



UNIVERSIDADE DE SÃO PAULO
FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE RIBEIRÃO PRETO



**Análise híbrida do perfil de compostos orgânicos voláteis em
amostras de fluido oral e urina como possível método de
diagnóstico do câncer**

Bruno Ruiz Brandão da Costa

**Ribeirão Preto
2022**

BRUNO RUIZ BRANDÃO DA COSTA

Análise híbrida do perfil de compostos orgânicos voláteis em amostras de fluido oral e urina como possível método de diagnóstico do câncer

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Toxicologia da Faculdade de Ciências Farmacêuticas de Ribeirão Preto/USP para obtenção do Título de Doutor em Ciências

Área de Concentração: Toxicologia.

Orientador: Prof. Dr. Bruno Spinosa De Martinis

Versão corrigida da Tese de Doutorado apresentada ao Programa de Pós-Graduação em Toxicologia no dia 16/12/2022. A versão original encontra-se disponível na Faculdade de Ciências Farmacêuticas de Ribeirão Preto/USP

Ribeirão Preto

2022

FICHA CATALOGRÁFICA

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Bruno Ruiz Brandão da Costa

Análise híbrida do perfil de compostos orgânicos voláteis em amostras de fluido oral e urina como possível método de diagnóstico do câncer, 2022.

222 p. : il. ; 30cm.

Tese de Doutorado, apresentada à Faculdade de Ciências Farmacêuticas de Ribeirão Preto/USP – Área de concentração: Toxicologia.

Orientador: Bruno Spinosa

1. VOCs. 2. Metabolômica. 3. Câncer. 4. Fingerprint
5. Headspace.

FOLHA DE APROVAÇÃO

Nome: Bruno Ruiz Brandão da Costa

Título: Análise híbrida do perfil de compostos orgânicos voláteis em amostras de fluido oral e urina como possível método de diagnóstico do câncer

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Toxicologia da Faculdade de Ciências Farmacêuticas de Ribeirão Preto/USP para obtenção do Título de Doutor em Ciências

Área de Concentração: Toxicologia.

Aprovado em:

Banca Examinadora

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Dedico este trabalho à minha querida vó Francisca que infelizmente se foi um pouco antes de poder ver um neto tornando-se Doutor. Dedico também ao meu filho (e melhor amigo) Arthur. Arthurzinho, cada palavra escrita neste documento sintetiza um pouco do meu esforço para poder te proporcionar um bom presente e um bom futuro. Te amo incondicionalmente.

AGRADECIMENTOS

Gostaria de expressar meus mais sinceros agradecimentos a muitas pessoas que foram de suma importância durante esses longos (e ao mesmo tempo tão curtos) quatro anos de doutorado.

Acima de tudo, agradeço aos meus pais, Mauro e Simone, por sempre colocarem a minha educação em primeiro lugar e por não pouparem esforços para que eu pudesse chegar até aqui.

Às minhas irmãs, Nathália e Marcela, por sempre torcerem pelo meu sucesso.

À toda família Gerçossimo, que sempre cuidou tão bem do nosso Arthurzinho. Estendo esses agradecimentos à Tia Lili e a todos(as) os(as) profissionais do colégio Alvorada.

Ao meu orientador, Prof. Dr. Bruno Spinosa De Martinis, por ter confiado em mim e me encorajado desde o início do projeto. Cada dia no laboratório foi um novo aprendizado e tenho certeza que saio um profissional muito mais preparado do que quando iniciei o doutorado.

Ao Jovem Pesquisador Dr. Ricardo Silva pela imensa contribuição no tratamento dos dados estatísticos. Muito obrigado pelos inúmeros cálculos, sugestões e pela paciência de, inclusive, me ensinar um pouco de programação, algo que nunca achei que iria conseguir fazer.

À Prof.^a Dr.^a Fernanda Peria por toda ajuda com os processos burocráticos relacionados ao HC-FMRP-USP e também por me ajudar a ter uma visão clínica do estudo. Estendo os agradecimentos também a todos os profissionais do HC que me auxiliaram na coleta e seleção dos voluntários.

Aos companheiros do Laboratório de Análises Toxicológicas Forenses. Agradeço Isabela Maximiano, Bruno Toledo, Wilson Júnior e Nayna Gomes por me receberem tão bem e terem facilitado minha adaptação ao laboratório. Agradeço Vitor Caleffo pelas inúmeras ajudas, principalmente aos assuntos relacionados ao planejamento experimental. Agradeço Lohanna Haddad por ter aceitado ser minha aluna de iniciação científica, por sempre ter acreditado em mim e por ter feito parte de tantos projetos junto comigo. Gostaria também de agradecer cada um dos companheiros citados também pelas colaborações nos seus respectivos trabalhos.

Essa troca de informações foi importantíssima para meu amadurecimento pessoal e científico. Por fim, além do âmbito “laboratorial”, gostaria de agradecer a todos por terem feito parte do meu dia a dia durante todo esse tempo e por terem deixado o ambiente de trabalho um lugar muito mais agradável.

A CAPES por conceder o aporte financeiro: o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Por fim, gostaria de agradecer imensamente cada um dos(as) voluntários(as) que fizeram parte deste estudo, principalmente aos indivíduos portadores de câncer. Mesmo em situação tão difícil, essas pessoas não hesitaram em ceder as amostras. Em diversas vezes ao longo das coletas ouvi dos pacientes que eles se sentiam satisfeitos em “pelo menos ajudar de alguma forma” com este estudo. Muito deles já se foram, mas a contribuição de cada um para esta pesquisa é algo inestimável.

Cabeça fria, coração quente.

RESUMO

da Costa, B. R. B. **Análise híbrida do perfil de compostos orgânicos voláteis em amostras de fluido oral e urina como possível método de diagnóstico do câncer.** 2022. 236f. Tese (Doutorado). Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2022.

O desenvolvimento de métodos simples, acessíveis e não invasivos para o diagnóstico do câncer é um dos maiores desafios científicos do século XXI. Atualmente, uma das técnicas mais promissoras é a análise de compostos orgânicos voláteis (VOCs) em amostras biológicas. Os VOCs são metabólitos gerados naturalmente no organismo, ao passo que, em um quadro patológico, como no câncer, alterações metabólicas levam à formação de novos VOCs e/ou modificam a concentração dos VOCs normalmente presentes. Desta maneira, foi levantada a hipótese de que os VOCs podem ser avaliados com o objetivo de determinar uma “impressão digital química” do câncer. Ainda, diagnósticos mais confiáveis seriam obtidos através da análise de duas amostras biológicas de um mesmo paciente (abordagem híbrida). Deste modo, o objetivo deste trabalho é avaliar os VOCs presentes no fluido oral e na urina de pessoas saudáveis e pacientes com câncer com o intuito de discriminar os dois grupos através do perfil de substâncias voláteis. Para tal, foi desenvolvido e otimizado um método baseado na extração por *headspace* estático (SHS) e posterior análise por cromatografia em fase gasosa com detector de ionização por chama (GC-FID). Esse método foi aplicado na análise de amostras de pessoas saudáveis ($n = 37$), pacientes com câncer de cabeça e pescoço (CCP, $n = 15$) e com câncer gastrointestinal (CG, $n = 19$). A análise discriminante por mínimos quadrados parciais ortogonais (OPLS-DA) foi utilizada para a classificação, onde o desempenho foi avaliado através do ajuste do modelo (R^2Y), da capacidade de predição (Q^2Y), e dos valores de sensibilidade e especificidade. Os melhores modelos individuais em termos de sensibilidade e especificidade foram o CCP urina (84,8 e 82,3%, respectivamente) e o CG fluido oral (78,6 e 87,5%). Ambos apresentaram R^2Y satisfatórios, próximos a 0,9, e valores aceitáveis (acima de 0,50) de Q^2Y (0,560 e 0,525, respectivamente). Já o modelo CCP fluido oral apresentou valor de sensibilidade mais baixo (70,2%) e o CG urina não foi adequado em nenhuma métrica avaliada. Posteriormente, a abordagem híbrida foi aplicada e observou-se um aumento dos valores de Q^2Y (0,623 para CCP e 0,562 para CG). Contudo, a sensibilidade foi maior para o modelo individual CCP urina (84,8%) do que para o modelo CCP híbrido (75,6%), e também foi maior para CG fluido oral (78,6%) do que para CG híbrido (69,8%). Esses resultados sugerem duas conclusões principais. Primeiro: a abordagem híbrida realmente tende a elevar a capacidade de predição do método. Contudo, para que isso aconteça, é preciso que os modelos individuais apresentem métricas adequadas de qualidade. Segundo: o método desenvolvido mostrou potencial para ser empregado como diagnóstico, porém estudos contando um tamanho amostral maior são essenciais para avaliar se uma técnica tão simples e prática pode, de fato, ser utilizada na rotina clínica como um método auxiliar no *screening* do câncer.

Palavras-chave: VOCs; metabolômica; câncer; *fingerprint*; *headspace*.

ABSTRACT

da Costa, B. R. B. **Hybrid analysis of volatile organic compounds in oral fluid and urine as a possible method to diagnose cancer**. 2022. 236f. Thesis (Doctoral). Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2022.

The development of simple, cost-effective and non-invasive methods for cancer diagnosis is one of the greatest scientific challenges of the 21st century. In this context, one of the most promising techniques is the analysis of volatile organic compounds (VOCs). VOCs are metabolites naturally generated in the body and exhaled in various biological matrices. In a pathological condition, such as cancer, metabolic alterations lead to the generation of new VOCs and/or change the concentrations of naturally present VOCs. Thus, two hypotheses were raised in this study. First, VOCs profile differences establishes a fingerprint that can be explored in the context of cancer diagnosis. Second: the data fusion from two biological samples (hybrid approach) leads to more accurate diagnoses. Thus, the objective of this work is to evaluate the VOCs present in oral fluid and urine of healthy people and cancer patients in order to discriminate these two groups through the VOCs profile. To this end, a method based on static headspace extraction (SHS) and subsequent analysis by gas chromatography with flame ionization detector (GC-FID) was developed and optimized. This method was applied in the analysis of samples from healthy people ($n = 37$), patients with head and neck cancer (HNC, $n = 15$) and with gastrointestinal cancer (GIC, $n = 19$). Orthogonal partial least squares discriminant analysis (OPLS-DA) was used for classification and the performance was evaluated through goodness of fit (R^2Y), predictive ability (Q^2Y), and sensitivity and specificity values. The best individual models in terms of sensitivity and specificity were HNC urine (84.8 and 82.3%, respectively) and GIC oral fluid (78.6 and 87.5%). Both methods showed satisfactory R^2Y values (close to 0.9), but lower Q^2Y values (0.560 and 0.525, respectively), thus showing evidence of overfitting. The HNC oral fluid model, on the other hand, had a lower sensitivity value (70.2%), and the urine GC was not adequate in any of the evaluated metrics. Then, the hybrid approach was applied and an increase in Q^2Y values was observed (0.623 for HNC and 0.562 for GIC). However, sensitivity was higher for the HNC urine model (84.8%) than for the hybrid CCP model (75.6%), and it was also higher for the GIC oral fluid (78.6%) than for the hybrid GC (69.8%). These results suggest two main conclusions. First: the hybrid approach really tends to increase the predictive capacity of the method; however, for this to happen, the individual models must present adequate quality metrics. Second: the SHS-GC-FID method showed interesting potential to be used to diagnose cancer; however, further studies with larger sample sizes are essential to assess whether such a practical technique can, in fact, be used in clinical routine as an auxiliary method in cancer screening.

Keywords: VOCs; metabolomics; cancer; fingerprinting; headspace.

LISTA DE FIGURAS

Figura 1 - Representação das três fases do processo de carcinogênese: iniciação, promoção e progressão. Adaptado de LIU et al., 2015.	5
Figura 2 - Capacidades biológicas características adquiridas pelas células tumorais (<i>Hallmarks of Cancer</i>). Adaptado de HANAHAN & WEINBERG, 2011.	7
Figura 3 - Porcentagem de cada classe de compostos orgânicos voláteis detectados em diferentes fluidos biológicos de pessoas saudáveis. Adaptado de DE LACY COSTELLO et al., 2014.	21
Figura 4 - Diagrama simplificado do detector de ionização por chama (FID). Adaptado de Förster et al., 2019.	30
Figura 5 - Representação da técnica de <i>headspace</i> estático.	32
Figura 6 - Possibilidades de condução de um experimento para 3 variáveis estudadas. Cada ponto representa um experimento e o cubo representa o espaço experimental. Adaptado de (RODRIGUES & IEMMA, 2005).	40
Figura 7 - Estrutura do Salivette®. 1 – tubo plástico; 2a – frasco coletor e 2b – filtro; 2C – rolo de algodão.	44
Figura 8 - Esquema do processo de coleta de fluido oral através do Salivette®. Adaptado de Gomes, 2018.	45
Figura 9 - Representação do procedimento experimental para a otimização do método de análise de VOCs em fluido oral e urina.	50
Figura 10 - Cromatograma da análise da mistura de fluido oral e urina utilizando a rampa de aquecimento adotada na Condição 1.	58
Figura 11 - Cromatograma da análise da mistura de fluido oral e urina utilizando a rampa de aquecimento adotada na Condição 2.	59
Figura 12 - Cromatograma da análise da mistura de fluido oral e urina realizada no modo <i>splitless</i>	60
Figura 13 - Cromatograma da análise da mistura de fluido oral e urina realizada no modo <i>split</i> 1:10.	61
Figura 14 - Diagrama de Pareto para os resultados obtidos através do delineamento de Plackett-Burman para a urina.	64

Figura 15 - Superfície de resposta em função da temperatura (°C) e tempo de incubação (min) para a média geométrica das áreas dos picos nas análises dos VOCs urinários por SHS-GC-FID.	69
Figura 16 - Gráfico de valores preditos versus os valores observados para o modelo gerado no DCCR para a urina.....	71
Figura 17 - Gráfico ilustrativo gerado pelo software Octave para a avaliação da significância do modelo gerado no DCCR para a urina.....	72
Figura 18 - Gráfico ilustrativo gerado pelo software Octave para a avaliação do erro puro e falta de ajuste do modelo gerado no DCCR para a urina.....	72
Figura 19 - Valor ótimo de resposta previsto para a extração por headspace dos VOCs urinários segundo o software statistica 7.0.	74
Figura 20 - Sobreposição dos cromatogramas referentes à análise do <i>pool</i> de urina realizada na condição ótima (em preto) determinada pelo planejamento experimental e em uma condição não otimizada (PB experimento 2 – Tabela 17).	75
Figura 21 - Diagrama de Pareto para os resultados obtidos através do delineamento de Plackett-Burman para o fluido oral.	77
Figura 22 - Superfície de resposta em função da temperatura (°C) e tempo de incubação (min) para a média geométrica das áreas dos picos nas análises dos VOCs presentes no fluido oral por SHS-GC-FID (adição de sal fixada em 1,0 g).	81
Figura 23 - Gráfico de valores preditos versus os valores observados para o modelo gerado no DCCR para a o fluido oral.	82
Figura 24 - Gráfico ilustrativo gerado pelo software Octave para a avaliação do da significância do modelo gerado no DCCR para o fluido oral.	83
Figura 25 - Gráfico ilustrativo gerado pelo software Octave para a avaliação do erro puro e falta de ajuste do modelo gerado no DCCR para a o fluido oral.	83
Figura 26 - Valor ótimo de resposta previsto para a extração por headspace dos VOCs presentes no fluido oral segundo o software Statistica.	85
Figura 27 - Sobreposição dos cromatogramas referentes à análise do <i>pool</i> de fluido oral realizada na condição ótima (em preto) determinada pelo planejamento experimental e em uma condição não otimizada (PB experimento 2 – Tabela 21).	85

Figura 28 - Cromatograma referente à análise de 3 mL de água deionizada com adição de 1 g de NaCl e 80 µL de ácido sulfúrico 0,1 M.....	86
Figura 29 - Cromatograma referente à análise dos possíveis interferentes oriundos do Salivette®.....	87
Figura 30 - Peroxidação lipídica: principal mecanismo proposto para a formação de VOCs vinculados ao câncer. Adaptado de MONEDEIRO <i>et al.</i> , 2020.	91
Figura 31 - Contribuição da microbiota intestinal para a formação de VOCs associados ao câncer. A) Catabolismo da metionina. B) Catabolismo da tirosina. Mecanismo proposto por MONEDEIRO <i>et al.</i> , 2020.	96
Figura 32 - Exemplo genérico de um gráfico de PCA (scores).	99
Figura 33 - Exemplo genérico de um gráfico de HCA.	100
Figura 34 - Exemplo genérico de uma árvore de decisão. Os números são separados de acordo com a cor e se estão sublinhados ou não.	101
Figura 35 - Visualização de um modelo genérico <i>Random Forest</i> na classificação de uma amostra entre os grupos “saudável” e “câncer”.....	102
Figura 36 – Fluxograma simplificado do procedimento utilizado no pré-tratamento dos dados.	106
Figura 37 - Principais sítios do câncer de cabeça e pescoço. Adaptado de NCI, 2021.	108
Figura 38 - Representação do valor da área do pico do padrão interno (PI - diidromircenol) nas amostras analisadas. Cada ponto significa o valor da área do PI (eixo y) em uma determinada amostra (eixo x). A linha central em azul representa a mediana, a linha superior em azul representa o percentil 75 e a linha inferior o percentil 25.....	119
Figura 39 - PCA Biplot referente às análises de fluido oral (FO) do grupo câncer de cabeça e pescoço (CCP) e do grupo controle. Em branco estão representadas as amostras de CCP e em azul as amostras do grupo controle (N).	121
Figura 40 - PCA referente às análises de fluido oral (FO) do grupo câncer gastrointestinal (CG) e do grupo controle (N). Em branco estão representadas as amostras de CG e em azul as amostras do grupo N.	122

Figura 41 - PCA referente às análises de urina (U) do grupo câncer de cabeça e pescoço (CCP) e do grupo controle (N). Em branco estão representadas as amostras de CCP e em azul as amostras do grupo N.	122
Figura 42 - PCA referente às análises de urina (U) do grupo câncer gastrointestinal (CG) e do grupo controle (N). Em branco estão representadas as amostras de CG e em azul as amostras do grupo N.	123
Figura 43 - PCA referente às análises de fluido oral (FO) do grupo câncer gastrointestinal (CG), câncer de cabeça e pescoço (CCP) e do grupo controle (N). Em branco estão representadas as amostras de CG, em azul as amostras do grupo CCP, e em rosa as amostras do grupo N.	123
Figura 44 - PCA referente às análises de urina (U) do grupo câncer gastrointestinal (CG), câncer de cabeça e pescoço (CCP) e do grupo controle (N). Em branco estão representadas as amostras de CG, em azul as amostras do grupo CCP, e em rosa as amostras do grupo N.	124
Figura 45 - PCA referente às análises de urina (U) e fluido oral (FO) do grupo câncer gastrointestinal (CG), câncer de cabeça e pescoço (CCP) e do grupo controle (N). À esquerda se encontram as amostras de fluido oral e à direita estão agrupadas as amostras de urina.	125
Figura 46 - HCA referente às análises de urina (U) e fluido oral (FO) do grupo câncer gastrointestinal (CG), câncer de cabeça e pescoço (CCP) e do grupo controle (N).	126
Figura 47 - PCoA referente às análises de urina. A escala de cores vai do mais claro (primeiras amostras analisadas) até o mais escuro (últimas amostras analisadas).	128
Figura 48 - PCoA referente às análises de urina. A escala de cores vai do mais claro (primeiras amostras analisadas) até o mais escuro (últimas amostras analisadas).	128
Figura 49 - OPLS-DA (gráfico de scores) referente às análises de fluido oral do grupo câncer de cabeça e pescoço (CCP) e do grupo controle (N). Em azul estão representadas as amostras de CCP e em vermelho as amostras do grupo N.	134
Figura 50 - OPLS-DA (gráfico de scores) referente às análises de fluido oral do grupo câncer gastrointestinal (CG) e do grupo controle (N). Em azul estão representadas as amostras de CG e em vermelho as amostras do grupo N.	134

Figura 51 - OPLS-DA (gráfico de scores) referente às análises de urina do grupo câncer de cabeça e pescoço (CCP) e do grupo controle (N). Em azul estão representadas as amostras de CCP e em vermelho as amostras do grupo N.....	135
Figura 52 - OPLS-DA (gráfico de scores) referente às análises de urina do grupo câncer gastrointestinal (CG) e do grupo controle (N). Em azul estão representadas as amostras de CG e em vermelho as amostras do grupo N.	136
Figura 53 - OPLS-DA (gráfico de scores) referente à análise híbrida das amostras de fluido oral e urina dos voluntários do grupo câncer de cabeça e pescoço (representado em azul) e do grupo controle (representado em vermelho).....	138
Figura 54 - OPLS-DA (gráfico de scores) referente à análise híbrida das amostras de fluido oral e urina dos voluntários do grupo câncer gastrointestinal (representado em azul) e do grupo controle (representado em vermelho).....	139
Figura 55 - Boxplots referentes aos cinco picos mais importantes para a construção do modelo de classificação para o câncer de cabeça e pescoço (CCP) segundo o score VIP. O grupo CCP está representado em azul e o grupo de voluntários saudáveis (N) está representado em amarelo. A) se refere ao pico de tempo de retenção 34,22 minutos; B) se refere ao pico 30,76; C) se refere ao pico 31,97; D) se refere pico 36,33; E) se refere ao pico 25,16.	145
Figura 56 - Boxplots referentes aos cinco picos mais importantes para a construção do modelo de classificação para o câncer gastrointestinal (CG) segundo o score VIP. O grupo CG está representado em azul e o grupo de voluntários saudáveis (N) está representado em amarelo. A) se refere ao pico de tempo de retenção 31,97 minutos; B) se refere ao pico 29,49; C) se refere ao pico 26,65; D) se refere pico 34,22; E) se refere ao pico 30,76.....	146
Figura 57 - Sobreposição dos picos de tempo de retenção 31,97 min. A) Comparação da intensidade deste pico nas amostras de câncer de cabeça e pescoço (CCP) e voluntários saudáveis (N) nas duas matrizes avaliadas: fluido oral (FO) e urina (U). B) Comparação da intensidade deste pico nas amostras de câncer gastrointestinal (CG) e voluntários saudáveis.....	147

LISTA DE TABELAS

Tabela 1 - Estimativa do número novos casos dos dez tipos de câncer mais incidentes no Brasil (triênio 2020-2022).....	3
Tabela 2 - Exemplos de estudos sobre a investigação de VOCs no contexto do câncer (em ordem alfabética).	13
Tabela 3 - Revisão dos principais estudos que exploram a análise dos VOCs presentes no fluido oral para o diagnóstico do câncer (em ordem alfabética).	20
Tabela 4 - Comparação entre as principais técnicas de extração baseadas no método de <i>headspace</i>	32
Tabela 5 - Representação da matriz experimental da metodologia Plackett-Burman de 12 experimentos para o estudo de 5 variáveis.	39
Tabela 6 - Matriz de experimentos do DCCR de 2 níveis.	41
Tabela 7 - Matriz de experimentos do DCCR de 3 níveis.	41
Tabela 8 - Programa de temperatura do forno para a análise de VOCs presentes no fluido oral e na urina através do GC-FID.....	46
Tabela 9 - Variáveis e níveis avaliados na metodologia Plackett-Burman para a urina.	51
Tabela 10 - Matriz PB de 12 experimentos para a urina.	52
Tabela 11 - Variáveis e níveis utilizados no DCCR de 2 variáveis para a otimização dos parâmetros de extração por HS para a urina.	53
Tabela 12 - Design experimental para o DCCR de 2 variáveis para a otimização dos parâmetros de extração por HS para a urina.	53
Tabela 13 - Variáveis e níveis utilizados na metodologia PB para o fluido oral.	54
Tabela 14 - Design de PB de 12 experimentos para o fluido oral.	55
Tabela 15 - Variáveis e níveis utilizados no DCCR para a otimização da extração por SHS dos VOCs presentes no fluido oral.	55
Tabela 16 - Design do DCCR de 3 variáveis para a otimização da extração por SHS dos VOCs presentes no fluido oral.....	56
Tabela 17 - Resultados do Delineamento de Plackett-Burman para urina.....	63

Tabela 18 - Resultados do DCCR de 2 fatores na otimização da extração dos VOCs urinários.	67
Tabela 19 – Análise de variância (ANOVA) dos resultados obtidos para o delineamento composto central rotacional (DCCR) aplicado na otimização dos parâmetros tempo e temperatura de incubação para a extração dos VOCs urinários através da técnica de <i>headspace</i> estático.	68
Tabela 20 - Valores ótimos para as variáveis avaliadas no processo de otimização da extração para a urina.	73
Tabela 21 - Resultados do Delineamento PB para o fluido oral.	76
Tabela 22 - Resultados do DCCR de 3 fatores para o fluido oral.	79
Tabela 23 - Análise de variância (ANOVA) dos resultados obtidos para o delineamento composto central rotacional (DCCR) aplicado na otimização dos parâmetros tempo de incubação, temperatura de incubação e adição de sal para a extração dos VOCs presentes no fluido oral através da técnica de <i>headspace</i> estático.	80
Tabela 24 - Valores ótimos para as variáveis avaliadas no processo de otimização para o fluido oral.	84
Tabela 25 - Informações dos voluntários diagnosticados com câncer (gênero, idade, tipo de neoplasia, classificação pelo sistema TNM, estágio clínico, uso crônico de álcool, uso crônico de tabaco, outros diagnósticos, uso de medicamentos, e outras observações).....	110
Tabela 26 - Resumo do perfil dos pacientes oncológicos participantes deste estudo levando em consideração a divisão dos dois grandes grupos avaliados (câncer de cabeça e pescoço - CP e câncer gastrointestinal - CG).....	113
Tabela 27 - Informações do grupo controle (gênero, idade, uso crônico de álcool, uso crônico de tabaco, diagnósticos médicos, uso de medicamentos, e outras observações).....	114
Tabela 28 - Perfil geral dos voluntários participantes deste estudo.	116
Tabela 29 - Matriz de confusão e estimativa do erro OOB da aplicação do modelo Random Forest referente às análises de fluido oral (FO) do grupo câncer de cabeça e pescoço (CCP) e do grupo controle.	130

Tabela 30 - Matriz de confusão e estimativa do erro OOB da aplicação do modelo Random Forest referente às análises de fluido oral (FO) do grupo câncer gastrointestinal (CG) e do grupo controle.....	130
Tabela 31 - Matriz de confusão e estimativa do erro OOB da aplicação do modelo Random Forest referente às análises de urina (U) do grupo câncer de cabeça e pescoço (CCP) e do grupo controle.	131
Tabela 32 - Matriz de confusão e estimativa do erro OOB da aplicação do modelo Random Forest referente às análises de urina (U) do grupo câncer gastrointestinal (CG) e do grupo controle.	131
Tabela 33 - Valores de R^2Y e Q^2Y para todos os modelos de OPLS-DA gerados.....	140
Tabela 34 – Comparação dos modelos de OPLS-DA utilizando como métrica os valores de R^2Y e Q^2Y obtidos na construção do modelo, bem como os valores de área sob a curva (AUC), sensibilidade e especificidade obtidos através das curvas ROC.	142
Tabela 35 - Valores de Score VIP para os cinco picos mais importantes para a construção de modelo de classificação para o CCP.	144
Tabela 36 - Valores de Score VIP para os cinco picos mais importantes para a construção de modelo de classificação para o CG.	144
Tabela 37 - Resumo de estudos semelhantes publicados nos últimos 10 anos (organizado em ordem alfabética - nome do primeiro autor, tipo de câncer, matriz biológica avaliada, número de voluntários, técnica de análise, tipo/finalidade do estudo; modelo estatístico empregado e capacidade de predição).	152

LISTA DE ABREVIATURAS E SIGLAS

ADC	Adenocarcinoma
ADH	Álcool desidrogenase
AFP	Alfa-fetoproteína
AJCC	<i>American Joint Committee on Cancer</i>
ALDH	Aldeído desidrogenase
ANN	<i>Artificial neural network</i>
AUC	Área sob a curva
CCP	Câncer de cabeça e pescoço
CEC	Carcinoma espinocelular
CG	Câncer gastrointestinal
CYP450	Citocromo P450
DCCR	Delineamento Composto Central Rotacional
DFA	<i>Discriminant factor analysis</i>
DOE	Planejamento experimental
<i>e-nose</i>	<i>Eletronic nose</i> (nariz eletrônico)
FAIMS	<i>High Field Asymmetric Waveform Ion Mobility Spectrometry</i>
FDA	<i>Fisher discriminant analysis</i>
FID	Detector de ionização por chama
FO	Fluido oral
GC	Cromatografia em fase gasosa
GC-FID	Cromatografia gasosa com detector de ionização por chama
GC-MS	Cromatografia em fase gasosa acoplada ao espectrômetro de massas
HCA	Análise por agrupamento hierárquico
HPV	Papilomavírus humano
HS	<i>Headspace</i>
IARC	Agência Internacional de Pesquisa em Câncer
INCA	Instituto Nacional de Câncer
K	Coeficiente de partição
MS	Espectrometria de massas

NaCl	Cloreto de sódio
NPD	Detector de nitrogênio e fósforo
OMS	Organização Mundial da Saúde
OOB	Estimativa de erro <i>out-of-bag</i>
OPLS-DA	Análise discriminante por mínimos quadrados parciais ortogonais
PB	Plackett-Burman
PC	Componentes principais
PCA	Análise de componentes principais
PCoA	Ánálise de coordenadas principais
PI	Padrão interno
PLS-DA	Análise discriminante por mínimos quadrados parciais
PSA	Antígeno Prostático Específico
PTR-MS	<i>Proton-transfer-reaction mass spectrometry</i>
PTV	Programmable temperature vaporizer
PUFA	Ácidos graxos poli-insaturados
RF	<i>Random Forest</i>
ROC	<i>Receiver Operating Characteristic Curve</i>
ROS	Espécies reativas de oxigênio
SHS	<i>Headspace</i> estático
SIFT-MS	<i>Selected-ion flow-tube mass spectrometry</i>
SPME	Microextração em fase sólida
SVM	<i>Support vector machine</i>
TFME	Thin film microextraction
TSD	Detector termiônico específico
UICC	União Internacional de Controle do Câncer
VIP	<i>Variable influence on projection</i>
VOCs	Compostos orgânicos voláteis
WHO	<i>World Health Organization</i>
β	Taxa de fase

SUMÁRIO

RESUMO	i
ABSTRACT	ii
LISTA DE FIGURAS	iii
LISTA DE TABELAS	viii
LISTA DE ABREVIATURAS E SIGLAS	xi
CAPÍTULO I – INTRODUÇÃO E ASPECTOS ÉTICOS DA PESQUISA	1
1 INTRODUÇÃO	2
1.1 Câncer: definição e incidência	2
1.2 Carcinogênese	4
1.3 Hallmarks of Cancer	5
1.4 Diagnóstico do câncer: a busca por novos métodos de análise	7
1.5 Análise de VOCs no diagnóstico do câncer	10
1.6 VOCs presentes no fluido oral: potencial aplicação na área clínica	15
1.7 Análise híbrida do perfil de VOCs	21
1.8 Análise de VOCs presentes em matrizes biológicas por cromatografia em fase gasosa com detector de ionização por chama	22
2 HIPÓTESE	24
3 JUSTIFICATIVA	24
4 OBJETIVOS	25
5 ASPECTOS ÉTICOS DA PESQUISA	25
CAPÍTULO II – DESENVOLVIMENTO E OTIMIZAÇÃO DE MÉTODO PARA ANÁLISE DOS VOCs PRESENTES NO FLUIDO ORAL E NA URINA	27
1 INTRODUÇÃO	28
2 REFERENCIAL TEÓRICO	28

2.1	Cromatografia em fase gasosa	28
2.2	<i>Headspace</i>	30
2.3	Fundamentos teóricos do <i>headspace</i> estático	33
2.4	Planejamento experimental.....	36
2.4.1	Plackett Burman	38
2.4.2	Delineamento Composto Central Rotacional.....	39
3	OBJETIVOS	42
4	MATERIAIS E MÉTODOS	42
4.1	Materiais	42
4.2	Reagentes, padrões e soluções.....	42
4.2.1	Preparo das soluções	42
4.3	Coleta das amostras	43
4.4	Análise instrumental.....	45
4.5	Desenvolvimento e otimização do método.....	46
4.5.1	Desenvolvimento do método cromatográfico.....	46
4.6	Planejamento experimental.....	48
4.6.1	Planejamento experimental para urina	48
4.6.2	Planejamento experimental para o fluido oral.....	54
4.6.3	Verificação de possíveis interferentes nas análises	56
5	RESULTADOS E DISCUSSÃO	57
5.1	Desenvolvimento do método cromatográfico	57
5.1.1	Escolha da rampa de aquecimento	57
5.1.2	Escolha da taxa de divisão do fluxo	59
5.2	Planejamento experimental.....	62

5.2.1	Otimização dos parâmetros da extração por SHS dos VOCs presentes na urina	62
5.2.2	Otimização dos parâmetros da extração por <i>headspace</i> estático dos VOCs presentes no fluido oral.....	75
5.2.3	DCCR de 3 fatores para o fluido oral.....	78
5.3	Verificação de possíveis interferentes nas análises.....	86
6	CONCLUSÃO	87
CAPÍTULO III – ANÁLISE DO PERFIL DE VOCS EM AMOSTRAS REAIS E A POTENCIAL APLICAÇÃO DO SHS-GC-FID NO DIAGNÓSTICO DO CÂNCER		
1	INTRODUÇÃO.....	89
1.1	Possíveis mecanismos que dão origem aos VOCs associados ao câncer	89
1.1.1	Estresse oxidativo	89
1.1.2	Indução enzimática.....	92
1.1.3	Efeito Warburg (glicólise aeróbica).....	93
1.1.4	Reprogramação do metabolismo de aminoácidos.....	94
1.1.5	Reprogramação do metabolismo de lipídeos	96
1.2	Análise estatística: embasamento teórico.....	97
1.2.1	Métodos não supervisionados.....	98
1.2.2	Métodos supervisionados.....	101
2	OBJETIVOS.....	103
3	CASUÍSTICA E MÉTODOS.....	103
3.1	Protocolo de coleta das amostras e método experimental.....	103
3.2	Análise estatística dos dados.....	104
4	RESULTADOS E DISCUSSÃO	107
4.1	Perfil dos voluntários da pesquisa.....	107
4.2	Análise estatística	118

4.2.1	Análise exploratória dos dados.....	120
4.2.2	Aplicação de métodos supervisionados de classificação	129
4.3	Comparação com estudos semelhantes presentes na literatura.....	148
5	CONCLUSÃO	155
	REFERÊNCIAS.....	156
	ANEXO 1 – Aprovação do Comitê de Ética em Pesquisa da FCFRP	194
	ANEXO 2 – Aprovação do Comitê de Ética em Pesquisa da FFCLRP	195
	ANEXO 3 – Aprovação do Comitê de Ética em Pesquisa da FMRP.....	196
	ANEXO 4 - Termo de consentimento livre e esclarecido	197

**CAPÍTULO I – INTRODUÇÃO E ASPECTOS
ÉTICOS DA PESQUISA**

1 INTRODUÇÃO

1.1 Câncer: definição e incidência

Câncer é um termo utilizado para designar um conjunto de mais de 100 tipos diferentes de doenças que possuem em comum o desordenado e acelerado crescimento de células anormais, as quais possuem a capacidade de invadir órgãos e tecidos vizinhos (INCA, 2020). O desenvolvimento dessa patologia é caracterizado por um processo que ocorre em múltiplas etapas. As células gradativamente passam a sofrer mudanças metabólicas e comportamentais, levando-as a proliferarem de maneira excessiva, a escaparem da vigilância realizada pelo sistema imunológico e, por fim, a alcançarem outras regiões do organismo através da corrente sanguínea ou do sistema linfático, propiciando a formação de novos tumores em um processo denominado como metástase (do termo grego *metastatis*, o qual significa “mudança de lugar”) (MURPHY, 2001; WHO, 2005).

Em termos estatísticos, estima-se que 1,7% da população mundial apresente alguma forma de câncer (ROSER & RITCHIE, 2019). Segundo dados da Organização Mundial da Saúde (OMS, ou do termo em inglês *World Health Organization*, WHO), esta doença corresponde a segunda principal causa de óbitos no mundo, com cerca de 10 milhões de mortes atribuídas em 2020 (atrás apenas de óbitos causados por doenças cardiovasculares - cerca de uma em cada seis mortes ocorre devido ao câncer) (ROSER & RITCHIE, 2019; WHO, 2022). Já no Brasil, o número de pessoas que faleceram devido ao câncer em 2020 foi de, aproximadamente, 260 mil (WHO, 2021). Além disso, segundo dados da Agência Internacional de Pesquisa em Câncer (IARC) em conjunto com a OMS, estima-se que o número de novos casos por ano dessa patologia deve exceder 27 milhões até 2040, um aumento de quase 50% em relação aos 18,1 milhões de novos casos registrados em 2018 (WHO, 2020).

No ano de 2020, os tipos de câncer mais incidentes no mundo, em número novos casos, foram: mama (2,26 milhões), pulmão (2,21 milhões), colorretal (1,93 milhão); próstata (1,41 milhão), pele (não melanoma; 1,2 milhão) e estômago (1,09 milhão). Por fim, os tipos de câncer que mais levaram indivíduos a óbito foram: pulmão (1,80 milhão de mortes), colorretal (935 mil), fígado (830 mil); estômago (769 mil) e mama (685 mil) (WHO, 2021a). Em relação especificamente ao Brasil, as estimativas

do INCA para o triênio 2020-2022 ilustram o impacto dessa doença na população brasileira (INCA, 2019) (Tabela 1).

Tabela 1 - Estimativa do número novos casos dos dez tipos de câncer mais incidentes no Brasil (triênio 2020-2022).

Tipos de câncer	Novos casos por ano		Risco estimado	
	2020 a 2022		(casos por 100 mil habitantes)	
	Homens	Mulheres	Homens	Mulheres
Próstata	65840	-	62,95	-
Mama	-	66280	-	61,61
Colorretal	20520	20470	19,63	19,03
Pulmão	17760	12440	16,99	11,56
Estômago	13360	7870	12,81	7,34
Colo do útero	-	16590	-	15,34
Cabeça e pescoço	11180	4010	10,69	3,71
Sistema nervoso central	5870	5220	5,61	4,85
Leucemia	5920	4890	5,67	4,56
Esôfago	8690	2700	8,32	2,49

Dados do INCA, 2019. Adaptado de Nascimento, 2020.

O Câncer é uma doença que pode ser originada por uma série de razões, como fatores ambientais, culturais, socioeconômicos, estilos de vida ou costumes, com destaque para os hábitos de ingerir bebidas alcóolicas e fumar; fatores genéticos e o próprio processo de envelhecimento (BROWN *et al.*, 2015; WHO, 2020). Como exemplo, estima-se que aproximadamente um terço das mortes por câncer sejam relacionadas a pelo menos um dos seguintes fatores: consumo de tabaco, alto índice de massa corporal, consumo de álcool, baixa ingestão de frutas e vegetais, ou falta de atividade física (WHO, 2022). Além disso, pesquisas recentes sugerem que há outro fator preponderante e imprevisível na ocorrência do câncer: a má sorte. Isto é: mutações aleatórias que surgem durante a replicação do DNA em células normais e que podem levar, futuramente, ao desenvolvimento da doença em questão. (TOMASETTI & VOGELSTEIN, 2015). Esses diversos fatores citados anteriormente podem agir em conjunto ou em sequência para iniciar ou promover a formação do câncer, em um processo denominado carcinogênese.

1.2 Carcinogênese

O mecanismo de formação do câncer é um processo complexo, lento e que envolve múltiplas etapas tanto a nível genético quanto a nível fenotípico, sendo resultado do acúmulo de mutações. Tal processo envolve três estágios imprescindíveis à instalação da doença: iniciação, promoção e progressão (KUMAR *et al.*, 2018).

A iniciação é caracterizada por alterações na sequência de bases do DNA celular provocadas pela exposição a um agente cancerígeno físico (como radiação ultravioleta), químico (como componentes do cigarro, aflatoxinas, álcool e arsênio), ou biológico (como o papilomavírus humano e o vírus da hepatite C) (WHO, 2022). A interação com esses agentes pode levar a danos irreversíveis em genes responsáveis pelo controle do ciclo celular, apoptose, e diferenciação celular. Conseqüentemente, essas ações podem culminar na ativação de oncogenes e/ou na inativação de genes supressores de tumor (GATENBY & GILLIES, 2008). Nesta etapa da carcinogênese, as células encontram-se geneticamente alteradas, ou seja, iniciadas, porém ainda não é possível detectar um tumor clinicamente (ALMEIDA *et al.*, 2005).

As células “iniciadas” podem permanecer neste estado durante tempos variados caso não sejam estimuladas à divisão. No entanto, as mesmas respondem de maneira aumentada a alguns estímulos, como aos fatores de crescimento (MORAES, 2016). Sendo assim, a iniciação deve ser seguida pela ação de um agente promotor (também conhecido como oncopromotor). Este agente não causa câncer por si só, mas sua colaboração é essencial para permitir que as células venham a se proliferar de maneira desorganizada, levando à expansão clonal das células iniciadas (BARRETO, 1998). Durante a promoção, a célula iniciada é transformada em célula maligna de forma lenta e gradual, onde é necessário um longo e continuado contato com o agente promotor para que ocorra essa transformação. É preciso destacar também que, durante este estágio, a suspensão do contato com o agente oncopromotor muitas vezes interrompe o crescimento tumoral (ALMEIDA *et al.*, 2005; SILVA *et al.*, 2004; VINCENT & GATENBY, 2008).

O estágio final da carcinogênese corresponde à progressão, processo considerado irreversível e que é caracterizado pelo aumento da instabilidade genômica e posterior progresso em direção à malignidade (metástase) e ao crescimento celular

autônomo. Ou seja, neste estágio o crescimento celular é acelerado, não havendo necessidade de contato com o agente oncopromotor, e as células tumorais apresentam capacidade de invadir outros órgãos e tecidos. Sendo assim, na progressão o câncer já está instalado, evoluindo até o surgimento das primeiras manifestações clínicas da doença (ALMEIDA *et al.*, 2005; LOUREIRO *et al.*, 2002). O processo de carcinogênese está ilustrado na Figura 1.

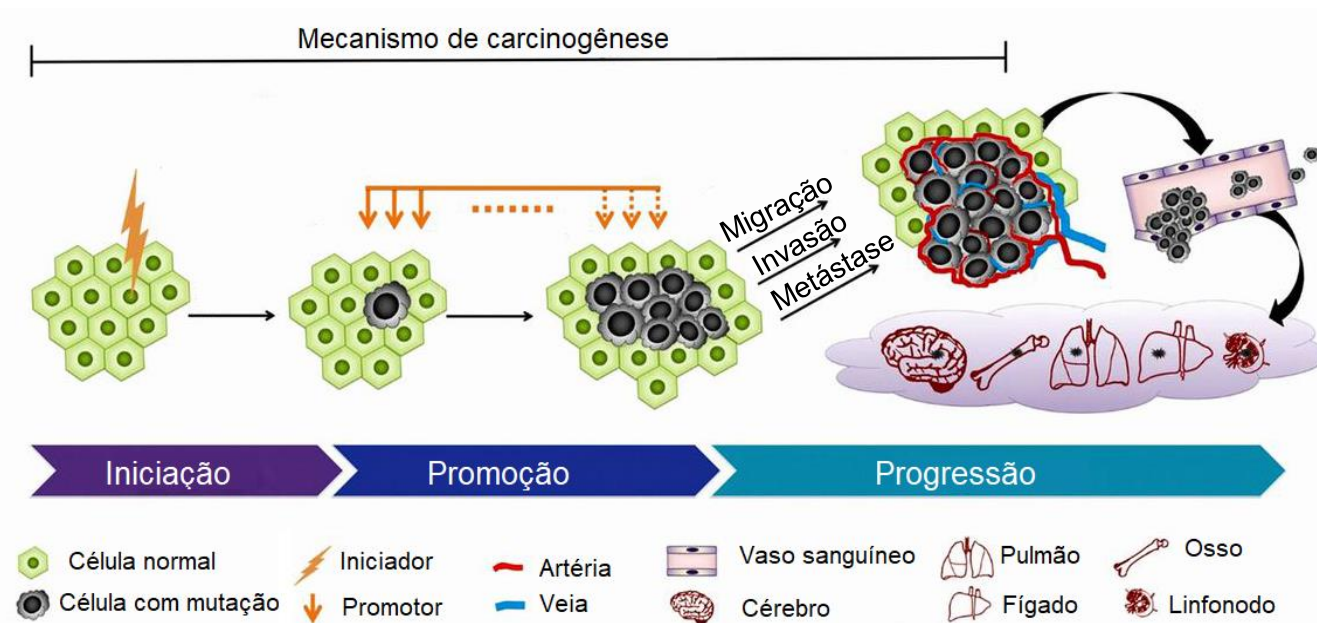


Figura 1 - Representação das três fases do processo de carcinogênese: iniciação, promoção e progressão. Adaptado de LIU *et al.*, 2015.

O uso da imagem foi autorizado pela AME Publishing Company (número de licença: 1203138-1).

1.3 Hallmarks of Cancer

O estudo do câncer é algo extremamente complexo e que envolve o conhecimento de diferentes áreas da ciência. Sendo assim, dois cientistas americanos, Hanahan e Weinberg, propuseram um princípio organizador, denominado *Hallmarks of Cancer*, para racionalizar a complexidade da doença em questão. Os *Hallmarks of Cancer* (em livre tradução, “marcas registradas do câncer”) correspondem a capacidades biológicas características que permitem o crescimento do tumor e a

metástase, levando ao desenvolvimento das doenças neoplásicas (HANAHAN & WEINBERG, 2000, 2011).

Segundo Hanahan e Weinberg, há ao menos seis alterações essenciais na fisiologia celular que ditam o crescimento tumoral. Essas alterações são compartilhadas pela maioria, ou talvez todos os tumores malignos, sendo elas: a autossuficiência em fatores de crescimento, insensibilidade a fatores inibitórios do crescimento; evasão à morte celular programada (apoptose), potencial replicativo ilimitado, angiogênese sustentada, e invasão tecidual e metástase. Cada uma dessas mudanças fisiológicas representa o rompimento de um mecanismo anticancerígeno de defesa (HANAHAN & WEINBERG, 2000). Além destas seis alterações, em 2011 Hanahan e Weinberg propuseram outras duas capacidades distintas e complementares que permitem o crescimento do tumor e a disseminação metastática: a reprogramação do metabolismo energético e a fuga do sistema imunológico (Figura 2). Ao lado destas alterações está a instabilidade do genoma, a qual gera a diversidade genética que acelera as alterações celulares; e também a inflamação, a qual pode contribuir para múltiplos *hallmarks*, fornecendo, por exemplo, fatores de crescimento que sustentam a sinalização proliferativa e também enzimas que facilitam a angiogênese, invasão e metástase (HANAHAN & WEINBERG, 2011). O reconhecimento da aplicabilidade generalizada destes conceitos influencia cada vez mais o desenvolvimento de novos meios para diagnosticar e tratar o câncer.

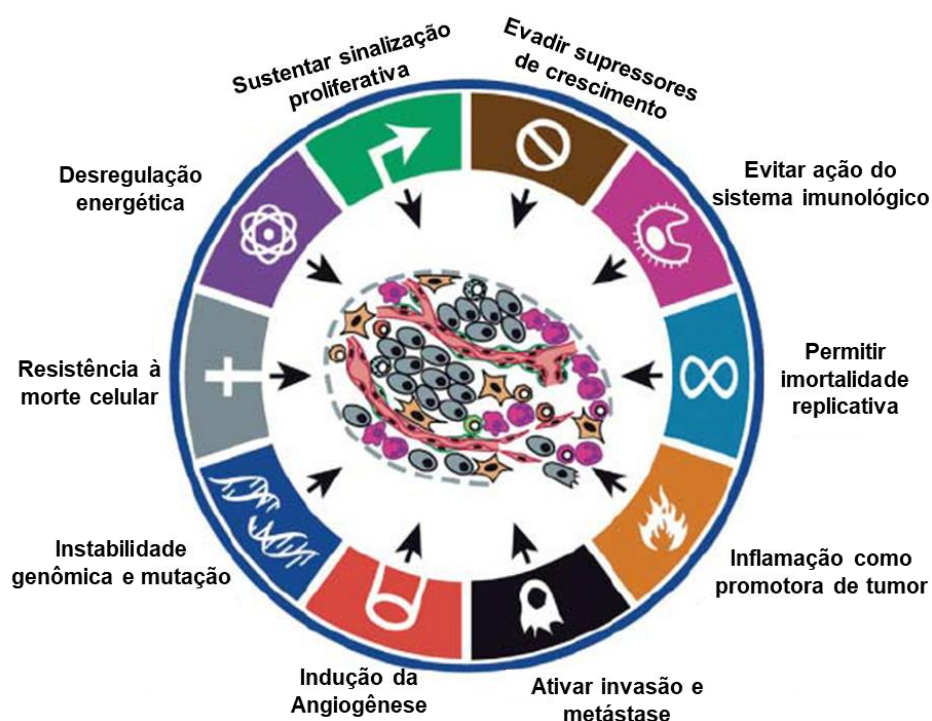


Figura 2 - Capacidades biológicas características adquiridas pelas células tumorais (*Hallmarks of Cancer*). Adaptado de HANAHAN & WEINBERG, 2011.

O uso da imagem foi autorizado pela Elsevier (número de licença: 5271371191019).

1.4 Diagnóstico do câncer: a busca por novos métodos de análise

O diagnóstico do câncer é geralmente realizado através de técnicas de diagnóstico por imagem, como ultrassom, radiografia, cintilografia, ressonância magnética, colonoscopia, tomografia entre outros. Esses exames auxiliam a identificar alterações e também a mensurar as dimensões do tumor. Essas técnicas normalmente são complementadas por exames cito e/ou histopatológicos. Neste caso, é realizada uma biópsia, onde é retirada uma pequena quantidade de tecido da região em que se suspeita que haja o tumor, material este que é posteriormente analisado laboratorialmente. Os exames cito e histopatológicos também podem ser utilizados como métodos preventivos para o rastreamento do câncer, como é o caso do exame Papanicolau (câncer do colo do útero) (INCA, 2021a; MONEDIRO, 2018; WHO 2022). Outros métodos diagnósticos que merecem destaque estão associados à investigação de marcadores tumorais. Esses marcadores são, geralmente, macromoléculas

(principalmente proteínas) presentes no tumor, no sangue, ou em outros fluidos biológicos, cujo aparecimento e/ou alterações em suas concentrações estão relacionados com a gênese e o crescimento de células cancerígenas (ALMEIDA *et al.*, 2007). Temos como exemplo de marcadores tumorais mais empregados atualmente no diagnóstico do câncer: a análise dos níveis sanguíneos de antígeno prostático específico (PSA) (câncer de próstata), dosagem sérica de alfa-fetoproteína (AFP) (câncer no testículo, ovários e, principalmente, no fígado), e o antígeno CA-27–29 (marcador tumoral do câncer de mama).

Outro processo de suma importância na avaliação da patologia em questão é o estadiamento do câncer. O estadiamento corresponde à determinação da extensão do câncer, ou seja, qual é o tamanho do tumor e qual é o grau de disseminação da doença (INCA, 2021b). Há diferentes tipos de sistemas de estadiamento; contudo, de maneira geral, os dois mais utilizados são o sistema TMN e o agrupamento em estágios. O sistema TNM é determinado pela American Joint Committee on Cancer (AJCC) e pela União Internacional de Controle do Câncer (UICC). Neste sistema, T se refere ao tamanho e à extensão do tumor principal, podendo receber números de 1 a 4, onde 1 representa um tumor pequeno e 4 um tumor grande. N se refere ao número de linfonodos adjacentes que apresentam células cancerígenas - podendo variar entre 0 (nenhum linfonodo contendo células cancerígenas) e 3 (muitos linfonodos contendo células cancerígenas). Por fim, M se refere à presença de metástase - podendo ser 0 (o câncer não se espalhou) ou 1 (o câncer se espalhou para outros órgãos). Assim, por exemplo, um tumor maligno pequeno que atingiu alguns gânglios linfáticos, mas que não apresenta metástase à distância, pode ser classificado como T1N1M0. Já no agrupamento por estágios, o câncer pode ser classificado através do uso de um número romano de I a IV, onde quanto maior o número, mais avançado e agressivo o câncer se encontra. No estágio I, o tumor é pequeno e não se espalhou. No estágio II, o câncer é maior, porém ainda não invadiu tecidos vizinhos. No estágio III, o tumor é grande e provavelmente começou a se espalhar para tecidos vizinhos e também para linfonodos adjacentes. No estágio IV, o câncer já se espalhou pelo organismo, invadindo outros órgãos (Mahul *et al.*, 2017).

