastante utilizado para o aprimoramento de separação entre grupos distintos em metabolômica (BOCCARD; VEUTHEY; RUDAZ, 2010; FARIA et al., 2011).

1.5 Espectrometria de massas

A espectrometria de massas (MS) é uma técnica analítica bastante utilizada em algumas ciências ômicas em combinação com métodos de separação e com o auxílio da bioinformática (ESPINDOLA et al., 2010). Essa possibilita identificar uma multiplicidade de moléculas simultaneamente através da relação massa/carga (m/z); permitindo assim a determinação da massa molecular, caracterização estruturalmente das moléculas, além da realização de análises qualitativas e quantitativas referentes aos compostos da mistura a ser estudada (BANACH et al., 2017).

O espectrômetro de massas é composto por uma fonte de íons, que é responsável pelo processo de ionização dos analitos; um analisador de massas, onde ocorre o processo de separação de íons através da relação massa/carga (m/z), mediante aplicações de campos elétricos e/ou magnéticos; e o detector, que é responsável pela detecção e amplificação dos íons para então serem transformados em sinais elétricos e assim processados no sistema de dados (LANÇAS, 2009), conforme mostra a Figura 5.

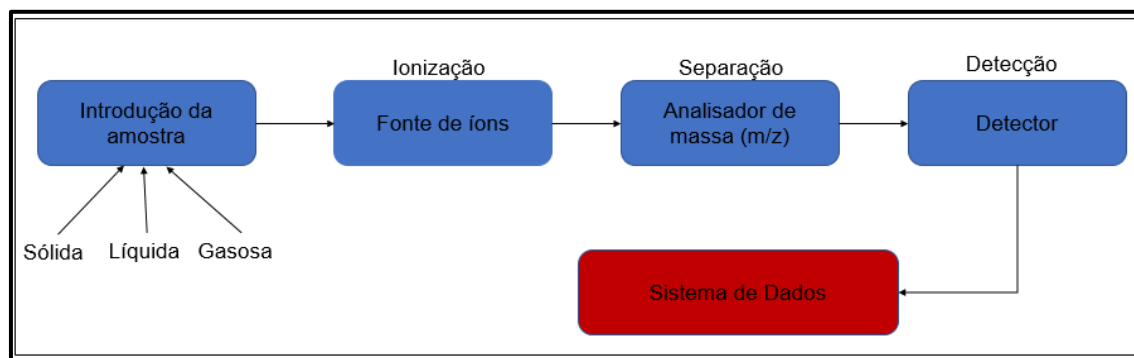


Figura 5 - Esquema dos principais componentes de um espectrômetro de massas.

Fonte: Adaptado de LANÇAS, 2009.

Atualmente, existe uma diversidade de fontes de ionização e analisadores de massas, os quais apresentam particularidades específicas que podem beneficiar a análise da classe de compostos desejados (ALBERICE, 2014). Em especial, para a análise de biomoléculas em espectrometria de massas, destaca-se as técnicas de ionização branda: ionização por *electrospray* – do inglês *Electrospray ionization* (ESI) e ionização/dessorção a laser auxiliada por matriz – do inglês *Matrix Assisted Laser Desorption Ionization* (MALDI) (FENSELAU; DEMIREV, 2001); além de analisadores de massas sequencial (MS/MS), os quais permitem a análise dos aspectos estruturais de compostos devido a sua resolução, acurácia e limite de detecção (ALBERICE, 2014; AQUINO, 2012).

A ionização por *electrospray* (ESI) opera a pressão atmosférica e, conforme demonstrado na Figura 6, a solução que contém a amostra a ser analisada é inserida no capilar metálico, sendo submetida a um campo elétrico bastante alto atribuído pela fonte de alimentação. Devido ao campo elétrico aplicado, há o acúmulo de íons eletrolíticos positivos na ponta do capilar metálico formando um cone (cone de Taylor). Um fluxo contracorrente de gás nebulizador faz com que o solvente evapore reduzindo o diâmetro das gotículas em carga constante e, conseqüentemente, eleva a repulsão eletrostática entre as espécies carregadas, ocasionando a fissão das gotículas denominada de “explosão coulômbica”. Tal fenômeno produz gotículas menores e posteriormente a liberação dos íons, formando um “spray” de partículas carregadas gerando assim uma corrente eletrolítica (FENN et al., 1989; KEBARLE; TANG, 1993).