As informações sobre o estadiamento do câncer são utilizadas para dar o prognóstico do paciente e também para o planejamento do tratamento. Neste sentido, quanto mais cedo o câncer for detectado, melhores são as chances de recuperação e, quanto mais cedo um tratamento adequado for iniciado, maior é a probabilidade de sua eficácia (INCA 2021c; WHO, 2022). No entanto, existem dois grandes obstáculos para o estabelecimento de um diagnóstico e tratamento precoce. De uma maneira geral, não há sinais e sintomas característicos do câncer que são facilmente identificáveis nos estágios iniciais desta doença. Além disso, mesmo quando o paciente apresenta sinais e sintomas, o diagnóstico frequentemente requer exames caros, demorados e invasivos (como discutido anteriormente). Desta maneira, muitas vezes os exames diagnósticos não são facilmente acessíveis a todos os tipos de pacientes (COSTA & DE MARTINIS 2020; ZHOU *et al.*, 2017). Por essas razões, há uma necessidade global para a pesquisa e desenvolvimento de metodologias rápidas, robustas, confiáveis, de baixo custo e não invasivas para a detecção precoce da patologia em questão (MONEDEIRO *et al.*, 2020).

Neste cenário, a busca por novos métodos e ferramentas diagnósticas que atendam às necessidades previamente citadas vem atraindo cada vez mais a atenção de pesquisadores de diversos campos da ciência, como a biomedicina, biofísica, bioinformática e química analítica (ULANOWSKA *et al.*, 2012). Nesta perspectiva, é crescente o interesse da aplicação de ferramentas “ômicas” (genômica, proteômica, e, principalmente, metabolômica) na investigação de métodos de *screening* (triagem) do câncer. Os métodos de *screening* se baseiam na análise de perfis de determinadas substâncias de interesse mediante diferentes circunstâncias biológicas, visando uma identificação presuntiva (não definitiva) de certa patologia (MONEDEIRO, 2018).

A metabolômica corresponde à análise abrangente dos metabólitos presentes em um determinado sistema biológico. Ou seja, é a investigação dos compostos de baixa massa molecular (<1500 Da) que constituem os produtos finais do metabolismo e que estão envolvidos nos processos de manutenção, crescimento e sinalização celular. Assim, os metabólitos formam a “impressão digital química” de um organismo em um determinado momento (CAMBIAGHI *et al.*, 2017). Em termos analíticos, a metabolômica emprega instrumentação analítica de última geração, em conjunto com

técnicas de bioinformática, para monitorar e descobrir alterações metabólicas relacionadas a uma determinada doença (BEGER, 2013).

Uma vez que os metabólitos representam a resposta final do organismo a fatores internos ou externos, os mesmos estão estritamente associados com as mudanças fisiológicas estabelecidas em um quadro patológico (BURTON & MA, 2019). Conseqüentemente, como o câncer é uma doença que leva a uma grande alteração do metabolismo celular (como discutido na seção 1.3), a metabolômica é um campo que tem um grande potencial de aplicação na descoberta de possíveis biomarcadores do câncer, podendo assim auxiliar no desenvolvimento de métodos diagnósticos e de tratamento desta patologia (BEGER, 2013; BURTON & MA, 2019; COSTA & DE MARTINIS, 2020). Neste contexto, os biomarcadores podem ser definidos como espécies químicas que são medidas e identificadas com precisão e reprodutibilidade em amostras biológicas, e que indicam repostas a determinadas condições (pato)fisiológicas (STRIMBU *et al.*, 2010).

Diferentes conjuntos de metabólitos podem ser estudados para esse determinado objetivo. Como exemplo, os lipídios desempenham funções críticas na sobrevivência, proliferação, interação e morte celular, processos os quais são fortemente relacionados à carcinogênese. Conseqüentemente, a lipidômica é um dos subconjuntos da metabolômica explorados na descoberta de possíveis biomarcadores do câncer (PERROTTI *et al.*, 2016). A avaliação do perfil de aminoácidos e ácidos nucleicos também é muito empregada para este fim (CHAN *et al.*, 2014; UBHI *et al.*, 2012; ZHANG *et al.*, 2021).

Neste cenário, uma das abordagens metabolômicas mais promissoras atualmente é a denominada “volatilômica”, ou seja, a análise dos compostos orgânicos voláteis (do inglês *Volatile Organic Compounds* - VOCs) presentes em diferentes matrizes biológicas (SCHMIDT & PODMORE, 2015).

1.5 Análise de VOCs no diagnóstico do câncer

Os VOCs correspondem a um diverso grupo de pequenas moléculas, geralmente com cadeia carbônica de C5 a C20, que possuem baixa massa molecular (até 500 Da), baixo ponto de ebulição (até 250 °C) e alta pressão de vapor

(MANSUROVA *et al.*, 2017). Na literatura, os VOCs estão ligados a diferentes temas. No contexto ambiental, por exemplo, são tratados como contaminantes presentes na água e na atmosfera (LIU *et al.*, 2020; LOMONACO *et al.*, 2020). No contexto ecológico, a liberação de VOCs corresponde a um dos principais mecanismos de defesa e comunicação de plantas (DOS SANTOS *et al.*, 2022). No âmbito ocupacional, consistem em agentes químicos presentes no ambiente e que estão associados a doenças do trabalho (ARI *et al.*, 2011). No que diz respeito aos VOCs como compostos endógenos (metabólitos), há uma diversidade de estudos que avaliam a presença dessas substâncias em diferentes matrizes biológicas, como no ar exalado, sangue, urina, fezes, fluido oral, suor, entre outras (ARGAWAL *et al.*, 2016; DE LACY COSTELLO *et al.*, 2014; SCHMIDT & PODMORE, 2015).

Nos dias atuais já é bem estabelecido na literatura que as doenças levam a alterações no estado fisiológico e metabólico do indivíduo. Em um quadro patológico, é esperado que as concentrações dos VOCs presentes normalmente no organismo possam ser alteradas, bem como novos VOCs possam ser gerados. Contudo, essa noção de relacionar substâncias voláteis a doenças vem desde a antiguidade, mais especificamente desde aproximadamente 400 AC, época em que viveu de Hipócrates, considerado por muitos como o pai da medicina. Hipócrates instruía seus alunos a cheirar o hálito de seus pacientes, bem como a despejar escarro em brasas para produzir um odor característico que seria indicativo de uma patologia (BOS *et al.*, 2013). A medicina tradicional chinesa também usou do poder do olfato para o diagnóstico de certas enfermidades. A diabetes, por exemplo, podia ser diagnosticada através do “cheiro de maçã podre” na urina (BULJUBASIC & BUCHBAUER, 2015).

Em termos mais recentes, o interesse da análise de VOCs em matrizes biológicas passou a ser crescente após trabalho publicado por Linus Pauling em 1971 (PAULING *et al.*, 1971). Neste trabalho, o vencedor do prêmio Nobel e seus colaboradores utilizaram a técnica de cromatografia em fase gasosa para identificar mais de 250 compostos voláteis presentes no ar exalado e mais de 280 na urina. Segundo os autores, essa técnica teria grandes aplicações nos princípios da medicina ortomolecular.

Os recentes avanços das técnicas analíticas e de bioinformática confirmaram que, de fato, a investigação do perfil de VOCs tem um potencial muito grande de aplicação na área clínica. Neste sentido, os VOCs vêm ganhando destaque em relação a outros tipos de análises principalmente devido à coleta das amostras, a qual é, geralmente, realizada de uma maneira rápida, simples e não invasiva (BULJUBASIC & BUCHBAUER, 2015; MONTEIRO *et al.*, 2017). Sendo assim, a volatilômica vem sendo explorada no contexto de diversas doenças, como a diabetes (BEHERA *et al.*, 2019), colite ulcerativa (SMOLINSKA *et al.*, 2017), asma (VAN VLIET *et al.*, 2017), síndrome do intestino irritável (BARANSKA *et al.*, 2016) e, principalmente, o câncer (SHIGEYAMA *et al.*, 2019; MARKAR *et al.*, 2019; RUDNICKA *et al.*, 2019).

O principal mecanismo de formação dos VOCs associados ao câncer é a peroxidação lipídica dos ácidos graxos poli-insaturados presentes nas membranas celulares, reação esta que é consequência principalmente do estresse oxidativo. Neste processo, hidrocarbonetos voláteis são formados e podem sofrer biotransformações, sendo oxidados através de reações enzimáticas (catalisadas, por exemplo, pelas enzimas do citocromo P450), ou podem permanecer com a estrutura inalterada, sendo incorporados posteriormente às diversas matrizes biológicas (MONEDEIRO *et al.*, 2018; VAN DE KANT *et al.*, 2012). Maiores detalhes sobre o mecanismo de formação dos VOCs estão presentes no Capítulo III.

Com relação aos trabalhos presentes na literatura, o maior número de estudos se refere às neoplasias de pulmão, analisando-se a presença de VOCs no ar exalado (KRILAVICIUTE *et al.*, 2015). Como exemplo, Hua e colaboradores (HUA *et al.*, 2018) realizaram uma revisão sistemática de estudos relacionados à análise de VOCs presentes no ar exalado como método de *screening* para o câncer de pulmão e identificaram 38 trabalhos realizados, envolvendo mais de 4800 pacientes. Os pesquisadores concluíram que os VOCs presentes no ar exalado têm potencial para serem utilizados para o diagnóstico da doença em questão, sendo que especialmente os VOCs das classes de alcanos e aldeídos precisam de maior atenção por parte de futuras pesquisas.

Contudo, os trabalhos presentes na literatura não se restringem somente ao câncer de pulmão e/ou à análise de ar exalado. Por exemplo, Wang e colaboradores (WANG *et al.*, 2014) analisaram os VOCs presentes no sangue de pessoas saudáveis e pessoas com câncer colorrectal e os resultados sugerem que, em comparação com indivíduos saudáveis, os pacientes oncológicos exibiram um perfil metabólico sanguíneo distinto com relação aos VOCs. Em revisão publicada por Costa e De Martinis (COSTA & DE MARTINIS, 2020), foram identificados 25 trabalhos relacionados à análise de VOCs urinários no contexto do diagnóstico do câncer entre os anos de 2011 e 2020. Mais de 10 diferentes tipos de câncer foram estudados e 180 diferentes VOCs foram reportados como possíveis biomarcadores da doença em questão. Os VOCs mais frequentemente citados foram o hexanal, dissulfeto de dimetila e fenol (este último sendo fortemente relacionado ao câncer de mama). A literatura apresenta outros diversos exemplos de trabalhos que utilizam a abordagem volatilômica para o desenvolvimento de métodos diagnósticos para o câncer (BOSCH *et al.*, 2019; GE *et al.*, 2022; JAHANGIRI-MANESH *et al.*, 2022; LIMA *et al.*, 2018; RODRIGUES *et al.*, 2018; VAN DE GOOR *et al.*, 2019). A Tabela 2 apresenta alguns dos principais estudos realizados nos últimos anos (de 2012 a 2022) que investigam os VOCs no contexto do câncer.

Tabela 2 - Exemplos de estudos sobre a investigação de VOCs no contexto do câncer (em ordem alfabética).

(Continua)

Autor (ano)	Técnica de análise	Amostra biológica	Tipo de câncer	Biomarcadores (Número de VOCs)
Altomare <i>et al.</i> (2013)	GC-MS	Ar exalado	Câncer colorretal	(5): 1,3-Dimetilbenzeno; 1,2-pentadieno; ciclohexano; metilciclohexano; 4-metiloctano
Bouza <i>et al.</i> 2017	GC-MS	Ar exalado	Câncer da cavidade oral	(7): Undecano; dodecano; decanal; 3,7-dimetilundecano; 4,5-dimetilnonano; 1-octeno; hexadecano
Buszewski <i>et al.</i> (2012)	GC-MS e cães farejadores	Ar exalado	Câncer de pulmão	(2): Acetato de etila; 2-pentanona
Filipiak <i>et al.</i> (2014)	GC-MS	Ar exalado e células de câncer de pulmão	Câncer de pulmão	(2): Etanol; n-octano
Hanai <i>et al.</i> (2012)	GC-MS	Urina	Câncer de pulmão	(4): 2-Etil-1-hexanol; 2-metilpirazina; 2-pentanona; tetraidrofurano.
Kumar <i>et al.</i> (2013)	SIFT-MS	Ar exalado	Câncer gastroesofágico	(4): Ácido hexanoico; fenol; 3-metilfenol; 3-etilfenol

(Conclusão)

Autor (ano)	Técnica de análise	Amostra biológica	Tipo de câncer	Biomarcadores (número de VOCs)
Kwak <i>et al.</i> (2013)	GC-MS	Melanócitos	Melanoma	(3): Álcool isopropílico; álcool isoamílico; ácido isovalérico
Lavra <i>et al.</i> (2015)	GC-MS	Células de câncer de mama	Câncer de mama	(3): 2,4-dimetilheptano; 1,3-di-tert-butilbenzeno; 2-xileno.
Monteiro <i>et al.</i> (2017)	GC-MS	Urina	Carcinoma de células renais	(2): 2-Oxopropanal; 2,5,8-Trimetil-1,2,3,4-(6): tetrahidronaftalene-1-o.
Silva <i>et al.</i> (2012)	GC-MS	Urina	Câncer de mama	(6): 3-Carene; 1,2,4-trimetilbenzeno; 2-metoxitiofeno; 3-heptanona; dissulfeto de dimetila; fenol
Van de Goor <i>et al.</i> (2018)	<i>e-nose</i>	Ar exalado	Câncer de pulmão	-
Xu <i>et al.</i> (2013)	Nanosensores	Ar exalado	Câncer de estômago	(5): Acrilonitrila; 2-butoxietanol; furfural; 6-metil-5-heptan-2-ona; isopreno

GC-MS: do termo em inglês *gas chromatography-mass spectrometry*;
 SIFT-MS: do termo em inglês *Selected-ion flow-tube mass spectrometry*;
e-nose: do termo em inglês *eletronic nose*.

As análises clínicas tradicionalmente são realizadas utilizando-se sangue e urina como amostras biológicas. Já no contexto da volatilômica, a matriz mais empregada é o ar exalado, como citado anteriormente. Contudo, nos últimos anos observa-se um crescente interesse na análise de VOCs em matrizes biológicas alternativas. Como exemplo, Barbosa e colaboradores (Barbosa *et al.*, 2019) analisaram amostras de cera de ouvido de 52 pessoas saudáveis e 50 pessoas com câncer (principalmente linfoma e leucemia). Vinte e sete VOCs foram selecionados como potenciais biomarcadores da doença e forneceram 100% de discriminação entre os grupos de pacientes oncológicos e indivíduos saudáveis. Ainda, Monedeiro e colaboradores (MONEDEIRO *et al.*, 2020) analisaram os VOCs presentes no suor e conseguiram distinguir com máxima sensibilidade e especificidade (100%) os pacientes do grupo saudável e com câncer em um modelo baseado na incidência de 2-etil-1-hexanol, hexanal e octanal.

Conforme exposto anteriormente, a pesquisa envolvendo VOCs é bastante diversificada, abordando diferentes tipos de câncer e diferentes tipos de amostra. Contudo, há uma matriz biológica alternativa que tem um potencial que ainda pode ser muito mais explorado na análise de VOCs no contexto do diagnóstico do câncer: o fluido oral.

1.6 VOCs presentes no fluido oral: potencial aplicação na área clínica

O fluido oral é o resultado da mistura do fluido produzido pelas glândulas salivares maiores (parótida, submandibular e sublingual) e glândulas salivares menores, com outros constituintes presentes na boca, como células epiteliais da mucosa, restos alimentares e de medicamentos (CROUCH *et al.*, 2004). Esta matriz é composta majoritariamente por água (99%), proteínas (como enzimas, imunoglobulinas e glicoproteínas da mucosa, representando cerca de 0,3%) e substâncias inorgânicas (0,2%) (COSTA *et al.*, 2021a; YOSHIZAWA *et al.*, 2013). Em média, um indivíduo normal saliva de 0,3 a 0,7 mL por minuto, produzindo cerca de 1,0 a 1,5 litros de saliva por dia. Este fluido tem a função de facilitar a digestão, a deglutição, aguçar o paladar, além de ser uma barreira contra patógenos (YOSHIZAWA *et al.*, 2013).

É preciso destacar que é necessário cuidado para não confundir o termo “fluido oral” com “saliva”, termo este que se refere ao fluido coletado de uma glândula salivar

específica e que é isento de outros materiais (CROUCH *et al.*, 2004). Embora esses dois termos sejam usados muitas vezes de forma intercambiável na literatura, “fluido oral” descreve com maior precisão as características biológicas da matriz utilizada neste trabalho. Portanto, este será o termo usado predominantemente neste manuscrito.

O uso do fluido oral como matriz analítica possui diversas vantagens em relação a outros tipos de amostra. Por exemplo, o fluido oral é menos complexo que as matrizes convencionais sangue e urina. Já a coleta é rápida, fácil e poder ser realizada praticamente a qualquer hora/local e em qualquer pessoa. Além disso, é um método barato e não invasivo (MARTÍ-ÁLAMO *et al.*, 2012; SEIDI *et al.*, 2019). Em relação às desvantagens, o volume de fluido oral para testagem é limitado e sua composição é afetada por diversos fatores, como idade, consumo de água, alimentos e medicamentos e o ciclo circadiano. Ainda, os analitos geralmente estão presentes em concentrações muito baixas, exigindo métodos de análise que apresentem grande capacidade de detecção (COSTA *et al.*, 2021a; ELMONGY & ABDEL-REHIM, 2016).

Tem sido demonstrado que o fluido oral apresenta fortes correlações com o plasma e com o soro em relação aos componentes celulares e moleculares, fazendo com que essa matriz possa atuar como um “espelho” que reflete o estado geral de saúde do indivíduo (MOTAMAYEL *et al.*, 2018). Desta maneira, o fluido oral vem sendo utilizado como matriz em avaliações de biomonitoramento humano e também na análise de diferentes tipos de doenças, como doenças cardiovasculares, renais, infecções e desordens metabólicas, incluindo diabetes e Síndrome de Cushing (KHURSHID *et al.*, 2019; VISWANATH *et al.*, 2017). Portanto, a análise de VOCs no fluido oral pode oferecer informações importantes sobre as condições fisiológicas e patológicas do organismo e, desta maneira, é uma matriz com grande potencial para ser explorada no âmbito da volatilômica.

A análise de VOCs na matriz em questão é baseada no fato de que os compostos orgânicos voláteis gerados nos processos celulares em diferentes regiões do organismo passam para o sangue e, em seguida, migram para o fluido oral principalmente através dos mecanismos de difusão passiva e ultrafiltração (AMANN *et al.*, 2014). Além dos VOCs de origem endógena, há outros fatores que podem afetar a

composição dos VOCs presentes no fluido oral. Por exemplo, já foram reportadas mais de 300 espécies de bactérias presentes na cavidade oral. Consequentemente, essa microbiota também contribui para o perfil de VOCs presentes nesta matriz (MONEDEIRO *et al.*, 2021). Ainda, outros fatores, como restos alimentares, produtos cosméticos e de higiene, exposição a poluentes do ar, o refluxo gastresofágico e também o ciclo circadiano (como citado anteriormente) podem influenciar a composição (HUDA *et al.*, 2013).

Apesar de ter grande potencial de aplicação na área clínica, ainda não há muitos trabalhos na literatura no que concerne a investigação dos VOCs presentes no fluido oral como possível método de diagnóstico do câncer. Neste sentido, cinco estudos recentes merecem destaque. Shigeyama e colaboradores (SHIGEYAMA *et al.*, 2019) identificaram 27 VOCs presentes no fluido oral em concentrações significativamente diferentes entre o grupo de controle (n = 50) e o grupo de pacientes com câncer de cavidade oral (n = 25). Desses 27 VOCs, 12 foram considerados possíveis biomarcadores da doença em questão. Taware e colaboradores (TAWARE *et al.*, 2018) identificaram 4 VOCs com forte evidências de estarem relacionados com mudanças metabólicas no câncer de cabeça e pescoço, sendo eles: 1,4-diclorobenzeno; 1,2-decanodiol; 2,5-bis(1,1-dimetil)etilfenol; e E-3-decen-2-ol. Em estudo semelhante, Cavaco e colaboradores (CAVACO *et al.*, 2018) analisaram amostras de fluido oral de pessoas saudáveis e pessoas com câncer de mama em Portugal e na Índia e identificaram um padrão muito diferente no perfil de VOCs entre a os voluntários dos dois países. Para os voluntários portugueses, ácido 3-metil-butanoico; ácido 4-metil-butanoico; fenol e 4-tert-butilfenol foram os VOCs estatisticamente relevantes para a discriminação entre o grupo de pessoas saudáveis e com câncer. Já para os voluntários indianos, os principais VOCs foram o ácido acético, ácido propanoico, ácido benzoico, 1,2-decanediol, 2-decanona e decanal. Antón e colaboradores (ANTÓN *et al.*, 2016) utilizaram sensores artificiais (mais conhecidos como *eletronic nose* ou *e-nose*) e discriminaram com 100% de eficiência amostras de fluido oral de pessoas saudáveis (n = 32) e amostras de pessoas com diferentes tipos de câncer (n = 23). Por fim, Monedeiro e colaboradores (MONEDEIRO *et al.*, 2021) analisaram o perfil de VOCs presentes no fluido oral de pessoas saudáveis (n = 15), de pessoas com lesões na

cavidade oral (n = 10) e pessoas com câncer da cavidade oral (n = 5). Os autores construíram um modelo capaz de prever os casos de câncer com 100% de sensibilidade e especificidade após a remoção dos VOCs de origem bacteriana. Os principais aspectos desses estudos estão sumarizados na Tabela 3.

Além do uso de fluido oral, há outro tipo de ferramenta que apresenta um grande potencial a ser explorado no âmbito da volatilômica: a análise híbrida do perfil de VOCs.

Tabela 3 - Revisão dos principais estudos que exploram a análise dos VOCs presentes no fluido oral para o diagnóstico do câncer (em ordem alfabética).

Autor (ano)	Técnica de análise	Tipo de câncer	Biomarcadores (número de VOCs)
Antón <i>et al.</i> (2016)	<i>e-nose</i>	Diferentes tipos (maioria pulmão)	-
Cavaco <i>et al.</i> (2018)	GC-MS	Câncer de mama	Amostras de Portugal (4): ácido 3-metil-butanoico, ácido 4-metil-butanoico, fenol; 4-tert-butilfenol Amostras da Índia (6): ácido acético; ácido propanoico; ácido benzoico; 1,2-decanediol; 2-decanona; decanal
Monedeiro <i>et al.</i> (2021)	GC-MS	Câncer de cavidade oral	(9): 1-octen-3-ol; ácido hexanoico; E-2-octenal; ácido octanoico; E-2-nonenal; ácido nonanoico; 2,4-decadienal; ácido 9-undecenoico
Shigeyama <i>et al.</i> (2019)	GC-MS	Câncer de cavidade oral	(12): Etanol; 2-pentanona; fenol; ácido hexadecanoico; undecano; 1-octanol; γ -butilrolactona; álcool benzílico, 3-heptanona; 1,3-butanodiol; 1,2 pentanodiol; 1-hexadecanol
Taware <i>et al.</i> (2018)	GC-MS	Câncer de cabeça e pescoço	(4): 1,4-diclorobenzeno; 1,2-decanodiol; 2,5-bis(1,1-dimetiletilfenol); E-3-decen-2-ol

1.7 Análise híbrida do perfil de VOCs

Em trabalho publicado por De Lacy Costello e colaboradores, foi realizada uma compilação dos VOCs presentes nos mais diversos fluidos biológicos de seres humanos saudáveis (DE LACY COSTELLO *et al.*, 2014). Neste estudo, um total de 1840 VOCs foram reportados, sendo distribuídos em: ar exalado (872), fluido oral (359), sangue (154), leite (256), secreções da pele (532), urina (279) e fezes (381). Essa compilação é de suma importância, uma vez que o conhecimento sobre o que é “normal” ajuda na avaliação sobre o que é “anormal” (condição patológica).

Cada matriz biológica apresenta um padrão de VOCs característico, com uma classe dominante diferente de compostos orgânicos. Por exemplo, o fluido oral apresenta uma quantidade considerável de hidrocarbonetos, cetonas e ácidos orgânicos, enquanto a urina apresenta um padrão mais rico em álcoois, furanos e éteres, assim como pode ser observado na Figura 3.

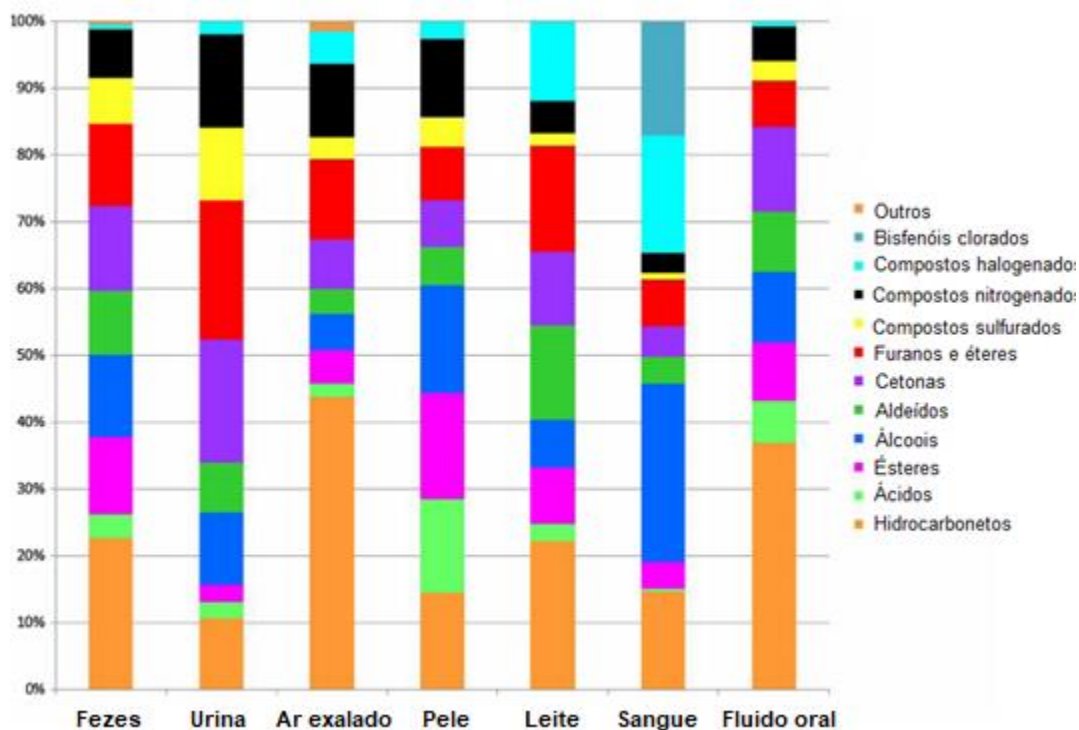


Figura 3 - Porcentagem de cada classe de compostos orgânicos voláteis detectados em diferentes fluidos biológicos de pessoas saudáveis. Adaptado de DE LACY COSTELLO *et al.*, 2014.

O uso da imagem foi autorizado pela International Association for Breath Research (número de licença: 1200794-1).

Na realidade, cerca de apenas 1% dos VOCs já identificados podem ser detectados em todos os fluídos biológicos, sendo que muitos dos estudos presentes na literatura visam justamente analisar essa grande diferença no perfil dos compostos orgânicos voláteis em cada matriz (BROZA *et al.*, 2015). Contudo, em estudo realizado por Broza e colaboradores, foi introduzido o conceito da análise híbrida de compostos orgânicos voláteis, uma abordagem que combina os perfis de VOCs obtidos de dois ou mais fluidos biológicos, com o objetivo de complementar a informação química sobre o estado fisiológico de um indivíduo (BROZA *et al.*, 2014). No trabalho em questão foi realizada uma análise para verificar se os VOCs presentes na pele e no ar exalado poderiam ser utilizados como biomarcadores do câncer de pulmão. Os dados obtidos sugerem que, apesar dessas matrizes serem pouco correlacionadas, a combinação dos dados das duas fontes leva a uma análise de um espectro mais amplo de substâncias, obtendo-se assim resultados com maior sensibilidade e especificidade.

Neste sentido, a análise híbrida, avaliando-se dois ou mais tipos de amostra de um único voluntário, pode ser uma ferramenta poderosa na determinação de VOCs como um método auxiliar no diagnóstico do câncer. Apesar deste grande potencial, até o presente momento ainda não há na literatura outros estudos que façam esta abordagem com relação à análise de VOCs.

1.8 Análise de VOCs presentes em matrizes biológicas por cromatografia em fase gasosa com detector de ionização por chama

A análise de VOCs em amostras biológicas no contexto do câncer é um grande desafio analítico. Até o presente momento, ainda não há um procedimento padrão de análise para nenhuma matriz avaliada e também não há uma técnica sendo usada, de fato, em uma rotina clínica. Há diversos métodos reportados na literatura, sendo que a grande maioria das análises é realizada através da técnica de cromatografia em fase gasosa (CG, ou do inglês *gas chromatography* - GC) acoplada ao espectrômetro de massas (EM, ou do inglês *mass spectrometry* - MS). Outros métodos que não dependem da separação cromatográfica também são comumente utilizados, como a espectrometria de massas de tubo de fluxo de íons selecionados (livre tradução do termo em inglês *Selected-ion flow-tube mass spectrometry*, SIFT-MS) e espectrometria

de massas de reação de transferência de prótons (livre tradução do termo em inglês *Proton-transfer-reaction mass spectrometry*, PTR-MS) (BAJTAREVIC *et al.*, 2009; KUMAR *et al.*, 2013). Além disso, os avanços da nanotecnologia e inteligência artificial também proporcionaram o advento de métodos de análise baseados em nanosensores e “narizes eletrônicos” (*e-noses*) (VAN DE GOOR *et al.*, 2018).

Através da espectrometria de massas é possível realizar a identificação dos compostos e, assim, determinar quais são as substâncias presentes em maior ou menor quantidade em um determinado grupo alvo (câncer ou saudável, por exemplo). Com essas informações é possível fazer uma investigação sobre possíveis biomarcadores e também analisar quais rotas metabólicas estão alteradas devido ao quadro patológico (COSTA & DE MARTINIS, 2020).

Contudo, a espectrometria de massas é uma técnica muito cara e que exige profissionais especializados para a utilização e manutenção do equipamento, sendo assim inviável em muitos laboratórios. Uma alternativa para a análise de VOCs neste contexto é a técnica de cromatografia em fase gasosa com detector de ionização por chama (DIC, ou do termo em inglês *flame ionization detector* – FID). O GC-FID é comumente disponível na maioria dos laboratórios analíticos, não é um equipamento tão caro, é fácil de operar e possui alta detectabilidade (semelhante ao GC-MS). Todavia, ao contrário da técnica de espectrometria de massas, onde é possível investigar a identidade de um composto através do seu espectro de fragmentação, com o FID só é possível obter informações sobre a intensidade e o tempo de retenção de um determinado pico. Consequentemente, a identificação dos compostos é realizada apenas através da comparação com padrões.

O GC-FID é frequentemente utilizado em estudos que visam estabelecer um *fingerprint*, ou seja, uma “impressão digital química” que é capaz de distinguir e/ou classificar as amostras de interesse sem necessariamente identificar os compostos presentes (CUADROS-RODRÍGUEZ *et al.*, 2016). Neste contexto, essa abordagem é comum na análise de produtos alimentícios e também na análise de petróleo e seus derivados. Como exemplo, essa técnica já foi aplicada na classificação de vinhos (JACOBSON *et al.*, 2013) e cervejas (JIAO *et al.*, 2011), na identificação sensorial de vinagre (DURANTE *et al.*, 2006), na predição da qualidade de chás (JUMTEE *et al.*,

2009), na diferenciação de amostras de diesel (KRAKOWSKA *et al.*, 2015) e também análises forenses de derramamento de óleo (NIELSEN *et al.*, 2012).

No contexto da análise de VOCs relacionados ao câncer, apesar de extensa procura na literatura, não foi encontrado nenhum artigo que aplicasse o GC-FID para realizar uma discriminação entre amostras de pessoas saudáveis e pacientes oncológicos através do estabelecimento de um *fingerprint*. Portanto, há uma lacuna na literatura que pode ser preenchida através deste trabalho.

2 HIPÓTESE

Os VOCs podem ser avaliados em amostras biológicas através da técnica de GC-FID com o objetivo de determinar uma “impressão digital química” do câncer. Ainda, ao utilizar duas amostras biológicas distintas de um mesmo paciente, um espectro mais amplo de substâncias é avaliado, obtendo-se assim diagnósticos mais confiáveis.

3 JUSTIFICATIVA

O mundo está passando por uma grande mudança demográfica: a população está cada vez mais envelhecendo e, com isso, está havendo uma transição dos principais tipos de doenças observados, com um declínio daquelas associadas a infecções, e o aumento das patologias crônicas (como é o caso do câncer). Neste sentido, o câncer representa um dos maiores problemas de saúde pública em todo o planeta e está entre as quatro principais causas de mortes prematuras (antes dos 70 anos de idade) na maioria dos países (INCA, 2019; ROSER & RITCHIE, 2019; WHO, 2020).

Essa grande incidência, aliada à elevada mortalidade, às dificuldades no diagnóstico, bem como ao fato de que quanto mais cedo a doença for detectada, melhores são as chances de recuperação, fazem com que o desenvolvimento de métodos não invasivos para a detecção do câncer seja um dos maiores desafios científicos do século XXI. Sendo assim, soluções inovadoras, simples e eficazes são necessárias para este fim. Nesta perspectiva, uma das técnicas mais promissoras atualmente é a análise de compostos orgânicos voláteis em amostras biológicas.

Desta maneira, este trabalho propõe uma abordagem inédita na literatura: a análise híbrida do perfil de VOCs através da técnica de GC-FID. Nesta abordagem são investigados os VOCs presentes em dois tipos de matrizes biológicas de um mesmo voluntário (fluido oral e urina – duas amostras não invasivas), utilizando-se uma técnica mais barata que a usual. É esperado que esse tipo de análise proporcione a combinação de dados complementares, gerando um *fingerprint* que pode ser utilizado, futuramente, como um método auxiliar para o diagnóstico do câncer.

4 OBJETIVOS

Avaliar os compostos orgânicos voláteis presentes no fluido oral e na urina de pessoas saudáveis e pacientes com câncer, com o intuito de discriminar os dois grupos através do perfil de VOCs.

5 ASPECTOS ÉTICOS DA PESQUISA

O presente trabalho possui aprovação do comitê de ética (CAAE 02519118.8.0000.5403). Atendendo à Resolução 466/12 do Conselho Nacional de Saúde, bem como às normas complementares à mesma, o projeto foi enviado para análise do Comitê de Ética em Pesquisa da Faculdade de Ciências Farmacêuticas de Ribeirão Preto (Anexos 1 e 2) e, após aprovação pela unidade, encaminhado ao Comitê de Ética em Pesquisa da Faculdade de Filosofia, Ciências e Letras e ao Comitê de ética em Pesquisa do Hospital das Clínicas e da Faculdade de Medicina de Ribeirão Preto (Anexos 3 e 4).

Esta pesquisa abrange a coleta de amostras de fluido oral e urina de indivíduos diagnosticados com câncer e a coleta de fluido oral e urina de pessoas saudáveis (que não foram diagnosticados com a patologia em questão). Os voluntários de ambos os grupos foram convidados a participar dessa pesquisa caso atendessem aos critérios de inclusão que estão descritos no próximo parágrafo. A fim de se estabelecer uma observação mais rigorosa entre os grupos envolvidos, as coletadas foram pareadas quanto à idade e gênero. Também foi realizado o registro do uso de medicamentos, consumo de álcool e tabaco. Além disso, tanto os indivíduos pertencentes ao grupo de

pacientes com câncer quanto os indivíduos saudáveis expressaram concordância com o Termo de Consentimento Livre e Esclarecido (Anexo 5).

Os critérios de inclusão para o primeiro grupo de interesse (indivíduos portadores de câncer) são: indivíduos com idade igual ou superior a 18 anos, diagnosticados com câncer e que ainda não iniciaram quimioterapia ou radioterapia junto aos Ambulatórios e Enfermaria de Oncologia Clínica do Hospital das Clínicas de Ribeirão Preto da FMRP-USP. Os critérios de inclusão para o segundo grupo de interesse (indivíduos não portadores de câncer) são: indivíduos com idade igual ou superior a 18 anos, sem histórico de qualquer diagnóstico positivo para câncer, e que, no caso de portarem diabetes, ou outra(s) doença(s) inflamatória(s), a(s) mesma(s) deve(m) estar sendo controlada(s) pelo uso de medicamentos adequados.

**CAPÍTULO II – DESENVOLVIMENTO E OTIMIZAÇÃO DE
MÉTODO PARA ANÁLISE DOS VOCS PRESENTES NO
FLUIDO ORAL E NA URINA**

1 INTRODUÇÃO

Conforme discutido no Capítulo I, os VOCs presentes em amostras biológicas podem ser analisados por diversos métodos analíticos, sendo mais comum a utilização de técnicas que envolvem a cromatografia em fase gasosa (GC). Além disso, como se tratam de compostos voláteis, os VOCs são extraídos das matrizes de interesse empregando-se, geralmente, técnicas de *headspace* (HS). Como os VOCs estão presentes normalmente em baixas concentrações, e o HS depende de uma série de fatores que influenciam na extração, é indicado o uso de ferramentas de planejamento experimental para estabelecer as melhores condições do processo em questão.

Sendo assim, neste capítulo será apresentado um sucinto referencial teórico sobre as técnicas de GC e HS, bem como sobre as ferramentas de planejamento experimental utilizadas neste presente estudo. Além disso, também será discutido o processo de desenvolvimento e otimização do método de análise dos VOCs presentes no fluido oral e na urina através da extração por *headspace* e posterior análise através da técnica de GC-FID.

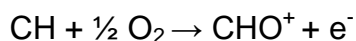
2 REFERENCIAL TEÓRICO

2.1 Cromatografia em fase gasosa

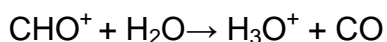
A GC é uma técnica analítica amplamente utilizada para a separação de gases ou de compostos volatilizáveis que sejam termicamente estáveis. Normalmente, as substâncias avaliadas são inferiores a 1000 Da, uma vez que é difícil vaporizar compostos maiores. Nesta técnica, a amostra normalmente é introduzida no sistema cromatográfico através de uma seringa, chegando primeiramente no injetor, onde é vaporizada no interior de um tubo de vidro inerte, mais conhecido como *liner*. Em seguida, um fluxo de gás de arraste é o responsável por conduzir essa amostra pela coluna capilar, a qual fica dentro de um forno que controla a temperatura da corrida cromatográfica. Na coluna ocorre a separação das substâncias de acordo com os diferentes pontos de ebulição e também de acordo com as diferentes interações com a fase estacionária. Após este processo, os analitos chegam ao detector, onde um sinal analítico é gerado, amplificado, e os dados são compilados na forma de um cromatograma (AQUINO NETO, 2003; COLLINS *et al.*, 2006).

Diversos sistemas de detecção podem ser associados ao cromatógrafo a gás. Dentre eles, os mais comumente utilizados são: o detector termiônico específico (do termo em inglês *thermionic specific detector*, TSD – o qual também é conhecido como *nitrogen phosphorus detector*, NPD), o espectrômetro de massas (MS), e, principalmente, o detector de ionização por chama (FID). O NPD é um detector específico para compostos que possuem nitrogênio e/ou fósforo em sua composição, sendo assim utilizado, por exemplo, na determinação de pesticidas e também na análise de substâncias psicoativas (WANG *et al.*, 2016; COSTA *et al.*, 2021b; Costa *et al.*, 2022). O GC-MS é um equipamento muito versátil, podendo ser empregado na investigação dos mais diversos tipos de compostos desde que os mesmos obedeçam aos princípios das análises por GC e também sejam ionizáveis. Sendo assim, o GC-MS apresenta uma vasta aplicação em diferentes campos da ciência e é considerado, muitas vezes, o padrão ouro nas análises toxicológicas e de controle de qualidade (MINA, 2020).

Já o FID é o detector mais utilizado no contexto da cromatografia em fase gasosa devido a sua simplicidade, aplicabilidade e custo. O princípio de funcionamento desta técnica é baseado no fato de que os compostos orgânicos presentes nas amostras sofrem uma reação de combustão e são ionizados por uma chama constituída por hidrogênio e ar sintético. Segue abaixo o principal mecanismo proposto para as reações que ocorrem no FID (VISENTAINER *et al.*, 2014):



Esta reação ocorre para cada 100000 átomos de carbono que são introduzidos na chama. Portanto, a resposta do FID é proporcional ao número de átomos de carbono que são queimados. Ainda, o íon CHO⁺ formado neste processo é instável e reage rapidamente com moléculas de água presentes na chama para produzir hidrônio, conforme a reação abaixo:



A formação desses cátions (CHO^+ e H_3O^+) faz com que a chama aumente sua condutividade elétrica, o que é convertido em um sinal elétrico (medido em picoamperes, pA) que é diretamente proporcional à concentração dos analitos na amostra (LANÇAS, 2016; MCMINN *et al.*, 2000). Um diagrama simplificado do detector FID se encontra na Figura 4.

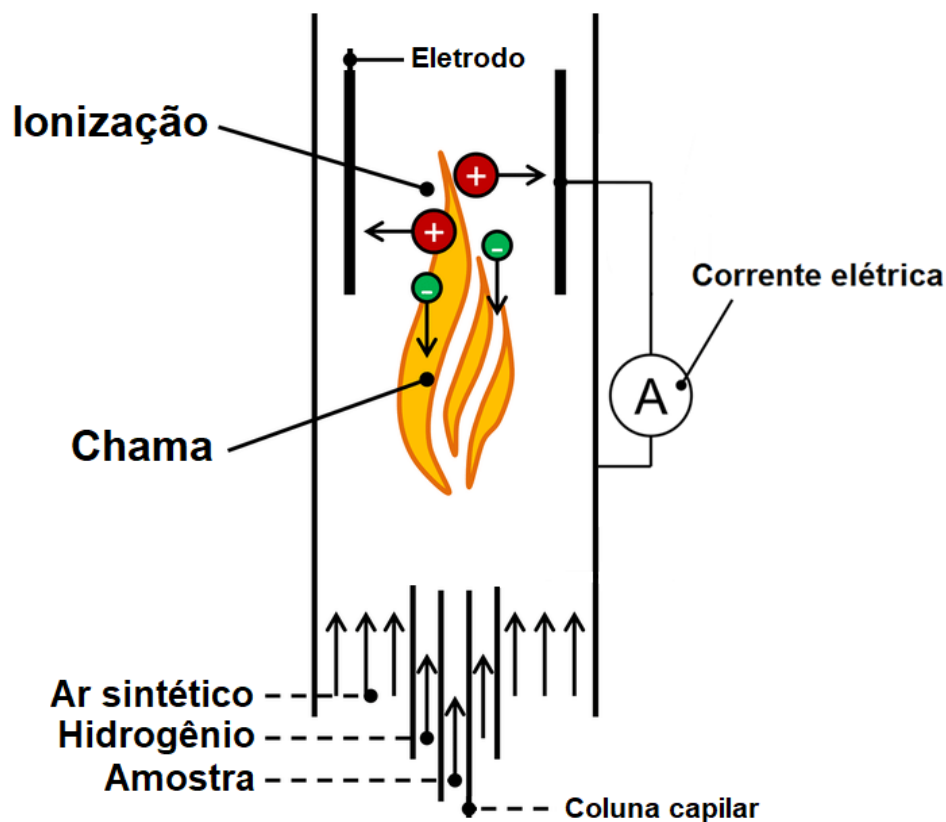


Figura 4 - Diagrama simplificado do detector de ionização por chama (FID). Adaptado de Förster *et al.*, 2019.

2.2 *Headspace*

Conforme relatado na seção anterior, geralmente os VOCs são extraídos das matrizes biológicas através de técnicas de *headspace* (termo que se refere ao espaço de ar confinado sobre uma amostra em um frasco fechado). Esta técnica consiste, sucintamente, na análise da fase de vapor que é formada após o aquecimento de uma amostra condicionada em um recipiente apropriado. Esse processo permite com que compostos de menor volatilidade não entrem no sistema cromatográfico, diminuindo o número de interferentes presentes no cromatograma (BORGES *et al.*, 2015).

As extrações baseadas no HS podem ser realizadas, principalmente, de três diferentes modos: HS estático (SHS), HS dinâmico, e com o auxílio de uma fibra de extração em fase sólida (do termo em inglês *solid phase microextraction* – SPME). No HS estático (Figura 5) a amostra (a qual pode ser sólida ou líquida) é colocada em um recipiente hermeticamente fechado e é aquecida durante um determinado tempo (denominado tempo de incubação). Em seguida, um volume adequado da fase de vapor é recolhido por uma seringa e é posteriormente injetado em um sistema de GC, onde é realizada a análise (RESTEK, 2000; SNOW & SLACK 2002). No HS dinâmico, o frasco é aquecido e um gás inerte arrasta a fase de vapor presente no HS da amostra, fazendo com que essa fase fique aprisionada em um material sorvente. Após o período de extração, os compostos voláteis são dessorvidos no injetor do GC, onde a análise é realizada. Devido a este mecanismo de ação, este modo também é chamado de *purge and trap* (do inglês “purga e aprisionamento”). Já na SPME, uma fibra revestida com material sorvente é exposta no HS da amostra, fazendo com que os analitos fiquem sorvidos no material em questão. Após certo tempo de equilíbrio, essa fibra é retirada do frasco de HS e inserida diretamente no injetor do cromatógrafo a gás, onde os analitos que foram previamente extraídos e pré-concentrados são dessorvidos e, posteriormente, são carregados para a coluna cromatográfica através da ação do gás de arraste (BORGES *et al.*, 2015).

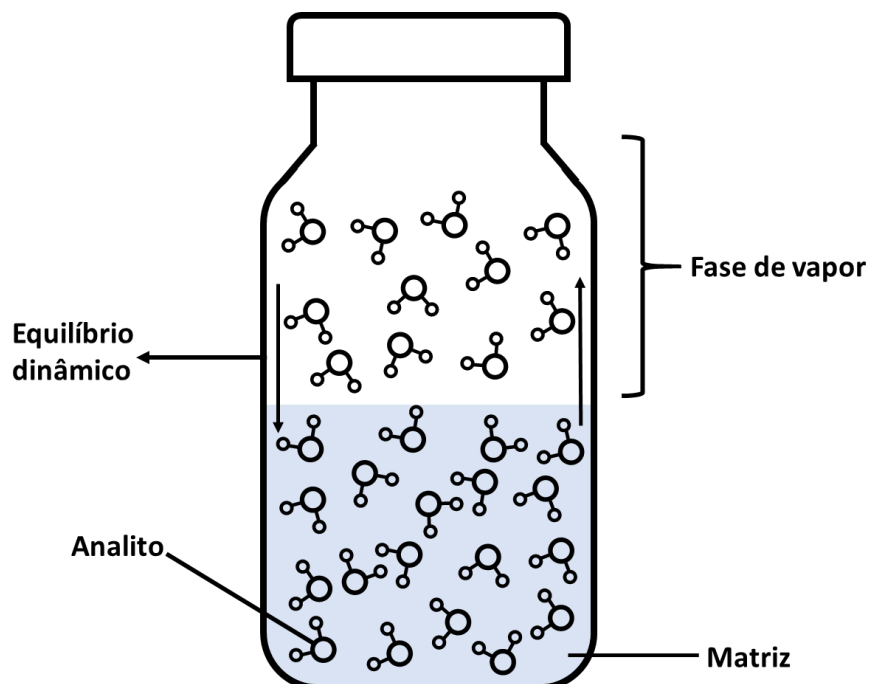


Figura 5 - Representação da técnica de *headspace* estático.

O SHS é consideravelmente rápido, simples e de baixo custo. Contudo, este modo não apresenta a mesma capacidade de extração do HS dinâmico e da SPME, métodos os quais possibilitam menores limites de detecção (SITHERSINGH & SNOW, 2021). A Tabela 4 apresenta uma comparação entre essas três técnicas discutidas.

Tabela 4 - Comparação entre as principais técnicas de extração baseadas no método de *headspace*.

Técnica	Concentração do analito	Principais vantagens	Principais desvantagens	Tempo de extração
<i>Headspace</i> estático	ppm	Baixo custo e simplicidade	Menor capacidade de extração	↑
<i>Headspace</i> dinâmico	ppb	Permite limites de detecção mais baixos	Alto custo; difícil automatização	↑↑
HS-SPME	ppt	Permite os menores limites de detecção; alta versatilidade; <i>solvent-free</i>	Alto custo; técnica laboriosa	↑↑↑

Como o SHS foi o modo utilizado neste presente estudo, na próxima seção serão discutidos os principais fundamentos teóricos deste método de extração.

2.3 Fundamentos teóricos do *headspace* estático

Os princípios que descrevem os fundamentos teóricos do SHS derivam de três leis físicas: a lei de Dalton, a lei de Henry e a lei de Raoult. A lei de Dalton afirma que a pressão total exercida por uma mistura gasosa é igual à soma das pressões parciais de cada componente individual dessa mistura. A lei de Henry afirma que, a uma temperatura constante, a quantidade de um gás dissolvido em um líquido é diretamente proporcional à pressão parcial desse gás. Já de acordo com lei de Raoult, a pressão de vapor de um composto acima de uma solução é diretamente proporcional à concentração deste composto nessa determinada solução (TIPLER, 2013).

De acordo com as leis acima, o princípio básico da teoria da extração por SHS é o estabelecimento de um equilíbrio dinâmico dos analitos entre as fases envolvidas no processo (matriz e fase de vapor). Neste sentido, em uma extração por SHS utilizando amostras líquidas, por exemplo, no estado de equilíbrio os compostos voláteis vaporizam para a fase de vapor na mesma proporção em que condensam e retornam para a fase líquida. A grandeza que relaciona a concentração do analito na amostra (C_A) e na fase de vapor (C_V) é chamada de coeficiente de partição (K) (Equação 1):

$$K = \frac{C_A}{C_V} \quad (\text{Equação 1})$$

Essa equação nos mostra que substâncias com baixo valor de K apresentam maior tendência de passagem para a fase de vapor. Deste modo, em análises baseadas no SHS, o ideal é garantir que os valores de K para os compostos de interesse sejam muito menores do que os valores de K para outros componentes “indesejados” presentes na matriz (TIPLER, 2013). O valor de K depende tanto da temperatura de incubação quanto da afinidade do analito com a matriz, como será discutido posteriormente.

Outra grandeza importante para ser avaliada é a denominada taxa de fase (β), a qual corresponde à razão entre o volume da fase de vapor (V_V) e o volume da fase condensada (V_A) (Equação 2):

$$\beta = \frac{V_V}{V_A} \quad (\text{Equação 2})$$

Segundo a Equação 2, valores mais baixos de β tendem a gerar melhores respostas analíticas para os compostos voláteis. Ou seja, um maior volume de amostra e/ou um menor volume de HS presente no frasco favorecem a uma amostragem mais concentrada das espécies em questão.

O coeficiente de partição e a taxa de fase são os princípios que norteiam a determinação da concentração final de compostos voláteis no HS. Sendo assim, a combinação de K e β dá origem à Equação 3, onde C_0 corresponde à concentração original do analito na amostra (RESTEK, 2000).

$$C_V = \frac{C_0}{(K + \beta)} \quad (\text{Equação 3})$$

A extração por SHS deve ser realizada, de maneira geral, de forma a maximizar a concentração dos componentes voláteis na fase de vapor, minimizando a contaminação indesejada de outros compostos na matriz. Sendo assim, como mostra a Equação 3, o desenvolvimento do método de extração deve ser pautado na obtenção de menores valores para K e β . Diferentes fatores devem ser avaliados para este fim, sendo que os parâmetros mais comumente avaliados são: temperatura e tempo de incubação, volume de amostra e a composição da matriz de análise (BORGES *et al.*, 2015).

Com relação à temperatura, quanto maior for o valor desta grandeza, maior é o grau de agitação das moléculas e, conseqüentemente, maior é a pressão de vapor de dos compostos presentes na amostra, fazendo com que uma maior quantidade dessas substâncias fique disponível para amostragem no HS. Contudo, é preciso notar que a temperatura apresenta um grau de influência diferente na extração dos compostos de acordo com os seus respectivos valores de K . Para substâncias que apresentam um baixo valor de K , ou seja, possuem baixa afinidade com a matriz (por exemplo, hexano em água), o aumento da temperatura tem pouco impacto no aumento da

disponibilidade desses analitos no HS, uma vez que os mesmos já apresentam alta tendência de passagem para a fase de vapor. Já para substâncias com alto valor de K , como na análise de etanol em uma matriz aquosa, a temperatura é um fator crítico na extração (TIPLER, 2013).

O tempo de incubação também é um fator relevante, pois leva certo tempo para que a partição dos analitos entre a matriz e a fase de vapor atinja o equilíbrio dinâmico. O aumento da temperatura e o auxílio da agitação podem fazer com que esse tempo seja reduzido.

Conforme discutido anteriormente, quanto maior o volume de amostra, menor o valor de β , o que aumenta a concentração dos compostos na fase de vapor. Contudo, assim como no caso da temperatura, a influência do volume de amostra também depende do valor de K dos compostos analisados. Para substâncias com alto valor de K , o efeito de alterar o volume de amostra faz pouca diferença na concentração da substância na fase de vapor, uma vez que esses compostos possuem alta afinidade com a matriz, tendendo a ficar retidos nesta fase. Portanto, para estes casos, volumes menores podem ser utilizados sem perda significativa de desempenho na extração. Já para compostos com baixo valor de K , uma pequena mudança no volume de amostra faz uma grande diferença na concentração desses compostos no HS (RESTEK, 2000).

A composição da matriz está diretamente ligada com as interações intermoleculares realizadas entre o soluto e o solvente. Sendo assim, este parâmetro pode ser alterado com o objetivo de diminuir a afinidade dos analitos com a matriz, diminuindo o valor de K e, conseqüentemente, aumentando a concentração dos compostos em questão no HS. Uma medida muito comum nas extrações por SHS é a adição de sais inorgânicos em amostras aquosas. Altas concentrações de sal nesse tipo de matriz aumentam a força iônica do meio, diminuindo a solubilidade de compostos com características polares (efeito *salting out*), facilitando assim a transferência desses analitos para a fase de vapor. No entanto, a magnitude deste efeito também é estritamente ligada ao valor de K . Compostos com valores de K que já são relativamente baixos apresentarão pouca mudança no coeficiente de partição após a adição de um sal. Por outro lado, compostos voláteis polares sofrerão um maior impacto, apresentando respostas mais pronunciadas após o efeito *salting out*. A

alteração do pH do meio também pode ser uma importante alternativa na extração por SHS, principalmente na análise de compostos de caráter ácido, básico, ou que estejam em sua forma conjugada (BORGES *et al.*, 2015).

Em suma, para compostos com alto valor de K, o ideal é diminuir a solubilidade do analito com o a matriz. Para tal, é possível aumentar a temperatura de incubação, promover o efeito *salting out* (no caso de moléculas polares em solventes polares, por exemplo), ou mudar o pH do meio (por exemplo, deixar o meio ácido em uma análise de ácidos voláteis em matriz aquosa). Já para compostos com baixo valor de K, a mudança no volume de amostra é o fator mais preponderante no aumento da concentração dos analitos na fase de vapor.

A discussão apresentada acima ilustra como a técnica de extração por SHS depende de uma série de fatores que devem ser otimizados para a obtenção dos melhores resultados no menor tempo possível. Conseqüentemente, para atingir esse determinado objetivo utilizando o menor número possível de experimentos, é preciso lançar mão de ferramentas de planejamento experimental.

2.4 Planejamento experimental

Quando um determinado processo é desenvolvido ou aperfeiçoado, é preciso realizar experimentos para avaliar os efeitos que as variáveis independentes (ou fatores) desse processo apresentam sobre o(s) resultados(s). Os fatores correspondem a entidades que podem ser mudadas e definidas em certo valor (nível). Essas variáveis podem ser quantitativas (tempo, temperatura, concentração, entre outros) ou qualitativas (tipo de solvente, tipo de reagente, tipo de enzima, etc). Parece lógico e instintivo avaliar um fator por vez, mantendo as outras variáveis fixadas para “controlar” o processo. Ou seja, determinar as condições experimentais que forneçam a melhor resposta (condição ótima) avaliando separadamente cada fator (RODRIGUES & IEMMA, 2005). Contudo, quando o número de variáveis é elevado, este procedimento “uma variável por vez” torna-se quase impraticável, pois seria necessário um número muito grande de experimentos. Além disso, essa estratégia não permite explorar um grande domínio experimental (variação entre os níveis dos fatores) e ainda não possibilita verificar o efeito das interações das variáveis nas respostas obtidas

(BENEDETTI *et al.*, 2020). Esses problemas podem ser solucionados utilizando-se ferramentas quimiométricas, como o planejamento experimental.

O planejamento experimental (do termo em inglês *Design of Experiments*, DOE) é uma abordagem multivariada que visa entender como os parâmetros envolvidos em um determinado processo estão relacionados às respostas estudadas. No DOE são utilizados modelos matemáticos para definir a magnitude dos efeitos das variáveis de um processo à fim de otimizar o desempenho do sistema e maximizar os resultados, utilizando um número reduzido de experimentos (MARRUBINI *et al.*, 2020; STOYANOV & WALMSLEY, 2006). Dentro do DOE existem diferentes estratégias que podem ser utilizadas para alcançar os objetivos descritos anteriormente. A seleção do melhor *design* experimental deve considerar vários aspectos, como os objetivos do estudo, número de fatores avaliados e a validade estatística dos resultados (BEZERRA *et al.*, 2008).

O primeiro passo da aplicação de ferramentas de DOE na otimização de um método consiste em entender quais são os objetivos dessa otimização e qual será a resposta avaliada nesse determinado processo. Por exemplo, a otimização de uma extração pode consistir em maximizar a recuperação de um analito, ao mesmo tempo minimizando o tempo e custo do processo. Já a otimização de uma corrida cromatográfica pode consistir na obtenção de picos bem resolvidos no menor tempo de análise possível (BENEDETTI *et al.*, 2020). Após o escopo da otimização ser bem definido, o segundo passo é detectar todos os possíveis fatores que podem influenciar o processo em questão. Essa seleção pode ser baseada tanto em testes preliminares quanto em estudos presentes na literatura. Se for observado que há um número elevado de variáveis que podem ser determinantes (geralmente 5 ou mais), é indicado a realização de um planejamento de triagem (RODRIGUES, & IEMMA, 2005).

O planejamento de triagem corresponde a delineamentos fracionários ou fatoriais, onde são realizadas combinações entre os níveis de dois ou mais fatores utilizando-se um número menor de ensaios (RODRIGUES & IEMMA, 2005). O resultado do planejamento de triagem é avaliado por meio de testes estatísticos para verificar quais fatores apresentam, de fato, efeito com significância estatística sobre a(s) resposta(s) avaliada(s). Após a seleção das variáveis mais importantes, são

realizados os experimentos de otimização, onde é possível obter um modelo matemático (por exemplo, uma superfície de resposta) que indica os valores ótimos teóricos para o experimento em questão (FUKUDA *et al.*, 2018; MARRUBINI *et al.*, 2020).

Há diferentes tipos de planejamento que podem ser aplicados para a realização da triagem de variáveis independentes, bem como para a otimização do processo. Dentre esses planejamentos, destacam-se o Plackett-Burman (PB, utilizado para triagem) e o Delineamento Composto Central Rotacional (DCCR, utilizado para otimização).

2.4.1 Plackett Burman

O *design* experimental mais eficiente para realizar a triagem de variáveis é o PB, o qual foi introduzido por Robin L. Plackett e J. P. Burman em 1946 (BENEDETTI *et al.*, 2020). Este delineamento consiste em um tipo de planejamento fatorial que permite estudar até N-1 variáveis através de um número N de experimentos, sendo N um número múltiplo de 4. Por exemplo, com apenas 8 experimentos é possível avaliar a influência de até 7 variáveis. Contudo, é recomendado que o número de experimentos seja, ao menos, o número de variáveis avaliadas +4. Isso evita a saturação do delineamento (RODRIGUES, & IEMMA, 2005). Como exemplo, com 12 experimentos é recomendado avaliar, no máximo, a influência de até 8 variáveis no processo estudado.

Uma matriz experimental de PB de 12 experimentos está representada na Tabela 5. É possível notar que no PB cada variável independente é estudada em dois níveis: um superior, codificado como +1, e um inferior (-1), sendo que o número de vezes em que cada variável está presente nos níveis superior e inferior é o mesmo. Além disso, é indicado realizar ao menos três experimentos no ponto central (“nível 0”), para verificar a repetibilidade do método e calcular o erro experimental (RODRIGUES & IEMMA, 2005).

Tabela 5 - Representação da matriz experimental da metodologia Plackett-Burman de 12 experimentos para o estudo de 5 variáveis.

Experimento	Variável 1	Variável 2	Variável 3	Variável 4	Variável 5
1	1	-1	1	-1	-1
2	1	1	-1	1	-1
3	-1	1	1	-1	1
4	1	-1	1	1	-1
5	1	1	-1	1	1
6	1	1	1	-1	1
7	-1	1	1	1	-1
8	-1	-1	1	1	1
9	-1	-1	-1	1	1
10	1	-1	-1	-1	1
11	-1	1	-1	-1	-1
12	-1	-1	-1	-1	-1

2.4.2 Delineamento Composto Central Rotacional

O DCCR corresponde a um planejamento experimental onde são realizados ensaios em 5 níveis, os quais são calculados para abranger toda a região de estudo (FUKUDA *et al.*, 2018; RODRIGUES & IEMMA, 2005). Esses níveis são codificados como $+\alpha$, $+1$, 0 , -1 , $-\alpha$, onde α corresponde aos “pontos axiais” ou “pontos-estrela”. O valor de α pode ser calculado através da Equação 4, onde K é igual ao número de variáveis avaliadas. Por exemplo, quando 2 fatores estão sendo avaliados, α é igual a 1,41; quando o planejamento é realizado com 3 fatores, α é igual a 1,68. O DCCR é um tipo de planejamento que gera um modelo quadrático para a resposta avaliada e é utilizado principalmente para explorar o maior número de possibilidades do domínio experimental para encontrar a região ótima.

$$\alpha = 2^{K/4} \quad (\text{Equação 4})$$

Para entender melhor esta metodologia, é possível supor que para a otimização de um determinado processo é preciso avaliar a influência de três diferentes variáveis: pH do meio, temperatura e concentração de um reagente, por exemplo. Como discutido

anteriormente, é possível estudar uma variável por vez, onde um parâmetro é avaliado em diferentes condições e os outros são fixados (Figura 6a). Neste caso, o método pode não ser o ideal, pois não é possível verificar a interação entre as variáveis e também o espaço experimental não é completamente explorado. Também é possível realizar todas as combinações possíveis de experimentos até a obtenção de uma solução final (Figura 6b). Este método, teoricamente, explora todo o espaço experimental, porém torna-se impraticável quando o número de variáveis é grande (geralmente 5 ou mais). Por fim, pode ser realizado um *design* experimental do tipo DCCR (Figura 6c), onde em 17 ensaios toda a região de estudo é contemplada com um número muito menor de experimentos (8 ensaios fatoriais, representados pelos vértices do cubo, mais 6 ensaios axiais – pontos fora do cubo -, mais 3 ensaios repetidos no ponto central). Este planejamento possibilita avaliar qual o efeito de cada variável envolvida e a interação entre elas. Além disso, também é possível elaborar um modelo matemático que gera uma superfície de resposta, possibilitando assim a determinação das condições ótimas do processo (RODRIGUES & IEMMA, 2005).

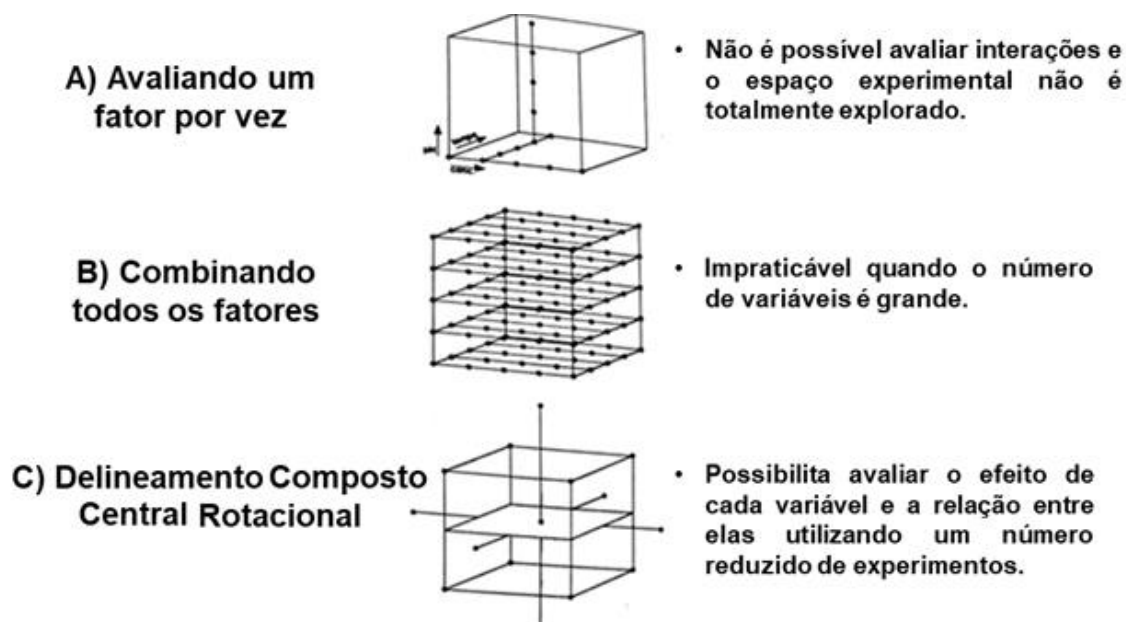


Figura 6 - Possibilidades de condução de um experimento para 3 variáveis estudadas. Cada ponto representa um experimento e o cubo representa o espaço experimental. Adaptado de (RODRIGUES & IEMMA, 2005).

A matriz de experimentos do DCCR para 2 variáveis está representada na Tabela 6 e a matriz de experimentos do DCCR para 3 níveis está representada na Tabela 7. Em ambos os casos está sendo considerada a realização de três experimentos no ponto central (onde todas as variáveis estão no nível 0).

Tabela 6 - Matriz de experimentos do DCCR de 2 níveis.

Experimento	Variável 1	Variável 2
1	-1	-1
2	1	-1
3	-1	1
4	1	1
5	-1.41	0
6	1.41	0
7	0	-1.41
8	0	1.41
9	0	0
10	0	0
11	0	0

Tabela 7 - Matriz de experimentos do DCCR de 3 níveis.

Experimento	Variável 1	Variável 2	Variável 3
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1
9	-1.68	0	0
10	1.68	0	0
11	0	-1.68	0
12	0	1.68	0
13	0	0	-1.68
14	0	0	1.68
15	0	0	0
16	0	0	0
17	0	0	0

3 OBJETIVOS

Desenvolver e otimizar um método analítico para a investigação de compostos orgânicos voláteis em amostras de fluido oral e urina.

4 MATERIAIS E MÉTODOS

4.1 Materiais

Para a coleta das amostras foram utilizados dispositivos coletores de fluido oral (Salivette®) adquiridos da Sarstedt (Hildesheim, Alemanha) e coletores universais estéreis de 80 mL obtidos da AzLabor (Ribeirão Preto, Brasil). Para a extração por HS foram utilizados frascos de *headspace* de 10 mL e tampas magnéticas com septo de PTFE obtidos da Sun Sri (Rockwood, EUA), bem como seringa *gastight* Hamilton de 2,5 mL adquirida da Sigma-Aldrich (San Luis, EUA). Por fim, para o preparo das amostras e também das soluções foram utilizados os seguintes equipamentos: balança analítica Sartorius BP211D com precisão de 0,1 mg (Gotinga, Alemanha), centrífuga modelo Universal 32 da Hettich Zentrifugen (Tuttlingen, Alemanha), agitador de tubos (*vortex*) modelo AP 56 da Phoenix (Araraquara, Brasil) e papéis indicadores de pH (1-14) modelo Whatman Papeha adquiridos da Sigma-Aldrich (San Luis, EUA).

4.2 Reagentes, padrões e soluções

Cloreto de sódio 99,5%, hidróxido de sódio 98,3%, bem como o padrão analítico de diidromircenol (2,6-dimetil-7-octen-2-ol, pureza >99, utilizado como padrão interno - PI) foram obtidos da Sigma-Aldrich (San Luis, EUA). Ácido sulfúrico 95% foi adquirido da Labsynth (Diadema, Brasil), água deionizada foi obtida através do sistema de purificação de água Milli-DI da Merck (Darmstald, Alemanha).

4.2.1 Preparo das soluções

A) Solução hidróxido de sódio 1 mol.L⁻¹

Utilizando-se uma balança analítica modelo Sartorius BP211D, uma massa de 2,0 g de NaOH foi pesada em frasco plástico PET de 100 mL. Em seguida, foram adicionados, lentamente, 50 mL de água deionizada. A solução foi homogeneizada

com o auxílio de um agitador magnético e, após a completa homogeneização, foi armazenada em local fresco e arejado a temperatura ambiente.

B) Solução de ácido sulfúrico 0,1 mol.L⁻¹

A um balão volumétrico de 50 mL contendo aproximadamente 25 mL de água deionizada foram adicionados 600 µL de H₂SO₄. Aguardou-se o resfriamento do frasco, então o volume foi completado com água deionizada. A solução foi homogeneizada e, posteriormente, armazenada em local fresco e arejado a temperatura ambiente.

C) Solução de diidromircenol 0,01% (v/v)

A um balão volumétrico de 50 mL contendo aproximadamente 25 mL de água deionizada foram adicionados 5 µL de diidromircenol (padrão analítico >99%). Em seguida, o volume foi completado com água deionizada. A solução foi homogeneizada com o auxílio de um agitador magnético e posteriormente armazenada em geladeira a 5 °C. O diidromircenol foi escolhido como PI por três motivos. Primeiro, é um VOC que apresenta boa resposta no FID. Segundo: após análise de 10 amostras de urina e fluido oral de pessoas saudáveis, este composto não foi encontrado em nenhuma das amostras em questão. Por fim, o pico referente a essa substância apresentou-se simétrico, com alta eficiência e resolução em todas as análises realizadas.

4.3 Coleta das amostras

O dispositivo Salivette[®] (Figura 7) foi utilizado para a coleta de fluido oral. Este dispositivo consiste em um tubo plástico que contém na porção interna um rolo de algodão que mede aproximadamente 1 cm de espessura e 3,5 cm de comprimento. Para a realização da coleta, é necessário abrir o tubo plástico, retirar o algodão e colocá-lo na boca durante um minuto para que o mesmo absorva o fluido oral. É indicado que o voluntário não morda o algodão e que o movimento pela boca, utilizando a língua, para que o maior volume de amostra possível seja coletado. Após a coleta, é preciso recuperar o fluido que ficou retido no algodão. Para tal, a amostra é centrifugada por 5 minutos a 3000 rpm, fazendo com que o fluido oral seja separado do algodão através de um filtro presente na porção interna do tubo plástico (Figura 8).

Com relação à amostragem da urina, as amostras foram coletadas utilizando um recipiente plástico estéril comum para análises de urina de rotina. Foi recomendado que o voluntário higienizasse as mãos antes do processo e que o volume de urina fosse pelo menos de 20 mL (marcação média do recipiente coletor).

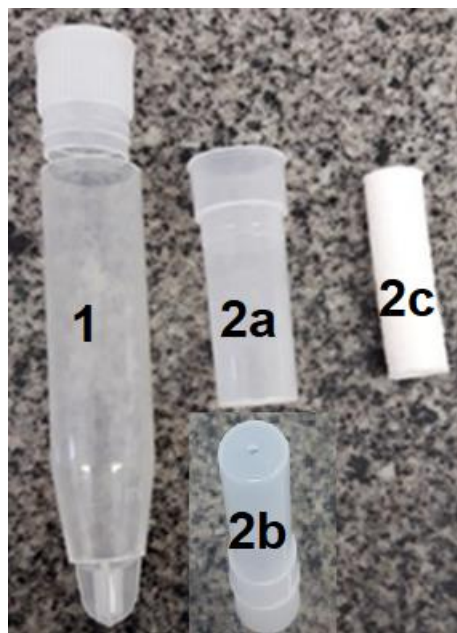


Figura 7 - Estrutura do Salivette®. 1 – tubo plástico; 2a – frasco coletor e 2b – filtro; 2C – rolo de algodão.

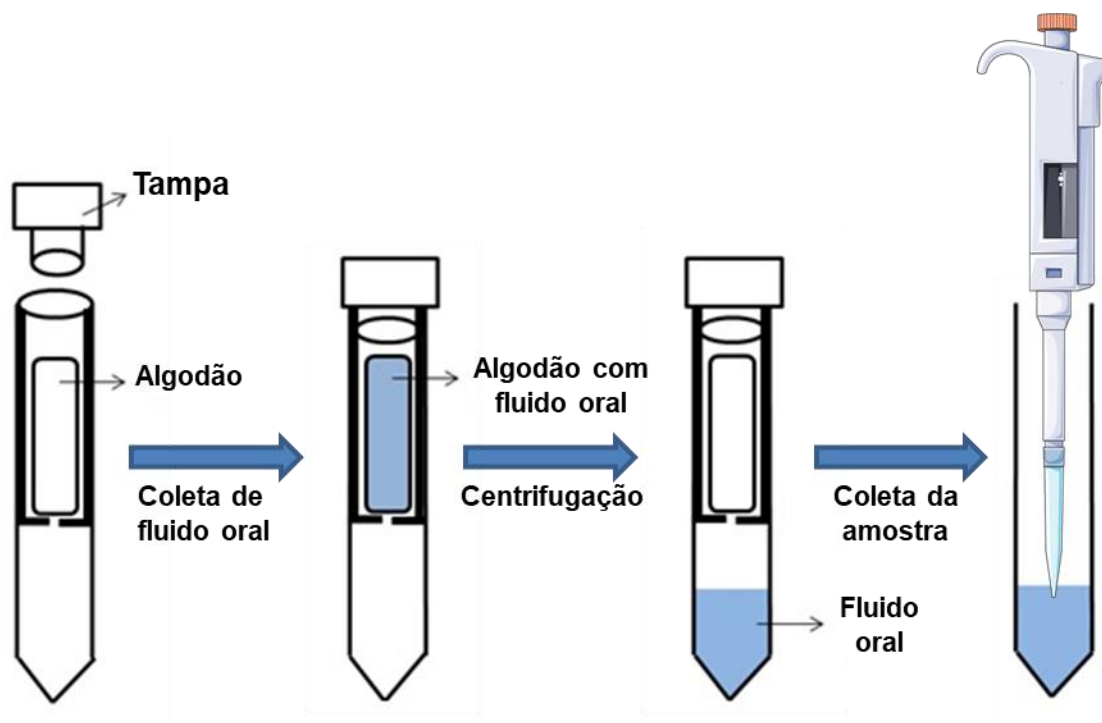


Figura 8 - Esquema do processo de coleta de fluido oral através do Salivette®. Adaptado de Gomes, 2018.

4.4 Análise instrumental

As análises foram realizadas utilizando-se a técnica de extração por *headspace* estático e posterior injeção em um sistema de cromatografia em fase gasosa com detector de ionização por chama (SHS-GC-FID). Os parâmetros de extração e análise descritos nesta seção são referentes ao processo já otimizado. Como dito anteriormente, ao longo deste capítulo serão discutidos os processos de desenvolvimento e otimização do método para chegar às condições especificadas abaixo.

Para o preparo das amostras, um volume adequado da matriz biológica (850 μL no caso do fluido oral e 3 mL com relação à urina) foi alocado em um frasco de HS de 10 mL. Em seguida, foram adicionados 30 μL de solução de diidromircenol 0,01% (v/v, em água) e também foi adicionada a devida quantidade de sal: 1,0 g de NaCl para as amostras de fluido oral e 0,75 g de NaCl para urina. Logo depois, o meio foi acidificado utilizando solução de ácido sulfúrico 0,1 M. Por fim, as amostras foram submetidas ao processo de incubação por 15 minutos (fluido oral) ou 20 minutos (urina) a 95 °C e

agitação de 500 rpm. Para tal, foi utilizado um amostrador automático Combipal GC Sampler 80 da Agilent Technologies (Santa Clara, EUA). Após a incubação, um volume de 2000 μL da fase de vapor foi injetado, através de uma seringa *gastight*, em um sistema GC-FID modelo 7890A da Agilent Technologies (Santa Clara, EUA).

Tanto para o fluido oral quanto para a urina as condições cromatográficas foram as mesmas. A coluna capilar utilizada foi de modelo CP-Wax 52CB (30 m x 0,25 mm, 0,25 μm de espessura de filme). Esta é uma coluna que suporta bem o vapor de água produzido durante o processo de extração por HS. A temperatura do injetor foi de 230 $^{\circ}\text{C}$ e do detector 300 $^{\circ}\text{C}$. O fluxo de gás de arraste (nitrogênio 5.0) manteve-se em 1,3 mL/min e a injeção foi no modo *split* 1:10. O tempo total de corrida foi de 50 minutos e o programa de temperatura da coluna está representado na Tabela 8.

Tabela 8 - Programa de temperatura do forno para a análise de VOCs presentes no fluido oral e na urina através do GC-FID.

Temperatura ($^{\circ}\text{C}$)	Taxa de aquecimento ($^{\circ}\text{C}/\text{min}$)	Tempo de permanência
40	-	3
80	4	5
200	5	5
230	10	-

4.5 Desenvolvimento e otimização do método

4.5.1 Desenvolvimento do método cromatográfico

4.5.1.1 Escolha da rampa de aquecimento

Neste presente estudo foram testadas duas rampas de aquecimento baseando-se em trabalhos semelhantes reportados na literatura (CAVACO *et al.*, 2018; MONEDEIRO *et al.*, 2020; SHIGEYAMA, *et al.*, 2019). A primeira rampa (Condição 1) iniciava-se em 40 $^{\circ}\text{C}$ (permanência de 1 minuto) e aumentava a uma taxa de 10 graus por minuto até 230 $^{\circ}\text{C}$ (permanência de 10 minutos). Tempo total de corrida: 30 min. A segunda rampa (Condição 2) está especificada na Tabela 8 (tempo total de corrida: 50 min).

Para esta avaliação foi utilizada como matriz uma mistura resultante da adição de 10 mL de um *pool* de fluido oral em 10 mL de um *pool* de urina. Tanto o *pool* de fluido oral quanto o *pool* de urina foram obtidos através da doação de 10 indivíduos saudáveis e a coleta foi realizada conforme descrito na seção 4.3. As análises foram realizadas utilizando-se um volume de amostra de 1,5 mL, adicionando-se 1 g de NaCl, 30 µL de solução de diidromircenol 0,01% v/v (padrão interno) e acidificando-se o meio (pH entre 1 e 2) com solução de ácido sulfúrico 0,1 M. As condições do SHS foram mantidas nas configurações padrão do equipamento: tempo de incubação de 10 minutos, temperatura de incubação de 80 °C e 300 rpm de agitação. Os outros parâmetros do GC-FID foram mantidos conforme o especificado na seção 4.4.

4.5.1.2 Escolha da taxa de divisão do fluxo

O modo de injeção de amostras no GC pode ser dividido, basicamente, de três maneiras principais: injeção com divisão de fluxo (*split*), sem divisão do fluxo (*splitless*) e diretamente na coluna (*on column*), este último sendo indicado apenas na análise de compostos com alto ponto de ebulição e/ou termossensíveis (AQUINO NETO, 2010).

Nas injeções com divisão de fluxo, parte da amostra injetada pela seringa é vaporizada e escapa pela purga do divisor. Ou seja, o fluxo do gás de arraste que passa pelo injetor é dividido entre duas porções: a coluna capilar (porção que é analisada no sistema cromatográfico) e a purga do divisor (porção eliminada). Já na injeção *splitless*, toda amostra é introduzida na coluna, não havendo assim a divisão do fluxo. As injeções *split* são indicadas para amostras mais concentradas e, geralmente, produzem picos mais eficientes e simétricos. Já as injeções *splitless* são indicadas para analitos presentes em baixas concentrações, aumentando consideravelmente o poder de detecção do método. Contudo, este modo de injeção pode levar a um alargamento dos picos, principalmente daqueles que estão na porção inicial da corrida (CHASTEEN, 2000).

Desta maneira, neste trabalho foi avaliado qual seria o melhor modo de injeção para a análise de VOCs no fluido oral e urina: no modo *split* ou *splitless*. Para tal, foi utilizada uma abordagem semelhante ao que foi descrito na seção anterior. Contudo, ao invés de variar a rampa de aquecimento, a mesma foi mantida nas condições

relatadas na Tabela 8, variando-se, portanto, apenas o modo de injeção: *splitless* ou *split* 1:10. Neste caso, um novo *pool* constituído de uma mistura de fluido oral e urina foi utilizado nas análises.

4.6 Planejamento experimental

Após a escolha da rampa de aquecimento e da taxa de divisão do fluxo de injeção, o método foi otimizado utilizando ferramentas de *design* experimental. Conforme discutido anteriormente, a técnica de extração por SHS depende de uma série de fatores que devem ser otimizados para a obtenção dos melhores resultados no menor tempo possível. Neste sentido, no presente trabalho foram avaliadas as seguintes variáveis: tempo (min) e temperatura de incubação (°C), velocidade de agitação (rpm), adição de sal (NaCl, g) e pH da amostra. Os níveis avaliados para cada uma dessas grandezas foram escolhidos de acordo com experimentos preliminares, com as especificações do equipamento, e também de acordo com a literatura. Sendo assim, com base no DOE, esses fatores foram variados, ao mesmo tempo, em um número reduzido de experimentos. Para este determinado fim, primeiro foi realizado um planejamento de triagem, através do método de PB, para selecionar os fatores mais importantes para o processo e, em seguida, as variáveis selecionadas foram otimizadas através do planejamento DCCR e da observação das superfícies de resposta obtidas. É preciso destacar que, em cada planejamento aplicado, os experimentos foram realizados de maneira randomizada, isto é, não foi seguida a ordem dos experimentos presentes na matriz experimental. Essa abordagem é normalmente utilizada para buscar diminuir a influência de erros sistemáticos. Além disso, ao longo do planejamento todos os experimentos foram realizados em uma única replicata, à exceção dos ensaios realizados no ponto central.

4.6.1 Planejamento experimental para urina

Primeiramente, um *pool* de urina foi obtido através da doação de 10 indivíduos saudáveis. Esse *pool* foi analisado em diferentes condições à luz do DOE (Figura 9) utilizando-se como critério de avaliação a média geométrica das áreas dos picos obtidos nas análises. Esse mesmo critério foi utilizado em todos os planejamentos

realizados neste estudo. Essa escolha foi baseada no fato de que, se a resposta avaliada fosse simplesmente a área total dos picos obtidos nos cromatogramas, os picos mais intensos acabariam tendo um impacto muito maior no resultado. Conseqüentemente, a otimização seria realizada em função apenas dessas substâncias presentes em maior concentração. Logo, o uso da média geométrica foi uma opção encontrada para ponderar as áreas obtidas e, assim, atribuir uma maior significância aos VOCs presentes em menor quantidade.

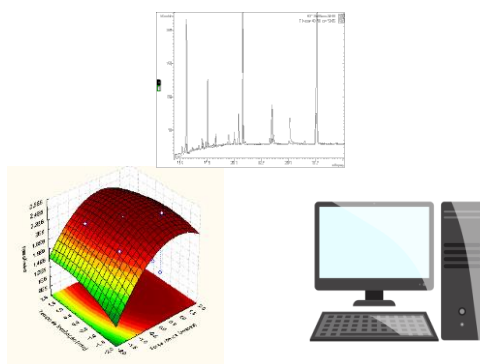
Além disso, também é preciso destacar que o único padrão adicionado às amostras foi o padrão interno (30 μ L de solução diidromircenol 0,01% v/v). Portanto, os picos presentes nos cromatogramas eram, majoritariamente, oriundos dos VOCs presentes nas amostras avaliadas. Este mesmo princípio é verdadeiro para as análises referentes à escolha da rampa de aquecimento e também da taxa de divisão do fluxo de injeção. O tratamento estatístico dos dados foi realizado através do *software* Statística 7.0 da Statsoft Inc. (Tulsa, EUA) e do *software* Octave 4.2.1 (Boston, EUA)



**Coleta de um pool de
urina/fluido oral**



Headspace (em diferentes condições)



Tratamento estatístico



GC-FID



Figura 9 - Representação do procedimento experimental para a otimização do método de análise de VOCs em fluido oral e urina.

. A Tabela 9 apresenta as variáveis avaliadas e os níveis utilizados na triagem das variáveis através da metodologia de PB. Para a acidificação do meio foi utilizada uma solução de ácido sulfúrico 0,1 M, enquanto que para deixar o meio básico foi utilizada uma solução de hidróxido de sódio 1 M. O pH foi verificado utilizando-se indicadores universais de pH (1-14). Por fim, o volume de amostra (neste caso, volume de urina) foi fixado em 3 mL. Como a maioria dos trabalhos da literatura utilizavam 2 ou 4 mL de urina, foram realizados testes preliminares nessa faixa de volume. Foi observado que o volume de amostra de 4 mL fornecia maiores valores de área para os analitos (assim como era esperado e foi discutido na seção 2.3). Porém, esse

determinado volume fazia com que fosse necessário um tempo de incubação muito grande para o estabelecimento do equilíbrio dinâmico dos analitos entre o HS e a fase condensada. Sendo assim, um volume de amostras 3 mL foi escolhido para as análises da matriz em questão.

Tabela 9 - Variáveis e níveis avaliados na metodologia Plackett-Burman para a urina.

Variáveis	-1	0	1
Adição de sal (g)	0	0,75	1,5
pH	Ácido (1-2)	Neutro (6-7)	Básico (10-11)
Agitação (rpm)	300	500	700
Tempo (min)	5	15	25
Temperatura (°C)	60	75	90

Os experimentos foram realizados segundo a matriz da metodologia PB de 12 experimentos para cinco variáveis (Tabela 5). Neste caso, foram realizados 16 ensaios ao todo (12 experimentos segundo o delineamento de PB, mais quatro experimentos com valores no ponto central). A Tabela 10 mostra o *design* experimental já com os valores reais utilizados. Por exemplo, no experimento de número 1, avaliou-se a resposta nas seguintes condições: adição de 1,5 g de NaCl, em pH ácido, com agitação de 700 rpm, tempo de incubação igual a 5 minutos e temperatura de 60 °C.

Tabela 10 - Matriz PB de 12 experimentos para a urina.

Experimento	Adição de sal (g)	pH	Agitação (rpm)	Tempo (min)	Temperatura (°C)
1	1,5	ácido	700	5	60
2	1,5	básico	300	25	60
3	0	básico	700	5	90
4	1,5	ácido	700	25	60
5	1,5	básico	300	25	90
6	1,5	básico	700	5	90
7	0	básico	700	25	60
8	0	ácido	700	25	90
9	0	ácido	300	25	90
10	1,5	ácido	300	5	90
11	0	básico	300	5	60
12	0	ácido	300	5	60
13	0,75	neutro	500	15	75
14	0,75	neutro	500	15	75
15	0,75	neutro	500	15	75
16	0,75	neutro	500	15	75

Após o tratamento estatístico dos resultados, foi realizada a escolha de quais variáveis seriam fixadas e quais seriam otimizadas através do DCCR (posteriormente será discutido o porquê dessas escolhas). As variáveis fixadas foram: agitação, adição de sal e pH. As variáveis escolhidas para serem otimizadas através do DCCR foram: tempo e temperatura de incubação. Sendo assim, os experimentos de otimização foram realizados segundo a matriz experimental para o DCCR de duas variáveis (Tabela 6). A Tabela 11 apresenta as variáveis avaliadas e os níveis utilizados (variando-se da magnitude mínima -1,41 até a máxima de +1,41).

Para o cálculo dos valores reais das variáveis, primeiramente é preciso estabelecer os valores extremos (neste caso, os valores reais para as magnitudes -1,41 e +1,41). Tendo como exemplo o tempo de incubação, os valores extremos foram escolhidos como 20 e 40 minutos. Logo, o valor real da variável em 0 deve ser a média desses dois valores, ou seja, 30 minutos. Os valores reais das variáveis em -1 e 1 podem ser determinados através de cálculos de interpolação linear.

Tabela 11 - Variáveis e níveis utilizados no DCCR de 2 variáveis para a otimização dos parâmetros de extração por HS para a urina.

Variáveis	-1,41	-1	0	1	1,41
Tempo de incubação (min)	20	23	30	37	40
Temperatura (°C)	65	69	80	91	95

A Tabela 12 mostra o design experimental já com os valores reais utilizados em cada experimento. Neste caso, foram realizados 8 experimentos à luz do DCCR de 2 variáveis, mais 3 repetições no ponto central.

Tabela 12 - Design experimental para o DCCR de 2 variáveis para a otimização dos parâmetros de extração por HS para a urina.

Experimento	Tempo (min)	Temperatura (° C)
1	23	69
2	37	69
3	23	91
4	37	91
5	20	80
6	40	80
7	30	65
8	30	95
9	30	80
10	30	80
11	30	80

Assim como exposto anteriormente, os resultados obtidos foram tratados no software Statistica, gerando uma superfície de resposta, a qual possibilita encontrar os valores ótimos para os experimentos. O mesmo raciocínio utilizado para o planejamento da urina foi empregado para a otimização envolvendo o fluido oral, como será demonstrado a seguir.

4.6.2 Planejamento experimental para o fluido oral

Assim como na avaliação da urina, no caso da otimização da extração dos VOCs presentes no fluido oral, um *pool* de amostras foi coletado e analisado à luz do DOE, realizando-se uma triagem das variáveis significativas através do PB e a otimização através do DCCR. Foram avaliadas as mesmas variáveis que na urina: tempo (min) e temperatura de incubação (°C), agitação (rpm), adição de sal (NaCl, g) e pH.

O volume de fluido oral obtido na amostragem pelo Salivette® é de, geralmente, 1,0 a 1,5 mL. Contudo, como pessoas de mais idade tendem a salivar menos, e é justamente a faixa etária que apresenta o maior número de casos de câncer, foi escolhido, por precaução, usar um volume menor do que 1,0 mL. Portanto, para as análises de fluido oral foi fixado um volume de amostra de 850 µL.

A Tabela 13 mostra as variáveis e níveis utilizados no PB e a Tabela 14 mostra o design completo do PB de 12 experimentos empregado nesta avaliação.

Tabela 13 - Variáveis e níveis utilizados na metodologia PB para o fluido oral.

Variáveis	-1	0	1
Adição de sal (g)	0	0,5	1,0
pH	Ácido (1-2)	Neutro (6-7)	Básico (10-11)
Agitação (rpm)	300	500	700
Tempo (min)	5	15	25
Temperatura (°C)	60	75	90

Tabela 14 - Design de PB de 12 experimentos para o fluido oral.

Experimento	Adição de sal (g)	pH	Agitação (rpm)	Tempo (min)	Temperatura (°C)
1	1,0	ácido	700	5	60
2	1,0	básico	300	25	60
3	0	básico	700	5	90
4	1,0	ácido	700	25	60
5	1,0	básico	300	25	90
6	1,0	básico	700	5	90
7	0	básico	700	25	60
8	0	ácido	700	25	90
9	0	ácido	300	25	90
10	1,0	ácido	300	5	90
11	0	básico	300	5	60
12	0	ácido	300	5	60
13	0,5	neutro	500	15	75
14	0,5	neutro	500	15	75
15	0,5	neutro	500	15	75
15	0,5	neutro	500	15	75

Assim como no caso do planejamento experimental da urina, após o PB foram escolhidas quais variáveis seriam fixadas e quais seriam otimizadas através do DCCR. Para o fluido oral, as variáveis avaliadas na otimização foram: tempo, temperatura de incubação e adição de sal. A Tabela 15 mostra os níveis dos fatores avaliados (variando-se da magnitude mínima -1,68 até a máxima de +1,68) e a Tabela 16 apresenta o *design* experimental completo utilizado no DCCR. A matriz de experimentos para o DCCR de três variáveis está encontra-se Tabela 7.

Tabela 15 - Variáveis e níveis utilizados no DCCR para a otimização da extração por SHS dos VOCs presentes no fluido oral.

Variáveis	-1.68	-1	0	1	1.68
T _{inc} (min)	5	8,3	12,5	16,7	20
T (° C)	65	71,1	80	88,9	95
NaCl (g)	0,2	0,36	0,6	0,84	1

Tabela 16 - Design do DCCR de 3 variáveis para a otimização da extração por SHS dos VOCs presentes no fluido oral.

Experimento	T _{inc} (min)	T (°C)	NaCl (g)
1	8,3	71,1	0,36
2	16,7	71,1	0,36
3	8,3	88,9	0,36
4	16,7	88,9	0,36
5	8,3	71,1	0,84
6	16,7	71,1	0,84
7	8,3	88,9	0,84
8	16,7	88,9	0,84
9	5,0	80	0,60
10	20	80	0,60
11	12,5	65	0,60
12	12,5	95	0,60
13	12,5	80	0,2
14	12,5	80	1,0
15	12,5	80	0,60
16	12,5	80	0,60
17	12,5	80	0,60

4.6.3 Verificação de possíveis interferentes nas análises

Para a investigação de possíveis picos interferentes nos cromatogramas obtidos durante o processo de desenvolvimento e otimização do método, foram realizadas análises em diferentes condições que buscaram avaliar a possível contribuição de outros compostos presentes no *headspace* (sem aqueles VOCs oriundos das matrizes biológicas avaliadas). Para tal, foram realizadas as seguintes avaliações:

- análise do frasco de HS vazio para identificar possíveis picos provenientes do próprio *vial*, da tampa, ou até mesmo da seringa utilizada no procedimento;
- análise contendo apenas 3 mL de água deionizada;
- análise de 3 mL de água deionizada com adição de 1 g de NaCl;
- análise de 3 mL de água deionizada com adição de ácido sulfúrico 0,1 M (pH 1-2)
- análise de 3 mL de água deionizada com adição de 1 g de NaCl e ácido sulfúrico 0,1 M (pH 1-2).

Além disso, para verificar possíveis interferentes oriundos do Salivette®, o algodão utilizado nesse dispositivo foi mergulhado em água deionizada por 1 minuto e posteriormente foi colocado de volta ao frasco coletor. Em seguida, esse frasco foi centrifugado e um volume de 850 µL da amostra foi transferido para um *vial* de HS, o qual foi analisado. Todos esses experimentos foram realizados em triplicata, utilizando-se os valores de tempo e temperatura de incubação, bem como de velocidade de agitação obtidos durante o processo de otimização.

5 RESULTADOS E DISCUSSÃO

5.1 Desenvolvimento do método cromatográfico

5.1.1 Escolha da rampa de aquecimento

Segundo os estudos presentes na literatura, as rampas de aquecimento utilizadas para a análise de VOCs em matrizes biológicas por GC apresentam, geralmente, duas características em comum: primeiro, a rampa é iniciada com uma temperatura baixa (geralmente 30 a 40 °C); além disso, a corrida é mais longa, durando de 30 a 90 minutos (CAVACO *et al.*, 2018; SHIGEYAMA, *et al.*, 2019). Isso ocorre devido ao fato de que os analitos são compostos voláteis (muitos dos quais apresentam pontos de ebulição semelhantes), e também o tipo de análise realizado não tem, normalmente, um (ou mais) composto específico de interesse (é uma análise *non-target*). Sendo assim, é preciso começar a análise com uma temperatura mais baixa e realizar uma corrida mais lenta para assim garantir que os VOCs sejam satisfatoriamente separados no processo cromatográfico.

Neste contexto, para o desenvolvimento do método utilizado neste trabalho foram testadas duas rampas de aquecimento descritas previamente: Condição 1, uma rampa mais rápida com tempo total de 30 minutos, e Condição 2, uma rampa mais lenta, com taxa de inclinação menor e tempo total de 50 minutos. Os cromatogramas resultantes das duas análises estão presentes nas Figuras 10 e 11.

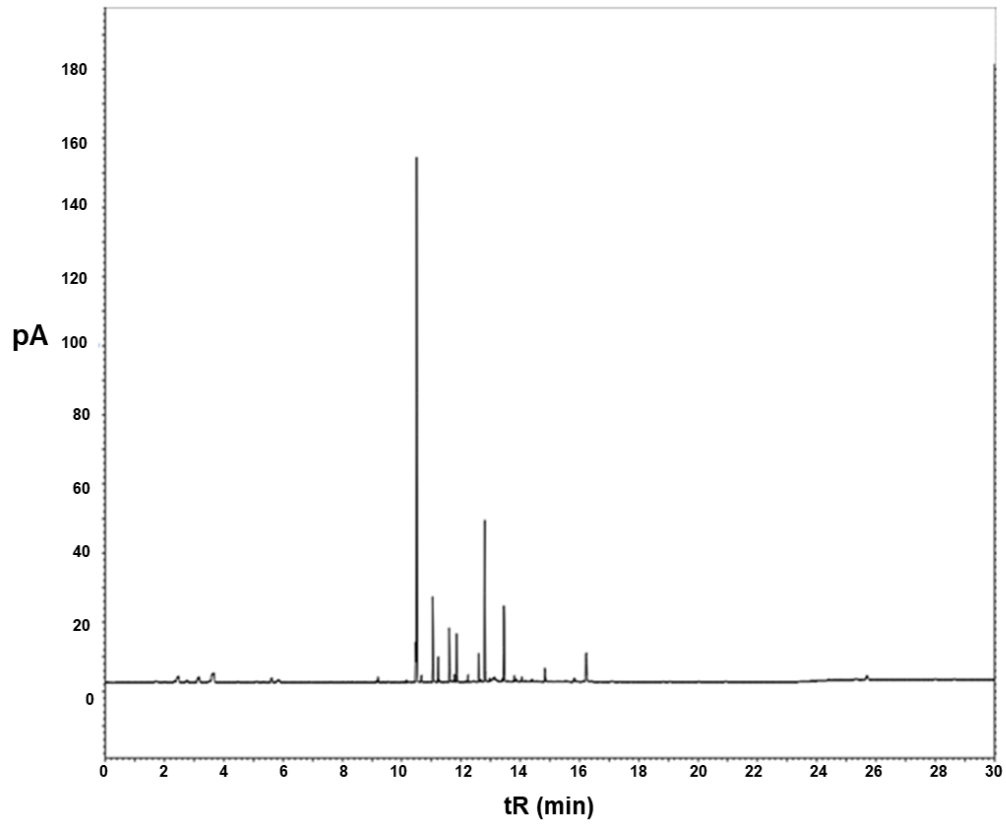


Figura 10 - Cromatograma da análise da mistura de fluido oral e urina utilizando a rampa de aquecimento adotada na Condição 1.

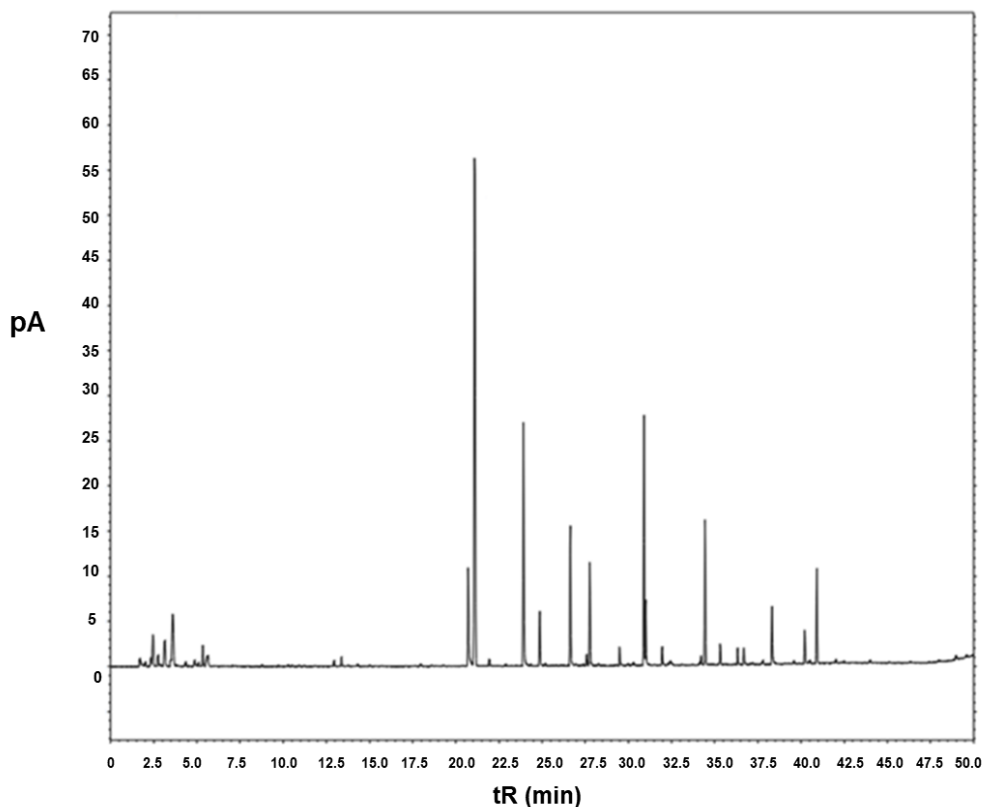


Figura 11 - Cromatograma da análise da mistura de fluido oral e urina utilizando a rampa de aquecimento adotada na Condição 2.

Através das análises das Figuras 10 e 11 é possível perceber que no cromatograma referente à Condição 2 há um maior número de picos e os mesmos apresentam uma melhor separação principalmente no início da corrida (entre 2,5 e 6 minutos) e nos tempos de retenção entre 20 e 41 minutos. Sendo assim, a Condição 2, apesar de ser uma corrida mais demorada, foi escolhida como a rampa de aquecimento a ser utilizada neste trabalho.

5.1.2 Escolha da taxa de divisão do fluxo

Os cromatogramas resultantes dos experimentos envolvendo injeções no modo *splitless* e *split* estão presentes nas Figuras 12 e 13, respectivamente.

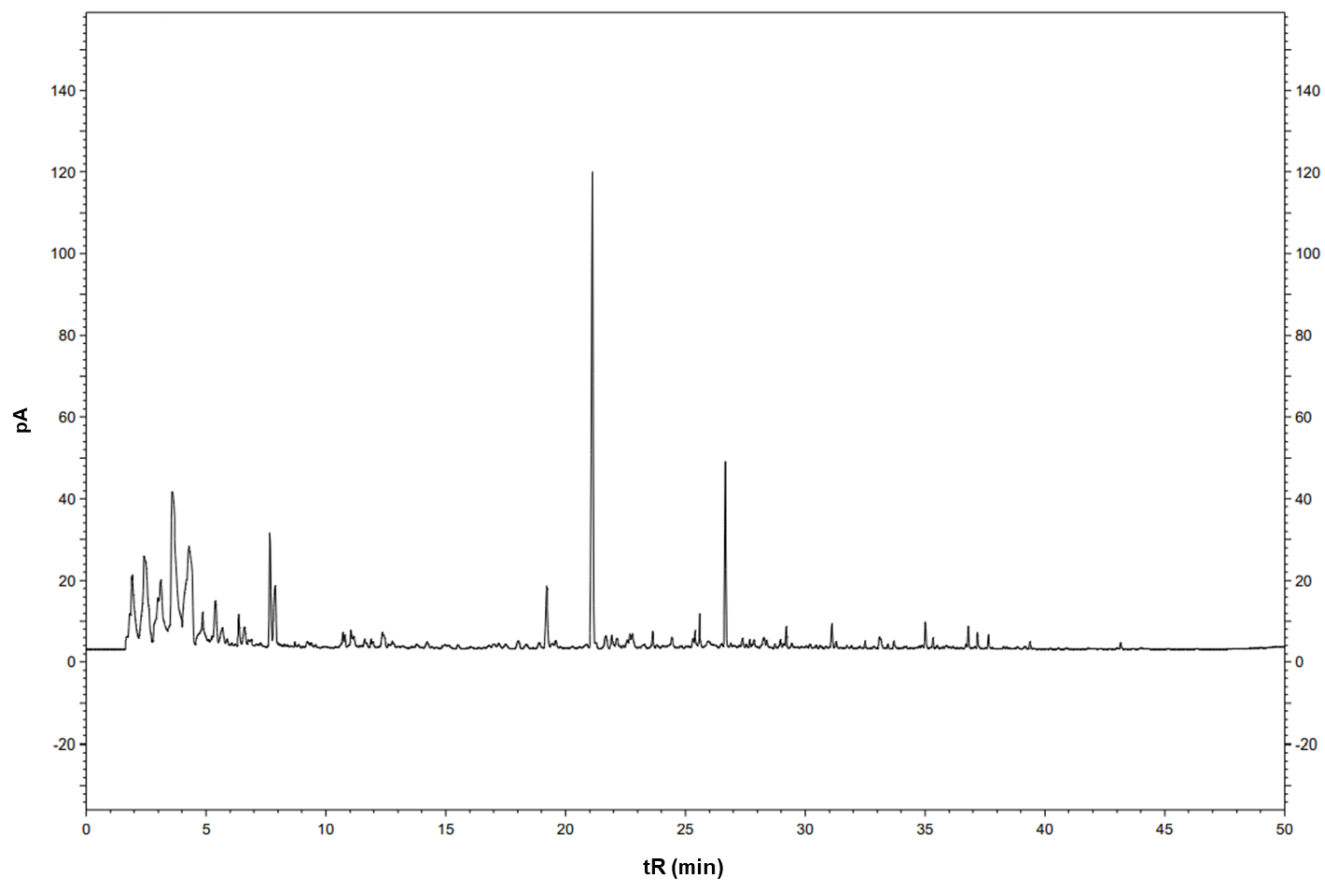


Figura 12 - Cromatograma da análise da mistura de fluido oral e urina realizada no modo *splitless*.