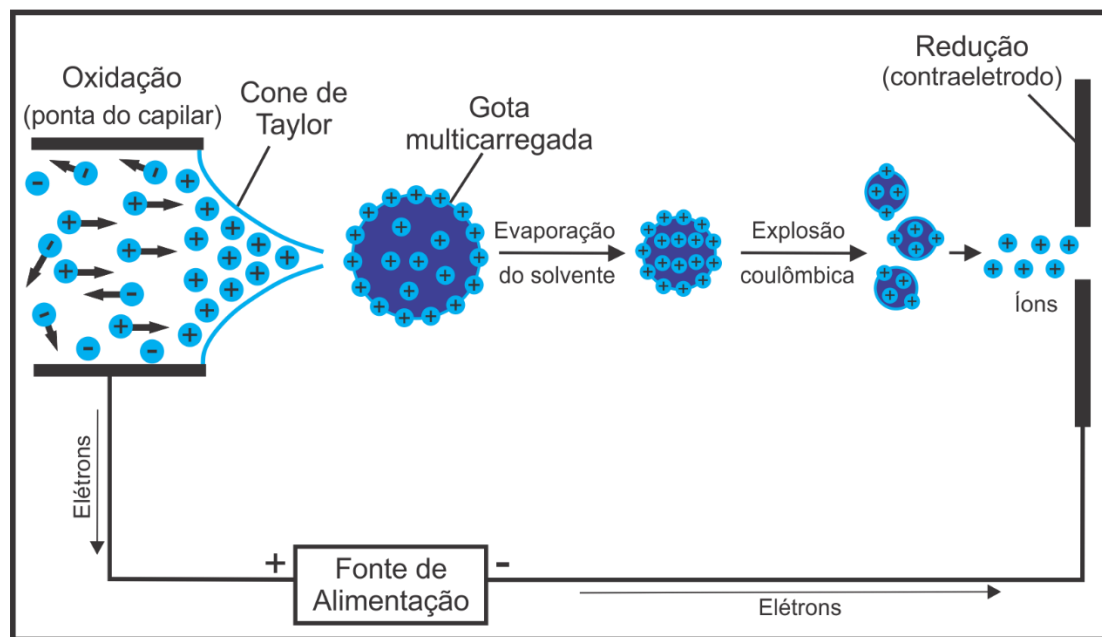


Figura 6 - Representação esquemática do funcionamento da fonte de ionização por ESI.
Fonte: Adaptado de CANTÚ et al. (2008) e ALBERICE (2014).

Um componente essencial em um espectrômetro de massas é o analisador de massas devido a sua função de separar os íons de acordo com a razão m/z . Atualmente, existem vários tipos de analisadores de massas, sendo os mais utilizados o *ion trap* (IT), quadrupolo (Q), tempo de voo (*Time-of-Flight* – TOF), FT-ICR (*Fourier-transform ion cyclotron resonance*) e *orbitrap* (OT). Cada um desses analisadores apresenta particularidades ou vantagens e desvantagens, que o tornam mais indicado para uma dada análise ou até mesmo para uma determinada classe de compostos (ALBERICE, 2014).