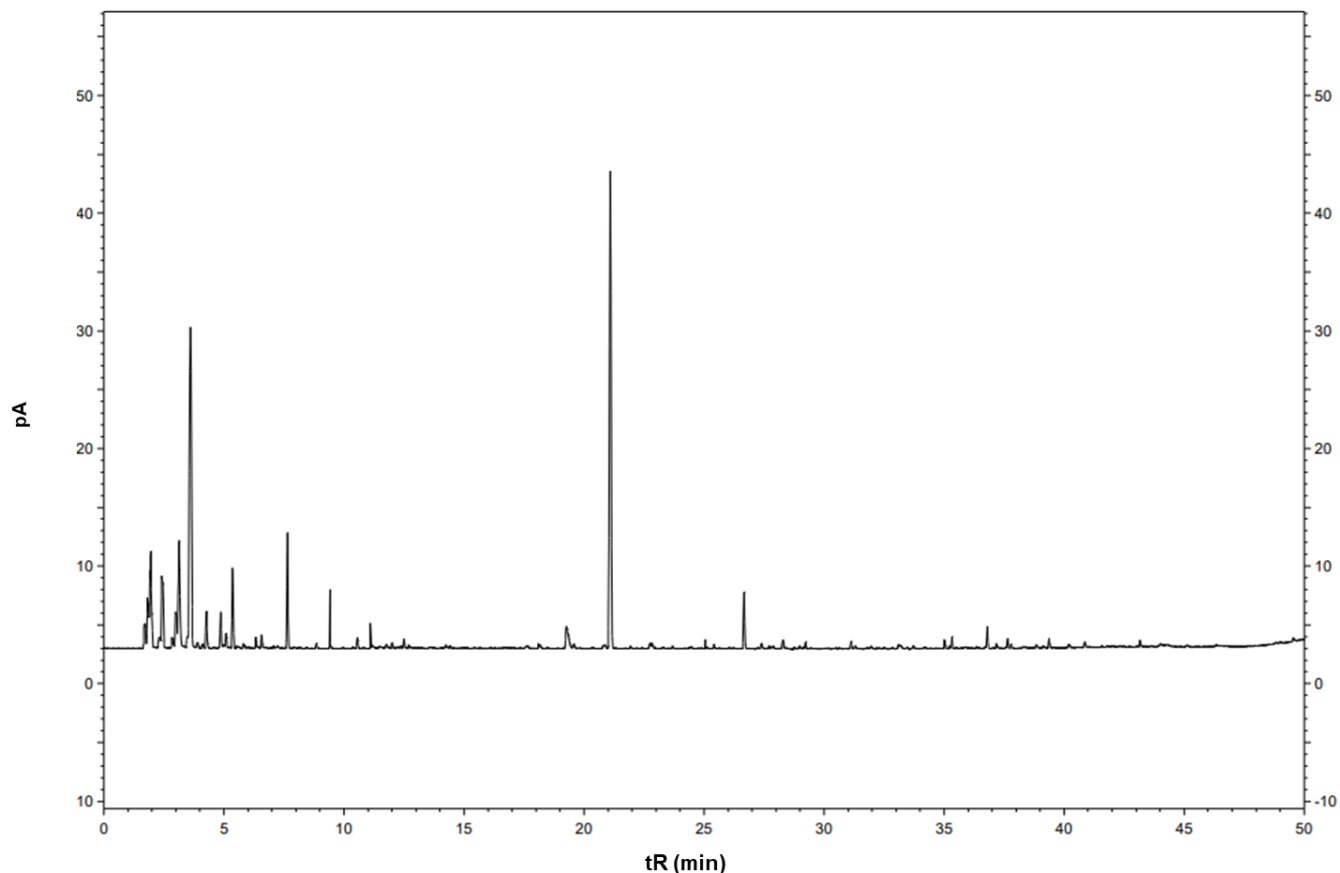


Figura 13 - Cromatograma da análise da mistura de fluido oral e urina realizada no modo split 1:10.

Conforme dito previamente, no modo *splitless* o volume de amostra é injetado de maneira integral na coluna. Conseqüentemente, os picos presentes no cromatograma da Figura 12 apresentam, de maneira geral, uma maior intensidade do que os picos representados na Figura 13. Sendo assim, é natural pensar que o modo *splitless* seria o ideal para a análise dos VOCs encontrados no fluido oral e urina através da técnica de SHS-GC-FID uma vez que os analitos estão presentes em baixas concentrações nessas amostras. Contudo, é preciso notar também que os picos iniciais do cromatograma (entre 3 e 10 minutos) referente à injeção no modo *splitless* são muito largos, assimétricos e com sobreposições. Isso ocorre principalmente devido a dois fatores. Primeiro, como o volume de amostra que entra na coluna capilar em uma injeção *splitless* é muito grande, pode ocorrer uma saturação da fase estacionária, o que “espalharia” as moléculas no início da coluna capilar, causando perda de resolução principalmente para os compostos que interagem menos com a coluna (AQUINO

NETO, 2010). Além disso, esse grande volume de amostra citado anteriormente também faz com que no modo *splitless* demore um tempo maior para que todo o conteúdo presente no injetor seja transferido para a coluna, o que também leva ao alargamento dos picos (processo conhecido como alargamento de banda no tempo, ou “*band broadening in time*”) (AQUINO NETO, 2010). Já na injeção no modo *split*, esses problemas são muito menos pronunciados já que o volume de amostra que chega, de fato, à coluna capilar é menor e todo conteúdo é transferido de maneira mais rápida. Sendo assim, neste modo de injeção os picos iniciais apresentaram uma melhora muito grande na eficiência e resolução. Portanto, mesmo apresentando picos menos intensos, optou-se pela análise no modo *split* visto que as informações químicas dos compostos com menor ponto de ebulição e/ou mais apolares (aqueles que saem mais no início da corrida neste método) poderiam ser perdidas na injeção *splitless* devido à baixíssima resolução dos picos iniciais.

Após a escolha da rampa de temperatura (Tabela 8) e do modo de injeção (*split* 1:10), as variáveis envolvidas no processo de extração por SHS foram otimizadas de acordo com o DOE, como será discutido nas próximas seções.

5.2 Planejamento experimental

5.2.1 Otimização dos parâmetros da extração por SHS dos VOCs presentes na urina

5.2.1.1 Triagem das variáveis: PB

Os resultados obtidos através do delineamento PB para urina estão presentes na Tabela 17.

Tabela 17 - Resultados do Delineamento de Plackett-Burman para urina.

Experimento	Adição de sal (g)	pH	Agitação (rpm)	Tempo (min)	Temperatura (°C)	Média geométrica das áreas
1	1,5	ácido	700	5	60	5142
2	1,5	básico	300	25	60	2330
3	0	básico	700	5	90	3760
4	1,5	ácido	700	25	60	5764
5	1,5	básico	300	25	90	5448
6	1,5	básico	700	5	90	3815
7	0	básico	700	25	60	2327
8	0	ácido	700	25	90	8382
9	0	ácido	300	25	90	7640
10	1,5	ácido	300	5	90	5509
11	0	básico	300	5	60	2199
12	0	ácido	300	5	60	2334
13	0,75	neutro	500	15	75	3287
14	0,75	neutro	500	15	75	2760
15	0,75	neutro	500	15	75	3367
16	0,75	neutro	500	15	75	2964

Como dito anteriormente, os cálculos dos efeitos de cada variável foram realizados através do software Statistica levando-se em consideração a média geométrica das áreas dos picos de cada experimento. Neste caso, na avaliação da influência das variáveis optou-se por utilizar um valor crítico (p) de significância do teste t de Student igual a 0,1 (ao invés de 0,05, o qual é mais comum). Essa escolha foi baseada no fato de que, como se trata de um método de triagem, é melhor incluir uma variável que não é significativa do que excluir uma que é (RODRIGUES & IEMMA, 2005). Os resultados podem ser visualizados de forma mais ilustrativa através do Diagrama de Pareto (Figura 14).

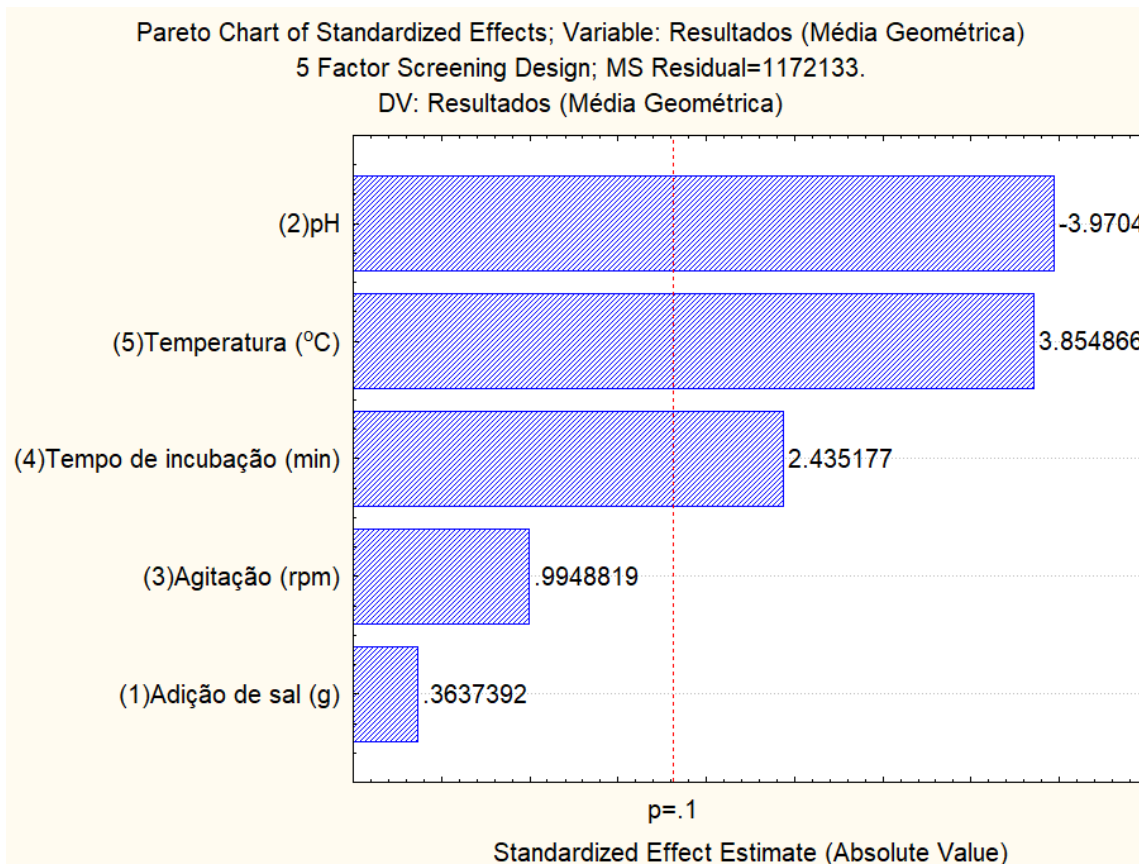


Figura 14 - Diagrama de Pareto para os resultados obtidos através do delineamento de Plackett-Burman para a urina.

No Diagrama de Pareto, as barras horizontais indicam a extensão da contribuição de cada variável para a obtenção dos resultados (as variáveis são apresentadas em ordem decrescente da importância dos efeitos) e a linha pontilhada indica o valor crítico de significância (p) utilizado no tratamento estatístico. Deste modo, as variáveis que apresentam um valor acima do crítico são aquelas onde existe uma significância estatística na contribuição dessa variável para a obtenção dos valores dos resultados. Além disso, o diagrama indica também qual o sinal da magnitude do efeito: se for um valor positivo, esse parâmetro é a favor da resposta avaliada – um aumento no valor deste determinado fator provoca um aumento no valor da resposta (se for um sinal negativo, o raciocínio é o contrário).

Analisando o diagrama acima é possível notar que as variáveis com os efeitos que mais contribuíram para o processo (ou seja, que mais contribuíram para que as médias geométricas das áreas fossem maiores) foram: pH, temperatura e tempo de

incubação. Por apresentarem a magnitude do efeito maior do que o valor de p , essas três variáveis são estatisticamente significativas. É interessante também observar que o pH tem um sinal negativo associado ao valor. Ou seja, menores valores de pH (ácido, entre 1 e 2) fazem com que as médias geométricas das áreas sejam maiores.

Conforme discutido na seção 2.3 deste capítulo, cada uma das variáveis avaliadas corresponde a um parâmetro que está relacionado à passagem do analito da matriz para a fase de vapor (*headspace*), fase esta que é posteriormente analisada pelo GC-FID. Neste sentido, o objetivo é fazer com que os coeficientes de partição (K) dos VOCs sejam os menores possível, aumentando a concentração desses compostos no HS. Os resultados obtidos para cada um destes parâmetros será discutido abaixo.

Primeiramente discutindo sobre o pH, este parâmetro é importante pois muitos compostos são eliminados na urina na sua forma conjugada. Quando o composto é conjugado, o mesmo apresenta grupos polares que interagem com maior intensidade com uma matriz polar (como o caso da urina, que é majoritariamente formada por água), diminuindo a concentração da substância no *headspace*. Além disso, a conjugação faz com que haja um aumento do peso molecular do composto, conseqüentemente o mesmo tende a ser menos volátil. Valores baixos de pH levam a reações de desconjugação, onde os analitos ficam na sua forma “livre”, tendo assim menor peso molecular e menor interação com a água (MONTEIRO *et al.*, 2014). Isso faz com que o número e também a concentração dos VOCs na fase de vapor aumente. Conseqüentemente, a média geométrica das áreas dos picos é maior. Portanto, escolheu-se por fixar o pH como ácido (entre 1-2) para as posteriores análises de urina.

A temperatura foi outro fator estatisticamente significativo para o processo. Quanto maior a temperatura, maior o grau de agitação das moléculas e maior é a tendência das mesmas passarem para o estado gasoso. Portanto, em temperaturas maiores é previsto uma maior quantidade de analitos na fase de vapor (RESTEK, 2000). Contudo, temperaturas muito altas podem fazer com que os analitos degradem, e também levam ao aumento da quantidade de vapor de água no *headspace*. Por esta razão, não foram avaliadas temperaturas acima de 100 °C. Sendo assim, a temperatura foi uma variável levada em consideração para a posterior otimização.

O tempo de incubação é avaliado no SHS pois leva um certo tempo para que os analitos passem para a fase de vapor, estabelecendo assim um equilíbrio dinâmico entre a amostra e o *headspace*. Segundo a Figura 14, quanto maior o tempo de incubação, maior tende a ser a resposta obtida. Contudo, um tempo de incubação muito elevado significa também uma análise muito mais lenta. Portanto, é preciso balancear esses dois pontos. Além disso, como discutido anteriormente, escolheu-se por fixar o pH como ácido para as posteriores análises de urina devido às reações de desconjugação, processo o qual também leva certo tempo para ocorrer. Sendo assim, devido a esses dois fatores, o tempo de incubação também foi um fator levado em consideração para ser otimizado pelo DCCR.

A agitação faz com que diminua o tempo necessário para estabelecer o equilíbrio dinâmico do analito entre fase líquida e de vapor (RESTEK 2000). Analisando a Figura 14 é possível perceber que, quanto maior é a agitação, maior tende a ser a resposta, porém esse aumento não é estatisticamente significativo. Portanto, foi optado por fixar este parâmetro no valor de 500 rpm já que este foi o valor médio analisado.

Por fim, a adição de sal é um fator importante no contexto das extrações por *headspace* uma vez que promove o *salting out*, conforme discutido previamente. Através da Figura 14 é possível observar que, assim como no caso da agitação, a adição de sal promove um aumento da média geométrica das áreas dos picos, porém esse efeito não é significativo. Isso pode ser devido ao fato de que a urina é uma matriz que contém naturalmente uma quantidade considerável de eletrólitos, fazendo com que o efeito da adição de sal não seja tão pronunciado nesse tipo de amostra. Desta forma, assim como na avaliação da agitação, o valor de adição de sal também foi fixado no nível médio avaliado (0,75 g de NaCl).

Em suma, as variáveis que tiveram os valores fixados através da avaliação do PB foram: agitação (500 rpm), pH (ácido – entre 1 e 2) e adição de NaCl (0,75 g). Já o tempo e temperatura de incubação foram escolhidos para serem otimizados. Como foi observado que, quanto maior era o tempo de incubação, maior era a média geométrica das áreas, para a avaliação do DCCR optou-se por analisar o comportamento da resposta em períodos maiores (entre 20 e 40 minutos, diferentemente do PB que foi entre 5 e 25 minutos). Com relação à temperatura, seguiu-se a mesma linha de

raciocínio: esta variável foi avaliada em um intervalo de 65 a 95 °C (no PB havia sido entre 60 e 90°C).

5.2.1.2 DCCR de 2 fatores para a urina: avaliando tempo e temperatura de incubação

Os resultados obtidos nos ensaios realizados à luz do DCCR estão presentes na Tabela 18.

Tabela 18 - Resultados do DCCR de 2 fatores na otimização da extração dos VOCs urinários.

Experimentos	Tempo (min)	Temperatura (°C)	Média geométrica das áreas
1	23	69	8028
2	37	69	10888
3	23	91	22509
4	37	91	24510
5	20	80	13336
6	40	80	16601
7	30	65	15236
8	30	95	27449
9	30	80	13379
10	30	80	14274
11	30	80	14691

Os cálculos das estimativas dos efeitos e a análise da variância (ANOVA) dos resultados foram realizados no *software* Statistica (Tabela 19). Essas análises foram determinadas através de um teste F e um nível de significância α , onde os valores de p menores que 0,05 indicavam que a variável independente avaliada apresentava efeito estatisticamente significativo no nível de confiança de 95%.

Tabela 19 – Análise de variância (ANOVA) dos resultados obtidos para o delineamento composto central rotacional (DCCR) aplicado na otimização dos parâmetros tempo e temperatura de incubação para a extração dos VOCs urinários através da técnica de *headspace* estático.

	SQ	GL	MQ	F	p-valor	R ²	R ² ajustado
Tempo de incubação (L)	11228994	1	11228994	2,71249	0,160483	0,9413	0,8827
Tempo de incubação (Q)	14811	1	14811	0,00358	0,954621		
Temperatura de incubação (L)	259526827	1	259526827	62,69158	0,000517		
Temperatura de incubação (Q)	55944179	1	55944179	13,51394	0,014351		
Tempo de incubação x Temperatura de incubação	184104	1	184104	0,04447	0,841303		
Falta de ajuste	19800149	3	6600050				
Erro puro	898549	2	449275				
Soma total de Quadrados	352984484	10					

SQ: soma dos quadrados; GL: graus de Liberdade; MQ: media dos quadrados; L: termo linear; Q: termo quadrático; R²: coeficiente de determinação do modelo.

Os valores representados em vermelho na Tabela 19 indicam as variáveis que foram significativas para os resultados obtidos no DCCR. Portanto, a temperatura foi o único fator significativo nesse conjunto de experimentos. Ou seja, o aumento da temperatura foi o único fator que contribuiu significativamente para o aumento da média geométrica das áreas. Consequentemente, é possível concluir que a partir de 20 minutos (menor valor de tempo avaliado no DCCR), o aumento do tempo não impacta significativamente nas áreas dos picos dos cromatogramas.

Com os esses resultados foi possível gerar uma de superfície de resposta, a qual corresponde a um modelo matemático em três dimensões que possibilita a obtenção dos valores ótimos de um processo (RODRIGUES, & IEMMA, 2005). A superfície de resposta para o DCCR da urina está representada na Figura 15.

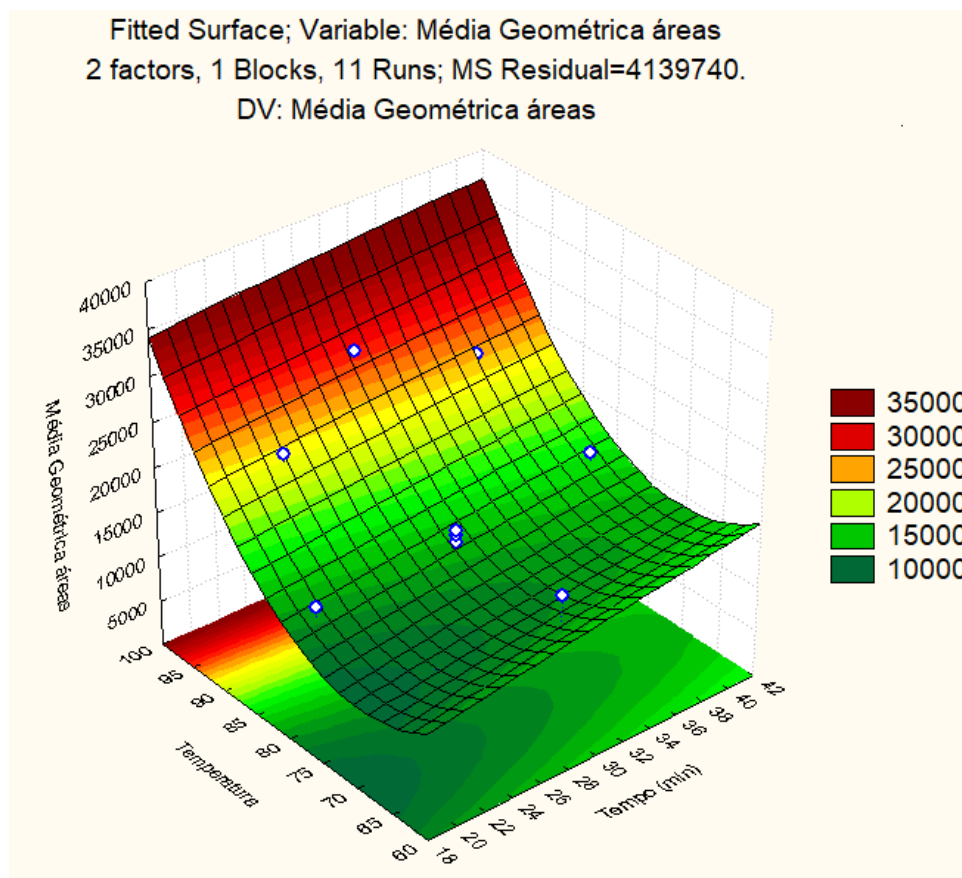


Figura 15 - Superfície de resposta em função da temperatura (°C) e tempo de incubação (min) para a média geométrica das áreas dos picos nas análises dos VOCs urinários por SHS-GC-FID.

A interpretação da superfície de resposta é a seguinte: quanto mais vermelha é a região do gráfico, maior é a resposta obtida. Sendo assim, é possível observar que a região que apresenta a maior resposta (região ótima) fica entre a temperatura de 90 e 95 °C. Além disso, a superfície de resposta confirma o que foi observado na Tabela 19 e discutido anteriormente: ela também nos mostra que o tempo pouco interfere nas respostas a partir de 20 minutos. Portanto, este é o valor ideal para que a extração dos VOCs urinários por SHS seja realizada com a maior resposta no menor intervalo possível.

A partir desses resultados foi gerado um modelo de regressão para estimar a resposta em função das variáveis estudadas dentro da região experimental. A equação que determina a superfície de resposta discutida anteriormente, ou seja, que descreve a resposta avaliada em função das variáveis codificadas é:

$$Y = 14150.11 + 1178.75 X_1 - 50.17 X_1^2 + 5798.43 X_2 - 3324.42 X_2^2 - 214.54 X_1 X_2 \text{ (Equação 5)}$$

Onde X_1 representa a variável tempo e X_2 a temperatura. Substituindo-se os valores das incógnitas pelas variáveis codificadas é possível determinar, de maneira teórica, qual é o valor de resposta obtido nas condições especificadas. Essa avaliação é importante para a validação do modelo, como será discutido posteriormente.

Após a escolha do ponto ótimo através da observação da superfície de resposta, foi realizada a verificação da qualidade do modelo matemático gerado. Os cálculos envolvidos nesse processo foram efetuados tanto pelo *software* Statistica quanto pelo *software* Octave e foram baseados no teste ANOVA apresentado anteriormente (Tabela 19).

O primeiro critério avaliado na análise da qualidade do modelo foi o valor do coeficiente de determinação (R^2). O valor de R^2 obtido neste processo foi de 0,9413. Deste modo, o modelo matemático que relaciona tempo e temperatura de incubação com a média geométrica das áreas foi capaz de contemplar 94,13% das informações dos dados (o modelo não foi capaz de explicar apenas 5,87% de todas as informações geradas). Esse número é considerado satisfatório uma vez que, segundo Joglekar e May (JOGLEKAR & MAY, 1987), um R^2 acima de, pelo menos, 0,80 é o ideal. Além disso, o valor obtido para R^2 ajustado foi próximo ao valor de R^2 , sugerindo que não há adição de coeficientes não significativos ao modelo. Além do valor de R^2 , também foi avaliado se a regressão do modelo foi significativa. Neste caso, foi utilizado um teste F: razão da média quadrática da regressão pela média quadrática do resíduo (15,2); dividido pelo valor de F tabelado (5,1). Como o resultado dessa divisão é maior que 1 (aproximadamente 3), o modelo apresenta uma regressão significativa. Isto é, a média quadrática da regressão é estatisticamente diferente da média quadrática do resíduo (erro). Essa tendência também pode ser observada através da Figura 16, a qual mostra o gráfico dos valores preditos versus observados. Esse gráfico demonstra que os valores previstos pelo modelo (reta em vermelho) são próximos aos obtidos experimentalmente (pontos em azul).

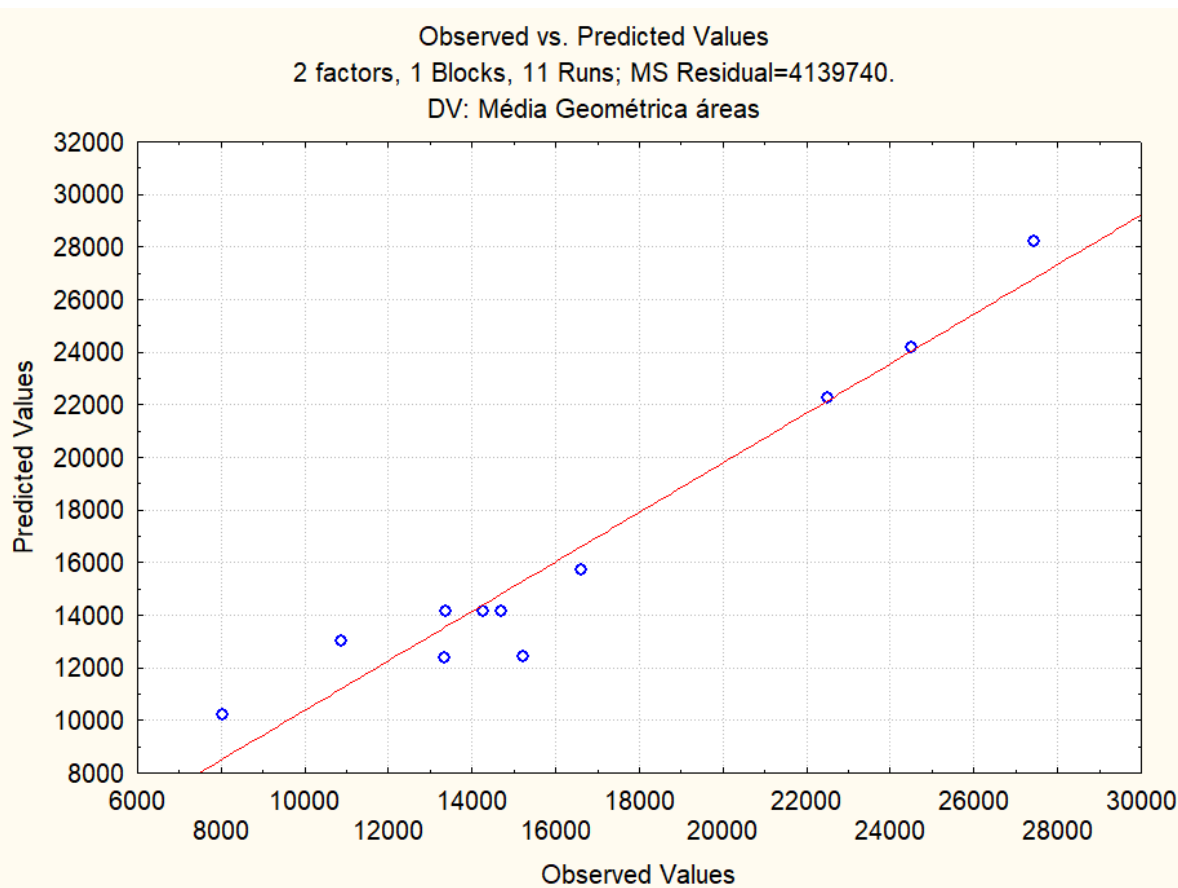


Figura 16 - Gráfico de valores preditos versus os valores observados para o modelo gerado no DCCR para a urina.

Por fim, foi avaliado se o modelo apresentava indícios de falta de ajuste. Para tal, foi realizado outro teste F: razão da variância da falta de ajuste pela variância do erro puro (15); dividido pelo valor de F tabelado (19). O valor dessa divisão foi de, aproximadamente, 0,79. Como este valor é menor que 1, podemos dizer que o erro puro (erro do laboratório) se confunde com a falta de ajuste (erro devido ao ajuste do modelo). Isso é justamente o desejado. Ou seja, não há diferença do ponto de vista estatístico entre o erro puro e a falta de ajuste. Em outras palavras, o erro cometido durante o experimento (erro puro) se confunde com o erro relacionado ao modelo gerado (erro da falta de ajuste). Como o valor de R^2 foi satisfatório, a regressão foi significativa, e não apresenta falta de ajuste, o modelo apresentou evidências de que era adequado àquilo que era proposto. Os gráficos ilustrativos (gerados no *software*

Octave) dos cálculos dos testes F utilizados na avaliação da qualidade do modelo estão presentes nas Figuras 17 e 18.

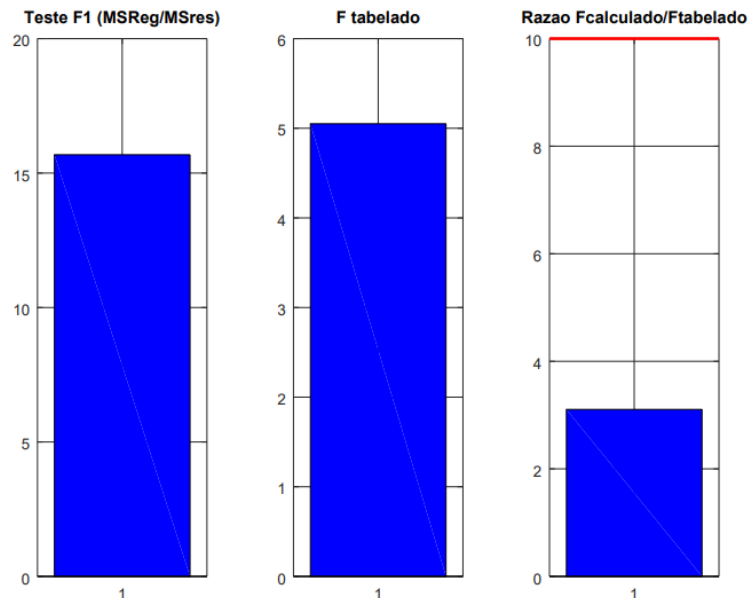


Figura 17 - Gráfico ilustrativo gerado pelo software Octave para a avaliação da significância do modelo gerado no DCCR para a urina

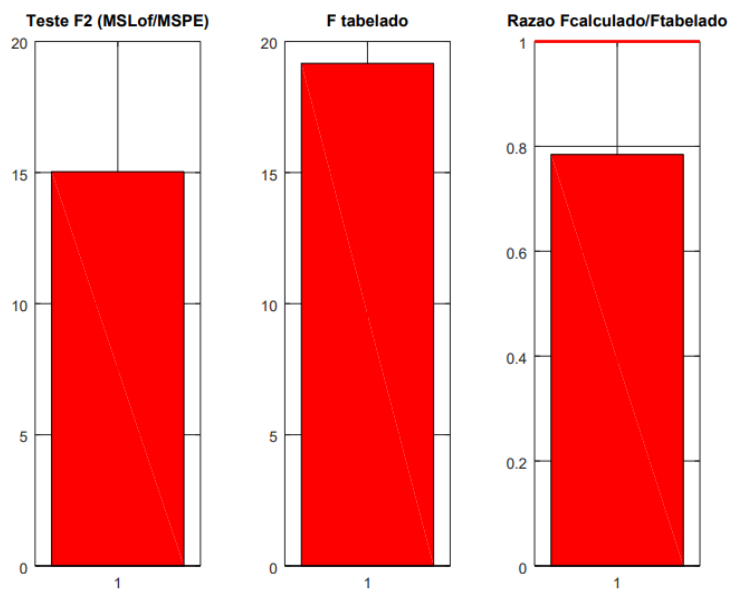


Figura 18 - Gráfico ilustrativo gerado pelo software Octave para a avaliação do erro puro e falta de ajuste do modelo gerado no DCCR para a urina.

Como o modelo mostrou evidências de ser adequado, o último passo foi realizar a validação deste modelo para confirmar a capacidade preditiva do mesmo. Para tal, foram realizados experimentos nas condições ótimas (Tabela 20) com o objetivo de verificar se a resposta obtida experimentalmente seria próxima ao valor previsto. Como dito anteriormente, substituindo os valores codificados das variáveis avaliadas na equação que determina o modelo gerado (Equação 5) é possível obter o valor previsto da resposta. Esse valor também pode ser obtido através do software Statistica (Figura 19).

Tabela 20 - Valores ótimos para as variáveis avaliadas no processo de otimização da extração para a urina.

Variável	Valor
Tempo (min)	20
Temperatura (° C)	95
NaCl (g)	0,75
Agitação (rpm)	500
pH	1-2
Volume de amostra (mL)	3

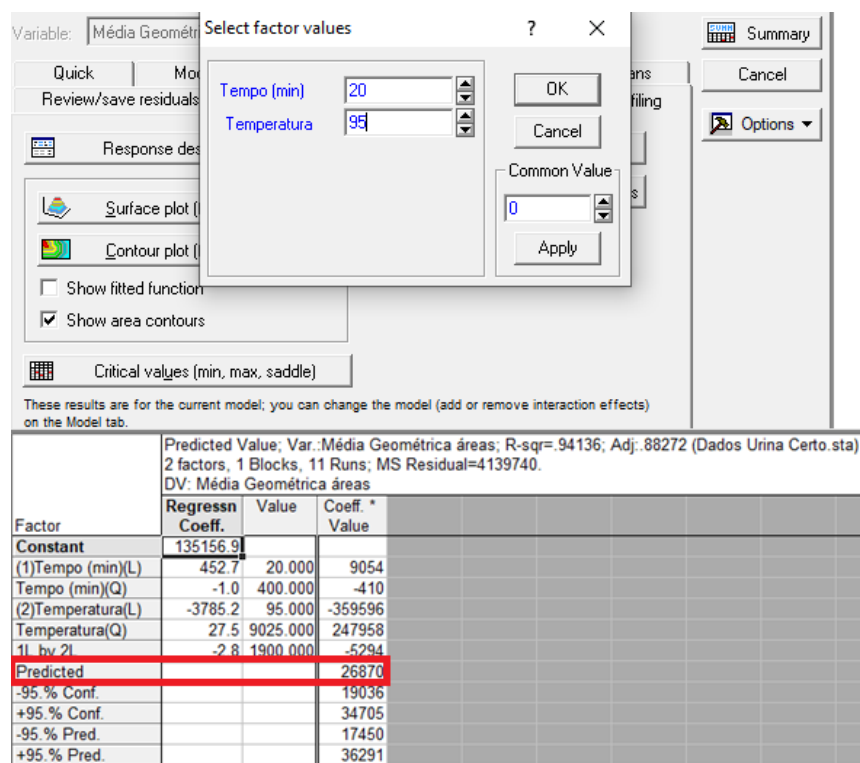


Figura 19 - Valor ótimo de resposta previsto para a extração por headspace dos VOCs urinários segundo o software estatística 7.0.

Para a validação, o mesmo *pool* de urina utilizado nos experimentos de otimização foi analisado em triplicata na condição teórica ideal. O valor médio obtido experimentalmente nesses ensaios foi de 27344 ± 374 (desvio padrão), representando um erro relativo de apenas 1,7% comparando-se com o valor previsto (26870). Portanto, o método para a análise de VOCs presentes na urina foi desenvolvido, otimizado, o modelo matemático gerado nessa otimização foi considerado adequado e, por fim, mostrou-se capaz de prever os dados (modelo preditivo). A Figura 20 representa uma comparação entre os cromatogramas das análises do *pool* de urina realizadas na condição ótima e em uma condição não otimizada (PB experimento 2 – Tabela 17).

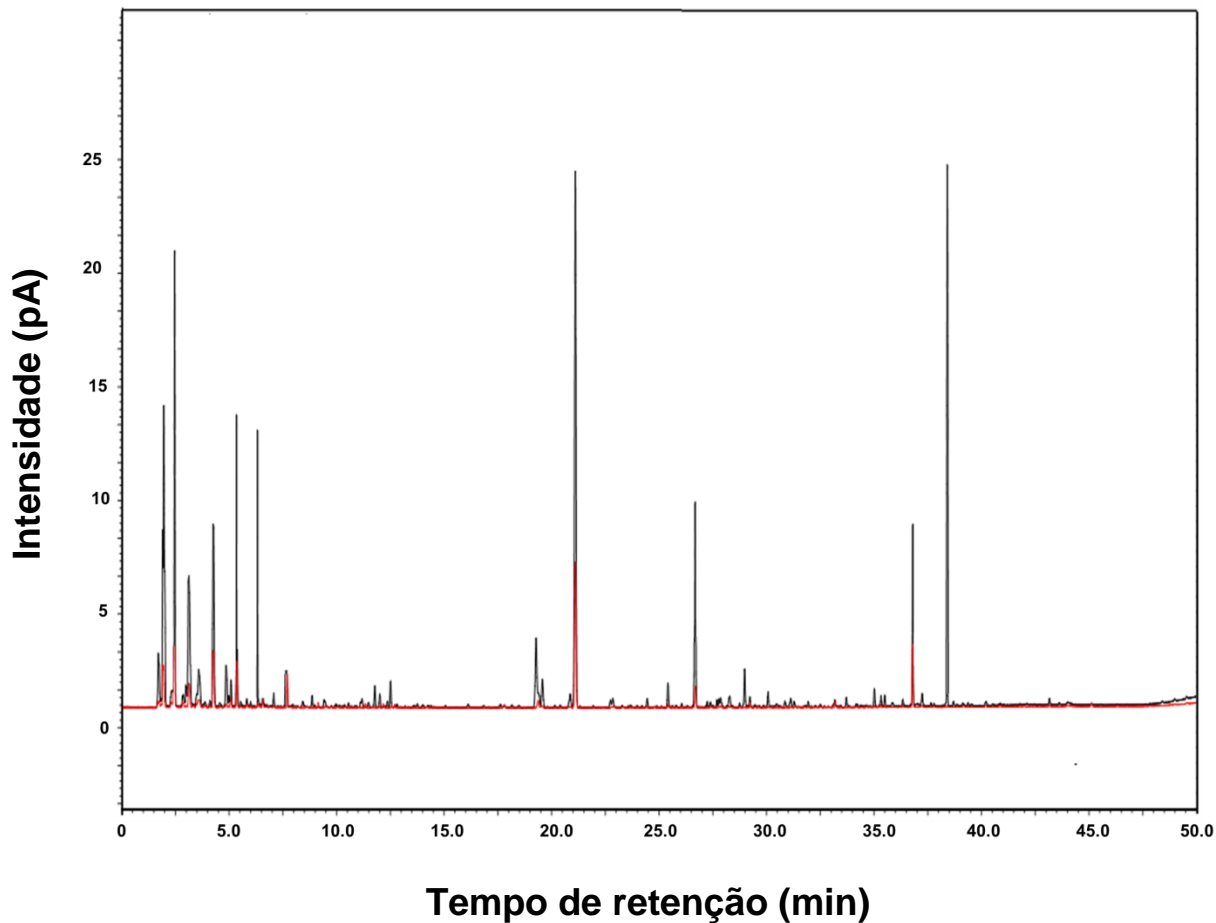


Figura 20 - Sobreposição dos cromatogramas referentes à análise do *pool* de urina realizada na condição ótima (em preto) determinada pelo planejamento experimental e em uma condição não otimizada (PB experimento 2 – Tabela 17).

5.2.2 Otimização dos parâmetros da extração por *headspace* estático dos VOCs presentes no fluido oral

Como especificado anteriormente, a otimização dos parâmetros envolvendo o fluido oral seguiu a mesma linha de raciocínio da otimização para a urina. Os resultados obtidos através do delineamento PB para o fluido oral estão presentes na Tabela 21.

Tabela 21 - Resultados do Delineamento PB para o fluido oral.

Experimento	Adição de sal (g)	pH	Agitação (rpm)	Tempo (min)	Temperatura (°C)	Média geométrica das áreas
1	1,0	ácido	700	5	60	4350
2	1,0	básico	300	25	60	2917
3	0	básico	700	5	90	2698
4	1,0	ácido	700	25	60	5309
5	1,0	básico	300	25	90	4331
6	1,0	básico	700	5	90	3153
7	0	básico	700	25	60	2170
8	0	ácido	700	25	90	4765
9	0	ácido	300	25	90	4573
10	1,0	ácido	300	5	90	4624
11	0	básico	300	5	60	1686
12	0	ácido	300	5	60	2291
13	0,50	neutro	500	15	75	4090
14	0,50	neutro	500	15	75	3743
15	0,50	neutro	500	15	75	3741
16	0,50	neutro	500	15	75	3435

Assim como no planejamento para a urina, os resultados podem ser visualizados de maneira mais ilustrativa na forma de um Diagrama de Pareto para avaliar as variáveis com efeito significativo sobre a resposta avaliada (Figura 21).

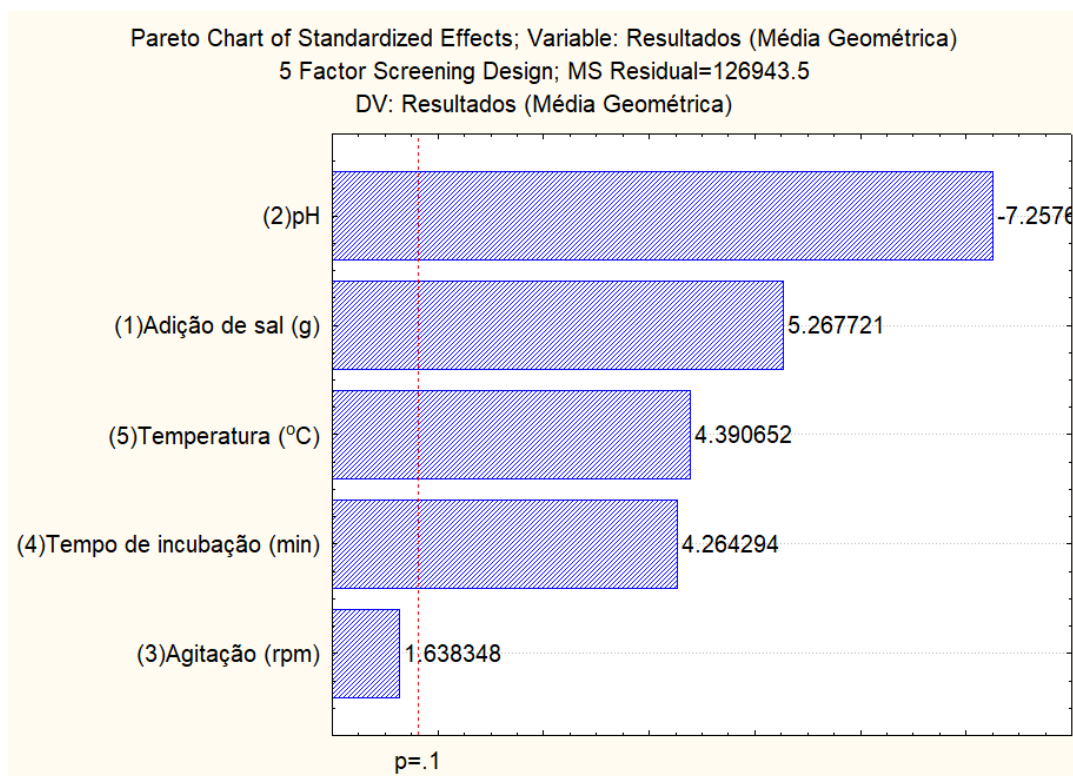


Figura 21 - Diagrama de Pareto para os resultados obtidos através do delineamento de Plackett-Burman para o fluido oral.

De modo semelhante ao PB da urina, as variáveis pH, temperatura e tempo de incubação foram significativas. Contudo, neste caso a adição de sal também foi estatisticamente significativa para a resposta avaliada. Isso pode ser devido ao fato do fluido oral não apresentar originalmente tantos eletrólitos quanto a urina. Sendo assim, ao contrário do que ocorreu na avaliação da urina, a adição de sal ao fluido oral levou, de fato, a um efeito *salting out* pronunciado. Portanto, este fator foi considerado na otimização por DCCR.

Com relação ao pH, apesar do fluido oral apresentar os metabólitos, geralmente, na sua forma livre (não conjugada, diferentemente da urina), foi observado que, assim como na otimização da urina, o meio ácido favoreceu a presença dos VOCs no *headspace*. Isso pode ser devido ao fato de que o fluido oral apresenta uma quantidade considerável de ácidos orgânicos voláteis (como o ácido acético, ácido propanoico e ácido butanoico), os quais ficam majoritariamente na forma molecular quando o pH do meio está entre 1 e 2. Conseqüentemente, em meio ácido tais compostos interagem

menos com a fase aquosa e apresentam maior tendência de difusão para a fase de vapor. Tal tendência já havia sido observada em trabalho realizado por Cavaco e colaboradores (CAVACO *et al.*, 2018).

Portanto, as variáveis que tiveram os valores fixados através da avaliação do PB para o fluido oral foram: pH (ácido – entre 1 e 2) e agitação em 500 rpm (variável esta que não foi significativa, logo foi fixada no valor médio avaliado). Já os valores de temperatura, tempo de incubação e adição de sal foram otimizados segundo o DCCR de 3 fatores.

Na otimização dos parâmetros para o fluido oral, optou-se por avaliar a mesma faixa de temperatura que foi utilizada para no DCCR para a urina (65 a 95 °C). Em relação ao tempo de incubação, como foi observado que para a urina após 20 minutos não havia diferença significativa na resposta, e o volume de amostra utilizado nas análises de urina (3 mL) é muito maior do que o volume de amostra de fluido oral (850 µL), decidiu-se por avaliar um tempo máximo 20 minutos para esta matriz em questão (faixa de tempo avaliada: entre 5 e 20 minutos). Por fim, no que se refere à adição de sal, foi observado que quando era adicionada uma massa de NaCl maior do que 1,0 g na amostra de fluido oral, a quantidade de sal era tão grande dentro do frasco de HS que acabava cobrindo praticamente todo o volume amostral. Sendo assim, optou-se por manter a avaliação da quantidade de NaCl na faixa de 0,2 e 1,0 g.

5.2.3 DCCR de 3 fatores para o fluido oral

Os resultados obtidos nos ensaios realizados segundo o DCCR de 3 fatores para o fluido oral estão presentes na Tabela 22.

Tabela 22 - Resultados do DCCR de 3 fatores para o fluido oral.

Experimentos	Tempo (min)	Temperatura (°C)	Adição de sal (g)	Média geométrica das áreas
1	8,3	71,1	0,36	3290
2	16,7	71,1	0,36	4849
3	8,3	88,9	0,36	4576
4	16,7	88,9	0,36	7296
5	8,3	71,1	0,84	4738
6	16,7	71,1	0,84	4153
7	8,3	88,9	0,84	7001
8	16,7	88,9	0,84	7704
9	5	80	0,6	3098
10	20	80	0,6	5412
11	12,5	65	0,6	3628
12	12,5	95	0,6	6890
13	12,5	80	0,2	3873
14	12,5	80	1,0	5439
15	12,5	80	0,6	4790
16	12,5	80	0,6	5219
17	12,5	80	0,6	4962

A análise da variância (ANOVA) dos resultados obtidos está representada na Tabela 23.

Tabela 23 - Análise de variância (ANOVA) dos resultados obtidos para o delineamento composto central rotacional (DCCR) aplicado na otimização dos parâmetros tempo de incubação, temperatura de incubação e adição de sal para a extração dos VOCs presentes no fluido oral através da técnica de headspace estático.

	SQ	GL	MQ	F	p-valor	R ²	R ² ajustado
Tempo de incubação (L)	5059713	1	5059713	17,64841	0,004031	0,9354	0,8523
Temperatura de incubação (L)	16544326	1	16544326	57,70705	0,000127		
Adição de sal (L)	2831166	1	2831166	9,87518	0,016323		
Tempo de incubação x Temperatura de incubação	749700	1	749700	2,61497	0,149893		
Tempo de incubação x Adição de sal	2164240	1	2164240	7,54893	0,028606		
Temperatura de incubação x Adição de sal	541320	1	541320	1,88814	0,211793		
Tempo de incubação (Q)	157256	1	157256	0,54851	0,483024		
Temperatura de incubação (Q)	696074	1	696074	2,42793	0,163151		
Adição de sal (Q)	18332	1	18332	0,06394	0,807633		
Falta de ajuste	1913641	5	382728				
Erro puro	93225	2	46612				
Soma total de quadrados	31081754	16					

SQ: soma dos quadrados; GL: graus de Liberdade; MQ: media dos quadrados; L: termo linear; Q: termo quadrático; R²: coeficiente de determinação do modelo.

Com relação à superfície de resposta, não é possível determinar uma única superfície para três variáveis ao mesmo tempo. Sendo assim é preciso fixar um determinado valor para uma variável e observar o comportamento das outras duas envolvidas. Neste sentido, foi observado que, ao utilizar a maior quantidade de sal avaliada (1,0 g), a região ótima da resposta fica localizada na temperatura de 95 °C e no tempo de 15 min. Como não há diferença significativa na resposta a partir desse intervalo de tempo, o mesmo foi escolhido como o ideal. A superfície de resposta para

a relação tempo e temperatura de incubação (com a adição de sal fixada em 1,0 g) está representada na Figura 22.

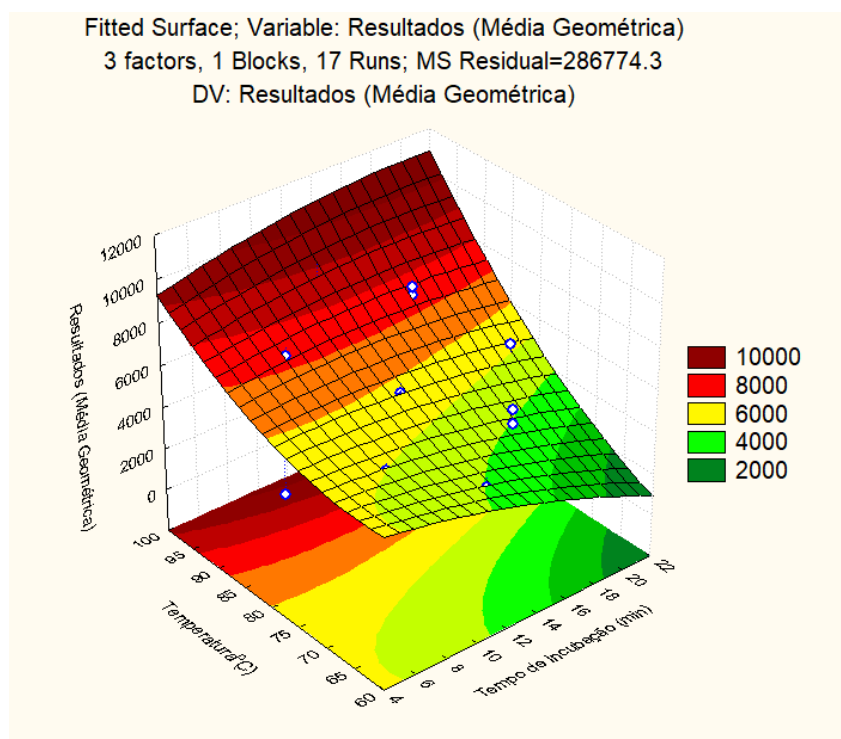


Figura 22 - Superfície de resposta em função da temperatura (°C) e tempo de incubação (min) para a média geométrica das áreas dos picos nas análises dos VOCs presentes no fluido oral por SHS-GC-FID (adição de sal fixada em 1,0 g).

A equação (onde X_1 representa a variável tempo, X_2 a temperatura, e X_3 a adição de sal) que determina a superfície de resposta discutida anteriormente em função das variáveis codificadas é:

$$Y = 4970.40 + 593.210 X_1 - 107.102 X_1^2 + 1099.711 X_2 + 249.319 X_2^2 + 456.933 X_3 + 41.120 X_3^2 + 305.866 X_1X_2 - 520.055 X_1X_3 + 260.088 X_2X_3 \text{ (Equação 6)}$$

Assim como foi feito para a urina, após a escolha do ponto ótimo através da observação da superfície de resposta, foi realizada a verificação da qualidade deste modelo matemático gerado.

Para o modelo de fluido oral, o valor de R^2 obtido foi de 0,9354 (com R^2 ajustado de 0,8523), o que é considerado satisfatório. Ainda, a razão da media quadrática da

regressão pela média quadrática do resíduo dividida pelo F tabelado foi de 2,91, representando assim uma regressão significativa. Com relação à avaliação da falta de ajuste, a razão da variância da falta de ajuste pela variância do erro puro dividida pelo valor de F tabelado foi de 0,45. Como discutido anteriormente, como esse valor é menor que 1, o modelo não demonstra indícios de falta de ajuste. Por fim, a Figura 23 demonstra que os valores previstos pelo modelo são próximos aos obtidos experimentalmente.

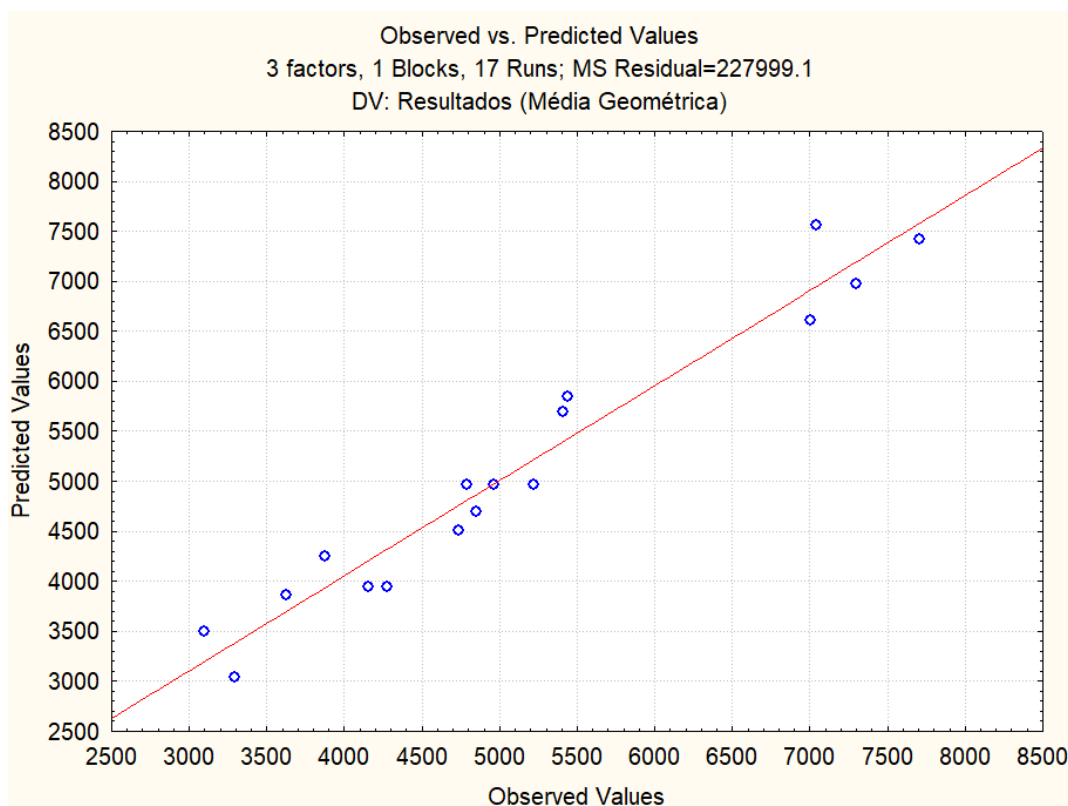


Figura 23 - Gráfico de valores preditos versus os valores observados para o modelo gerado no DCCR para a o fluido oral.

Portanto, assim como na otimização para a urina, o modelo gerado para a otimização do fluido oral apresentou valor de R^2 adequado, regressão significativa, e não apresentou falta de ajuste, sugerindo assim que o modelo é adequado. Os gráficos ilustrativos gerados no *software* Octave estão presentes nas Figuras 24 e 25.

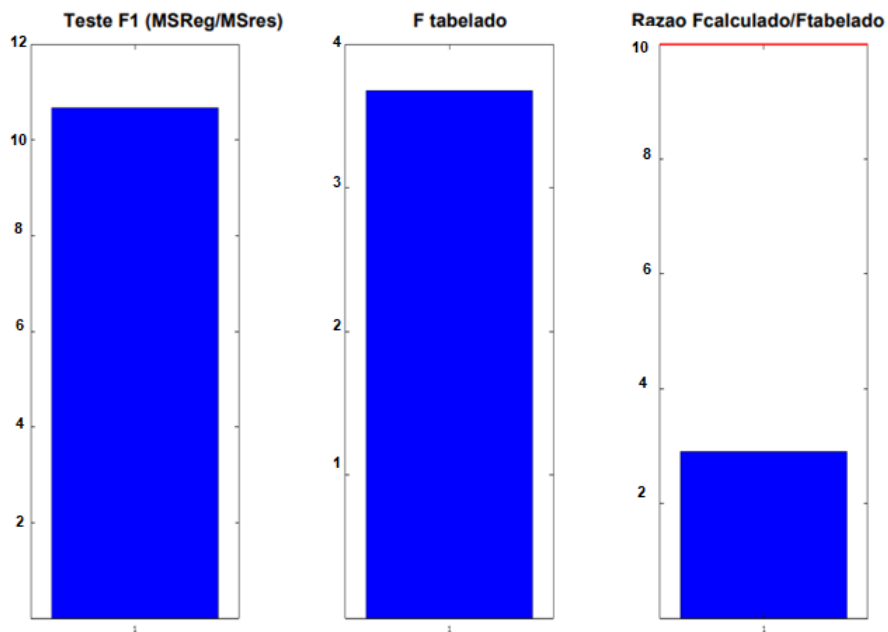


Figura 24 - Gráfico ilustrativo gerado pelo software Octave para a avaliação do da significância do modelo gerado no DCCR para o fluido oral.

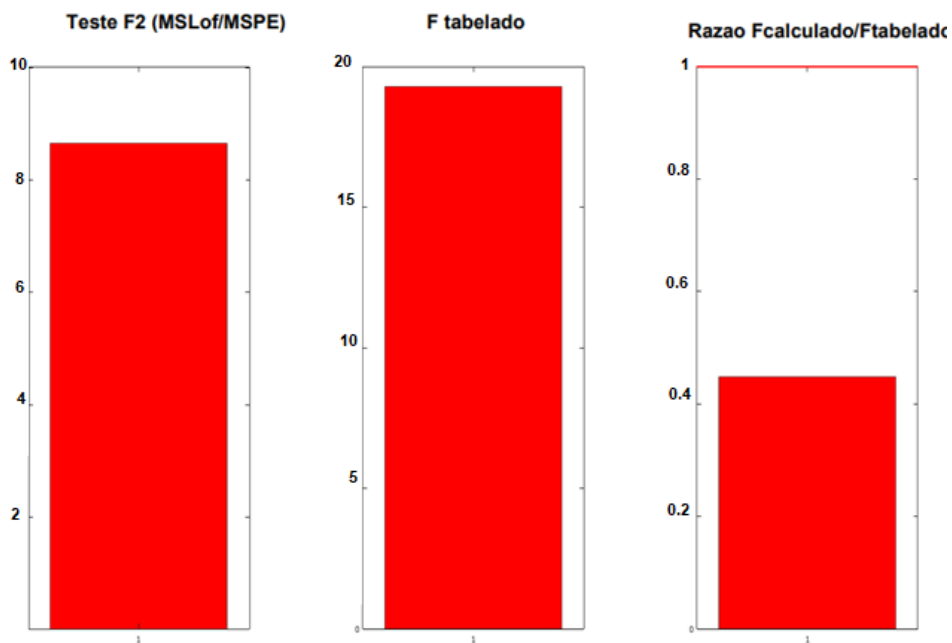


Figura 25 - Gráfico ilustrativo gerado pelo software Octave para a avaliação do erro puro e falta de ajuste do modelo gerado no DCCR para a o fluido oral.

Para a validação do modelo, o mesmo *pool* de fluido oral utilizado nos experimentos de otimização foi analisado em triplicata nas condições ideais (Tabela 24). A média obtida nos ensaios foi de $8668,3 \pm 443$ (desvio padrão) um erro relativo de 6,23% comparando-se com o valor previsto (9244,5 – Figura 26). Portanto, o método para a análise de VOCs presentes na o fluido oral foi desenvolvido, otimizado e o modelo obtido foi considerado adequado e, por fim, validado. A Figura 27 representa uma comparação entre os cromatogramas das análises do *pool* de fluido oral realizadas na condição ótima determinada pelo DoE e em uma condição não otimizada (PB experimento 2 – Tabela 17).

Tabela 24 - Valores ótimos para as variáveis avaliadas no processo de otimização para o fluido oral.

Variável	Valor
T_{inc} (min)	15
T (°C)	95
NaCl (g)	1,0
A (rpm)	500
pH	1-2
Volume de amostra (µL)	850

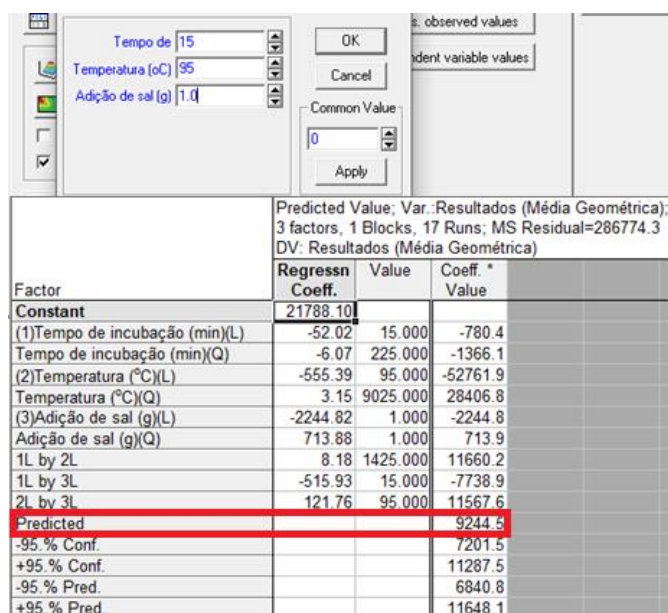


Figura 26 - Valor ótimo de resposta previsto para a extração por headspace dos VOCs presentes no fluido oral segundo o software Statistica.

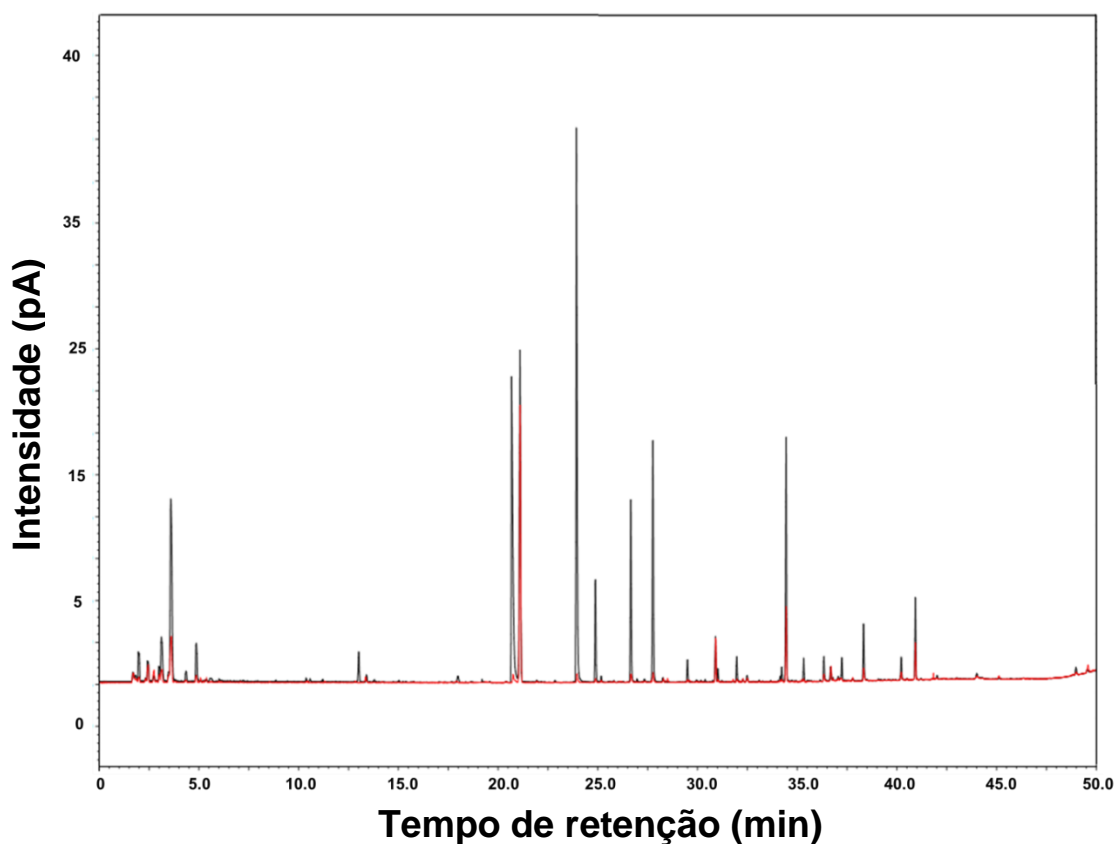


Figura 27 - Sobreposição dos cromatogramas referentes à análise do *pool* de fluido oral realizada na condição ótima (em preto) determinada pelo planejamento experimental e em uma condição não otimizada (PB experimento 2 – Tabela 21).

5.3 Verificação de possíveis interferentes nas análises

A contribuição dos possíveis interferentes na composição dos cromatogramas pode ser considerada mínima. A Figura 28 ilustra a possível contribuição da água deionizada, do NaCl e do ácido sulfúrico, os quais foram adicionados nas mesmas condições do método otimizado. Já na Figura 29 observa-se o cromatograma referente à análise dos interferentes presentes no Salivette®. Como os cromatogramas em questão são muito pouco complexos, pode-se concluir que os possíveis interferentes não apresentam potencial em influenciar as análises das amostras reais.

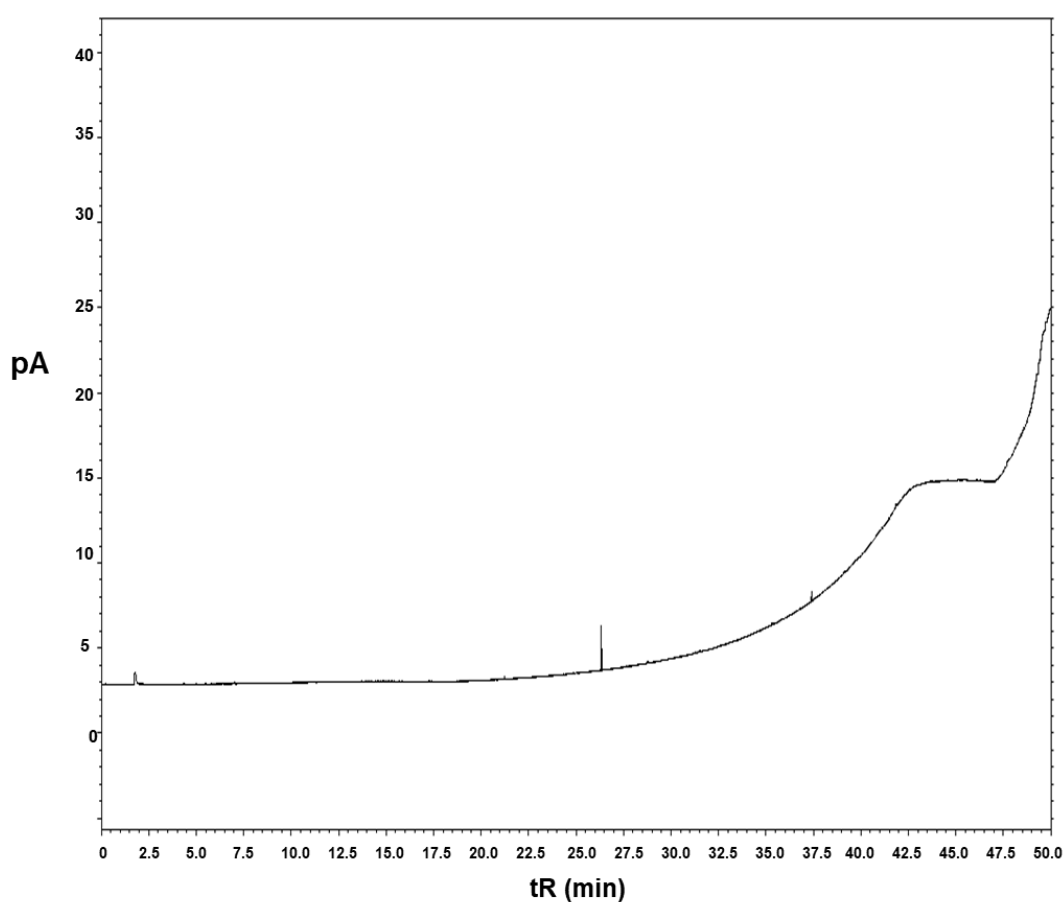


Figura 28 - Cromatograma referente à análise de 3 mL de água deionizada com adição de 1 g de NaCl e 80 μ L de ácido sulfúrico 0,1 M.

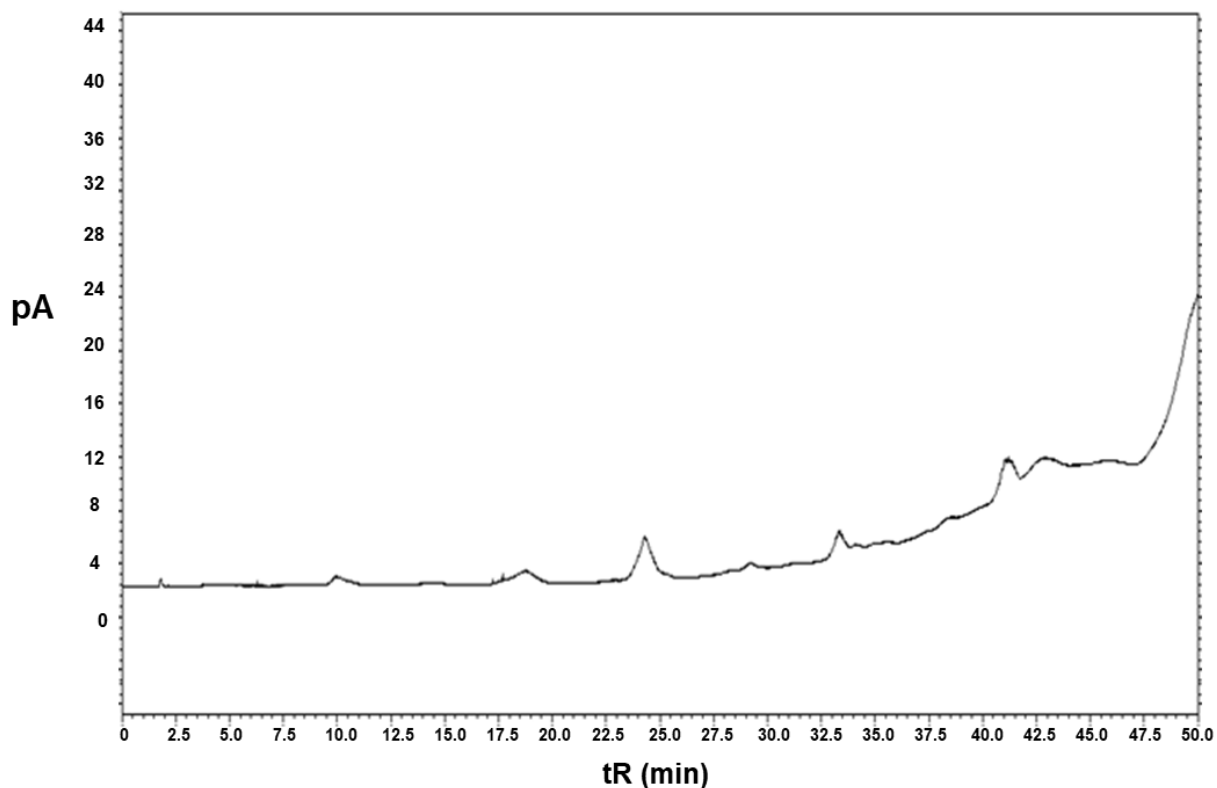


Figura 29 - Cromatograma referente à análise dos possíveis interferentes oriundos do Salivette®.

6 CONCLUSÃO

Um método para a análise de VOCs presentes no fluido oral e na urina utilizando a técnica de SHS-GC-FID foi desenvolvido e otimizado. O DOE possibilitou a otimização dos parâmetros da extração dos VOCs por SHS utilizando um número muito menor de experimentos para chegar às condições ótimas, diminuindo assim o tempo e os gastos relacionados ao processo experimental.

As condições ótimas para a extração dos VOCs urinários através da técnica de SHS foram: volume de amostra de 3 mL, temperatura de 95 °C, tempo de incubação de 20 minutos, pH ácido (entre 1-2), adição de 0,75 g de NaCl e agitação de 500 rpm. Já para o fluido oral, as condições ótimas foram: volume de amostra de 850 µL, temperatura de 95 °C, tempo de incubação de 15 minutos, pH ácido (entre 1-2), adição de 1,0 g de NaCl e agitação de 500 rpm. Essas condições foram empregadas na análise de amostras de pessoas saudáveis e pacientes oncológicos, como será discutido no próximo capítulo (Capítulo III).

**CAPÍTULO III – ANÁLISE DO PERFIL DE VOCS EM
AMOSTRAS REAIS E A POTENCIAL APLICAÇÃO DO
SHS-GC-FID NO DIAGNÓSTICO DO CÂNCER**

1 INTRODUÇÃO

No capítulo I foi discutido que uma das principais características do câncer é que esta doença leva a uma série de mudanças metabólicas, o que acarreta na formação de novos VOCs e/ou na alteração da concentração de VOCs presentes naturalmente. Além disso, também foi levantada a hipótese de que essas mudanças gerariam uma “impressão digital química”, a qual poderia ser explorada para o desenvolvimento de um método diagnóstico para o câncer através da técnica de GC-FID aliada a métodos de bioinformática.

Sendo assim, neste capítulo será apresentada uma discussão mais detalhada sobre os principais mecanismos propostos na literatura para associação entre as alterações metabólicas decorrentes do câncer e a formação de VOCs, bem como será apresentado um breve embasamento teórico sobre as ferramentas de bioinformática empregadas neste trabalho. Por fim, será discutido sobre a aplicação desses métodos estatísticos com o objetivo de discriminar amostras de pessoas saudáveis e pacientes oncológicos.

1.1 Possíveis mecanismos que dão origem aos VOCs associados ao câncer

1.1.1 Estresse oxidativo

O câncer apresenta uma grande flexibilidade em alterar diferentes vias metabólicas para se adaptar ao microambiente tumoral e também para atender a alta demanda de proliferação celular. Por esses motivos, os exatos mecanismos responsáveis pela formação dos VOCs vinculados à doença em questão ainda não estão devidamente elucidados. Contudo, as principais evidências presentes na literatura apontam que os VOCs associados ao câncer estão relacionados ao estresse oxidativo (BROZA *et al.*, 2015; MIEKISCH *et al.*, 2004; WEN *et al.*, 2020).

O estresse oxidativo representa o desequilíbrio entre a presença de espécies reativas de oxigênio (EROS, ou do termo em inglês *reactive oxygen species* – ROS) e a capacidade do sistema biológico em eliminar os intermediários reativos e/ou reparar os danos resultantes da ação dessas espécies no organismo. As ROS são moléculas ou íons derivados do O₂ que apresentam um elétron desemparelhado na camada mais externa, o que faz com que essas substâncias sejam muito instáveis e,

consequentemente, como o próprio nome já diz, muito reativas. As principais ROS incluem o íon peróxido (O_2^{2-}), o ânion superóxido ($O_2^{\cdot-}$), o peróxido de hidrogênio (H_2O_2), o radical hidroxila (OH^{\cdot}) e o oxigênio singlete (1O_2) (LIOU & STORZ, 2010). No contexto biológico, as ROS são constantemente produzidas, principalmente nas mitocôndrias, como parte do processo de respiração celular. Essas espécies são importantes para fisiologia basal, atuando principalmente como sinalizadores celulares. Todavia, quando as mesmas estão presentes em excesso, podem levar a danos irreversíveis a proteínas, lipídeos e ao DNA. Este processo pode acarretar no desenvolvimento de diversas doenças, como a doença pulmonar obstrutiva crônica, a doença de Parkinson (e outras desordens degenerativas) e, notadamente, o câncer (CHEUNG & VOUSDEN, 2022; DIAS *et al.*, 2013).

É bem estabelecido na literatura que taxas elevadas de ROS estão presentes em praticamente todos os tipos de neoplasias (LIOU & STORZ, 2010). Essas espécies estão presentes em concentrações mais elevadas em consequência do aumento da atividade metabólica e também aumento do dano mitocondrial (MONEDEIRO *et al.*, 2020). Neste contexto, o estresse oxidativo pode contribuir para o desenvolvimento do tumor de múltiplas maneiras, podendo atuar nos três principais estágios da carcinogênese. No que diz respeito à iniciação, como mencionado anteriormente, essas espécies reativas podem levar a danos irreversíveis ao DNA, acarretando no acúmulo de mutações que iniciam o desenvolvimento do câncer. Com relação à promoção, as ROS podem atuar como sinalizadores celulares que auxiliam na proliferação e sobrevivência celular, bem como na inflamação. Já na propagação, há fortes evidências na literatura que sugerem que os ROS estão vinculados ao aumento da angiogênese e também na viabilização da metástase (LIOU & STORZ, 2010; MONEDEIRO, 2018).

Em relação aos VOCs vinculados ao câncer, o principal mecanismo proposto na literatura é que o estresse oxidativo leva a um aumento das reações de peroxidação lipídica (Figura 30), processo no qual as ROS reagem com os ácidos graxos poli-insaturados (do termo em inglês *polyunsaturated fatty acids* – PUFA) presentes nas membranas celulares, levando à formação de hidrocarbonetos (como etano, pentano e dodecano), aldeídos (como pentanal, hexanal e heptanal) e, em menor extensão,

álcoois voláteis (BROZA *et al.*, 2015; WEN *et al.*, 2020). Neste sentido, vários estudos sugerem um aumento significativo da presença de hidrocarboneto e aldeídos em amostras biológicas de pacientes oncológicos em comparação com controles saudáveis (BULJUBASIC & BUCHBAUER, 2015; GUADAGNI *et al.*, 2011; HOPKINSON & SCHOFIELD 2018; JANFAZA *et al.*, 2019).

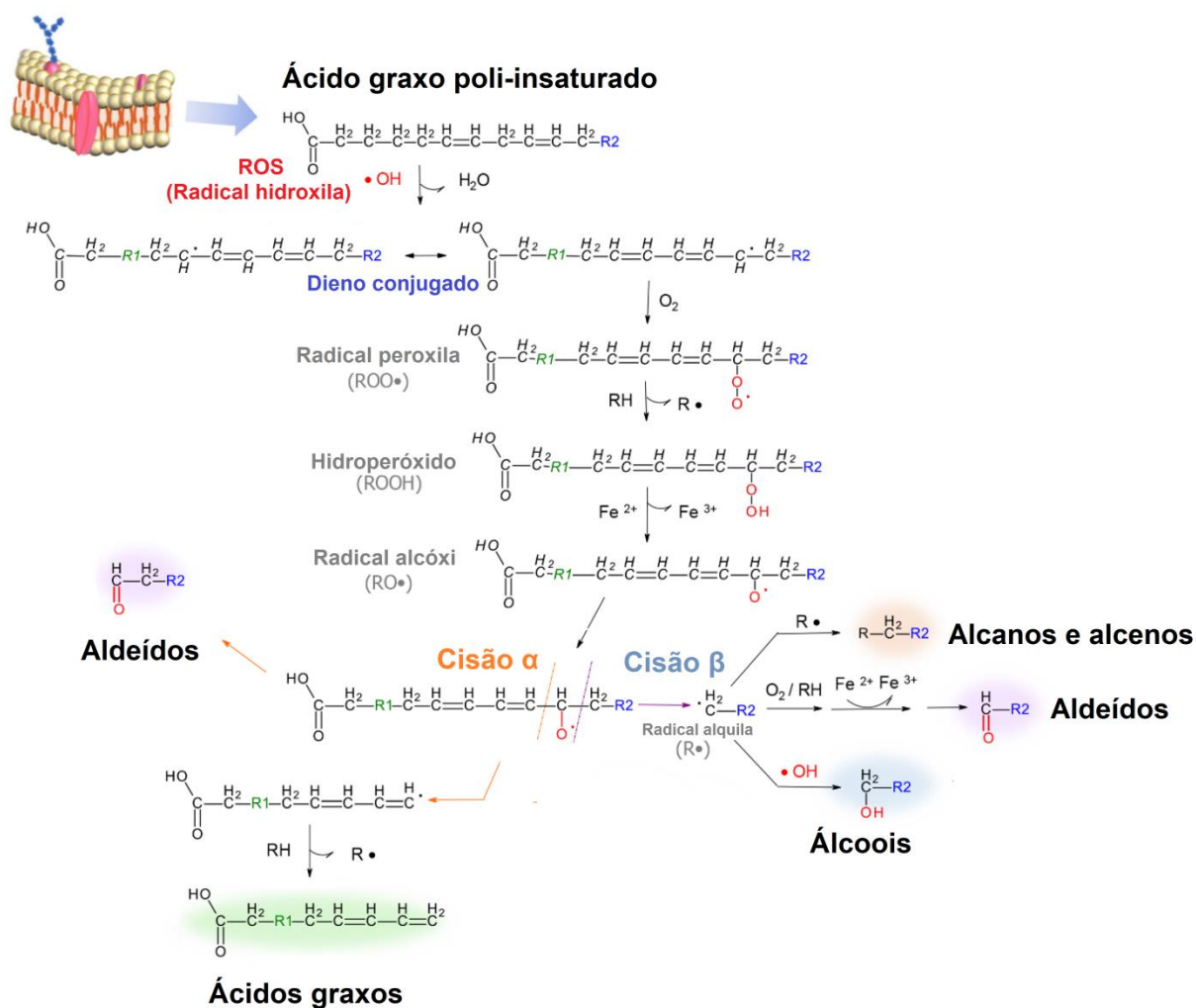


Figura 30 - Peroxidação lipídica: principal mecanismo proposto para a formação de VOCs vinculados ao câncer. Adaptado de MONEDIRO *et al.*, 2020.

Os alcanos, aldeídos e álcoois voláteis gerados a partir das reações de peroxidação lipídica podem ser biotransformados (principalmente oxidados) através de

ações enzimáticas. Esse processo também é influenciado pelo câncer, como será discutido a seguir.

1.1.2 Indução enzimática

O Citocromo P450 (CYP450) é uma superfamília de enzimas responsáveis pela biotransformação, mais precisamente pelas reações de oxidação, de xenobióticos e substâncias endógenas. A atividade dessas enzimas pode ser modulada por fatores intrínsecos e extrínsecos, podendo aumentar (indução enzimática) ou diminuir (inibição enzimática) a biotransformação dos compostos em questão (STIPP & ACCO, 2021). Um dos moduladores da ação das enzimas da CYP é justamente o câncer, o qual acarreta, geralmente, no aumento da expressão enzimática das isoformas que constituem essa superfamília. Uma das consequências dessa indução enzimática é o aumento das reações de oxidação (hidroxilação) de hidrocarbonetos voláteis, o que leva à formação de álcoois (HAKIM *et al.*, 2012).

A álcool desidrogenase (ADH), responsável pela oxidação de álcoois em aldeídos, é outro conjunto de enzimático que se apresenta, geralmente, com maior atividade nos indivíduos portadores de neoplasias com relação aos controles saudáveis (JELSKI & SZMITKOWSKI, 2008). Desta forma, como os aldeídos podem ser gerados tanto pela peroxidação lipídica decorrente do estresse oxidativo, como pela ação da ADH, é presumível que essa classe de compostos orgânicos possa apresentar VOCs candidatos a biomarcadores do câncer (FUCHS *et al.*, 2010). De fato, muitos estudos apontam que o hexanal, heptanal, nonanal e decanal podem se enquadrar nesse quesito (BOUZA *et al.*, 2020; GUADAGNI *et al.*, 2011; OENNING *et al.*, 2017).

Um detalhe interessante é que o mesmo aldeído pode estar presente em quantidades significativamente maiores em um tipo de câncer e em concentrações menores em outro, como reportado por Costa & De Martinis em revisão publicada recentemente (COSTA & DE MARTINIS 2020). Uma hipótese para essa discrepância é que, em alguns tipos de tumor, a atividade da aldeído desidrogenase (ALDH) também está mais pronunciada, fazendo com que essa classe de compostos seja mais extensivamente oxidada em ácidos carboxílicos (JANSSENS *et al.*, 2020). Devido toda essa complexidade bioquímica, ainda são necessários mais estudos para compreender

em maior profundidade como os aldeídos voláteis poderiam ser explorados para o desenvolvimento de um possível método de diagnóstico para o câncer.

O estresse oxidativo e a indução enzimática são os principais fatores associados à formação de VOCs vinculados ao câncer. Contudo, como dito anteriormente, a doença em questão apresenta uma grande flexibilidade em alterar vias metabólicas. Neste sentido, nos próximos tópicos serão discutidos outros importantes mecanismos bioquímicos afetados pelo câncer, porém que apresentam uma extensão menor de trabalhos na literatura que vinculam essas mudanças à formação de VOCs.

1.1.3 Efeito Warburg (glicólise aeróbica)

Em condições normais, as células saudáveis transformam a glicose em duas moléculas de piruvato através da glicólise, realizada no citosol, em processo que tem um saldo líquido de 2 ATPs. Em seguida, as moléculas de piruvato são totalmente metabolizadas na mitocôndria através do ciclo do Krebs e da fosforilação oxidativa, gerando 36 ATPs. Já em situações de hipóxia (baixa concentração de oxigênio), as células normais tendem a mudar sua estratégia metabólica primária: há uma diminuição da produção de ATP através da fosforilação oxidativa e um aumento do uso da glicólise seguida da fermentação láctica como fonte energética. Neste processo, além dos 2 ATPs, também é produzido o lactato. As células cancerígenas, por outro lado, apresentaram uma elevada taxa de glicólise mesmo quando o oxigênio é abundante, o que é conhecido como Efeito Warburg, ou glicólise aeróbica (POTTER *et al.*, 2016).

A realização da glicólise aeróbica parece ser contraditória do ponto de vista energético uma vez que o número de ATPs gerados é menor do que quando a fosforilação oxidativa acontece. Na realidade, ainda não está muito bem elucidado na literatura quais são as reais vantagens que o Efeito Warburg confere para as células cancerígenas. Contudo, as duas principais hipóteses são de que os metabólitos intermediários produzidos na glicólise são utilizados como “matéria prima” para a síntese de componentes celulares para as células neoplásicas, e que o aumento de lactato no microambiente tumoral torna o meio mais ácido, o que favorece a invasão tumoral e a supressão da ação de agentes anticancerígenos (VANDER *et al.*, 2009).

O número de trabalhos na literatura que demonstram uma ligação direta entre a glicólise aeróbica e formação de VOCs é limitado. Como exemplo, Silva e colaboradores (SILVA *et al.*, 2019) e Taunk e colaboradores (TAUNK *et al.*, 2018) analisaram os VOCs urinários no contexto do câncer de mama e ambos os trabalhos reportaram que o perfil de VOCs obtido indicava que tanto o metabolismo do piruvato quanto a via da glicólise pareciam estar alterados nos pacientes oncológicos em relação aos controles. Contudo, nenhum dos artigos apresentou um possível mecanismo que vinculasse esses fatores. Portanto, a relação entre o Efeito Warburg e a alteração da concentração de VOCs presentes em amostras biológicas de pacientes oncológicos é um campo que pode ser mais explorado futuramente nas pesquisas.

1.1.4 Reprogramação do metabolismo de aminoácidos

O metabolismo de aminoácidos apresenta efeitos extremamente extensos e cruciais tanto para células saudáveis quanto para células cancerígenas. Esses efeitos incluem (mas não se limitam) a: estabelecer um pool de aminoácidos como “blocos construtores”, especialmente com relação à produção de aminoácidos para a biossíntese de proteínas, conversão em glicose, lipídios e precursores de metabólitos contendo nitrogênio; fornecer suporte energético através da produção de α -cetoácidos, os quais são oxidados pelo ciclo de Krebs e fosforilação oxidativa; realizar a desintoxicação da amônia por conversão em uréia; e realizar a manutenção do status redox intracelular através da síntese da glutatona (que é o principal antioxidante celular não enzimático) (WEI *et al.*, 2021).

No contexto das neoplasias, é necessário um aporte abundante de aminoácidos para que as células cancerígenas sustentem seu impulso proliferativo e também sua grande demanda energética. Desta maneira, no câncer há uma reprogramação do metabolismo de aminoácidos para suprir essas necessidades. A consequência biológica dessa reprogramação está revisada por diversos trabalhos na literatura (LIEU *et al.*, 2020; VETTORE *et al.*, 2020; WEI *et al.*, 2021).

Todavia, no contexto da volatilômica ainda é reduzido o número de artigos que discute a relação entre a reprogramação do metabolismo de aminoácidos e a produção de VOCs vinculados ao câncer. Neste sentido, dois trabalhos realizados por Silva e

colaboradores (SILVA *et al.*, 2011, 2019) discorrem que pacientes oncológicos apresentaram maiores concentrações de metabólitos urinários voláteis que contém enxofre, como o dissulfureto de dimetilo e o metanotiol, e que isso seria decorrente de alterações do metabolismo do aminoácido metionina. De fato, os VOCs que contém enxofre parecem ter papel importante na diferenciação de amostras de pessoas saudáveis e com câncer, uma vez que muitos compostos sulfurados são apontados como possíveis biomarcadores da doença em questão (BATTY *et al.*, 2015).

Outro ponto que precisa ser levantado é que é bem estabelecido na literatura que a composição da microbiota pode ser significativamente diferente em pessoas saudáveis e indivíduos com câncer e que isso pode ser, inclusive, um dos fatores para o desenvolvimento e evolução dessa doença (GARRET, 2015). Neste sentido, as bactérias intestinais são responsáveis pelo catabolismo de uma série de aminoácidos, incluindo a metionina, o que também pode contribuir para a alteração na concentração de metabólitos sulfurados. Além da metionina, outro aminoácido metabolizado pelas bactérias intestinais e que é importante no contexto dos VOCs associados ao câncer é a tirosina. O aumento do catabolismo da tirosina leva ao aumento da concentração de p-cresol, o qual é um VOC frequentemente reportado em concentrações elevadas em amostras de urina de pacientes oncológicos (COSTA & DE MARTINIS 2020; MONEDEIRO *et al.*, 2020; PINTO *et al.*, 2021). Os mecanismos propostos para a formação de metanotiol, dissulfureto de dimetilo e p-cresol estão presentes na Figura 31.

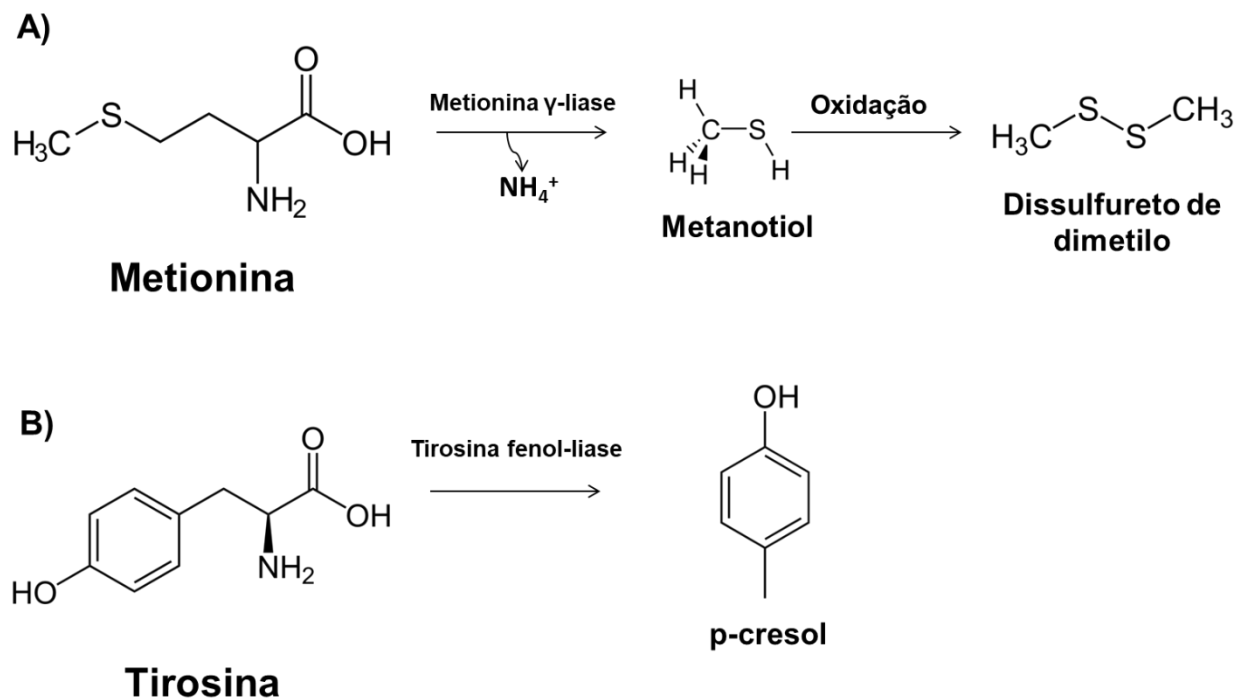


Figura 31 - Contribuição da microbiota intestinal para a formação de VOCs associados ao câncer. A) Catabolismo da metionina. B) Catabolismo da tirosina. Mecanismo proposto por MONEDEIRO et al., 2020.

1.1.5 Reprogramação do metabolismo de lipídeos

A desregulação do metabolismo lipídico está entre as alterações metabólicas mais proeminentes no câncer. As células cancerígenas aproveitam o metabolismo lipídico para a obtenção de energia, síntese de componentes das membranas biológicas, e também síntese de moléculas de sinalização necessárias para a proliferação, sobrevivência, invasão e metástase (BIAN *et al.*, 2021). Neste sentido, tanto as reações de lipogênese quanto de lipólise são mais elevadas no câncer. A lipogênese gera materiais usados como “blocos construtores” para as membranas celulares, enquanto lipólise é explorada pelas células cancerígenas como fonte energética e também como uma maneira de usar ácidos graxos exógenos para aumentar seu próprio crescimento (ZAIDI *et al.*, 2013).

Com relação aos VOCs, os principais relatos da literatura reportam que a intensificação das reações de β -oxidação (quebra de ácidos graxos) gera um aumento na concentração de corpos cetônicos. Desta forma, a acetona é um dos VOCs mais

explorados atualmente na literatura no contexto do câncer, principalmente em amostras de ar exalado. Contudo, é preciso ter cautela nessa avaliação, pois o aumento da concentração de corpos cetônicos e seus derivados pode também ter origem através de outras doenças, como a diabetes mellitus, bem como pode ser fruto da alimentação (CHATTERJEE *et al.*, 2013; OPITZ & HERBARTH, 2018; RUZSÁNYI & KALAIPOS, 2017).

1.2 Análise estatística: embasamento teórico

A metabolômica visa realizar uma análise abrangente do subconjunto de metabólitos (aqueles que podem ser extraídos/detectados) presentes em uma amostra biológica em um determinado momento. De uma maneira geral, os experimentos metabolômicos são caracterizados pela alta dimensionalidade, ou seja, o número de variáveis avaliadas por amostra excede amplamente o número de amostras avaliadas no estudo. Essa grande quantidade e complexidade de dados representa um grande desafio do ponto de vista analítico e estatístico: o tratamento manual dos resultados é praticamente impossível, fazendo com que seja necessário o uso de ferramentas de bioinformática para a extração de informações biologicamente relevantes a partir dos dados obtidos. Neste contexto, a principal ferramenta utilizada para dar sentido ao conjunto de dados metabolômicos é a estatística multivariada (WORLEY, & POWERS, 2013).

Os métodos multivariados analisam simultaneamente as variáveis que caracterizam um grupo de amostras, levando em consideração as possíveis interações entre essas variáveis presentes. No âmbito da metabolômica, esses métodos geralmente se enquadram em duas categorias: 1) métodos não supervisionados de reconhecimento de padrões, onde a avaliação da existência de similaridades entre as amostras é sugerida pelo próprio método estatístico, ou seja, nenhuma informação prévia é fornecida ao sistema; 2) métodos supervisionados, os quais se baseiam em informações a priori sobre a pertinência de cada amostra a uma classe específica. Neste caso, os dados existentes são previamente rotulados: por exemplo, é especificado que uma amostra X pertence a um paciente com câncer e uma amostra Y é de uma pessoa saudável. Posteriormente, essas informações são utilizadas de modo

que possam servir para o aprendizado do comportamento de novos dados, visando assim a classificação das amostras de acordo com suas respectivas classes (GOODACRE *et al.*, 2007).

Há uma série de métodos multivariados (tanto supervisionados quanto não supervisionados) que podem ser aplicados em estudos metabolômicos. Deste modo, os próximos tópicos apresentarão um embasamento teórico, sem entrar em detalhes matemáticos, sobre os métodos estatísticos que foram empregados neste trabalho com o objetivo de discriminar amostras de pessoas saudáveis e de pacientes com câncer.

1.2.1 Métodos não supervisionados

1.2.1.1 *Análise de Componentes Principais (PCA)*

A análise de componentes principais (do termo em inglês *principal component analysis* – PCA) é o método mais utilizado nos estudos metabolômicos para explorar inicialmente a distribuição dos dados (GOWDA & DJUKOVIC, 2014). Na PCA é realizada uma transformação linear, onde as variáveis originais são substituídas por variáveis latentes (componentes principais – PC), reduzindo assim a dimensionalidade dos dados, porém preservando ao máximo a informação original. Os novos componentes latentes são derivados dos dados originais de tal forma que a maior variância dos dados é capturada no primeiro grupo (primeiro componente, ou PC1), a segunda maior variância é capturada no PC2, e assim por diante. Desta forma, as variações mais importantes são explicadas nos primeiros PCs, enquanto o ruído experimental e as variações aleatórias estão contidos nos últimos PCs (COURANT *et al.*, 2014).

A PCA é particularmente útil como um primeiro passo na visualização de tendências, presença de *outliers* e também possíveis efeitos de batelada (influências do dia/momento da análise que não estão relacionados às variáveis biológicas ou científicas de um estudo). A PCA fornece duas ferramentas principais para análise de dados: o gráfico de *scores*, o qual representa as coordenadas das amostras no espaço dado pelos PCs, e o gráfico de *loadings*, o qual representa os pesos das variáveis originais em cada PC. Com relação à interpretação, amostras semelhantes ficam próximas uma das outras no gráfico de *scores*, já a observação do correspondente

gráfico de *loadings* permite a identificação das variáveis que são responsáveis por essas similaridades (ou diferenças) (COURANT *et al.*, 2014; WORLEY & POWERS, 2013). A Figura 32 é uma representação genérica de um gráfico de PCA (*scores*). Nesse exemplo, houve uma separação do grupo de amostras em amarelo do grupo das amostras em verde de acordo com suas respectivas similaridades.

Uma discussão mais profunda sobre os fundamentos matemáticos e estatísticos da PCA pode ser encontrada em Abdi & Williams, 2010.

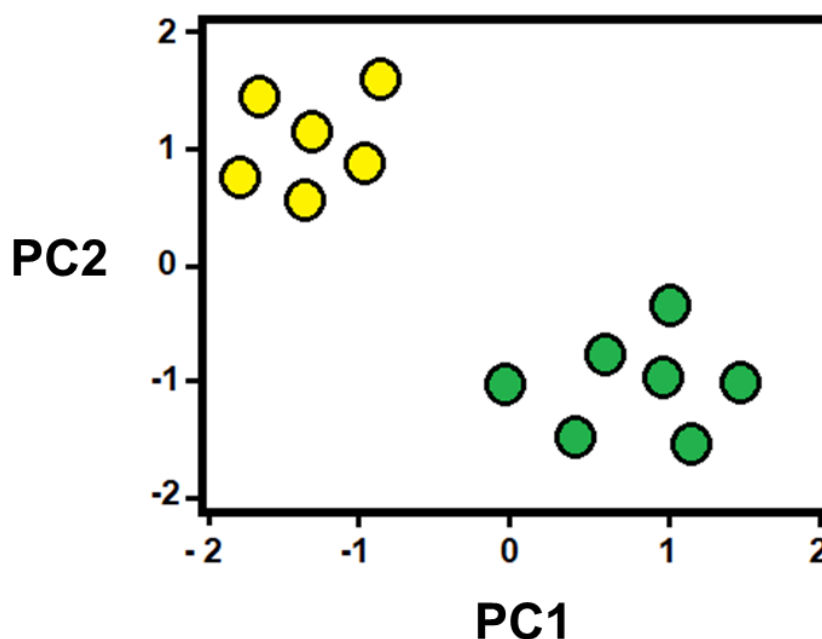


Figura 32 - Exemplo genérico de um gráfico de PCA (*scores*).

1.2.1.2 Análise por agrupamento hierárquico (HCA)

A análise por agrupamento hierárquico (do termo em inglês *hierarchical cluster analysis* – HCA) é um método multivariado que visa estruturar um conjunto complexo de dados através do agrupamento de indivíduos com características semelhantes. O HCA é normalmente representado na forma de um dendograma, ou diagrama de árvore, onde as menores distâncias entre as ramificações representam amostras com maiores similaridades, formando assim uma hierarquia de clusters (ou agrupamentos).

De maneira geral, é utilizada a distância euclidiana ou a distância Bray-Curtis como métrica para calcular o grau de similaridade entre as amostras (MARGARITIS *et al.*, 2020). Além disso, muitas vezes os gráficos de HCA são associados com gráficos de mapas de calor (*heatmap*). A Figura 33 apresenta um exemplo genérico de HCA, onde 15 amostras diferentes (nomeadas de 0 a 14) estão agrupadas de acordo com o grau de similaridade entre elas. É possível notar, por exemplo, que as amostras 1 e 14 apresentam elevada similaridade e se distanciam consideravelmente das amostras 10 e 6.

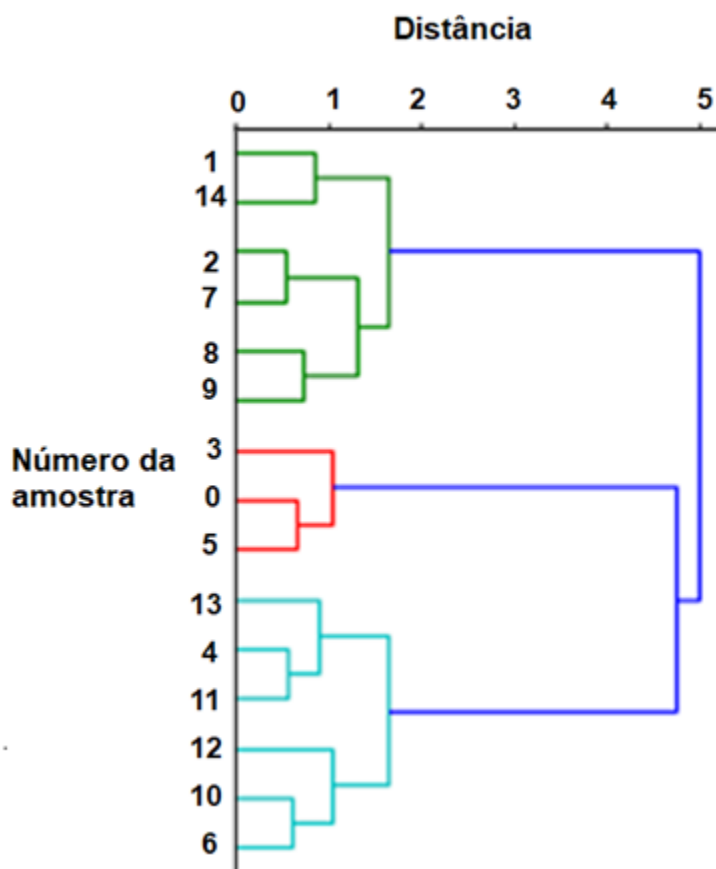


Figura 33 - Exemplo genérico de um gráfico de HCA.

Os métodos não supervisionados, como dito anteriormente, são utilizados para explorar inicialmente os dados, procurando padrões de similaridade e identificando possíveis *outliers*. Contudo, esses métodos não podem ser utilizados para a classificação e predição dos dados (REN *et al.*, 2015). Para tal, é preciso utilizar métodos supervisionados, conforme será discutido nos próximos tópicos.