Dentre os mais amplamente utilizados, encontram-se os tipos quadrupolo e o *ion trap*. O tipo quadrupolo é constituído por quatro barras metálicas sólidas (eletrodos) paralelas ao feixe de íons. Este dispositivo utiliza campos elétricos oscilantes, produzidos pelas barras, para seleccionar os íons através da sua estabilidade de acordo com a razão m/z , separando-os no decorrer do percurso no centro da região dos quatro cilindros do quadrupolo (Figura 7). Um campo elétrico oscilante é produzido entre as barras através da aplicação da voltagem de corrente direta (DC – *direct-current*) e rádio frequência (RF). Dependendo das variações dos valores de DC e RF, os íons adquirem uma certa oscilação eletrostática que influencia em sua trajetória, podendo ser estabilizada ou desestabilizada. Sendo assim, apenas íons com uma dada m/z e uma determinada condição de DC e RF terão “oscilações estáveis” e atravessarão o quadrupolo (íon ressonante). Os demais íons com “oscilações instáveis” entrarão em contato com os polos e serão descartados (íon não ressonante). Com isso, o quadrupolo funciona como um filtro, permitindo que os íons com diferentes valores de m/z em seus respectivos tempos sejam detectados (HOFFMANN; STROOBANT, 2007; PAVIA et al., 2010; SILVERSTEIN; BASSLER; MORRILL, 1994).

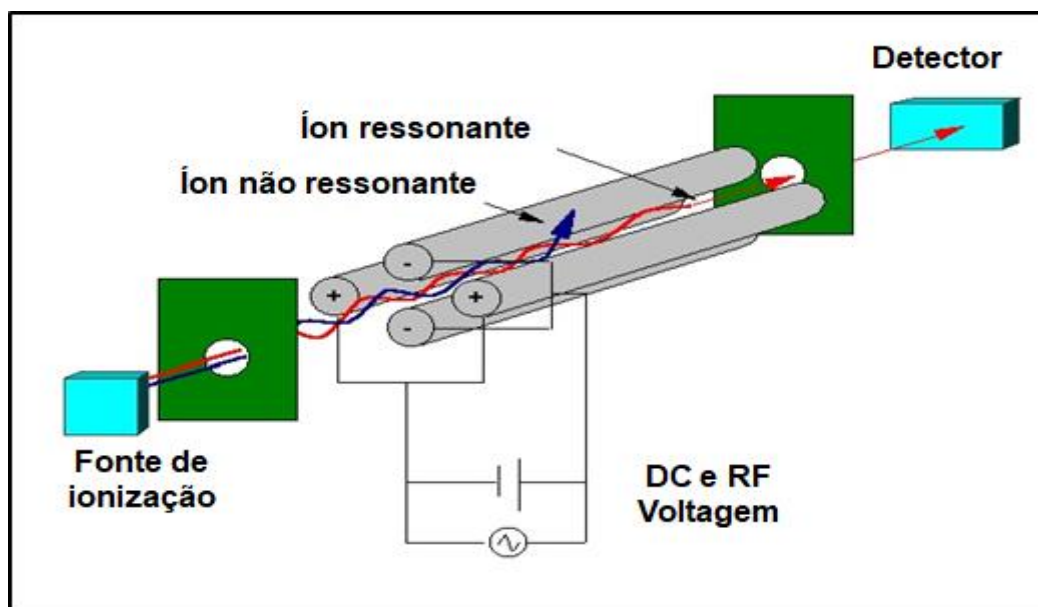


Figura 7 - Representação esquemática do analisador quadrupolo.

Fonte: Adaptado de nthu.edu.tw

Já os analisadores do tipo armadilha de íons (do inglês *ion trap*) possuem um princípio de funcionamento semelhante ao quadrupolo; entretanto, os íons não são descritos pela trajetória a partir do ambiente quadrupolar e sim são aprisionados dentro

dele. Neste, tem-se um campo elétrico no interior fazendo com que os íons permaneçam em uma órbita estável. O potencial de RF é empregado e com isso os íons são desestabilizados e liberados para fora do analisador, através dos valores de m/z (SCHALLEY, 2000).