1.2.2 Métodos supervisionados

1.2.2.1 *Random Forest (RF)*

O *Random Forest (RF)* é um algoritmo de aprendizagem de máquina (*machine learning*) baseado em um grande número de árvores de decisão (*decision trees*) individuais que operam como em conjunto. A árvore de decisão é um fluxograma que apresenta uma estrutura de árvore, onde cada nó denota um teste em um atributo, cada ramo representa um resultado do teste e cada nó terminal (“folha”) contém um rótulo de classe. Desta maneira, a classificação de um indivíduo (ou de uma amostra) é realizada de acordo com o conjunto de respostas obtidas pelo conjunto de perguntas formuladas (CHEN *et al.*, 2013). A Figura 34 apresenta um exemplo genérico de modelo de árvore de decisão, onde os números são classificados de acordo com a cor e se estão ou não sublinhados.

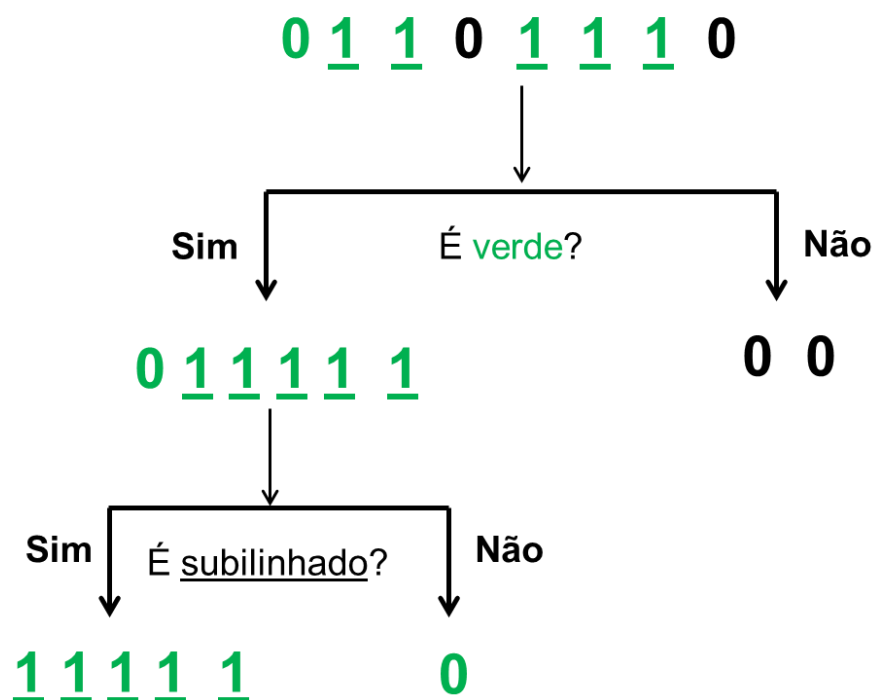


Figura 34 - Exemplo genérico de uma árvore de decisão. Os números são separados de acordo com a cor e se estão sublinhados ou não.

O RF se baseia no fato de que um grande número de árvores de decisão relativamente não correlacionadas atuando em conjunto irá apresentar um poder de predição mais robusto do que quando esses modelos atuam individualmente. Neste

sentido, a construção do RF pode ser dividida em quatro passos (CHEN & ISHWARAN, 2012):

- 1 – Seleção aleatória de amostras de um determinado conjunto de dados;
- 2 – Construção de uma árvore de decisão para cada amostra e obtenção da predição de cada árvore;
- 3 – Realização de uma votação para cada resultado predito;
- 4 – Seleção do resultado da predição mais votado como sendo o resultado final.

A Figura 35 apresenta um modelo de RF onde uma amostra é classificada entre “saudável” e “câncer” de acordo com a predição de nove árvores de decisão.

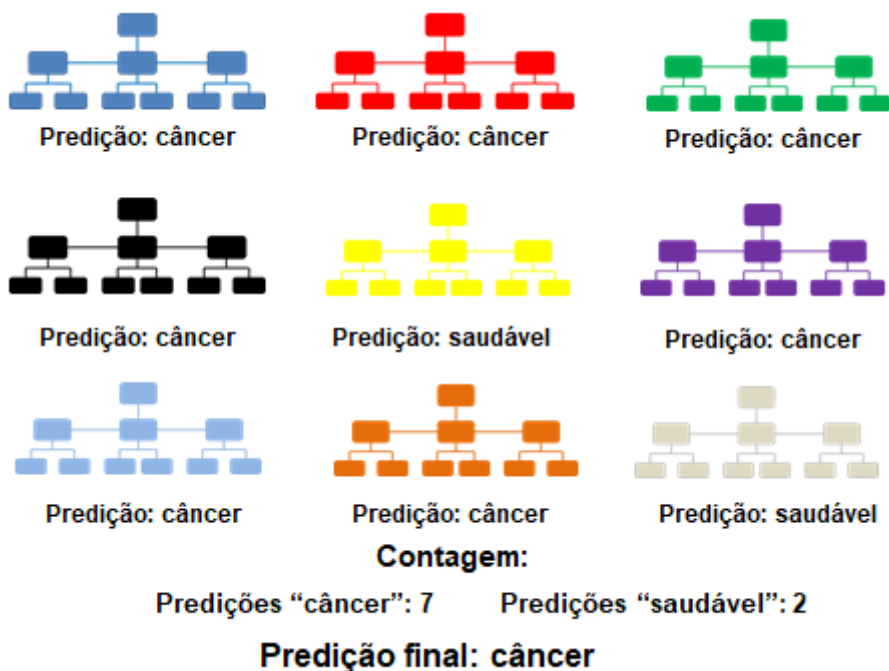


Figura 35 - Visualização de um modelo genérico *Random Forest* na classificação de uma amostra entre os grupos “saudável” e “câncer”.

1.2.2.2 *Análise discriminante por mínimos quadrados parciais ortogonais (OPLS-DA)*

Um dos métodos supervisionados mais utilizados na metabolômica é a análise discriminante por mínimos quadrados parciais (ou, do termo em inglês, *partial least squares-discriminant analysis* – PLS-DA). No PLS-DA, primeiramente a técnica de PCA

(discutida anteriormente) é aplicada para a redução da dimensionalidade do conjunto de dados. Em seguida, é construído um modelo de classificação (regressão multivariada) que busca uma relação direta entre uma resposta instrumental (por exemplo, a intensidade dos picos do cromatograma - matriz X) e uma propriedade de interesse (por exemplo, uma variável categórica que descreve a atribuição de classe da amostra, assim como é o caso de “câncer” e “saudável” - matriz Y) (SANTANA *et al.*, 2020). A análise discriminante por mínimos quadrados parciais ortogonais (do termo em inglês, *orthogonal projections to latent structures discriminant analysis* – OPLS-DA) é uma análise derivada do PLS-DA, onde é aplicado um filtro (correção de sinal ortogonal) que remove a variação da matriz X que não está correlacionada com a matriz Y (BYLESJÖ *et al.*, 2006). Consequentemente, o OPLS-DA gera modelos mais simples, o que facilita a interpretação da discriminação das amostras avaliadas e também a previsão da pertinência da classe de amostras desconhecidas (BOCCARD & RUTLEDGE, 2013).

2 OBJETIVOS

- Aplicar o método desenvolvido e otimizado (Capítulo II) na análise de amostras de urina e fluido oral coletadas de voluntários saudáveis e com câncer;
- Realizar a análise multivariada dos dados referentes aos resultados obtidos;
- Avaliar se a abordagem híbrida aumenta a capacidade do método em discriminar as amostras de pessoas saudáveis e amostras de indivíduos com câncer.

3 CASUÍSTICA E MÉTODOS

3.1 Protocolo de coleta das amostras e método experimental

Os materiais, reagentes e soluções empregados nas análises de amostras de pessoas saudáveis e pacientes com câncer estão presentes nas seções 4.1 e 4.2 do Capítulo II. Os dispositivos utilizados para a coleta das amostras de fluido oral e urina estão descritos na seção 4.3 do Capítulo II. Como dito anteriormente, a composição do fluido oral é influenciada pelo ritmo circadiano e pelos hábitos alimentares e de higiene. Desta maneira, as coletas de fluido oral tanto dos indivíduos saudáveis quanto dos pacientes oncológicos foram realizadas todas no período da manhã (entre 8:30 e 11:30

h). É preciso destacar também que foi solicitado para que o voluntário não comesse ou escovasse os dentes em um período de pelo menos uma hora antes da coleta. Além disso, foi pedido para que o indivíduo enxaguasse a boca com água cinco minutos antes do procedimento. Já a amostragem de urina foi realizada logo após a coleta de fluido oral. Logo após o procedimento de coleta de ambas as matrizes de estudo, as amostras foram identificadas, condicionadas em caixa de isopor contendo gelo e foram transferidas do Ambulatório do Hospital das Clínicas até o Laboratório de Análises Toxicológicas Forenses. As amostras foram armazenadas em refrigerador a -80°C até o momento das análises.

Por fim, o método de preparo de amostras por SHS e as condições cromatográficas da análise por GC-FID estão descritos na seção 4.4 do Capítulo II.

3.2 Análise estatística dos dados

Os cromatogramas foram obtidos através do *software* Agilent OpenLab CDS versão 04.04 da Agilent Technologies (Santa Clara, EUA). Os arquivos referentes às análises foram convertidos do formato .dat para o formato .csv (*comma-separated values*) através do *software* OpenChrom (<https://lablicate.com/platform/openchrom>) para possibilitar o tratamento dos dados. As análises estatísticas foram realizadas através do ambiente R versão 3.6.1 (R CORE TEAM, 2020).

Apesar de ser uma técnica muito robusta, o GC-FID pode apresentar algumas pequenas variações no tempo de retenção para uma mesma substância ao longo de diferentes análises. Por exemplo, em uma corrida cromatográfica um determinado composto pode eluir no tempo de retenção 13,50 minutos e, na corrida seguinte, o mesmo composto pode apresentar um tempo de retenção 13,55. Isso pode ocorrer devido à idade da coluna cromatográfica, a pequenas variações nas condições ambientais e na vazão do gás de arraste, bem como devido à interação de diferentes componentes presentes na amostra (LI & WANG, 2019). Mesmo sendo muito pequenas, essas diferenças no tempo de retenção podem fazer com que os algoritmos utilizados nos tratamento estatísticos interpretem de maneira errônea os dados obtidos (seguindo o exemplo citado acima, os picos em 13,50 e 13,55 poderiam ser interpretados como sendo oriundos de substâncias diferentes). Desta maneira, um

passo fundamental no pré-processamento dos dados metabolômicos obtidos pelo GC-FID é o alinhamento dos picos, processo que é realizado com o objetivo de corrigir esses pequenos desvios. Portanto, os picos dos cromatogramas obtidos neste presente estudo foram alinhados através do pacote de dados “MALDIquant” versão 1.21 (<https://strimmerlab.github.io/software/maldiquant/index.html>) (GIBBI & STRIMMER, 2012) . O pico do padrão interno (diidromircenol, t_R : 21,05-21,20 min) foi utilizado para a inspeção manual do alinhamento.

Antes da aplicação da estatística multivariada, é preciso também que os dados passem por um pré-tratamento para corrigir aspectos que dificultam a interpretação biológica dos resultados (VAN DEN BERG *et al.*, 2006). Neste sentido, com o objetivo de diminuir a variação relacionada ao procedimento de amostragem (por exemplo, efeitos de diluição na coleta da urina, ou da influência do ciclo circadiano no caso do fluido oral), cada amostra foi normalizada pelo valor da área total dos picos presentes em seu respectivo cromatograma (COURANT *et al.*, 2014). Além da normalização, os dados também foram autoescalados (centrados na média e divididos pelo desvio padrão). Ao escalar os dados, o peso relativo dado a cada variável, qualquer que seja sua abundância no conjunto de dados original, pode ser ajustado, a fim de evitar, por exemplo, que uma determinada substância presente em maior concentração tenha uma influência muito maior nos resultados. Isso faz com que os metabólitos se tornem igualmente importantes antes do tratamento dos dados (VAN DEN BERG *et al.*, 2006). A Figura 36 representa um fluxograma simplificado do procedimento utilizado no pré-tratamento dos dados deste estudo.

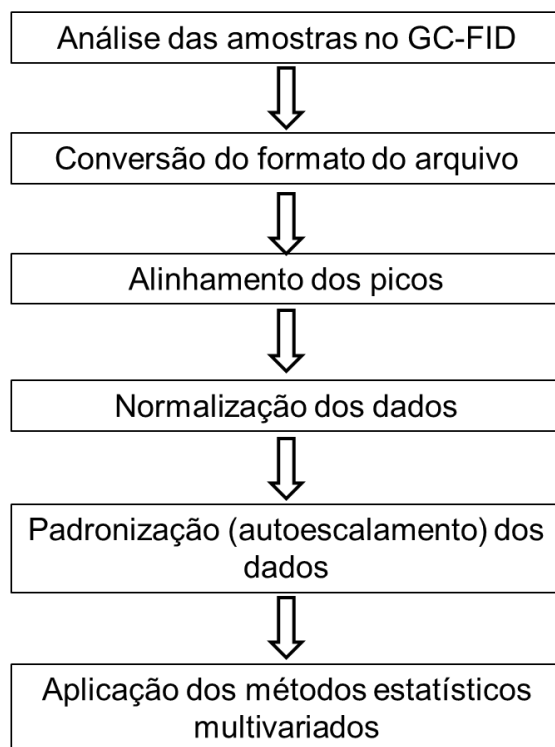


Figura 36 – Fluxograma simplificado do procedimento utilizado no pré-tratamento dos dados.

Com relação à aplicação dos métodos multivariados de análise de dados, para a construção dos gráficos de PCA e HCA foi utilizado o pacote de dados “*vegan*” (OKNASEN *et al.*, 2013) (<https://github.com/vegandevs/vegan>). Para a construção dos modelos de RF foi utilizado o pacote “*randomForest*” (LIAW & WIENER, 2002) (<https://www.stat.berkeley.edu/~breiman/RandomForests/>). Os modelos de OPLS-DA foram construídos utilizando o pacote “*ropls*” (GENTLEMAN *et al.*, 2004) (<https://www.bioconductor.org/packages/devel/bioc/vignettes/ropls/inst/doc/ropls-vignette.html>). As curvas ROC foram construídas utilizando-se o pacote “*pROC*” (ROBIN *et al.*, 2011) (<https://web.expasy.org/pROC/>). O *script* utilizado para o tratamento dos dados está disponível publicamente no link <https://github.com/computational-chemical-biology/gc-fid>.

4 RESULTADOS E DISCUSSÃO

4.1 Perfil dos voluntários da pesquisa

Os critérios de inclusão dos voluntários e o protocolo de coleta das amostras estão descritos, respectivamente, na seção 5, do Capítulo I. Com relação aos pacientes oncológicos, foram coletadas amostras de dois grandes grupos de câncer: câncer de cabeça e pescoço (CCP), e câncer gastrointestinal (CG).

O CCP corresponde a um vasto grupo de doenças oncológicas, incluindo principalmente tumores da mucosa bucal, nasal, glândulas salivares, faringe e laringe (CHOW, 2020) (Figura 37). Esse grupo de neoplasias tem origem, geralmente, nas células escamosas (células finas e planas que revestem os tratos respiratório e digestivo) das regiões citadas anteriormente, resultando assim em um carcinoma espinocelular (CEC). Normalmente, os tumores no cérebro, nos olhos, esôfago, e na glândula tireoide não são classificados dentro do grande grupo “câncer de cabeça e pescoço” (NCI, 2021). Com relação à incidência, o CCP é um dos tipos de câncer mais prevalentes, correspondendo a mais de 600 mil novos casos e 300 mil óbitos por ano (Sung *et al.*, 2021). Os principais fatores de risco para o desenvolvimento da doença em questão são: o alcoolismo, o tabagismo, e a infecção com o papilomavírus humano (HPV) (CHOW, 2020). Este grupo de neoplasias foi escolhido para ser avaliado no presente estudo principalmente pelo fato de que muitos tumores desse tipo de câncer estão próximos (ou na região) da boca e/ou das glândulas salivares, o que pode auxiliar ainda mais na diferenciação do perfil de VOCs presentes no fluido oral (TAWARE *et al.*, 2018).

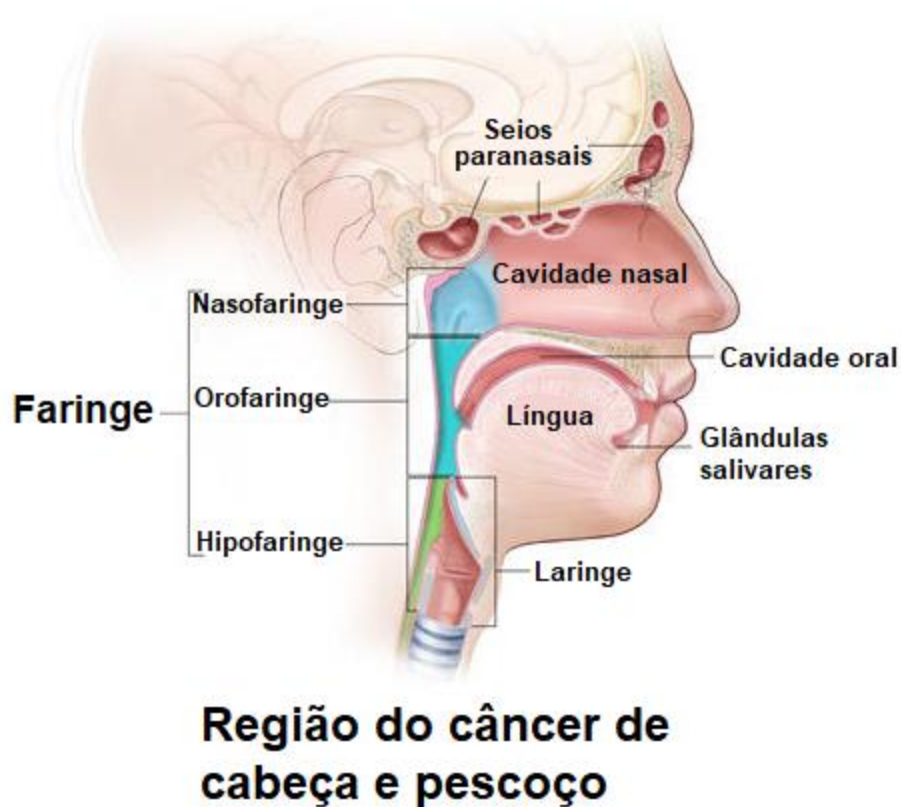


Figura 37 - Principais sítios do câncer de cabeça e pescoço. Adaptado de NCI, 2021.

O câncer gastrointestinal corresponde a um grande grupo que abrange os tumores incidentes no sistema digestório (do esôfago ao ânus). Os tipos mais comuns incluem o câncer colorretal, câncer gástrico, câncer do fígado (principalmente hepatocarcinoma), câncer esofágico e câncer pancreático (BUCKLEY *et al.*, 2020). Geralmente, os tumores desse grande grupo são classificados como adenocarcinomas, um subtipo de carcinoma onde o tumor é formado no tecido epitelial das glândulas presentes nas regiões citadas anteriormente. De maneira conjunta, essas neoplasias são responsáveis por mais de 25% dos novos casos de câncer e por mais de 30% dos óbitos devido a essa doença no mundo (Sung *et al.*, 2021). A escolha da análise desse grande grupo de neoplasias se deu justamente por essa alta incidência, o que facilitaria no número de amostras coletadas.

As informações dos voluntários diagnosticados com câncer estão presentes na Tabela 25. Os dados foram obtidos a partir do banco de dados do Hospital das Clínicas

da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo e foram referentes ao dia da coleta das amostras. A Tabela 26 resume as informações dos pacientes oncológicos levando em consideração a divisão dos dois grandes grupos avaliados (CCP e CG). As informações dos voluntários do grupo controle (“saudáveis”) se encontram na Tabela 27. Já a Tabela 28 resume o perfil geral de todos os voluntários que participaram deste estudo (neste caso, considerando os voluntários CCP e CG como sendo apenas um grupo: pacientes oncológicos).

Tabela 25 - Informações dos voluntários diagnosticados com câncer (gênero, idade, tipo de neoplasia, classificação pelo sistema TNM, estágio clínico, uso crônico de álcool, uso crônico de tabaco, outros diagnósticos, uso de medicamentos, e outras observações).

(CCP = câncer de cabeça e pescoço; CG = câncer gastrointestinal; CEC = carcinoma espinocelular; ADC = adenocarcinoma)										(Continua)
Voluntário (a)	Gênero	Idade	Neoplasia	TNM	Estágio clínico	Etilismo	Tabagismo	Outros diagnósticos	Medicamentos	Outras observações
1 CCP	Masculino	41	CEC de nasofaringe	T3N1M0	III	Sim	Sim	-	-	-
2 CCP	Masculino	59	CEC de laringe	T4N2M0	IVa	Ex-etilista	Ex-tabagista	-	-	-
3 CCP	Masculino	76	CEC de órbita	T3N0M0	III	Não	Não	-	Daforin	-
4 CCP	Masculino	53	CEC de rinofaringe	T4N3M0	IV	Ex-etilista	Ex-tabagista	Hiperplasia prostática benigna	Duomo, Enalapril, Finasterida, Tramadol, Dipirona	-
5 CCP	Masculino	70	CEC de úvula	T1N2M0	IVa	Sim	Sim	-	Tramadol	-
6 CCP	Masculino	52	CEC de hipofaringe	T4N3M0	IVb	Ex-etilista	Ex-tabagista	-	Dipirona, Amoxicilina	Sem análise de urina
7 CCP	Masculino	59	CEC supraglótida esquerda	T1N0M0	III	Ex-etilista	Sim	Depressão	Omeprazol, Diazepan, Amitriptilina, Gabapentina, Fluoxetina	-
8 CCP	Masculino	53	CEC de trígono retromolar	T4N3M0	III	Ex-etilista	Sim	-	-	-
9 CCP	Masculino	69	CEC de base de língua	TXN+M0	IVa	Ex-etilista	Sim	Diabetes, hipertensão	Losartana, Hidroclorotiazida, Glifage	-
10 CCP	Masculino	60	CEC de amígdala	T4N1M0	IVa	Ex-etilista	Sim	Hipertensão	Sinvastatina, AAS, Oxocodona	-

(CCP = câncer de cabeça e pescoço; CG = câncer gastrointestinal; CEC = carcinoma espinocelular; ADC = adenocarcinoma)

(Continuação)

Voluntário (a)	Gênero	Idade	Neoplasia	TNM	Estágio clínico	Etilismo	Tabagismo	Outros diagnósticos	Medicamentos	Outras observações
11 CCP	Feminino	46	Carcinoma Mioepitelial de glândula salivar	Sem informação	III	Não	Não	-	-	-
13 CCP	Masculino	49	CEC de rinofaringe	T3N0M0	III	Sim	Não	Diabetes, HIV positivo	Glifage, Enalapril, Coquetel anti-ídios	-
14 CCP	Feminino	59	CEC de base de língua	T3N1M0	III	Sim	Sim	-	Enalapril	-
15 CCP	Masculino	55	Carcinoma primário oculto	TXNXM1	IV	Sim	Ex-tabagista	-	Dipirona	-
16 CCP	Masculino	54	CEC de hipofaringe	T3N2M0	IVa	Ex-etilista	Ex-tabagista	-	Dipirona	-
1 CG	Masculino	85	ADC gástrico	TXN0M0	IV	Não	Sim	-	-	-
2 CG	Masculino	70	ADC pancreático	T3N0M0	Ila	Ex-etilista	Ex-tabagista	-	-	-
3 CG	Feminino	64	ADC gástrico	T2N0M0	Ila	Não	Não	Diabetes	Glifage	-
4 CG	Feminino	68	ADC de duodeno	T4N1M0	III	Não	Não	-	-	Sem análise de fluido oral
5 CG	Feminino	49	ADC gástrico	T2N1M0	III	Não	Não	Hipertensão	Losartana	-
6 CG	Feminino	78	ADC pancreático	T3N0M0	Ila	Não	Sim	Diabetes, hipertensão	Glifage, Losartana, Rivotril, Levotiroxina, Omeprazol, Propranolol	Sem análise de fluido oral
7 CG	Masculino	70	ADC esofágico	T4N1M0	IIIb	Sim	Ex-tabagista	Hipertensão	Enalapril	-
8 CG	Masculino	71	Câncer primário oculto	Sem informação	Sem informação	Ex-etilista	Ex-tabagista	-	Dipirona, Morfina	-

(CCP = câncer de cabeça e pescoço; CG = câncer gastrointestinal; CEC = carcinoma espinocelular; ADC = adenocarcinoma)

(Conclusão)

Voluntário (a)	Gênero	Idade	Neoplasia	TNM	Estágio clínico	Etilismo	Tabagismo	Outros diagnósticos	Medicamentos	Outras observações
9 CG	Masculino	57	Colangiocarcinoma extra-hepático	T4N0M0	IIIb	Ex-etilista	Ex-tabagista	Hipertensão	Monocardil, Atenolol	-
10 CG	Masculino	47	Tumor neuroecodérmico primitivo da vesícula	TXN1M1	IV	Não	Não	-	Bromoprida, Paracetamol, Codeína	-
11 CG	Feminino	59	Colangiocarcinoma intra-hepático	T4N0M1	IV	Ex-etilista	Ex-tabagista	-	Furosemida	Sem análise de urina
12 CG	Feminino	72	ADC de vias biliares	TXNXM1	IV	Ex-etilista	Ex-tabagista	Asma crônica	Lodipen, Alenia	-
13 CG	Masculino	44	CEC de esôfago	T2N0M0	III	Ex-etilista	Ex-tabagista	-	Hidroxizina	-
14 CG	Masculino	40	ADC gástrico	T4N3M0	III	Ex-etilista	Ex-tabagista	-	-	-
15 CG	Feminino	69	ADC gástrico	T4N3M0	IIIb	Ex-etilista	Não	Hipertensão	Atenolol, Predinisona	-
16 CG	Masculino	56	ADC gástrico	T2N1M0	Ila	Ex-etilista	Ex-tabagista	Hipertensão	Losartana, Xarelto, Propanolol, Amlodipina	-
17 CG	Masculino	68	ADC de pâncreas	T2N0M1	IV	Não	Não	-	-	-
18 CG	Masculino	52	CEC de esôfago	T4NXM1	III	Ex-etilista	Ex-tabagista	HIV positivo	Coquetel antiaids	Sem análise de urina
19 CG	Masculino	72	ADC gástrico	T2N0M0	Ila	Não	Ex-tabagista	-	-	-

*A amostra 12 CCP não foi analisada devido à desistência do voluntário em participar do estudo.

Tabela 26 - Resumo do perfil dos pacientes oncológicos participantes deste estudo levando em consideração a divisão dos dois grandes grupos avaliados (câncer de cabeça e pescoço - CP e câncer gastrointestinal - CG).

CCP (n = 15)			CG (n = 19)		
Gênero	Média de idade	Desvio padrão	Gênero	Média de idade	Desvio padrão
Masculino 13	57	9,3	Masculino 12	62,7	12,2
Feminino 2			Feminino 7		
*Estágio clínico					
I	-		I	-	
II	-		II	5 (26%)	
III	7 (47%)		III	8 (42%)	
IV	8 (53%)		IV	5 (26%)	
Etilismo					
Sim	5 (33,3%)		Sim	1 (5,3%)	
Ex-etilismo	8 (53,3%)		Ex-etilismo	10 (52,7%)	
Não	2 (13,3%)		Não	8 (42%)	
Tabagismo					
Sim	7 (46,7%)		Sim	2 (10,5%)	
Ex-tabagismo	5 (33,3%)		Ex-tabagismo	11 (58%)	
Não	3 (20%)		Não	6 (31,5%)	

*O estágio clínico de um dos voluntários com câncer gastrointestinal não está especificado no banco de dados.

Tabela 27 - Informações do grupo controle (gênero, idade, uso crônico de álcool, uso crônico de tabaco, diagnósticos médicos, uso de medicamentos, e outras observações).

N = amostras do grupo controle							(Continua)
Voluntário (a)	Sexo	Idade	Etilismo	Tabagismo	Diagnósticos médicos	Medicamentos	Outras observações
1 N	Masculino	53	Não	Não	Pré-diabetes	-	-
2 N	Feminino	47	Não	Não	Artrose	Duloxetina, Hidroxicloroquina, Desogestrel	-
3 N	Masculino	69	Não	Não	Hipertensão	Diovan	-
4 N	Feminino	75	Não	Não	Hipertensão	Fluxocor, Trezor	-
5 N	Feminino	70	Não	Não	Diabetes	Glifage, AS	-
6 N	Masculino	62	Sim	Sim	Hipertensão	Losartana, Propranolol, As, Fenofibrato	-
7 N	Feminino	80	Não	Não	Hipertensão	Somalgin, Rusovas	-
8 N	Masculino	61	Não	Não	Hipertensão	Benicar, Amoxicilina	-
9 N	Masculino	91	Não	Não	Hipertensão	Levotiroxina, Zodel, Apevitin, AAS, Bisoprolol, Lixiana, Sinvastatina	-
10 N	Masculino	61	Não	Não	-	-	-
11 N	Feminino	60	Não	Não	-	-	-
12 N	Masculino	58	Não	Não	-	Propranolol	-
13 N	Masculino	68	Não	Não	Diabetes, hipertensão	Trayenta, Naprix	-
14 N	Feminino	49	Não	Não	-	-	-
15 N	Feminino	58	Não	Não	-	Benicar	-
16 N	Masculino	67	Não	Não	Hipertensão	Ésio, Digidrat, Sinvastina, Alopurinol	Sem análise de fluido oral
17 N	Masculino	60	Não	Não	-	-	-
18 N	Masculino	58	Não	Não	Asma crônica	Symbicort	-
19 N	Masculino	68	Não	Não	Glaucoma	Sinvastatina, Britens, Lacrifilm	-
20 N	Masculino	63	Não	Não	Hipertensão	Micardis, Atenolol, Lipanon	-

N = amostras do grupo controle							(Conclusão)
Voluntário (a)	Sexo	Idade	Etilismo	Tabagismo	Diagnósticos médicos	Medicamentos	Outras observações
21 N	Masculino	63	Não	Não	Arritmia cardíaca	Ritmonorm, As, Aprovel, Pitavastatina	-
22 N	Masculino	51	Não	Não	-	-	-
23 N	Feminino	81	Não	Não	Hipertensão	Puran T4, Clortalidona, Ancoron, Lixiana	-
24 N	Masculino	55	Não	Não	Hipertensão	Venzer	Desconforto com o Salivette®
25 N	Masculino	66	Não	Não	Diabetes, hipertensão	Diovan, Glifage, Atenolol, Glicasida, Lipidil	-
26 N	Feminino	55	Não	Não	-	Fenoston	-
27 N	Masculino	51	Não	Ex-tabagista	-	-	-
28 N	Masculino	52	Não	Não	-	Levoid	-
29 N	Masculino	56	Sim	Não	-	-	-
30 N	Masculino	54	Não	Não	-	Finasterida	-
31 N	Masculino	66	Ex-etilista	Não	-	-	-
32 N	Masculino	52	Não	Não	Artrite reumatoide	Diovan, Humira	-
33 N	Masculino	53	Não	Não	-	-	-
34 N	Masculino	57	Não	Não	-	Losartana	-
36 N	Masculino	59	Sim	Sim	Hipertensão, artrite reumatoide	Olmecor, Natrilix, Alopurinol, Lipanon, Rosucor, Anlodipina	-
37 N	Masculino	56	Não	Não	-	Rosucor, Brintellix, Bupium	-
38 N	Masculino	65	Ex-etilista	Não	Asma crônica, esteatose hepática	Fostair Dpi, Xigduo Xr, Stanglit	-

*A amostra 35 N não foi analisada devido à desistência do voluntário em participar do estudo.

Tabela 28 - Perfil geral dos voluntários participantes deste estudo.

Grupo controle (n = 37)			Pacientes oncológicos (n = 34)		
Gênero	Média de idade (anos)	Desvio padrão	Gênero	Média de idade (anos)	Desvio padrão
Masculino	28	61,3	Masculino	25	60,2
Feminino	9		Feminino	9	
		9,6			11,2
Etilismo					
Sim	3 (8%)		Sim	6 (18%)	
Ex-etilismo	2 (5,4%)		Ex-etilismo	18 (53%)	
Não	32 (86,6%)		Não	10 (29%)	
Tabagismo					
Sim	2 (5,4%)		Sim	9 (26,5%)	
Ex-tabagismo	1 (2,6%)		Ex-tabagismo	16 (47%)	
Não	34 (92%)		Não	9 (26,5%)	

Apesar de o número de voluntários deste presente estudo não ser elevado, as Tabelas 25 a 28 reforçam alguns pontos que estão bem estabelecidos na literatura: há evidências epidemiológicas significativas que ligam o desenvolvimento do câncer ao consumo crônico de álcool e tabaco (JIANG *et al.*, 2018). Neste sentido, como discutido anteriormente, os principais fatores de risco para o desenvolvimento do CCP são, justamente, o alcoolismo e o tabagismo. Isso fica muito bem ilustrado na Tabela 26, uma vez que 86,6% (n = 13) dos voluntários do grupo CCP eram etilistas ou ex-etilistas e 80% (n = 12) eram tabagistas ou ex-tabagistas.

Os hábitos de beber e fumar em excesso estão vinculados a outros tipos de neoplasias, como o câncer de estômago, colorretal, pâncreas, fígado, entre outros. Sendo assim, ainda em referência à Tabela 26, apesar de a prevalência ser menor em comparação ao CCP, o grupo CG também apresenta uma alta incidência de indivíduos que apresentavam esses determinados costumes. Por fim, a Tabela 28 completa o raciocínio com relação às evidências expostas anteriormente: apenas 13,4% (n = 5) dos indivíduos do grupo de pessoas saudáveis eram etilistas ou ex-etilistas e 8% eram tabagistas ou ex-tabagistas (n = 3). Uma porcentagem muito menor em comparação ao grupo de pacientes oncológicos (71% eram etilistas ou ex-etilistas, e 74% tabagistas ou ex-tabagistas).

Com relação a esses resultados, foi aplicado o Teste Exato de Fisher para confirmar estatisticamente a hipótese de que, no conjunto de dados avaliado neste

presente estudo, havia a associação entre o etilismo/ex-etilismo e a formação do câncer, bem como também havia a associação entre o tabagismo/ex-tabagismo e a doença em questão. Para ambos os casos, a associação foi estatisticamente significativa no intervalo de 95% de confiança ($p = 2.74 \times 10^{-7}$ e $p = 2.17 \times 10^{-9}$, respectivamente). Portanto, esses dados reforçam as evidências de que há uma chance muito maior do indivíduo desenvolver neoplasias em idades mais avançadas (por exemplo, 60 anos, como é aproximadamente a média dos dois grupos avaliados) se o mesmo apresenta o hábito de consumir álcool e tabaco em grandes quantidades e por muito tempo.

Outra questão preocupante que foi discutida previamente, e que também pode ser visualizada nas Tabelas (mais especificamente na Tabela 26), é o fato de o câncer ser diagnosticado, normalmente, em estágios mais avançados. Por exemplo, todos os voluntários do grupo CCP apresentavam tumores que foram enquadrados nos estágios III ou IV. O atraso no diagnóstico do CCP é algo muito comum e que está reportado em diversos trabalhos na literatura (LEE *et al.*, 2011; SCHUTTE *et al.*, 2020). Esse atraso se dá principalmente devido ao fato dos primeiros sinais dessa doença não serem tão específicos e/ou alarmantes, o que faz com que os pacientes procurem ajuda médica somente quando a mesma já está avançada, levando assim a um grande impacto negativo no prognóstico do indivíduo. Tendência semelhante pode ser observada no grupo CG, onde quase 70% dos casos se enquadram nos estágios clínicos III e IV. Sendo assim, uma das limitações deste estudo é o fato de que os achados não podem ser generalizados para uma possível “detecção precoce” do câncer (o que seria considerado como o cenário “ideal”). Essa é uma limitação muito comum na maioria dos trabalhos presentes na literatura justamente devido à dificuldade em conseguir amostras que se enquadrem em estágios mais iniciais da doença em questão.

Por fim, é interessante observar que apenas um voluntário (24 N) sentiu desconforto ao utilizar o Salivette® para fazer a coleta do fluido oral. O mesmo reportou que ficou com um pouco de náusea, porém que é algo comum para ele, uma vez que sente incômodo semelhante quando vai ao dentista. Contudo, apesar do desconforto, o volume de fluido oral coletado foi suficiente para realizar as análises. Neste sentido, três voluntários (4 CG, 6 CG e 16 N) não produziram o volume de fluido oral mínimo

necessário (850 μL). O voluntário 16 N relatou que tem, de fato, um fluxo salivar mais baixo. Já no caso dos voluntários 4 CG e 6 CG, acredita-se que os fatores preponderantes foram: a diminuição da produção salivar decorrente da idade avançada e também a debilitação em decorrência da doença. Com relação às amostras de urina, não foi possível fazer a coleta de dois pacientes que se encontravam acamados, e de um voluntário que alegou motivos particulares.

4.2 Análise estatística

Conforme discutido na seção 3.2 deste capítulo, antes da aplicação dos modelos multivariados é preciso realizar um pré-tratamento dos dados, processo este que é crucial para garantir uma boa qualidade na interpretação dos resultados.

Após o alinhamento dos picos, normalmente os dados são normalizados. A normalização consiste na divisão de cada componente presente em uma mesma amostra (neste caso, as áreas dos picos presentes no cromatograma) por um fator de normalização. Diferentes fatores de normalização podem ser aplicados e a escolha deste fator pode ter grande impacto nas análises estatísticas e, conseqüentemente, na interpretação dos resultados.

De maneira geral, dois fatores de normalização são comumente utilizados: a divisão pela área do PI e a divisão pela área total dos picos presentes no cromatograma. Ambos os fatores foram testados neste presente estudo. Na avaliação da divisão pela área do padrão interno, foi observado que havia grande variação da área do PI de acordo com as diferentes corridas cromatográficas (Figura 38). Desta maneira, ao invés de agir como um fator de normalização, a divisão pela área do PI levaria à introdução de mais um fator de variabilidade entre as amostras. Além disso, a divisão pelo PI não é muito recomendada nas análises metabômicas *non-target*. Isso é devido ao fato deste tipo de normalização não conseguir representar a diversidade química dos compostos analisados, sendo capaz de “corrigir” apenas as variações indesejadas que estão relacionadas com o processo de extração e com a introdução da amostra no instrumento. Nas análises metabômicas *non-target* também é preciso levar em conta outras fontes de variação, como os fatores de diluição, algo que o uso do PI não é capaz de normalizar (KARAMAN, 2017). Sendo assim, o fator de

normalização escolhido foi a divisão pela área total do cromatograma. De fato, os modelos multivariados gerados após a normalização pela área total foram considerados de melhor qualidade (melhor capacidade de distinção entre amostras de pessoas saudáveis e com câncer).

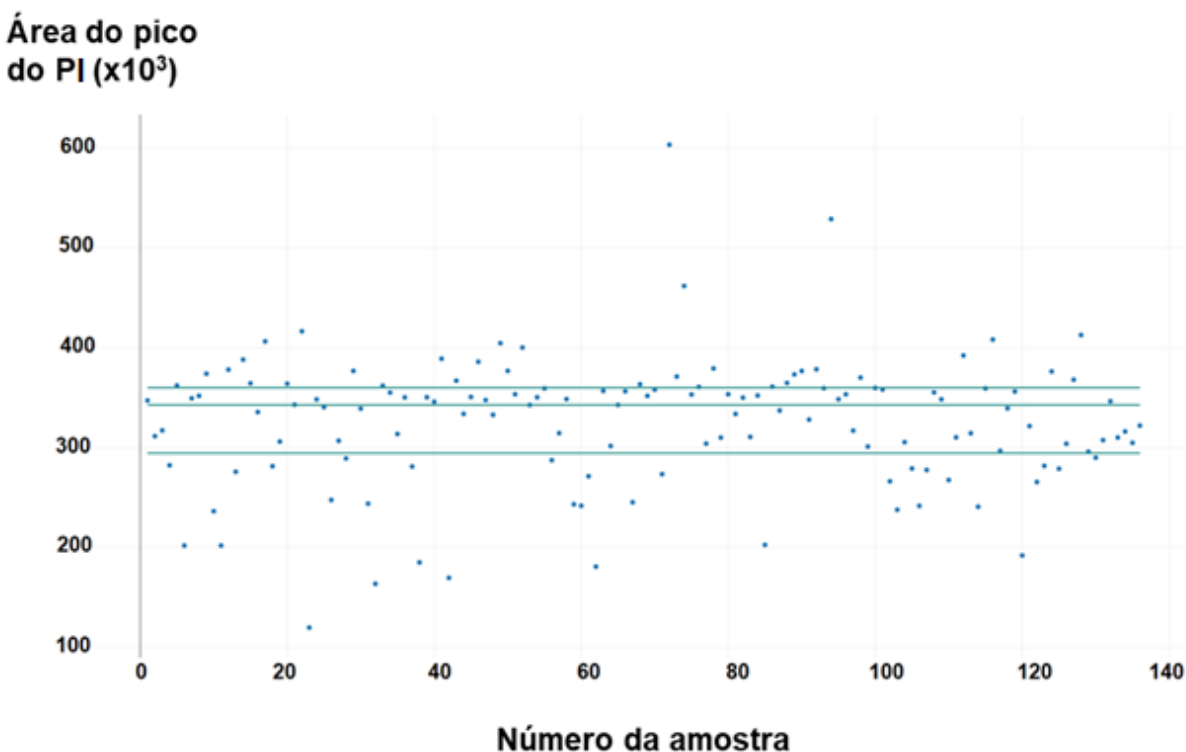


Figura 38 - Representação do valor da área do pico do padrão interno (PI - diidromircenol) nas amostras analisadas. Cada ponto significa o valor da área do PI (eixo y) em uma determinada amostra (eixo x). A linha central em azul representa a mediana, a linha superior em azul representa o percentil 75 e a linha inferior o percentil 25.

Após a normalização pela área total, os dados foram centrados na média. Centrar na média significa calcular o valor médio de uma variável e subtrair este determinado valor de cada elemento. Centralizar os dados é uma medida padrão antes da aplicação de técnicas baseadas em projeções de variáveis latentes (como é o caso da PCA e OPLS), onde o objetivo deste tratamento é focar na influência das variações de origem biológica, diminuindo a influência de variações de origem instrumental e de preparo de amostra (KARAMAN, 2017). Depois de centrados na média, os dados foram escalados com o objetivo de dimensionar todos os metabólitos de modo que os mesmos fossem igualmente importantes e com uma escala equivalente antes da

aplicação das análises multivariadas, como foi explicado anteriormente. Assim como no caso da normalização, há diferentes métodos disponíveis para escalar os dados. Os mais comuns incluem o autoescalamamento (centrar na média e dividir pelo desvio padrão), método de pareto (centrar na média e dividir pela raiz quadrada do desvio padrão), o método da escala de intervalo (centrar na média e dividir pela diferença do maior e menor número observado), e o método da escala vasta (centrar na média e dividir pelo coeficiente de variação). Seguindo a mesma linha de raciocínio da normalização, a autoescala foi escolhida pelo fato de ser o método padrão e também por gerar os modelos com melhor capacidade de classificação.

4.2.1 Análise exploratória dos dados

O primeiro método que foi aplicado para explorar inicialmente os dados foi a PCA. Para tal, foram testadas diferentes combinações: foram realizadas PCAs para verificar a distribuição dos dados referentes ao CCP e grupo controle; CG e grupo controle; e CCP, CG e grupo controle. Isso foi realizado para ambas as matrizes avaliadas (fluido oral e urina). Como exemplo, a Figura 39 mostra a PCA do grupo controle e do grupo CCP para a matriz fluido oral. É possível observar que a primeira PC explica apenas 22,2% da variância e a segunda 10,4%, o que indica que a redução de dimensionalidade conseguiu explicar apenas 32,6% da variância total dos dados. Além disso, não foi possível observar uma boa tendência de separação entre os grupos das amostras dos voluntários saudáveis e com câncer.

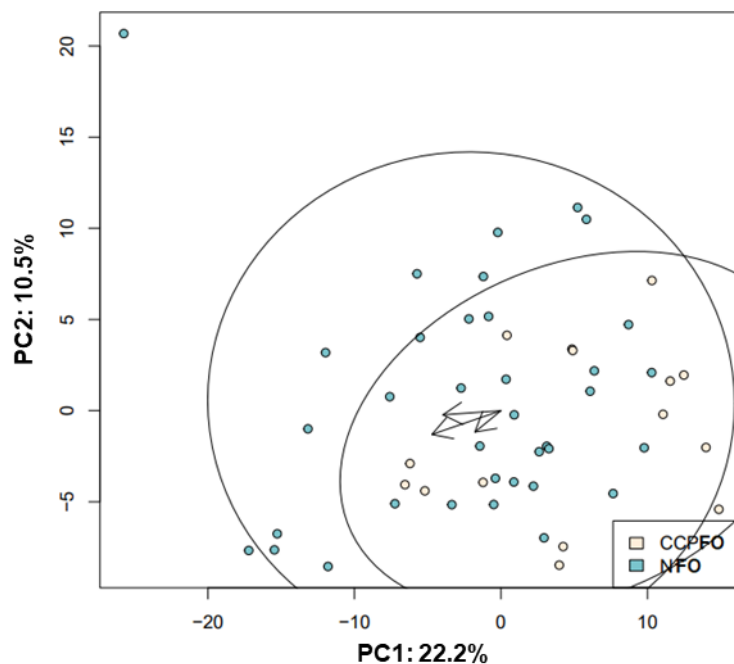


Figura 39 - PCA Biplot referente às análises de fluido oral (FO) do grupo câncer de cabeça e pescoço (CCP) e do grupo controle. Em branco estão representadas as amostras de CCP e em azul as amostras do grupo controle (N).

Todos os modelos de PCA construídos seguiram a mesma tendência: não foi possível observar separação entre o grupo controle e o grupo de pacientes oncológicos. As Figuras 39 a 44 mostram as representações gráficas dos *scores* referentes às diferentes combinações avaliadas pelas análises quimiométricas por PCA.

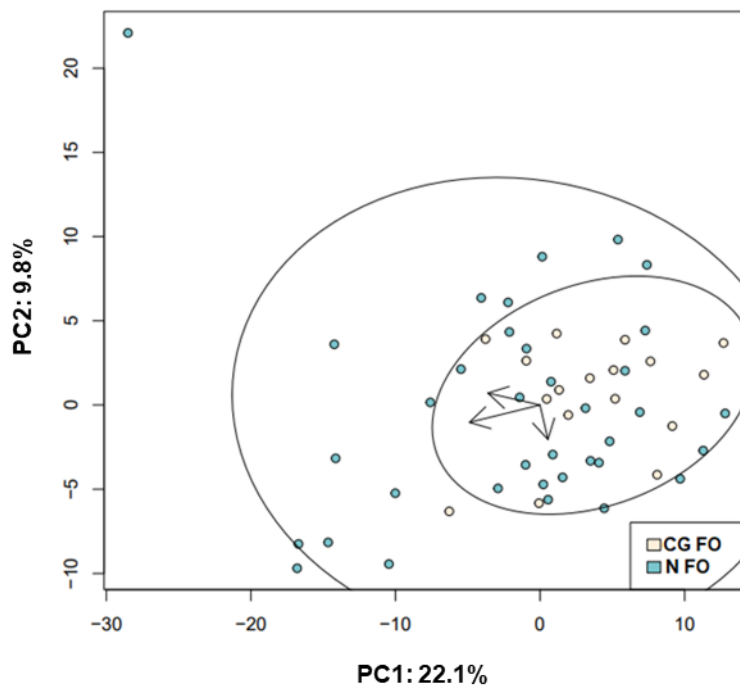


Figura 40 - PCA referente às análises de fluido oral (FO) do grupo câncer gastrointestinal (CG) e do grupo controle (N). Em branco estão representadas as amostras de CG e em azul as amostras do grupo N.

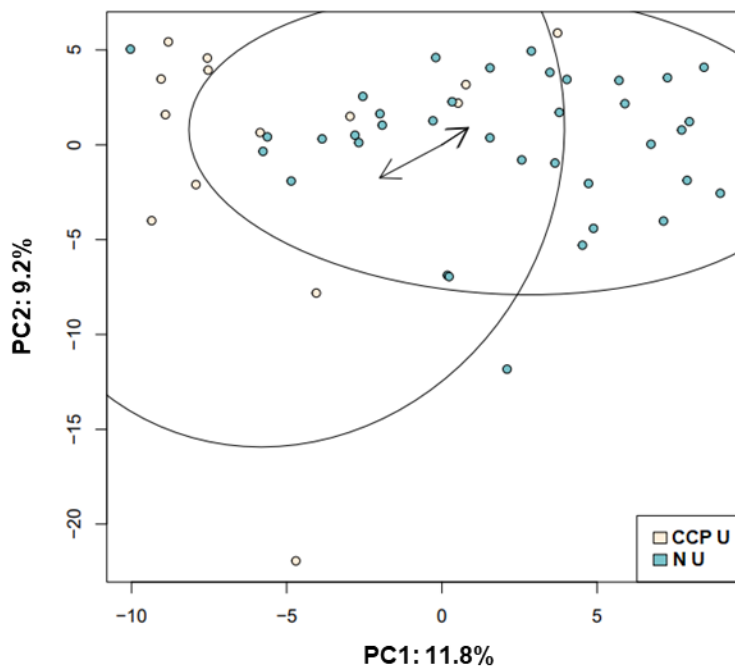


Figura 41 - PCA referente às análises de urina (U) do grupo câncer de cabeça e pescoço (CCP) e do grupo controle (N). Em branco estão representadas as amostras de CCP e em azul as amostras do grupo N.

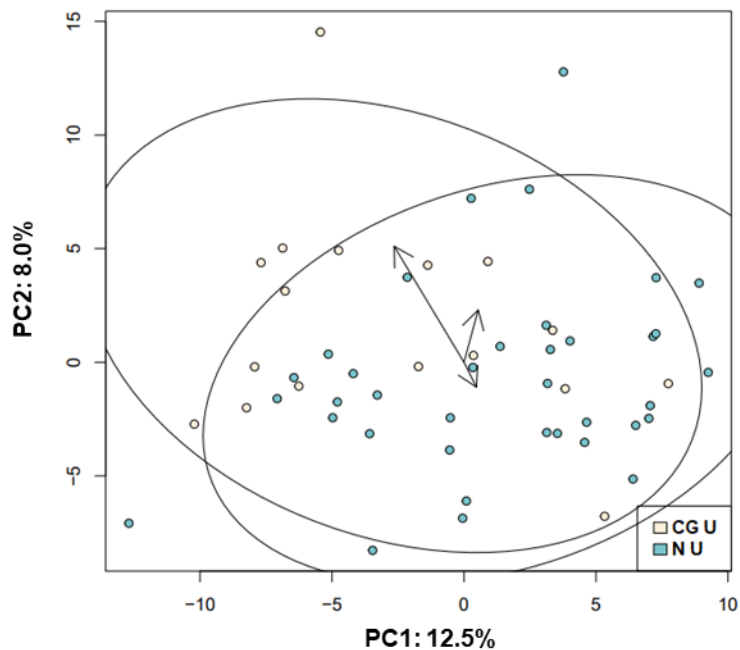


Figura 42 - PCA referente às análises de urina (U) do grupo câncer gastrointestinal (CG) e do grupo controle (N). Em branco estão representadas as amostras de CG e em azul as amostras do grupo N.

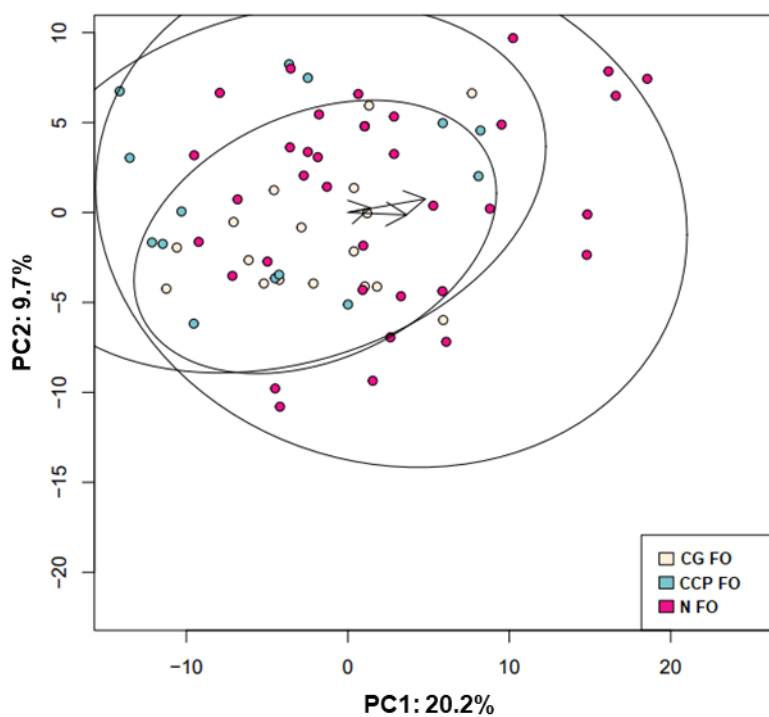


Figura 43 - PCA referente às análises de fluido oral (FO) do grupo câncer gastrointestinal (CG), câncer de cabeça e pescoço (CCP) e do grupo controle (N). Em branco estão representadas as amostras de CG, em azul as amostras do grupo CCP, e em rosa as amostras do grupo N.

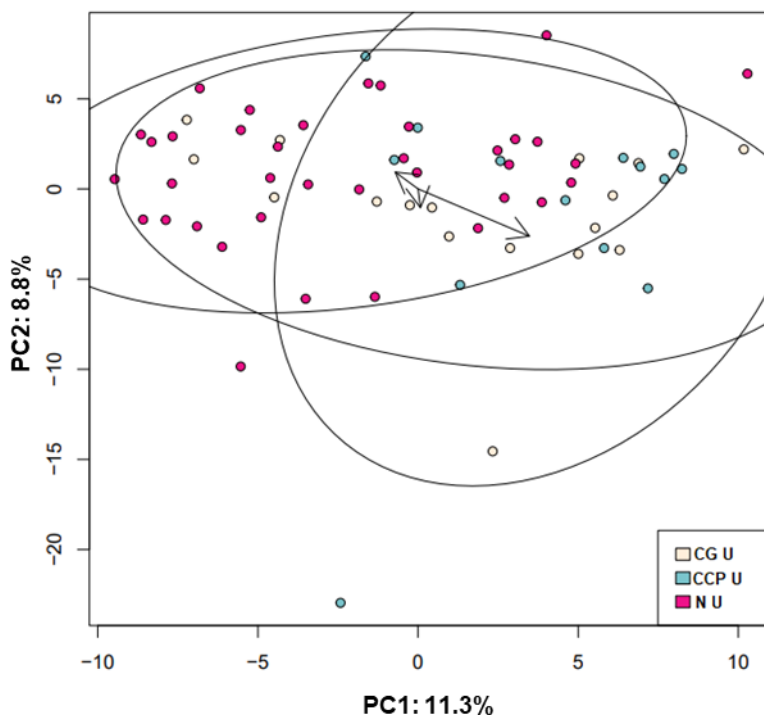


Figura 44 - PCA referente às análises de urina (U) do grupo câncer gastrointestinal (CG), câncer de cabeça e pescoço (CCP) e do grupo controle (N). Em branco estão representadas as amostras de CG, em azul as amostras do grupo CCP, e em rosa as amostras do grupo N.

Após a realização das PCA referentes às combinações discutidas anteriormente, foi realizada uma nova análise por PCA, só que neste caso envolvendo todas as amostras, incluindo as duas matrizes avaliadas. Conforme discutido no Capítulo I, é bem estabelecido na literatura que o fluido oral e a urina apresentam padrões característicos (e distintos) de VOCs. Portanto, o principal objetivo desta PCA era justamente verificar se o modelo gerado seria capaz de agrupar as amostras de acordo com a matriz avaliada. Como é possível observar na Figura 45, as amostras de urina e fluido oral ficaram bem agrupadas (como era esperado), porém não foi possível separar as amostras de acordo com suas respectivas classes. Neste sentido, a figura em questão resume bem os resultados obtidos neste estudo com relação às análises por PCA: com essa ferramenta estatística foi possível observar uma tendência de separação entre as amostras considerando as duas matrizes biológicas, porém nem o fluido oral nem a urina forneceram informações que permitiram observar uma tendência de separação entre os grupos de voluntários (isto é, CCP, CG e controle) através da aplicação deste método não supervisionada de análise.

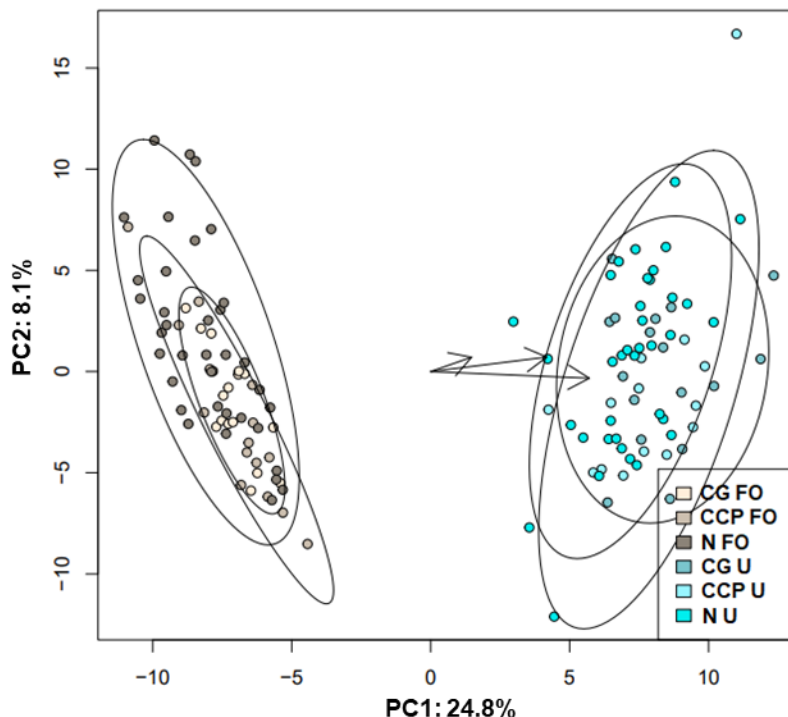


Figura 45 - PCA referente às análises de urina (U) e fluido oral (FO) do grupo câncer gastrointestinal (CG), câncer de cabeça e pescoço (CCP) e do grupo controle (N). À esquerda se encontram as amostras de fluido oral e à direita estão agrupadas as amostras de urina.

Além do PCA, a HCA foi outra ferramenta estatística não supervisionada aplicada ao conjunto de dados para verificar tendências de agrupamento entre as amostras. De maneira geral, os resultados foram muito semelhantes em relação aos obtidos nas PCAs: boa tendência de separação apenas no que diz respeito ao conjunto de amostras de matrizes biológicas distintas. A Figura 46 representa a aplicação da HCA associado ao *heatmap* com relação a todos os grupos e matrizes avaliados. É possível notar que houve o agrupamento das amostras de urina na parte superior, enquanto que as amostras de fluido oral ficaram na parte inferior (eixo y). Contudo, não houve a formação de agrupamentos (*clusters*) entre as amostras dos diferentes grupos avaliados.

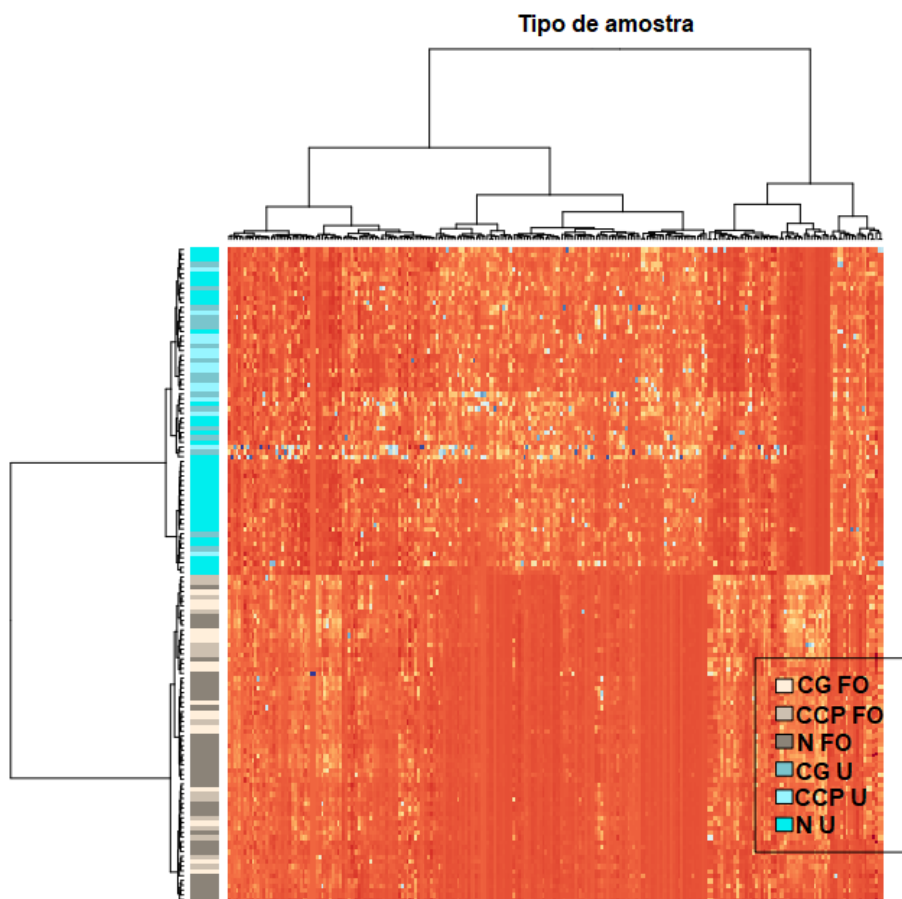


Figura 46 - HCA referente às análises de urina (U) e fluido oral (FO) do grupo câncer gastrointestinal (CG), câncer de cabeça e pescoço (CCP) e do grupo controle (N).

O objetivo deste estudo não é distinguir as amostras de acordo com as diferentes matrizes avaliadas. Contudo, a tendência de agrupamento entre as amostras de urina e de fluido oral mostra que, de fato, essas duas matrizes apresentam perfis de VOCs distintos e que isso foi captado através dos picos presentes nos cromatogramas. Se as amostras de urina e fluido oral ficassem agrupadas, isso iria significar que, muito provavelmente, as variações aleatórias de origem não-biológica (por exemplo as variações devido a extração e/ou introdução da amostra no equipamento) estariam se sobrepondo em relação as variações biológicas (oriundas das diferenças dos perfis de VOCs das matrizes avaliadas e das condições patofisiológicas dos voluntários). Conseqüentemente, seria levantada uma incerteza quanto à eficácia do método de análise aplicado neste projeto. Como isso não aconteceu, foi dado seguimento ao

tratamento dos dados, agora com a aplicação de métodos supervisionados para a construção de modelos capazes de distinguir os grupos de amostras estudados (próximas seções).

Contudo, antes de aplicar os métodos supervisionados, foi verificado se por um acaso o dia/horário da análise das amostras poderia ter tido algum impacto nos resultados (efeito batelada introduzindo vieses não biológicos, como citado anteriormente). Para tal, foi aplicado um método denominado análise de coordenadas principais (em inglês PCoA), que pode ser considerado uma generalização da PCA, permitindo a utilização de outras medidas de dissimilaridade entre as amostras, além da distância euclidiana. As Figuras 47 e 48 representam, respectivamente, a PCoA aplicado para as análises de fluido oral e para urina. A escala de cores vai do mais claro (primeiras amostras analisadas) até o mais escuro (últimas amostras analisadas). Se as cores mais claras (ou mais escuras) estivessem agrupadas, ou seguissem uma ordem, como em um gradiente, haveria uma evidência de efeito do tempo de análise (ou efeito batelada). Como as amostras não mostram agrupamento por ordem, não há uma evidência deste efeito experimental.

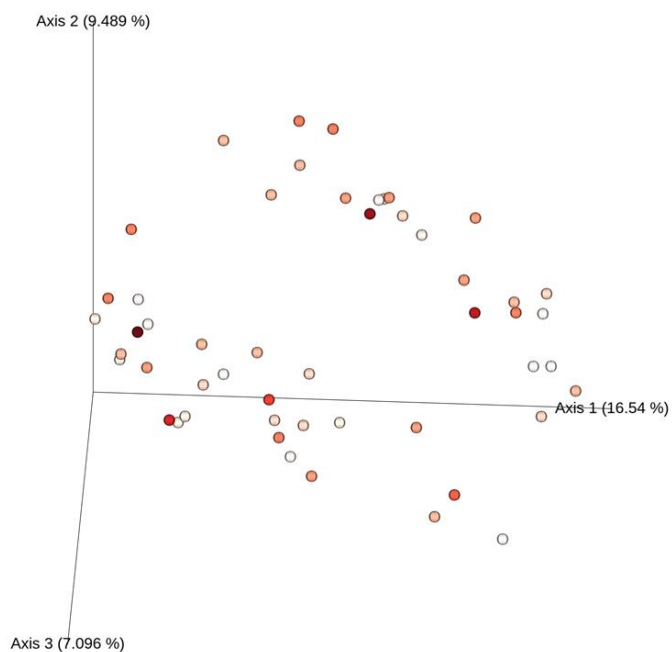


Figura 47 - PCoA referente às análises de urina. A escala de cores vai do mais claro (primeiras amostras analisadas) até o mais escuro (últimas amostras analisadas).

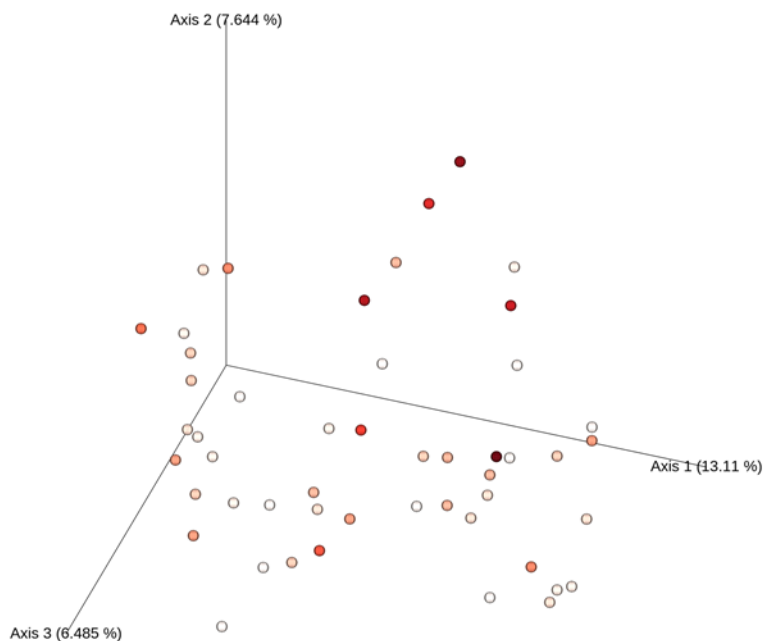


Figura 48 - PCoA referente às análises de urina. A escala de cores vai do mais claro (primeiras amostras analisadas) até o mais escuro (últimas amostras analisadas).

4.2.2 Aplicação de métodos supervisionados de classificação

O primeiro método supervisionado aplicado para a classificação das amostras analisadas neste estudo foi o RF, onde em todas as avaliações deste modelo foram construídas 5000 árvores de decisão. A capacidade de predição dos modelos gerados foi avaliada através da estimativa de erro *out-of-bag* (OOB) e também da análise da matriz de confusão.

O termo OOB refere-se às amostras que não foram utilizadas para a construção de determinada árvore de decisão. Sendo assim, a estimativa do erro OOB é importante, pois se leva em consideração dados que não necessariamente foram utilizados na construção do modelo (JANITZA & HORNUNG, 2018). Já a matriz de confusão é uma maneira de expressar quantas previsões de um classificador estavam corretas e, quando incorretas, onde o classificador errou a previsão. Esta matriz é organizada da seguinte maneira: as colunas representam os rótulos verdadeiros e as linhas indicam os rótulos previstos pelo modelo. Através da matriz de confusão é possível calcular os valores de sensibilidade, especificidade e a taxa de erro. A sensibilidade corresponde ao percentual de resultados positivos dentre os indivíduos que, de fato, apresentam a doença em questão, ou seja, é a capacidade do método detectar a doença em um indivíduo que esteja realmente doente. A especificidade é capacidade de um método dar um resultado negativo em quem realmente não possui a doença. Por fim, a taxa de erro em cada classe é calculada dividindo-se o número de erros pelo número total de amostras da determinada classe. As Tabelas 29 a 32 mostram matriz de confusão para cada uma das análises realizadas.

Tabela 29 - Matriz de confusão e estimativa do erro OOB da aplicação do modelo Random Forest referente às análises de fluido oral (FO) do grupo câncer de cabeça e pescoço (CCP) e do grupo controle.

FO = fluido oral; CCP = câncer de cabeça e pescoço			
Rótulos previstos →	CCP FO	Controle FO	Erro de classificação (%)
Rótulos verdadeiros ↓			
CCP FO (n = 15)	7	8	53,3
Controle FO (n = 36)	1	35	2,8
Sensibilidade: 46,7%			
Especificidade: 97,2%			
Estimativa do erro OOB: 17,65%			

Tabela 30 - Matriz de confusão e estimativa do erro OOB da aplicação do modelo Random Forest referente às análises de fluido oral (FO) do grupo câncer gastrointestinal (CG) e do grupo controle.

FO = fluido oral; CG = câncer gastrointestinal			
Rótulos previstos →	CG FO	Controle FO	Erro de classificação (%)
Rótulos verdadeiros ↓			
CG FO (n = 17)	8	9	52,9
Controle FO (n = 36)	2	34	5,6
Sensibilidade: 47,1%			
Especificidade: 94,4%			
Estimativa do erro OOB: 20,75%			

Tabela 31 - Matriz de confusão e estimativa do erro OOB da aplicação do modelo Random Forest referente às análises de urina (U) do grupo câncer de cabeça e pescoço (CCP) e do grupo controle.

U = urina; CCP = câncer de cabeça e pescoço			
Rótulos previstos →	CCP U	Controle U	Erro de classificação (%)
Rótulos verdadeiros ↓			
CCP U (n = 14)	7	7	50,0
Controle U (n = 37)	1	35	2,8
Sensibilidade: 50,0%			
Especificidade: 97,8%			
Estimativa do erro OOB: 15,69%			

Tabela 32 - Matriz de confusão e estimativa do erro OOB da aplicação do modelo Random Forest referente às análises de urina (U) do grupo câncer gastrointestinal (CG) e do grupo controle.

U = urina; CCP = câncer de cabeça e pescoço			
Rótulos previstos →	CG U	Controle U	Erro de classificação (%)
Rótulos verdadeiros ↓			
CG U (n = 17)	8	9	52,9
Controle U (n = 37)	2	35	5,4
Sensibilidade: 47,1%			
Especificidade: 94,6%			
Estimativa do erro OOB: 20,37%			

Usando como exemplo os resultados do modelo de RF para a classificação das amostras de CCP e controle para o fluido oral (Tabela 29), podemos notar que a estimativa do erro OOB foi de 17,65%, ou seja, o RF acertou a classificação de 82,35% das amostras que não foram utilizadas para construir o modelo. Este é um número aceitável, porém através da análise da matriz de confusão é possível perceber uma alta taxa de erro para a classificação das amostras dos pacientes oncológicos: mais da metade das amostras de CCP (53,3%) foram erroneamente classificadas como sendo de indivíduos saudáveis, correspondendo a uma sensibilidade muito baixa (46,7%). Pensando na aplicação do método desenvolvido neste presente estudo como sendo um possível método de *screening* do câncer, este seria um cenário muito desfavorável, pois o número de resultados falso-negativos seria muito grande. Por outro lado, o

modelo apresentou grande capacidade de classificação das amostras do grupo controle, uma vez que apenas uma amostra deste grupo foi erroneamente classificada como CCP (erro de 2,8%), correspondendo a uma especificidade de 97,2%. Seguindo esta linha de raciocínio, o número de resultados falso-positivos seria baixo. É interessante observar que os outros modelos de RF construídos (Tabelas 30 a 32) seguiram a mesma tendência: valores aceitáveis da estimativa de erro OOB, elevada capacidade de classificação das amostras controle, porém alta taxa de erro na classificação das amostras de câncer (baixa sensibilidade). Desta maneira, o modelo de classificação gerado através do RF não foi considerado satisfatório para o conjunto de dados apresentados neste presente estudo. Portanto, foi realizada a avaliação de outro modelo supervisionado, o OPLS-DA.

O OPLS-DA é um método supervisionado de classificação que é aplicado em diversas áreas da metabolômica uma vez que é capaz de discriminar diferentes grupos mesmo na presença de elevado ruído experimental e/ou fatores de confusão (KIM *et al.*, 2019). Um modelo de classificação deve fornecer poder de predição satisfatório para ser considerado apropriado para a seleção de biomarcadores e/ou para aplicação na área clínica. Portanto, é crucial validar o método desenvolvido através da avaliação do seu desempenho de previsão (DEJA *et al.*, 2014). Neste sentido, a avaliação da qualidade do modelo de OPLS-DA normalmente é descrita através dos valores de R^2Y e Q^2Y . O R^2Y indica quão bem o modelo se ajustou aos dados, já o Q^2Y indica quão bem o método previu os resultados de experimentos que não foram utilizados para a construção do modelo.

Um método comum para a avaliação do Q^2Y é a validação cruzada (do termo em inglês *cross-validation*). No contexto do desenvolvimento de métodos de previsão, um modelo geralmente é alimentado com um conjunto de dados conhecidos, chamado conjunto de dados de treinamento, e um conjunto de dados desconhecidos contra os quais o modelo é testado, conhecido como conjunto de dados de teste. O objetivo é ter um conjunto de dados para testar o modelo na fase de treinamento e, em seguida, fornecer informações sobre como o modelo se adapta a um conjunto de dados independentes. Uma rodada de validação cruzada compreende na divisão dos dados em subconjuntos complementares e, em seguida, na análise desses subconjuntos.

Posteriormente, a análise é validada em outros subconjuntos (conjuntos de teste). Para reduzir a variabilidade, muitas rodadas de validações cruzadas são realizadas usando divisões diferentes e, em seguida, uma média dos resultados é obtida (LIU & ÖZSU, 2009). Neste presente trabalho foi utilizado o método de validação cruzada denominado *7-fold cross-validation*, onde o conjunto de dados foi dividido em sete subconjuntos diferentes e, posteriormente, submetido à validação.

Na avaliação da qualidade da classificação, ambos os valores de R^2Y e Q^2Y variam entre 0 e 1,0. Um valor R^2Y mais alto indica um melhor ajuste do modelo, sendo considerado ideal um valor acima de 0,9. Com relação ao Q^2Y , valores dentro da faixa de 0,5 a 0,9 são considerados bons indicadores de previsibilidade, enquanto aqueles de 0,9 a 1,0 indicam excelente previsibilidade (LEE *et al.*, 2017). Além disso, é preciso ter cuidado com modelos que apresentam diferença entre R^2Y e Q^2Y acima de 0,3, pois isso é uma indicação de sobreajuste (ou, do termo em inglês, *overfitting*), condição na qual o modelo gerado se comporta de maneira adequada apenas para o conjunto de dados utilizado na sua construção, podendo não apresentar boa capacidade de predição para novos dados (KIRALJ & FERREIRA, 2009).

As Figuras 49 a 52 mostram os modelos de OPLS-DA (*scores*) gerados para as condições avaliadas em ambas as matrizes estudadas: CCP e grupo controle, e CG e grupo controle. No OPLS-DA, a direção horizontal do gráfico capta a variação entre os grupos, enquanto a direção vertical capta a variação dentro do próprio grupo. Assim como no caso da PCA, nesse tipo de análise é desejável observar o agrupamento de amostras que pertencem à mesma classe.

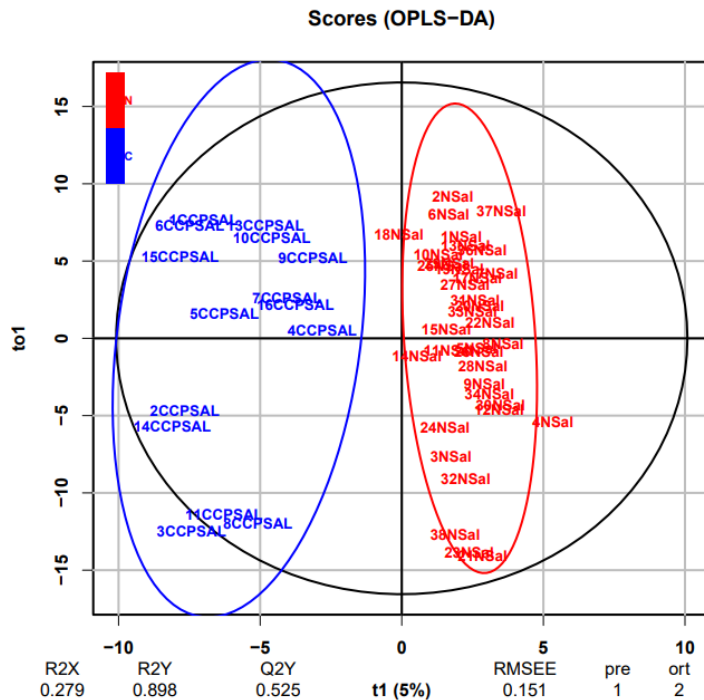


Figura 49 - OPLS-DA (gráfico de scores) referente às análises de fluido oral do grupo câncer de cabeça e pescoço (CCP) e do grupo controle (N). Em azul estão representadas as amostras de CCP e em vermelho as amostras do grupo N.

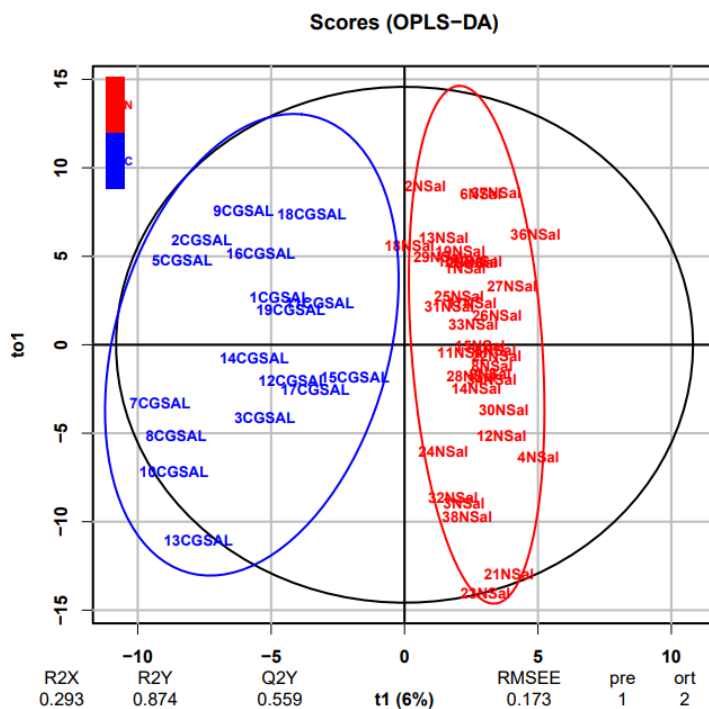


Figura 50 - OPLS-DA (gráfico de scores) referente às análises de fluido oral do grupo câncer gastrointestinal (CG) e do grupo controle (N). Em azul estão representadas as amostras de CG e em vermelho as amostras do grupo N.

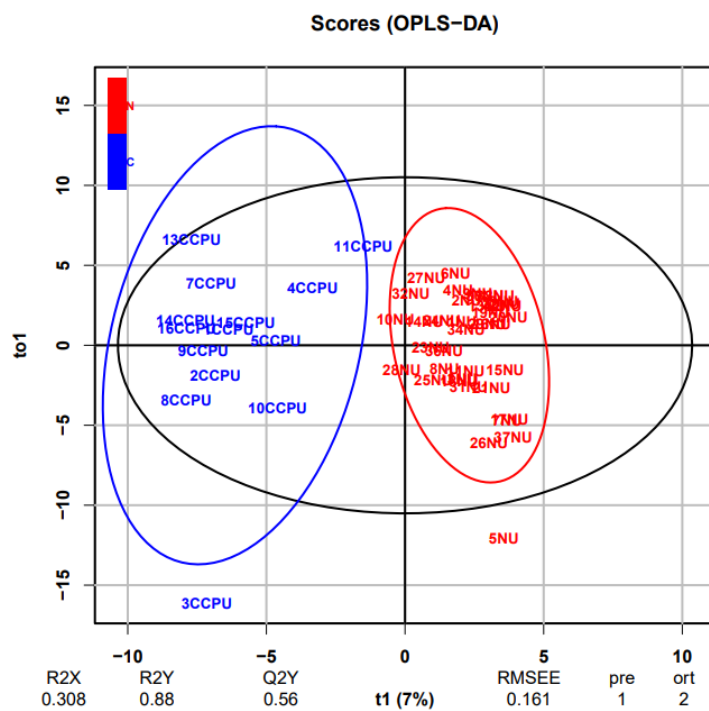


Figura 51 - OPLS-DA (gráfico de scores) referente às análises de urina do grupo câncer de cabeça e pescoço (CCP) e do grupo controle (N). Em azul estão representadas as amostras de CCP e em vermelho as amostras do grupo N.

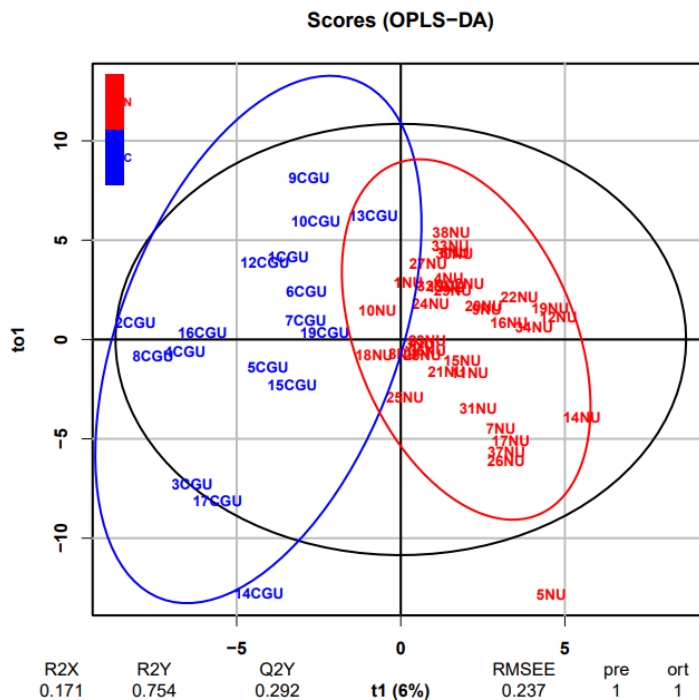


Figura 52 - OPLS-DA (gráfico de scores) referente às análises de urina do grupo câncer gastrointestinal (CG) e do grupo controle (N). Em azul estão representadas as amostras de CG e em vermelho as amostras do grupo N.

Esses resultados mostram que, exceto para análises de urina dos grupos CG e controle (Figura 52), os modelos de OPLS-DA apresentaram uma clara separação e um padrão bem agrupado para amostras dos diferentes grupos avaliados. Com relação à qualidade dos modelos gerados, as condições CCP versus grupo controle (Figuras 49 e 51) e CG versus grupo controle (apenas para a matriz fluido oral, Figura 50) apresentam tendências semelhantes: os modelos apresentaram valores de R^2Y próximos de 0,9, o que é considerado satisfatório, porém valores de Q^2Y relativamente mais baixos, ficando na faixa entre 0,525 e 0,56. Como dito anteriormente, valores de Q^2Y acima de 0,5 são considerados bons indicadores de previsibilidade, porém os três modelos citados apresentam uma diferença entre R^2Y e Q^2Y acima de 0,3, o que é um indicador de *overfitting*. Desta forma, esses três modelos não podem ser considerados como “ideais” para realizar a classificação na análise de novas amostras, porém podem ser considerados melhores que os modelos de RF apresentados anteriormente. Já com relação às análises do grupo CG versus controle (matriz urina, Figura 52), o modelo

não foi considerado adequado para a classificação dos dados uma vez que o mesmo apresenta baixos valores de R^2Y e Q^2Y (0,754 e 0,292, respectivamente).

Como a maioria dos modelos de OPLS-DA gerados apresentou uma boa tendência de agrupamento e valores aceitáveis de R^2Y e Q^2Y , o passo seguinte foi aplicar a abordagem híbrida, ou seja, realizar o processo de fusão dos dados das análises de fluido oral e urina para avaliar como os novos modelos de OPLS-DA iriam se comportar. Existem diferentes estratégias que podem ser utilizadas para fundir os dados oriundos de diferentes técnicas de análise e/ou de diferentes matrizes biológicas (BORRÀS *et al.*, 2015; SMOLINSKA *et al.*, 2014). Neste presente trabalho, os dados foram fundidos admitindo-se que as análises de fluido oral e as análises de urina poderiam apresentar compostos diferentes eluindo em um mesmo tempo de retenção. Sendo assim, para cada voluntário foi gerado um novo conjunto de dados através da junção dos dados obtidos em cada matriz biológica avaliada. Neste novo conjunto, o número total de picos avaliados correspondia à soma do número de picos presentes na análise de fluido oral com os picos presentes na análise de urina. É preciso destacar também que, para alguns voluntários, não foi possível realizar as análises das duas amostras biológicas (assim como está especificado na coluna “outras observações” das Tabelas 25 e 27). Portanto, as amostras 6 CCP, 11 CG e 18 CG (somente análise de fluido oral), e 4 CG, 11 CG e 16 N (somente análise de urina) não foram incluídas na fusão de dados. Os modelos de OPLS-DA gerados a partir da abordagem híbrida estão presentes nas Figuras 53 e 54.

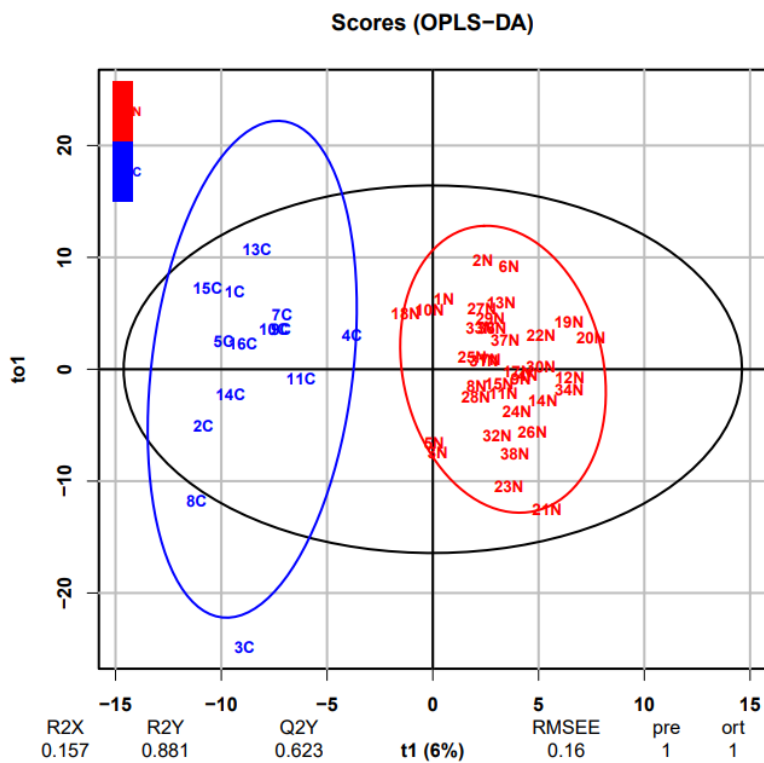


Figura 53 - OPLS-DA (gráfico de scores) referente à análise híbrida das amostras de fluido oral e urina dos voluntários do grupo câncer de cabeça e pescoço (representado em azul) e do grupo controle (representado em vermelho).

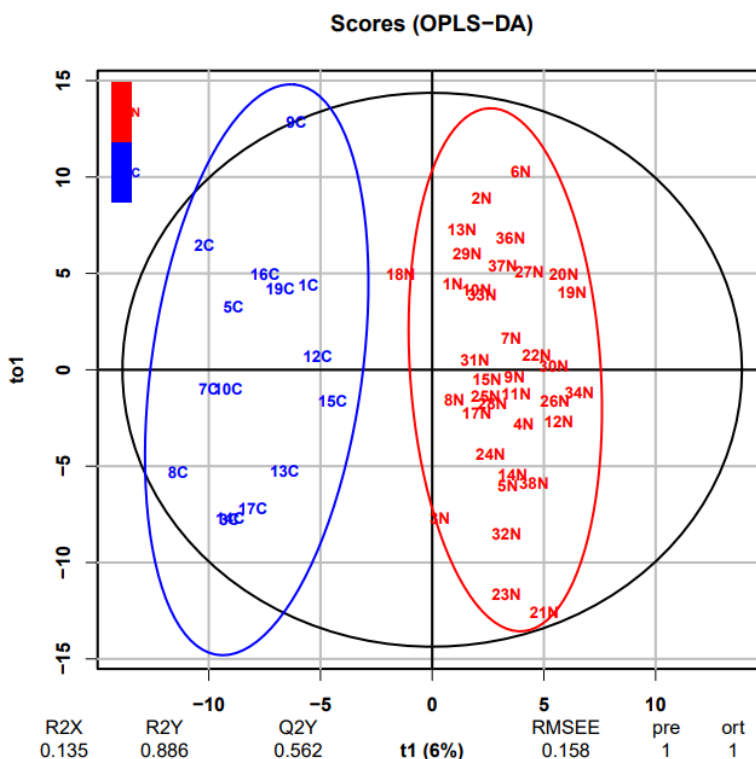


Figura 54 - OPLS-DA (gráfico de scores) referente à análise híbrida das amostras de fluido oral e urina dos voluntários do grupo câncer gastrointestinal (representado em azul) e do grupo controle (representado em vermelho).

Conforme discutido no Capítulo I, a ideia por trás da abordagem híbrida era que a junção dos dados oriundos da análise de duas matrizes biológicas de um mesmo indivíduo poderia fornecer informações químicas complementares sobre o estado (pato)fisiológico de um paciente, aumentando a capacidade de classificação do método aplicado. É interessante notar que, tanto para a análise de CCP quanto para análise de CG, os valores de Q^2Y obtidos após aplicação da abordagem híbrida foram maiores do que valores de Q^2Y obtidos nos modelos gerados quando cada matriz foi avaliada individualmente (Tabela 33). Ou seja, levando-se em conta este fator em questão, é possível concluir que a abordagem híbrida levou, de fato, a um aumento no poder de distinção dos métodos, corroborando com uma das hipóteses levantadas no início do projeto.

Tabela 33 - Valores de R^2Y e Q^2Y para todos os modelos de OPLS-DA gerados.

Modelo	R^2Y	Q^2Y
CCP fluido oral	0,898	0,525
CCP urina	0,880	0,560
CCP híbrido	0,881	0,623
CG fluido oral	0,874	0,559
CG urina	0,752	0,292
CG híbrido	0,886	0,562

Com relação às análises de CCP, a abordagem híbrida aumentou o valor de Q^2Y e ainda diminuiu a diferença entre R^2Y e Q^2Y . Sendo assim, o indicador da capacidade de predição melhorou, enquanto o indicador de *overfitting* diminuiu. Contudo, mesmo com essas melhorias, o valor de R^2Y continuou menor que 0,9 e o valor de Q^2Y se manteve relativamente abaixo. Portanto, ainda não é indicado que este modelo seja aplicado na clínica como um método de *screening* para o câncer, porém o mesmo apresenta um grande potencial de aplicação. Neste sentido, uma das maiores limitações deste presente estudo é o número reduzido de pacientes do grupo CCP. A análise de um número maior de voluntários poderia levar à construção de modelos mais robustos e com maior capacidade de classificação.

No que tange as análises de CG, a abordagem híbrida levou apenas a uma ligeira melhora no valor de Q^2Y em relação ao modelo gerado na avaliação somente do fluido oral. Isso possivelmente ocorreu devido ao fato de que o modelo de OPLS-DA referente às análises de urina do grupo CG não foi satisfatório, apresentando um Q^2Y baixo. Logo, ao realizar a fusão dos dados, as informações da matriz urina pouco contribuíram para o aumento da capacidade de predição do método. Portanto, comparando os resultados da análise híbrida para o grupo CCP e para o grupo CG, é possível inferir que a melhora do poder de classificação do modelo híbrido tende a ser significativa apenas se os modelos individuais de cada matriz biológica avaliada apresentarem valores adequados de Q^2Y .

Por fim, para um estudo mais robusto da capacidade de predição dos modelos gerados através do OPLS-DA, foi realizada a construção de curvas ROC dos modelos

avaliados. A curva ROC (do inglês *Receiver Operating Characteristic Curve*) é um gráfico que é construído plotando-se a taxa de resultados verdadeiramente positivos (sensibilidade) versus os resultados verdadeiramente negativos (especificidade) considerando vários limiares (ou, do termo em inglês *thresholds*), onde é avaliado o quanto um modelo é capaz de distinguir duas classes (o gráfico pode ser construído em função também dos resultados falsos negativos, ou $1 - \text{especificidade}$) (MARTÍNEZ-CAMBLOR *et al.*, 2016). Neste sentido, a área sob a curva (ou, do termo em inglês, *area under the curve* - AUC) é uma medida global dessa capacidade de predição. Uma AUC de 0,5 representa um teste sem capacidade discriminativa (ou seja, não melhor do que o acaso), enquanto uma AUC de 1,0 representa um teste com uma perfeita discriminação (todos os resultados positivos são verdadeiramente positivos e todos os resultados negativos são verdadeiramente negativos, ou seja, 100% de sensibilidade e especificidade) (HOO *et al.*, 2017). Mais especificamente, valores de AUC acima de 0,85 são considerados como uma “convincente evidência” da precisão de classificação, valores entre 0,75 e 0,85 representam uma “evidência parcialmente convincente”, e menos de 0,75 corresponde a “evidências não convincentes” de classificação.

Assim como na construção dos modelos de OPLS-DA, foi escolhido por realizar a construção das curvas ROC utilizando a validação cruzada (*7-fold cross-validation*, dividindo os dados em 70% para treinamento e 30% para teste) com o objetivo de obter resultados mais realistas. Sendo assim, neste caso é obtida uma média para os valores de AUC, sensibilidade e especificidade de acordo com os respectivos valores obtidos para cada uma das sete rodadas de validação. Os valores de AUC, sensibilidade, especificidade e obtidos nesta avaliação estão presentes na Tabela 34. É importante ressaltar que esses resultados estão apresentados de maneira conjunta com os dados presentes na Tabela 33, pois essas informações são complementares, como será discutido abaixo. Já as Curvas ROC obtidas nesta avaliação estão compiladas no seguinte repositório online:

https://drive.google.com/drive/folders/16mEqE_933PhgQuNRbwLo2qSIHR4pe86B?usp=sharing

Tabela 34 – Comparação dos modelos de OPLS-DA utilizando como métrica os valores de R^2Y e Q^2Y obtidos na construção do modelo, bem como os valores de área sob a curva (AUC), sensibilidade e especificidade obtidos através das curvas ROC.

Modelo	R^2Y	Q^2Y	AUC média	Sensibilidade média (%)	Especificidade média (%)
CCP fluido oral	0,898	0,525	0,850	70,2	81,5
CCP urina	0,880	0,560	0,814	84,8	82,3
CCP híbrido	0,881	0,623	0,932	75,6	88,3
CG fluido oral	0,874	0,559	0,926	78,6	87,5
CG urina	0,752	0,292	0,694	66,1	77,0
CG híbrido	0,886	0,562	0,862	69,8	87,0

A análise da Tabela 34 completa o raciocínio da discussão relacionada à Tabela 33. Conforme apresentado anteriormente, de maneira geral, quanto maior é o valor de AUC, maior tende a ser a capacidade do modelo em distinguir as duas classes avaliadas. Desta maneira, analisando os modelos de CCP é possível observar que o maior valor de AUC é obtido justamente no modelo híbrido, o que está de acordo com a avaliação da qualidade dos modelos (discutida anteriormente e ilustrada na Tabela 33), onde a capacidade de predição (valor de Q^2Y) foi maior ao realizar a fusão dos dados do que quando as matrizes biológicas foram avaliadas individualmente. Contudo, é preciso ponderar que a sensibilidade foi maior para o modelo CCP urina (84,8%) do que para o modelo híbrido (75,6%). Isso pode ser decorrente do fato da sensibilidade do modelo CCP fluido oral não ser tão elevada (70,2%); conseqüentemente, ao realizar a fusão de dados, houve na realidade uma diminuição na predição dos resultados verdadeiramente positivos em relação às análises unicamente de urina. Neste caso, uma questão delicada quanto à aplicação clínica do método poderia ser levantada: seria melhor a utilização da abordagem híbrida, já que a capacidade geral de predição do método é melhor e o modelo apresenta menor tendência de *overfitting*, ou seria melhor utilizar somente a urina, já que é mais simples coletar e analisar apenas uma matriz biológica, e o modelo apresenta maior sensibilidade (o que é desejável em métodos de *screening*)? Mais uma vez, a análise de um número maior de amostras

seria fundamental para observar se essa tendência iria se manter (ou não) e, assim, possibilitaria uma melhor tomada de decisão.

Com relação aos modelos de CG, pode ser observada uma tendência semelhante ao discutido anteriormente: o modelo construído para a análise de urina não foi satisfatório, conseqüentemente, a fusão de dados levou à construção de um modelo híbrido que apresenta valores de AUC, sensibilidade e especificidade menores em relação ao modelo que avalia unicamente fluido oral. Deste modo, para o conjunto de dados avaliados, no caso especificamente do CG seria mais vantajosa a avaliação somente de uma única matriz (fluido oral).

Em suma, levando-se em consideração tanto os valores de Q^2Y , quanto a avaliação dos valores de AUC, sensibilidade e especificidade, é possível concluir que, assim como havia sido levantada a hipótese anteriormente, a abordagem híbrida realmente tende a elevar a capacidade de predição dos métodos em relação às análises de uma única matriz biológica; contudo, para que isso aconteça é preciso que os modelos individuais apresentem métricas adequadas de qualidade.

4.2.2.1 Avaliação do score VIP

Além da avaliação da qualidade dos modelos gerados, outro aspecto interessante para ser analisado é o score VIP. O VIP (do termo em inglês *variable influence on projection*) é a medida da importância de uma variável no modelo de OPLS-DA: quanto maior o score, maior é a contribuição daquela variável na construção do modelo. Em estudos metabolômicos que visam a identificação de biomarcadores, normalmente os principais candidatos começam a ser escolhidos justamente através da avaliação do score VIP. Geralmente, um score acima de 1 é usado como critério para indicação desses possíveis marcadores.

No caso deste presente estudo, o principal objetivo era obter um *fingerprint* capaz de diferenciar os grupos avaliados através dos cromatogramas gerados nas análises por SHS-GC-FID. Desta maneira, a identificação dos compostos referentes aos picos presentes nos cromatogramas não estava no escopo deste trabalho. Apesar disso, optou-se por avaliar quais seriam as principais variáveis (ou seja, quais seriam os principais picos obtidos nas análises) responsáveis pela diferenciação dos grupos

de pacientes oncológicos e voluntários saudáveis. Essa opção foi feita uma vez que essas informações poderiam ser interessantes para futuros trabalhos que seguissem linha de pesquisa semelhante a este estudo.

Deste modo, a avaliação do score VIP foi realizada tanto para o grupo CCP quanto para o grupo CG. As Tabelas 35 e 36 contêm os cinco picos mais importantes para a construção do modelo de classificação para os grupos CCP e CG, respectivamente. A significância de cada um dos picos foi verificada através do teste de Wilcoxon (nível de confiança de 95%, onde valores de $p < 0,05$ foram considerados significativos) e também foram construídos gráficos do tipo Boxplot (Figuras 55 e 56) para observar as diferenças de comportamento desses picos em relação às amostras dos voluntários saudáveis e pacientes oncológicos.

Tabela 35 - Valores de Score VIP para os cinco picos mais importantes para a construção de modelo de classificação para o CCP.

Tempo de retenção (min)	Score VIP	Valor de p*
34,22	2,58	0,11
30,76	2,56	0,57
31,97	2,51	0,034**
36,33	2,48	0,34
25,16	2,48	0,84

*Segundo teste de Wilcoxon, intervalo de confiança de 95%.

** Valor significativo ($p < 0,05$).

Tabela 36 - Valores de Score VIP para os cinco picos mais importantes para a construção de modelo de classificação para o CG.

Tempo de retenção (min)	Score VIP	Valor de p*
31,97	2,62	0,035**
29,49	2,60	0,035**
26,65	2,56	0,52
34,22	2,54	0,045**
30,76	2,53	0,088

*Segundo teste de Wilcoxon, intervalo de confiança de 95%.

** Valores significativos ($p < 0,05$).

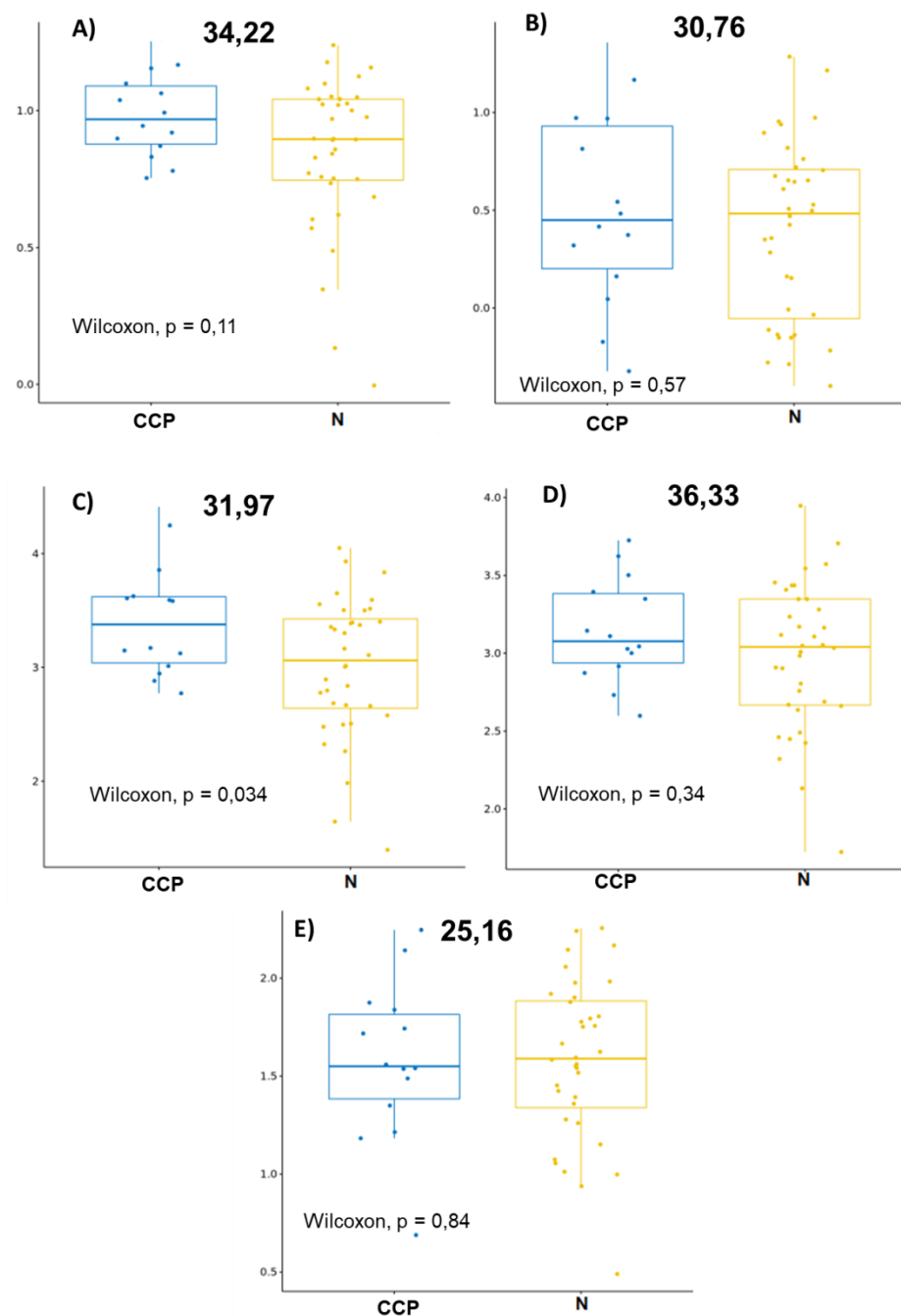


Figura 55 - Boxplots referentes aos cinco picos mais importantes para a construção do modelo de classificação para o câncer de cabeça e pescoço (CCP) segundo o score VIP. O grupo CCP está representado em azul e o grupo de voluntários saudáveis (N) está representado em amarelo. A) se refere ao pico de tempo de retenção 34,22 minutos; B) se refere ao pico 30,76; C) se refere ao pico 31,97; D) se refere ao pico 36,33; E) se refere ao pico 25,16.