As áreas de pesquisas como a proteômica e a metabolômica têm empregado espectrômetros de massas que cada vez mais necessitam ter parâmetros associados a características como a acurácia, resolução, sensibilidade, precisão entre as diferentes análises (HU et al., 2005). Nesse sentido, um analisador que vem se destacando nesses princípios é o *Orbitrap* (Figura 8), que possui em sua estrutura um eletrodo externo no formato de barril e um central coaxial fusiforme. Estes eletrodos são conectados com fontes elétricas independentes, que fornecem um campo com potencial do tipo quadro-logarítmica (HU et al., 2005; KAWAHARA, 2015). Após injetados, os íons são aprisionados em órbita, onde iniciam a oscilação em movimento harmônico no decorrer do eletrodo axial. A movimentação dos íons entre os eletrodos provoca uma corrente elétrica que consegue ser medida e por conseguinte, transformada em espectro de massas a partir da equação da transformada de Fourier (HU et al., 2005; KAWAHARA, 2015).

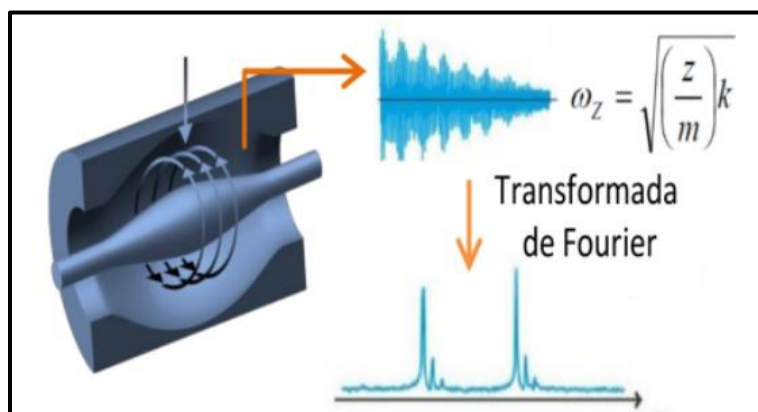


Figura 8 – Representação do analisador *Orbitrap*, onde a detecção da corrente elétrica gerada pelo movimento dos íons é convertida em um espectro de massas através da transformada de Fourier.

Fonte: KAWAHARA (2015)

Adicionalmente, é possível combinar dois ou mais analisadores em um dado equipamento de massas, denominado de híbridos, para a melhoria de propriedades entre as análises das amostras estudadas (VESSECCHI et al., 2011). A técnica que usa esse tipo de sistema é a espectrometria de massas sequencial ou em *tandem* (MS/MS),

utilizada com frequência para fragmentar metabólitos, proteínas e peptídeos, fornecendo informações sobre os compostos ou biomoléculas analisadas (DASS, 2007; HOFFMANN; STROOBANT, 2007).

Finalizando a descrição do espectrômetro de massas, tem-se o detector cuja função é a detecção e amplificação do sinal da corrente de íons proveniente do analisador, para assim transferir esse sinal para um sistema de processamento de dados (KAWAHARA, 2015). Os mais utilizados são os multiplicadores de elétrons e os fotomultiplicadores. No primeiro, há diversas placas coletoras de íons em série mantidas em potencial crescente. Quando o íon atinge à superfície do multiplicador, há emissão de elétrons. Estes, por sua vez, são atraídos por uma segunda placa e posteriormente, novos elétrons são gerados. O processo continua até a chegada na extremidade do multiplicador, formando uma cascata de elétrons que resulta no ganho total de uma corrente elétrica que é analisada e registrada pelo sistema de dados (DASS, 2007; HOFFMANN; STROOBANT, 2007). Já a placa de conversão fotomultiplicadora possui o princípio parecido com a multiplicadora de elétrons, onde os íons se chocam na placa, obtendo a emissão dos elétrons. Porém, quando há esse choque no detector resulta na liberação de fóton e então são amplificados. Assim, os elétrons são detectados a partir de uma fotomultiplicadora, operando também em cascata de elétrons semelhante à multiplicadora (DASS, 2007; HOFFMANN; STROOBANT, 2007; KAWAHARA, 2015).