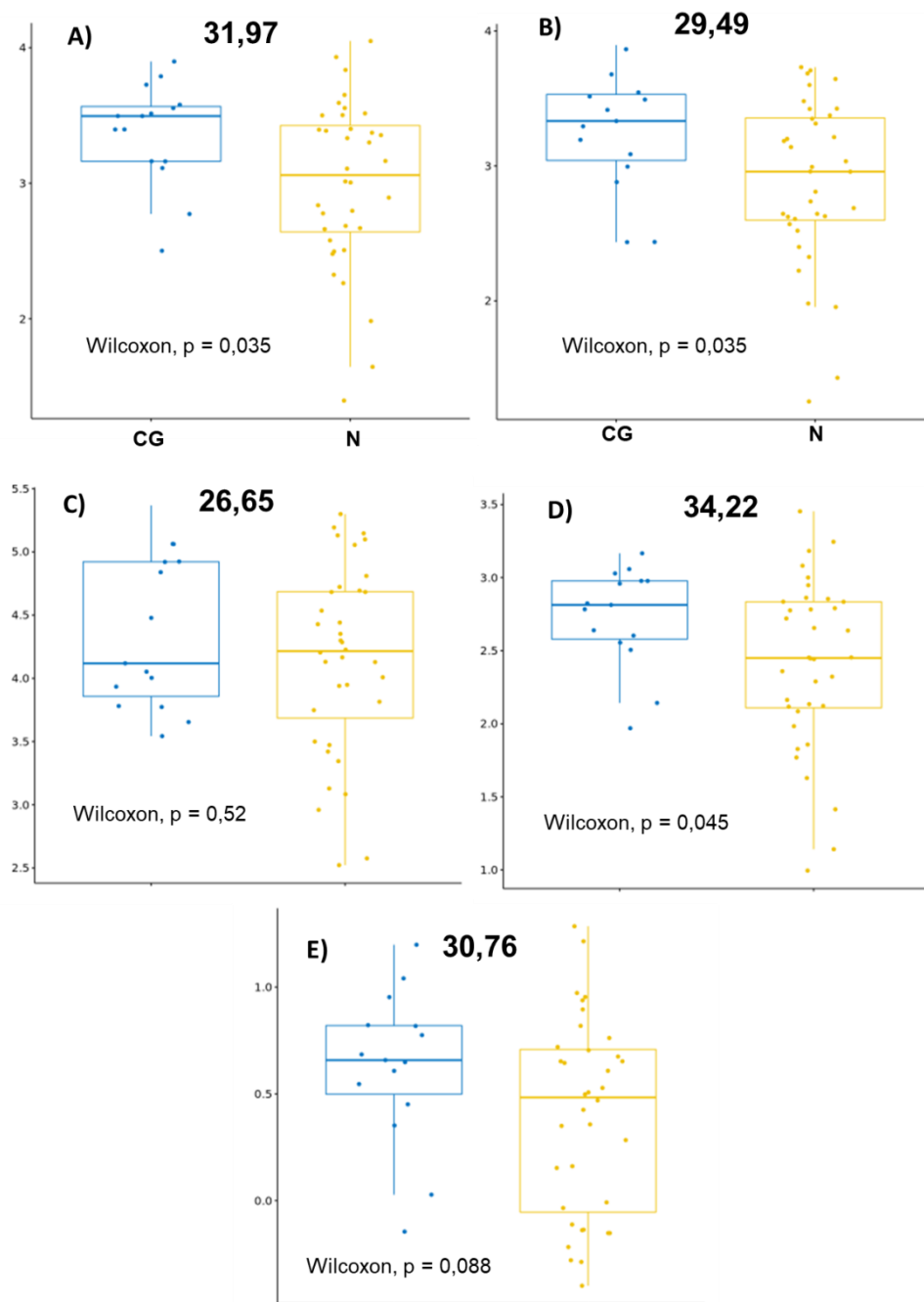


Figura 56 - Boxplots referentes aos cinco picos mais importantes para a construção do modelo de classificação para o câncer gastrointestinal (CG) segundo o score VIP. O grupo CG está representado em azul e o grupo de voluntários saudáveis (N) está representado em amarelo. A) se refere ao pico de tempo de retenção 31,97 minutos; B) se refere ao pico 29,49; C) se refere ao pico 26,65; D) se refere ao pico 34,22; E) se refere ao pico 30,76.

A análise do score VIP revela algumas informações interessantes. No caso do CCP, apenas um pico dos considerados mais importantes segundo o VIP apresentou valores de área estatisticamente significativos em comparação ao grupo controle (tempo de retenção 31,97). Este fato mostra que, acima de um único possível “biomarcador”, o mais importante para a diferenciação dos grupos estudados pode ser a análise de todos com os compostos em conjunto. O modelo do CG apresenta tendência semelhante: dos cinco picos mais importantes segundo o VIP, apenas três foram estatisticamente significativos (tempo de retenção 31,97; 29,49 e 34,22).

Outro detalhe importante é que dois picos estiveram entre os cinco primeiros no score VIP tanto do grupo CCP quanto do grupo CG: os picos 31,97 e 34,22. Neste sentido, o pico 31,97 é o mais interessante para ser analisado em futuros trabalhos uma vez que apresenta valores de área estatisticamente maiores nos dois grupos de neoplasias em comparação ao grupo controle (Figuras 55 C e 56 A). Por fim, foi avaliada a integridade deste pico de interesse (Figuras 57). Ou seja, foi verificado se o pico 31,97 correspondia, de fato, a um pico cromatográfico, ou se o modelo havia interpretado erroneamente um ruído como sendo um pico relevante.

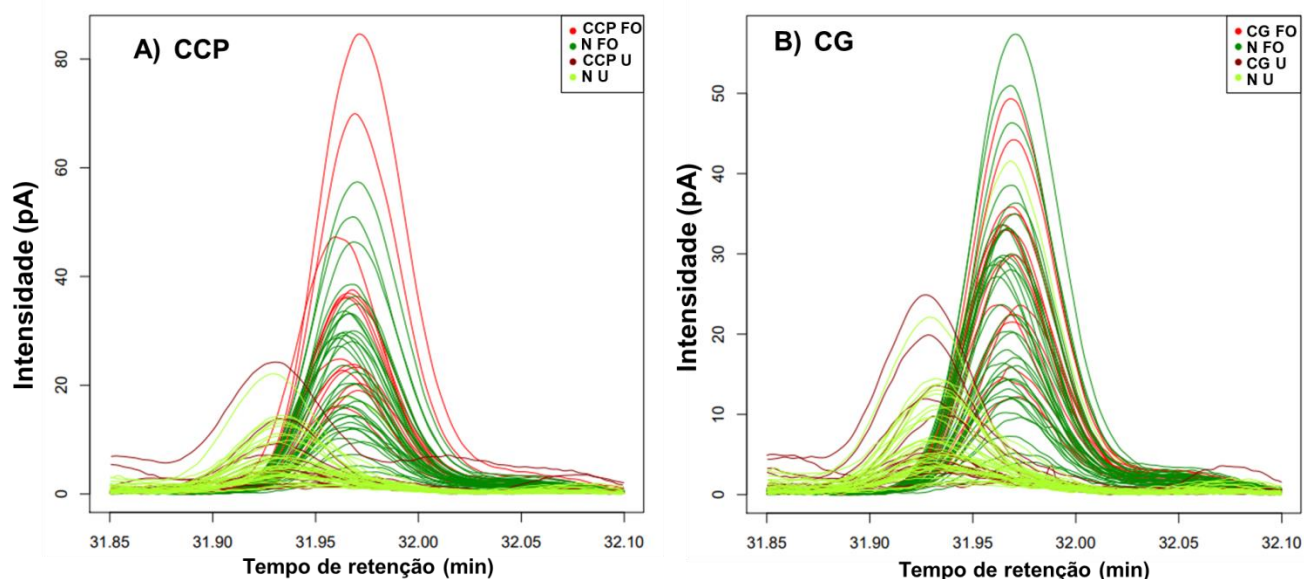


Figura 57 - Sobreposição dos picos de tempo de retenção 31,97 min. A) Comparação da intensidade deste pico nas amostras de câncer de cabeça e pescoço (CCP) e voluntários saudáveis (N) nas duas matrizes avaliadas: fluido oral (FO) e urina (U). B) Comparação da intensidade deste pico nas amostras de câncer gastrointestinal (CG) e voluntários saudáveis.

Através da análise da sobreposição dos picos de tempo de retenção 31,97 minutos nas amostras avaliadas é possível perceber que este se trata realmente de um pico cromatográfico relevante. Se fosse um ruído experimental, a intensidade seria muito pequena (a intensidade mínima observada é de 10 pA) e também a variação na intensidade seria muito menor. Inclusive, na parte A da Figura 57 é possível observar visualmente uma tendência que já havia sido ilustrada no tratamento estatístico: a intensidade do pico 31,97 tende a ser maior nas amostras do grupo CCP. Tal tendência não é observada visualmente no grupo CG (Figura 58 B).

4.3 Comparação com estudos semelhantes presentes na literatura

Conforme apresentado no Capítulo I, este é o primeiro estudo que aplica a técnica de GC-FID no contexto do desenvolvimento de um método de *fingerprint* capaz de distinguir amostras de pessoas saudáveis e pacientes oncológicos, bem como também é o primeiro estudo que avalia a capacidade de predição de modelos híbridos no contexto da volatilômica do câncer. Contudo, é possível comparar o desempenho do método desenvolvido neste trabalho com outros estudos similares presentes na literatura (alguns já foram apresentados no Capítulo I), principalmente aqueles que analisam VOCs no contexto do CCP e/ou CG, que analisam VOCs nas mesmas matrizes avaliadas neste trabalho, e/ou que empregam métodos de *fingerprint* com objetivos semelhantes a este presente estudo.

Neste sentido, dois trabalhos recentes de Taware e colaboradores se enquadram muito bem nesses quesitos. No primeiro trabalho, Taware e colaboradores (TAWARE *et al.*, 2017) avaliaram os VOCs urinários no contexto do CCP através da extração por SPME e análise por GC-MS. Ao todo, 29 amostras de pacientes com CCP e 31 amostras voluntários saudáveis foram analisadas e o método desenvolvido foi capaz de diferenciar os grupos com 80 a 90% de sensibilidade e especificidade. Já no segundo estudo, Taware e colaboradores (TAWARE *et al.*, 2018) utilizaram uma abordagem muito semelhante para avaliar os VOCs presentes no fluido oral também no contexto do CCP. Ao todo, 32 amostras de pacientes com CCP e 27 amostras voluntários saudáveis foram analisadas e método supervisionado de classificação utilizado foi o OPLS-DA (assim como neste presente estudo), onde o modelo gerado

apresentou valor de R^2Y de 0,941 e Q^2Y de 0,867. Interessantemente, ambos os estudos destacam que um dos principais responsáveis pelas diferenças nos perfis de VOCs é a reprogramação do metabolismo da glicose que ocorre nos indivíduos portadores de câncer, assim como foi discutido na seção 1.1.3 deste Capítulo.

Comparando-se com este presente estudo, os dois trabalhos citados acima apresentaram resultados relativamente mais favoráveis com relação à capacidade de predição. Este fato pode ser devido ao número de voluntários avaliados e também à técnica de extração. Com relação ao N amostral, este presente trabalho contou com 15 voluntários do grupo CCP, enquanto os outros dois tiveram praticamente o dobro. Como discutido anteriormente, um número maior de voluntários poderia ter levado à construção de modelos mais robustos. No que concerne à técnica de extração, neste presente trabalho foi utilizado o *headspace* estático, enquanto que nos outros dois foi utilizada a técnica de SPME. Como discutido na seção 2.2 do Capítulo II, a SPME é uma técnica de extração e pré-concentração que permite uma maior detectabilidade dos analitos, proporcionando a detecção de compostos em menores concentrações. Considerando que no âmbito da metabolômica muitas vezes a diferenciação entre dois grupos é devido à avaliação de compostos presentes em concentrações muito baixas, essa maior detectabilidade proporcionada pela SPME pode ter sido um fator chave no elevado poder de classificação dos modelos construídos em ambos os estudos descritos anteriormente.

Em relação a trabalhos que visam estabelecer um *fingerprint* dos VOCs relacionados ao câncer sem necessariamente identificar esses compostos, há outros dois estudos relativamente recentes, e de um mesmo grupo de pesquisas, que avaliam os voláteis presentes no fluido oral e na urina. Antón e colaboradores (ANTÓN *et al.*, 2016) desenvolveram um método de *fingerprint* para diferenciar amostras de pessoas com câncer (foram avaliados diferentes tipos de neoplasias) e amostras de indivíduos saudáveis através do perfil de VOCs presentes no fluido oral. Diferentemente da técnica aplicada neste presente trabalho, no método desenvolvido por Antón e parceiros, o *fingerprint* não era baseado em diferentes perfis cromatográficos. Pelo contrário, o método desenvolvido era não separativo: o *fingerprint* era baseado nos diferentes espectros obtidos através do MS. Mais especificamente, os voláteis eram

extraídos através de *headspace* dinâmico e, em seguida, eram injetados diretamente no MS utilizando injetor com programação de temperatura de vaporização (ou, do termo em inglês *programmable temperature vaporizer* – PTV). Com relação ao tratamento estatístico, foram utilizados três métodos de classificação distintos para avaliar 23 amostras de pessoas com câncer e 32 amostras de pessoas saudáveis. O método KNN (método dos vizinhos mais próximos, ou *K nearest neighbor*) forneceu classificação com 78% de sensibilidade e 92% de especificidade, o método Distância de Mahalanobis forneceu 100% de sensibilidade e 83% de especificidade; já o método SVM (máquina de vetores de suporte, ou *support vector machine*) apresentou 100% de sensibilidade e especificidade. Em trabalho muito semelhante do mesmo grupo de pesquisas, Ramos e colaboradores (Ramos *et al.*, 2017) utilizaram o mesmo método para avaliar 14 amostras de urina de pessoas com câncer e 24 amostras de indivíduos saudáveis. Foram avaliados dois métodos de classificação: o PLS-DA e o SVM. O primeiro forneceu 100% de sensibilidade e 91% de especificidade, já o segundo 100% de sensibilidade e especificidade. Esses dois métodos de *fingerprint* citados também apresentam melhor capacidade de classificação do que o método desenvolvido neste presente estudo. Mais uma vez, a chave para essa diferença pode estar na detectabilidade, já que ambos os trabalhos utilizam o método de extração por *headspace* dinâmico, o qual proporciona limites de detecção mais baixos em relação ao *headspace* estático (seção 2.2 do Capítulo II). Considerando esses dois estudos citados anteriormente, também pode ser levantada a hipótese de que o *fingerprint* gerado a partir através dessas técnicas não separativas (neste caso, o espectro de fragmentação gerado no MS) pode ser mais efetivo na discriminação das amostras de indivíduos saudáveis e pessoas com câncer do que o *fingerprint* gerado através dos cromatogramas da análise por GC-FID.

Por fim, foi encontrado um único estudo na literatura que analisa os VOCs no contexto do câncer utilizando o GC-FID. Contudo, esse trabalho utilizou uma abordagem *target* visando a presença de acetaldeído, acetona, acetato de etila, metanol, 2-propanol, 2-butanol, 1-propanol e etanol em amostras de fluido oral. Foram avaliadas 29 amostras de pessoas com câncer gastrointestinal e 16 amostras de indivíduos saudáveis. Quando foram levadas em consideração todas as substâncias

citadas acima, o modelo gerado não foi capaz de distinguir satisfatoriamente os dois grupos. Contudo, quando realizada a seleção de variáveis, levando-se em consideração apenas acetaldeído, acetona, 2-propanol e etanol, o método apresentou 95,9% de sensibilidade e 90,9% de especificidade. Contudo, é preciso destacar também que o método de classificação utilizado foi o de árvores de decisão, método este que não é tão robusto e apresenta grande tendência a *overfitting*, assim como discutido na seção 1.2.2.1 deste capítulo.

A Tabela 37 resume os cinco trabalhos citados acima, bem como apresenta outros exemplos estudos semelhantes publicados nos últimos 10 anos (a lista não é exaustiva). Esta tabela está organizada de acordo com nome do primeiro autor, ano de publicação, tipo de câncer avaliado, matriz biológica, número de voluntários, técnica de análise, tipo/objetivo do estudo (*target*, *untarget*, possível determinação de biomarcadores, ou *fingerprint*), modelo estatístico empregado e, por fim, capacidade de predição. Além disso, é importante ressaltar que a Tabela 37 difere da Tabela 2 e da Tabela 3, as quais estão presentes no Capítulo I. A Tabela 2 foi criada para ilustrar de maneira geral alguns estudos sobre a investigação de VOCs no contexto do câncer e a Tabela 3 ilustra exclusivamente estudos mais importantes dos VOCs presentes no fluido oral. Já a tabela abaixo resume estudos semelhantes ao presente trabalho e é focada na discussão sobre o desempenho de classificação dos métodos.

Tabela 37 - Resumo de estudos semelhantes publicados nos últimos 10 anos (organizado em ordem alfabética - nome do primeiro autor, tipo de câncer, matriz biológica avaliada, número de voluntários, técnica de análise, tipo/finalidade do estudo; modelo estatístico empregado e capacidade de predição).

CCP = câncer de cabeça e pescoço; CG= cancer gastrointestinal							(Continua)
Primeiro autor (ano)	Tipo de câncer	Matriz biológica	N.º de voluntários (câncer/controle)	Técnica de análise	Tipo/finalidade do estudo	Modelo estatístico	Capacidade de predição
Antón (2016)	Diferentes tipos	Fluido oral	23/32	HS-PTV-MS	<i>Untarget/fingerprint</i>	SVM	100% de sensibilidade 100% de especificidade
Arasaradnam (2014)	Colorretal	Urina	83/50	FAIMS	<i>Untarget/fingerprint</i>	DFA	88% de sensibilidade 60% de especificidade
Bel'skaya (2020)	CG	Fluido oral	29/16	GC-FID	<i>Target/determinação de biomarcadores</i>	Árvores de decisão; curva ROC	95,9% de sensibilidade 90,9% de especificidade
Chandran (2019)	CCP	Ar exalado	23/21	SIFT-MS	<i>Target/determinação de biomarcadores</i>	Mann-Whitney; curva ROC	91% de sensibilidade 76% de especificidade
Gruber (2014)	CCP	Ar exalado	22/19	<i>e-nose</i>	<i>Untarget/fingerprint</i>	DFA	77% de sensibilidade 90% de especificidade
Huang (2013)	Gastresofágico	Urina	17/13	SIFT-MS	<i>Target/determinação de biomarcadores</i>	Mann-Whitney; curva ROC	AUC de 0,904
Leunis (2014)	CCP	Ar exalado	36/23	<i>e-nose</i>	<i>Untarget/fingerprint</i>	Regressão logística; curva ROC	90% de sensibilidade 80% de especificidade
Monedeiro (2021)	Cavidade oral	Fluido oral	5/15	SPME-GC-MS	<i>Untarget/determinação de biomarcadores</i>	ANN	100% de sensibilidade 80% de especificidade
Ramos (2017)	Diferentes tipos	Urina	14/24	HS-PTV-MS	<i>Untarget/fingerprint</i>	SVM	100% de sensibilidade 100% de especificidade

CCP = câncer de cabeça e pescoço; CG= cancer gastrointestinal

(Continua)

Primeiro autor (ano)	Tipo de câncer	Matriz biológica	N.º de voluntários (câncer/controle)	Técnica de análise	Tipo/finalidade do estudo	Modelo estatístico	Capacidade de predição
Shigeyama (2019)	Cavidade oral	Fluido oral	24/50	TFME-GC- MS	<i>Untarget</i> /determinação de biomarcadores	Árvores de decisão	95,8% de sensibilidade 94,0% de especificidade
Taware (2017)	CCP	Urina	29/31	SPME-GC- MS	<i>Untarget</i> /determinação de biomarcadores	PLS-DA	>80% de sensibilidade >80% de especificidade
Taware (2018)	CCP	Fluido oral	32/27	SPME-GC- MS	<i>Untarget</i> /determinação de biomarcadores	OPLS-DA	R2Y 0,941 e Q2Y de 0,867; >80% de sensibilidade >80% de especificidade
Tyagi (2021)	Colorretal	Urina	58/36	<i>e-nose</i>	<i>Untarget/fingerprint</i>	RF e RN	91% de sensibilidade 55% de especificidade

FAIMS: do termo em inglês *High Field Asymmetric Waveform Ion Mobility Spectrometry*; SIFT-MS: do termo em inglês *Selected-ion flow-tube mass spectrometry*; PTV: do termo em inglês *programmable temperature vaporizer* e-nose: do termo em inglês *eletronic nose*; TFME: do termo em inglês *Thin film microextraction*; SVM: do termo em inglês *support vector machine*; DFA: do termo em inglês *discriminant factor analysis*; FDA: do termo em inglês *Fisher discriminant analysis*; ANN: do termo em inglês *artificial neural network*; RF: do termo em inglês *Random Forest*.

Ao analisar a Tabela 37 é possível fazer ponderações semelhantes às realizadas anteriormente: o número de voluntários e a capacidade de detecção proporcionada pelo método HS-GC-FID podem ter sido fatores limitantes deste presente estudo, fazendo com que a predição do modelo gerado não fosse a ideal. Contudo, mesmo com essas limitações, os modelos híbridos apresentaram capacidade de predição semelhante a outros estudos reportados na literatura, como os trabalhos de Arasaradnam e colaboradores, Huang e colaboradores, Leunis e colaboradores, e Tyagi e colaboradores (ARASARADNAM *et al.*, 2014; HUANG *et al.*, 2013; LEUNIS *et al.*, 2014; TYAGI *et al.*, 2021).

Além disso, ainda é preciso ressaltar que se fosse utilizado outro sistema de extração, como a SPME, ou outro sistema de análise, como o GC-MS, o presente estudo iria perder o sentido em alguns quesitos fundamentais, como a praticidade e os custos. Neste trabalho, a extração por SHS e a injeção no GC-FID foram realizadas de maneira automática através do sistema Combipal, como citado no Capítulo II. O uso da SPME demandaria outro sistema de aquecimento/agitação para extrair os VOCs, além do que a injeção seria realizada de maneira manual, o que tornaria o método muito menos prático e muito mais demorado. Com relação aos custos, as fibras utilizadas na SPME são relativamente frágeis e caras. Sendo assim, a utilização dessa técnica de extração também demandaria uma maior quantidade de recursos. Ainda, o uso do GC-FID foi realizado justamente como uma alternativa mais barata ao GC-MS, conforme discutido no Capítulo I.

Em suma, o método desenvolvido neste presente trabalho é simples, prático e de menor custo em relação a outros métodos disponíveis na literatura. A capacidade de distinção dos modelos OPLS-DA gerados foram, de maneira geral, considerados satisfatórios. Neste sentido, há outros estudos similares que apresentam modelos com maior poder de predição. Contudo, esses métodos são, de maneira geral, mais laboriosos e mais caros. Ademais, seria interessante se futuros trabalhos realizassem uma abordagem semelhante a este presente estudo (volatilômica no contexto do câncer utilizando o SHS-GC-FID), só que com um número maior de voluntários. Desta maneira, poderia ser analisado se um N amostral maior poderia compensar algumas

das limitações da técnica em questão. Essa continuação seria muito importante na avaliação da potencial aplicação deste método simples e prático na área clínica.

5 CONCLUSÃO

Os resultados deste estudo corroboram com as evidências da literatura de que os perfis de compostos voláteis presentes no fluido oral e na urina são diferentes em pessoas saudáveis e em pessoas com doenças neoplásicas. Isso ressalta a importância da investigação dos VOCs nessas matrizes biológicas não invasivas como potencial método de *screening* do câncer. Contudo, algumas ponderações precisam ser feitas quanto ao uso do SHS-GC-FID no desenvolvimento de um método de *fingerprint* para esse determinado fim.

No que diz respeito aos métodos de classificação empregados, o modelo RF apresentou uma grande capacidade de predição em relação as amostras do grupo controle (alta especificidade), porém apresentou uma baixa capacidade de classificação das amostras de indivíduos portadores de câncer (baixa sensibilidade). Este fato levaria a uma grande quantidade de resultados falso negativos, o que é justamente o que precisa ser evitado no desenvolvimento de métodos de *screening*. Já os resultados do modelo de OPLS-DA foram mais satisfatórios, onde a maioria das avaliações apresentou valores aceitáveis de R^2Y , Q^2Y AUC, sensibilidade e especificidade. Além disso, foi constatado que a abordagem híbrida realmente tende a elevar a capacidade de predição dos métodos, desde que os modelos individuais apresentem valores adequados dos parâmetros de qualidade. Esses resultados são inéditos na literatura e, talvez, constituam a maior contribuição deste presente estudo.

Por fim, acredita-se que a análise de um número maior de voluntários poderia levar à construção de modelos mais robustos, diminuindo a tendência de *overfitting* e melhorando a capacidade de predição dos métodos, auxiliando assim a compensar algumas limitações do SHS-GC-FID. Sendo assim, novos estudos contando com um tamanho amostral maior são essenciais para avaliar se uma técnica simples, prática, tão presente em diversos laboratórios, e que teve o potencial demonstrado neste presente trabalho, pode, de fato, um dia ser aplicada na rotina clínica como um método auxiliar no diagnóstico do câncer.

REFERÊNCIAS

ABDI, H., & WILLIAMS, L. J. Principal component analysis. **Wiley interdisciplinary reviews: computational statistics**, v. 2, n. 4, p. 433-459, 2010.

ALMEIDA, J. R. C., de Lima Pedrosa, N., Leite, J. B., do Prado Fleming, T. R., de Carvalho, V. H., & Cardoso, A. D. A. A. **Revista Brasileira de Cancerologia**, v. 53, n. 3, p. 305-316, 2007.

ALMEIDA, V. D., Leitão, A., Reina, L. D. C. B., Montanari, C. A., Donnici, C. L., & Lopes, M. T. P. Câncer e agentes antineoplásicos ciclo-celular específicos e ciclo-celular não específicos que interagem com o DNA: uma introdução. **Química Nova**, v. 28, n. 1, p. 118-129, 2005.

ALTOMARE, D. F., DI LENA, M., PORCELLI, F., TRIZIO, L., TRAVAGLIO, E., TUTINO, M., & DE GENNARO, G. Exhaled volatile organic compounds identify patients with colorectal cancer. **Journal of British Surgery**, v. 100, n. 1, p. 144-150, 2013.

AMANN, A., COSTELLO, B., MIEKISCH, W., SCHUBERT, J., BUSZEWSKI, B., PLEIL, J., & RISBY, T. The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. **Journal of breath research**, v. 8, n. 3, p. 034001, 2014.

ANTÓN, A.P., DEL NOGAL SÁNCHEZ, M., POZAS, Á.P.C., PAVÓN, J.L.P., & CORDERO, B.M. Headspace-programmed temperature vaporizer-mass spectrometry and pattern recognition techniques for the analysis of volatiles in saliva samples. **Talanta**, v. 160, p. 21-27, 2016.

AQUINO NETO, F. R. D. Uma Abordagem Simples e Abrangente para Dominar as Variáveis que Afetam a Injeção em Cromatografia Gasosa de Alta Resolução. **Sci. Chromatogr.**, v. 2, n. 4, p. 51-67, 2010.

ARASARADNAM, R. P., MCFARLANE, M. J., RYAN-FISHER, C., WESTENBRINK, E., HODGES, P., THOMAS, M. G., & COVINGTON, J. A. Detection of colorectal cancer (CRC) by urinary volatile organic compound analysis. **PloS one**, v. 9, n. 9, p. e108750, 2014.

ARGAWAL, M.S., SHARMA, M., SHARMA, M., & FATIMA, S. VOCC: a database of volatile organic compounds in cancer. **RSC Advances**, v. 6, n. 115, p. 114783-114789, 2016.

ARI, A., ARI, P. E., YENISOY-KARAKAŞ, S., & GAGA, E. O. Source characterization and risk assessment of occupational exposure to volatile organic compounds (VOCs) in a barbecue restaurant. **Building and Environment**, v. 174, p. 106791, 2020.

BAJTAREVIC, A., AGER, C., PIENZ, M., KLIEBER, M., SCHWARZ, K., LIGOR, M., & AMANN, A. Noninvasive detection of lung cancer by analysis of exhaled breath. **BMC cancer**, v. 9, n. 1, p. 1-16, 2009.

BARANSKA, A., MUJAGIC, Z., SMOLINSKA, A., DALLINGA, J. W., JONKERS, D. M. A. E., TIGCHELAAR, E. F., & VAN SCHOOTEN, F. J. Volatile organic compounds in breath as markers for irritable bowel syndrome: a metabolomic approach. **Alimentary pharmacology & therapeutics**, v. 44, n. 1, p. 45-56, 2016.

BARBOSA, J. M. G., PEREIRA, N. Z., DAVID, L. C., DE OLIVEIRA, C. G., SOARES, M. F. G., AVELINO, M. A. G., & SHOKRY, E. Cerumenogram: a new frontier in cancer diagnosis in humans. **Scientific reports**, v. 9, n. 1, p. 1-9, 2019.

BARRETO, M. L., ALMEIDA FILHO, N. D., VERAS, R. P., & BARATA, R. B. **Epidemiologia, serviços e tecnologias em saúde**. Editora FIOCRUZ, 1998. p 139.

BATTY, C. A., CAUCHI, M., LOURENCO, C., HUNTER, J. O., & TURNER, C. Use of the analysis of the volatile faecal metabolome in screening for colorectal cancer. *PloS one*, v. 10, n. 6, p. e0130301, 2015.

BEGER, R. D. A review of applications of metabolomics in cancer. **Metabolites**, v. 3, n. 3, p. 552-574, 2013.

BEHERA, B., JOSHI, R., VISHNU, G.A., BHALERAO, S., & PANDYA, H. J. Electronic nose: A non-invasive technology for breath analysis of diabetes and lung cancer patients. **Journal of breath research**, v. 13, n. 2, p. 024001, 2019.

BEL'SKAYA, L. V., SARF, E. A., SHALYGIN, S. P., POSTNOVA, T. V., & KOSENOK, V. K. Identification of salivary volatile organic compounds as potential markers of stomach and colorectal cancer: A pilot study. **Journal of Oral Biosciences**, v. 62, n. 2, p. 212-221, 2020.

BENEDETTI, B., CAPONIGRO, V., & ARDINI, F. Experimental design step by step: a practical guide for beginners. **Critical Reviews in Analytical Chemistry**, p. 1-14, 2020.

BEZERRA M.A., SANTELLI R.E., OLIVEIRA E. P., VILLAR L.S., & ESCALEIRA L.A. Response surface methodology (RSM) as a tool for optimization in analytical chemistry. **Talanta**. 2008;76(5):965-977.

BIAN, X., LIU, R., MENG, Y., XING, D., XU, D., & LU, Z. **Journal of Experimental Medicine**, v. 218, n. 1, 2021.

BOCCARD, J., & RUTLEDGE, D. N. A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. **Analytica chimica acta**, v. 769, p. 30-39, 2013.

BORGES, K. B., FIGUEIREDO, E. C.; & QUEIROZ, M. E. C. Preparo de amostras para análise de compostos orgânicos. **GEN, editor. Rio de Janeiro**: Editora LTC, 2015.

BORRÀS, E., FERRÉ, J., BOQUÉ, R., MESTRES, M., ACEÑA, L., & BUSTO, O. Data fusion methodologies for food and beverage authentication and quality assessment—A review. **Analytica Chimica Acta**, v. 891, p. 1-14, 2015.

BOS, L. D. J., STERK, P. J., & SCHULTZ, M. J. Volatile metabolites of pathogens: a systematic review. **PLoS pathogens**, v. 9, n. 5, p. e1003311, 2013.

BOSCH, S., BERKHOUT, D. J., LARBI, I. B., DE MEIJ, T.G., & DE BOER, N. K. Fecal volatile organic compounds for early detection of colorectal cancer: where are we now?. **Journal of cancer research and clinical oncology**, v. 145, n. 1, p. 223-234, 2019.

BOUZA, M., GONZALEZ-SOTO, J., PEREIRO, R., DE VICENTE, J. C., & SANZ-MEDEL, A. Exhaled breath and oral cavity VOCs as potential biomarkers in oral cancer patients. **Journal of breath research**, v. 11, n. 1, p. 016015, 2017.

BROWN, K. F., RUMGAY, H., DUNLOP, C., RYAN, M., QUARTLY, F., COX, A., & PARKIN, D. M. The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. **British journal of cancer**, v. 118, n. 8, p. 1130-1141, 2018.

BROZA, Y. Y., MOCHALSKI, P., RUZSANYI, V., AMANN, A., & HAICK, H. Hybrid volatolomics and disease detection. **Angewandte Chemie International Edition**, v. 54, n. 38, p. 11036-11048, 2015.

BROZA, Y.Y., ZURI, L., & HAICK, H. Combined volatolomics for monitoring of human body chemistry. **Scientific reports**, v. 4, p. 4611, 2014.

BUCKLEY, A. M., LYNAM-LENNON, N., O'NEILL, H., & O'SULLIVAN, J. Targeting hallmarks of cancer to enhance radiosensitivity in gastrointestinal cancers. **Nature reviews Gastroenterology & hepatology**, v. 17, n. 5, p. 298-313, 2020.

BULJUBASIC, F., & BUCHBAUER, G. Scent of human diseases: a review on specific volatile organic compounds as diagnostic biomarkers. **Flavour and Fragrance Journal**, 2015.

BURTON, C., & MA, Y. Current trends in cancer biomarker discovery using urinary metabolomics: achievements and new challenges. **Current medicinal chemistry**, v. 26, n. 1, p. 5-28, 2019.

BUSZEWSKI, B., LIGOR, T., JEZIERSKI, T., WENDA-PIESIK, A., WALCZAK, M., & RUDNICKA, J. Identification of volatile lung cancer markers by gas chromatography–mass spectrometry: comparison with discrimination by canines. **Analytical and bioanalytical chemistry**, v. 404, n. 1, p. 141-146, 2012.

BYLESJÖ, M., RANTALAINEN, M., CLOAREC, O., NICHOLSON, J. K., HOLMES, E., & TRYGG, J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. **Journal of Chemometrics: A Journal of the Chemometrics Society**, v. 20, n. 8-10, p. 341-351, 2006.

CAMBIAGHI, A., FERRARIO, M., & MASSEROLI, M. Analysis of metabolomic data: tools, current strategies and future challenges for omics data integration. **Briefings in bioinformatics**, v. 18, n. 3, p. 498-510, 2017.

CAVACO, C., PEREIRA, J. A., TAUNK, K., TAWARE, R., RAPOLE, S., NAGARAJARAM, H., & CÂMARA, J. S. Screening of salivary volatiles for putative breast cancer discrimination: an exploratory study involving geographically distant populations. **Analytical and bioanalytical chemistry**, v. 410, n. 18, p. 4459-4468, 2018.

CHAN, A. W., GILL, R. S., SCHILLER, D., & SAWYER, M. B. Potential role of metabolomics in diagnosis and surveillance of gastric cancer. **World journal of gastroenterology: WJG**, v. 20, n. 36, p. 12874, 2014.

CHANDRAN, D., OOI, E. H., WATSON, D. I., KHOLMURODOVA, F., JAENISCH, S., & YAZBECK, R. The use of selected ion flow tube-mass spectrometry technology to identify breath volatile organic compounds for the detection of head and neck squamous cell carcinoma: a pilot study. **Medicina**, v. 55, n. 6, p. 306, 2019.

CHASTEEN, T.G. **Split/Splitless gas chromatography injection**. 2000. Disponível em https://www.shsu.edu/~chm_tgc/primers/pdf/GC.pdf. Acesso em 24 Fev. 2022.

CHATTERJEE, S., CASTRO, M., & FELLER, J. F. An e-nose made of carbon nanotube based quantum resistive sensors for the detection of eighteen polar/nonpolar VOC biomarkers of lung cancer. **Journal of Materials Chemistry B**, v. 1, n. 36, p. 4563-4575, 2013.

CHEN, T., CAO, Y., ZHANG, Y., LIU, J., BAO, Y., WANG, C., & ZHAO, A. Random forest in clinical metabolomics for phenotypic discrimination and biomarker selection. **Evidence-Based Complementary and Alternative Medicine**, v. 2013, 2013.

CHEN, X., & ISHWARAN, H. Random forests for genomic data analysis. **Genomics**, v. 99, n. 6, p. 323-329, 2012.

CHEUNG, E. C., & VOUSDEN, K. H. The role of ROS in tumour development and progression. **Nature Reviews Cancer**, p. 1-18, 2022.

CHOW, L. M. Head and neck cancer. **New England Journal of Medicine**, v. 382, n. 1, p. 60-72, 2020.

COLLINS, C. H.; BRAGA, G.L.; BONATO, P. S.; Fundamentos de Cromatografia. 1. ed. Campinas: Editora da UNICAMP, 2006.

COSTA, B. R. B, SANTOS JÚNIOR, W. J. R., MAXIMIANO, I. F., GOMES, N. C.; FREITAS, B. T., & DE MARTINIS, B. S. Application of microextraction techniques in alternative biological matrices with focus on forensic toxicology: a review. **Bioanalysis**, v. 13, n. 1, p. 45-64, 2021a.

COSTA, B. R. B., & DE MARTINIS, B.S. Analysis of urinary VOCs using mass spectrometric methods to diagnose cancer: A review. **Clinical Mass Spectrometry**, 2020.

COSTA, B. R. B., EL HADDAD, L. P., FREITAS, B. T., MARINHO, P. A., & DE MARTINIS, B. S. Pre-workout supplements marketed in Brazil: Caffeine quantification and caffeine daily intake assessment. **Drug Testing and Analysis**, 2021b.

DA COSTA, B. R. B., EL HADDAD, L. P., & DE MARTINIS, B. S. Caffeine content of energy drinks marketed in Brazil. **Drug Testing and Analysis**, 2022.

COURANT, F., ANTIGNAC, J. P., DERVILLY-PINEL, G., & LE BIZEC, B. Basics of mass spectrometry based metabolomics. **Proteomics**, 14(21-22), 2369-2388, 2014.

CROUCH, D. J., DAY, J., BAUDYS, J., & FATAH, A. A. **Evaluation of saliva/oral fluid as an alternate drug testing specimen**. US Department of Commerce, Technology Administration, National Institute of Standards and Technology, 2004.

CUADROS-RODRÍGUEZ, L., RUIZ-SAMBLÁS, C., VALVERDE-SOM, L., PÉREZ-CASTAÑO, E., & GONZÁLEZ-CASADO, A. Chromatographic fingerprinting: An innovative approach for food identification and food authentication—A tutorial. **Analytica Chimica Acta**, v. 909, p. 9-23, 2016.

DE LACY COSTELLO, B., AMANN, A., AL-KATEB, H., FLYNN, C., FILIPIAK, W., KHALID, T., & RATCLIFFE, N. M. A review of the volatiles from the healthy human body. **Journal of breath research**, v. 8, n. 1, p. 014001, 2014.

DEJA, S., POREBSKA, I., KOWAL, A., ZABEK, A., BARG, W., PAWELCZYK, K., & MLYNARZ, P. Metabolomics provide new insights on lung cancer staging and discrimination from chronic obstructive pulmonary disease. **Journal of pharmaceutical and biomedical analysis**, v. 100, p. 369-380, 2014.

DIAS, V., JUNN, E., & MOURADIAN, M. M. The role of oxidative stress in Parkinson's disease. **Journal of Parkinson's disease**, v. 3, n. 4, p. 461-491, 2013.

DOS SANTOS, T. C., DOMINUTTI, P., PEDROSA, G. S., COELHO, M. S., NOGUEIRA, T., BORBON, A., SOUZA, S. R., & FORNARO, A. Isoprene in urban

Atlantic forests: Variability, origin, and implications on the air quality of a subtropical megacity. **Science of the Total Environment**, 824, 153728, 2022.

DURANTE, C., COCCHI, M., GRANDI, M., MARCHETTI, A., & BRO, R. Application of N-PLS to gas chromatographic and sensory data of traditional balsamic vinegars of Modena. **Chemometrics and Intelligent Laboratory Systems**, v. 83, n. 1, p. 54-65, 2006.

ELMONGY, H., & ABDEL-REHIM, M. Saliva as an alternative specimen to plasma for drug bioanalysis: A review. **TrAC Trends in Analytical Chemistry**, v. 83, p. 70-79, 2016.

FILIPIAK, W., FILIPIAK, A., SPONRING, A., SCHMID, T., ZELGER, B., AGER, C., & AMANN, A. Comparative analyses of volatile organic compounds (VOCs) from patients, tumors and transformed cell lines for the validation of lung cancer-derived breath markers. **Journal of breath research**, v. 8, n. 2, p. 027111, 2014.

FÖRSTER, J., KUIPERS, W., LENZ, C., ZIESCHE, S., & BECHTOLD, F. An autonomous flame ionization detector for emission monitoring. **Journal of Sensors and Sensor Systems**, v. 8, n. 1, p. 67-73, 2019.

FUCHS, P., LOESEKEN, C., SCHUBERT, J. K., & MIEKISCH, W. Breath gas aldehydes as biomarkers of lung cancer. **International Journal of Cancer**, v. 126, n. 11, p. 2663-2670, 2010.

FUKUDA, I. M., PINTO, C. F. F., MOREIRA, C. D. S., SAVIANO, A. M., & LOURENÇO, F. R. Design of Experiments (DoE) applied to pharmaceutical and analytical Quality by Design (QbD). **Brazilian Journal of Pharmaceutical Sciences**, 54(SPE), 2018.

GARRETT, W. S. Cancer and the microbiota. **Science**, v. 348, n. 6230, p. 80-86, 2015.

GATENBY, R. A. & GILLIES, R. J. A microenvironmental model of carcinogenesis. **Nature Reviews Cancer**, v. 8, n. 1, p. 56, 2008.

GE, D., ZHOU, J., CHU, Y., LU, Y., ZOU, X., XIA, L., & CHU, Y. VOCs and control their potential impact on breath biomarkers. **Analytical and bioanalytical chemistry**, p. 1-10, 2022.

GENTLEMAN, R. C., CAREY, V. J., BATES, D. M., BOLSTAD, B., DETTLING, M., DUDOIT, S., & ZHANG, J. 2004. Bioconductor: open software development for computational biology and bioinformatics. **Genome biology**, v. 5, n. 10, p. 1-16, 2004.

GIBB, S., STRIMMER, K. MALDIquant: a versatile R package for the analysis of mass spectrometry data. 2012. *Bioinformatics* **28**: 2270-2271

GOMES, Nayna Cândida. **Saliva: uma matriz alternativa para determinação de biomarcadores do cigarro em gestantes**. Dissertação de mestrado. Universidade de São Paulo. Ribeirão Preto. 2018.

GOODACRE, R., BROADHURST, D., SMILDE, A. K., KRISTAL, B. S., BAKER, J. D., BEGER, R., & WULFERT, F. Proposed minimum reporting standards for data analysis in metabolomics. **Metabolomics**, v. 3, n. 3, p. 231-241, 2007.

GOWDA, G. A., & DJUKOVIC, D. Overview of mass spectrometry-based metabolomics: opportunities and challenges. **Mass Spectrometry in Metabolomics**, p. 3-12, 2014.

GRUBER, M., TISCH, U., JERIES, R., AMAL, H., HAKIM, M., RONEN, O., & HAICK, H. Analysis of exhaled breath for diagnosing head and neck squamous cell carcinoma: a feasibility study. **British journal of cancer**, v. 111, n. 4, p. 790-798, 2014.

GUADAGNI, R., MIRAGLIA, N., SIMONELLI, A., SILVESTRE, A., LAMBERTI, M., FEOLA, D., & SANNOLO, N. Solid-phase microextraction–gas chromatography–mass spectrometry method validation for the determination of endogenous substances: urinary hexanal and heptanal as lung tumor biomarkers. **Analytica Chimica Acta**, v. 701, n. 1, p. 29-36, 2011.

HAICK, H. Volatile organic compounds of lung cancer and possible biochemical pathways. **Chemical reviews**, v. 112, n. 11, p. 5949-5966, 2012.

HANAHAN, D., & WEINBERG, R.,A. The hallmarks of cancer. **cell**, v. 100, n. 1, p. 57-70, 2000.

HANAHAN, D., & WEINBERG, R.,A. Hallmarks of cancer: the next generation. **cell**, v. 144, n. 5, p. 646-674, 2011.

HANAI, Y., SHIMONO, K., MATSUMURA, K., VACHANI, A., ALBELDA, S., YAMAZAKI, K., & OKA, H. **Bioscience, biotechnology, and biochemistry**, v. 76, n. 4, p. 679-684, 2012.

HOO, Z. H., CANDLISH, J., & TEARE, D. What is an ROC curve?. **Emergency Medicine Journal**, v. 34, n. 6, p. 357-359, 2017.

HOPKINSON, R. J., & SCHOFIELD, C. J. Deciphering functions of intracellular formaldehyde: linking cancer and aldehyde metabolism. **Biochemistry**, v. 57, n. 6, p. 904-906, 2018.

HUA, Q., ZHU, Y., & LIU, H. Detection of volatile organic compounds in exhaled breath to screen lung cancer: A systematic review. **Future Oncology**, v. 14, n. 16, p. 1647-1662, 2018.

HUANG, B., SONG, B. L., & XU, C. Cholesterol metabolism in cancer: mechanisms and therapeutic opportunities. **Nature metabolism**, v. 2, n. 2, p. 132-141, 2020.

HUANG, J., KUMAR, S., ABBASSI-GHADI, N., ŠPANĚL, P., SMITH, D., & HANNA, G. B. Selected ion flow tube mass spectrometry analysis of volatile metabolites in urine

headspace for the profiling of gastro-esophageal cancer. **Analytical Chemistry**, v. 85, n. 6, p. 3409-3416, 2013.

HUDA, A. K., DE LACY COSTELLO, B., & RATCLIFFE, N. An investigation of volatile organic compounds from the saliva of healthy individuals using headspace-trap/GC-MS. **Journal of breath research**, v. 7, n. 3, p. 036004, 2013.

INCA, Instituto Nacional de Câncer. **O que é câncer?**. 2020. Disponível em <<https://www.inca.gov.br/o-que-e-cancer>>. Acesso em 02 Fev. 2021.

INCA, Instituto Nacional de Câncer. **Deteção Precoce do Câncer**. 2021c. Disponível em <<https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/deteccao-precoce-do-cancer.pdf>> Acesso em 04 Fev. 2022.

INCA, Instituto Nacional de Câncer. **Deteção precoce**. 2021a. Disponível em <<https://www.inca.gov.br/controlado-cancer-do-colo-do-utero/acoes-de-controlado-deteccao-precoce>> Acesso em 04 Fev. 2022.

INCA, Instituto Nacional de Câncer. **Estadiamento**. 2021b. Disponível em <<https://www.inca.gov.br/estadiamento>> Acesso em 04 Fev. 2022.

INCA, Instituto Nacional de Câncer. **Estimativa 2020 – Incidência de Câncer no Brasil**. 2019. Disponível em

<<https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//estimativa-2020-incidencia-de-cancer-no-brasil.pdf>> Acesso em 02 Fev. 2022.

JACOBSON, D., MONFORTE, A. R., & FERREIRA, A. C. S. Untangling the chemistry of port wine aging with the use of GC-FID, multivariate statistics, and network reconstruction. **Journal of agricultural and food chemistry**, v. 61, n. 10, p. 2513-2521, 2013.

JAHANGIRI-MANESH, A., MOUSAZADEH, M., NIKKHAH, M., ABBASIAN, S., MOSHAI, A., MASROOR, M. J., & NOROUZI, P. Molecularly imprinted polymer-based chemiresistive sensor for detection of nonanal as a cancer related biomarker. **Microchemical Journal**, v. 173, p. 106988, 2022.

JANFAZA, S., KHORSAND, B., NIKKHAH, M., & ZAHIRI, J. Digging deeper into volatile organic compounds associated with cancer. **Biology Methods and Protocols**, v. 4, n. 1, p. bpz014, 2019.

JANITZA, S., & HORNUNG, R. On the overestimation of random forest's out-of-bag error. **PloS one**, v. 13, n. 8, p. e0201904, 2018.

JANSSENS, E., VAN MEERBEECK, J. P., & LAMOTE, K. (2020). Volatile organic compounds in human matrices as lung cancer biomarkers: A systematic review. **Critical Reviews in Oncology/Hematology**, v. 153, p. 103037, 2020.

JELSKI, W., & SZMITKOWSKI, M. Alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH) in the cancer diseases. **Clinica Chimica Acta**, v. 395, n. 1-2, p. 1-5, 2008.

JIANG, H., LIVINGSTON, M., ROOM, R., CHENHALL, R., & ENGLISH, D. R. Temporal associations of alcohol and tobacco consumption with cancer mortality. **JAMA network open**, v. 1, n. 3, p. e180713-e180713, 2018.

JIAO, J., DING, N., SHI, T., CHAI, X., CONG, P., & ZHU, Z. Study of chromatographic fingerprint of the flavor in beer by HS-SPME-GC. **Analytical letters**, v. 44, n. 4, p. 648-655, 2011.

JOGLEKAR, A.M.; MAY, A.T. Product excellence through experimental design. **Food product and development: From concept to the marketplace**, v. 211, 1987.

JUMTEE, K., BAMBA, T., & FUKUSAKI, E. Fast GC-FID based metabolic fingerprinting of Japanese green tea leaf for its quality ranking prediction. **Journal of separation science**, v. 32, n. 13, p. 2296-2304, 2009.

Karaman, I. (2017). Preprocessing and pretreatment of metabolomics data for statistical analysis. **Metabolomics: From Fundamentals to Clinical Applications**, 145-161.

KHURSHID, Z., Zafar, M., Khan, E., Mali, M., & Latif, M. Human saliva can be a diagnostic tool for Zika virus detection. **Journal of infection and public health**, v. 12, n. 5, p. 601-604, 2019.

KIM, E. R., KWON, H. N., NAM, H., KIM, J. J., PARK, S., & KIM, Y. H. Urine-NMR metabolomics for screening of advanced colorectal adenoma and early stage colorectal cancer. **Scientific Reports**, v. 9, n. 1, p. 1-10, 2019.

KIRALJ, R., & FERREIRA, M. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. *Journal of the Brazilian Chemical Society*, v. 20, n. 4, p. 770-787, 2009.

KRAKOWSKA, B., STANIMIROVA, I., ORZEL, J., DASZYKOWSKI, M., GRABOWSKI, I., ZALESZCZYK, G., & SZNAJDER, M. Detection of discoloration in diesel fuel based on gas chromatographic fingerprints. **Analytical and Bioanalytical Chemistry**, v. 407, n. 4, p. 1159-1170, 2015.

KRILAVICIUTE, A., HEISS, J. A., LEJA, M., KUPCINSKAS, J., HAICK, H., & BRENNER, H. Detection of cancer through exhaled breath: a systematic review. **Oncotarget**, v. 6, n. 36, p. 38643, 2015.

KUMAR, S., HUANG, J., ABBASSI-GHADI, N., ŠPANĚL, P., SMITH, D., & HANNA, G. B. (2013). Selected ion flow tube mass spectrometry analysis of exhaled breath for

volatile organic compound profiling of esophago-gastric cancer. **Analytical chemistry**, v. 85, n. 12, p. 6121-6128, 2013.

KUMAR, V., ABBAS, A.K., & ASTER, J.C. **Robbins patologia básica**. Elsevier Brasil. São Paulo, 2018.

KWAK, J., GALLAGHER, M., OZDENER, M. H., WYSOCKI, C. J., GOLDSMITH, B. R., ISAMAH, A., & PRETI, G. Volatile biomarkers from human melanoma cells. **Journal of Chromatography B**, v. 931, p. 90-96, 2013

LANÇAS, F. M. Fundamentos da Cromatografia Gasosa. Campinas: Átomo, 2016.

LAVRA, L., CATINI, A., ULIVIERI, A., CAPUANO, R., BAGHERNAJAD SALEHI, L., SCIACCHITANO, S., & DI NATALE, C. Investigation of VOCs associated with different characteristics of breast cancer cells. **Scientific reports**, v. 5, n. 1, p. 1-12, 2015.

LEE, B. J., KIM, H. Y., LIM, S. R., HUANG, L., & CHOI, H. K. Discrimination and prediction of cultivation age and parts of *Panax ginseng* by Fourier-transform infrared spectroscopy combined with multivariate statistical analysis. **PloS one**, v. 12, n. 10, p. e0186664, 2017.

LEE, S. C., TANG, I. P., AVATAR, S. P., AHMAD, N., SELVA, K. S., TAY, K. K., TAN, T. Y. Head and neck cancer: possible causes for delay in diagnosis and treatment. **Med J Malaysia**, v. 66, n. 2, p. 101, 2011.

Leunis, N., Boumans, M. L., Kremer, B., Din, S., Stobberingh, E., Kessels, A. G., & Kross, K. W. Application of an electronic nose in the diagnosis of head and neck cancer. **The Laryngoscope**, v. 124, n. 6, p. 1377-1381, 2014.

LI, M., & WANG, X. R. Peak alignment of gas chromatography–mass spectrometry data with deep learning. **Journal of Chromatography A**, v. 1604, p. 460476, 2019.

LIAW A, WIENER M. “Classification and Regression by randomForest.” *R News*, **2(3)**, 18-22. 2022. <https://CRAN.R-project.org/doc/Rnews/>.

LIEU, E. L., NGUYEN, T., RHYNE, S., & KIM, J. Amino acids in cancer. **Experimental & molecular medicine**, v. 52, n. 1, p. 15-30, 2020.

LIMA, A. R., ARAÚJO, A. M., PINTO, J., JERÓNIMO, C., HENRIQUE, R., BASTOS, M. D. L., & GUEDES DE PINHO, P. Discrimination between the human prostate normal and cancer cell exometabolome by GC-MS. **Scientific reports**, v. 8, n. 1, p. 1-12, 2018.

LIOU, G., & STORZ, P. Reactive oxygen species in cancer. **Free radical research**, v. 44, n. 5, p. 479-496, 2010.

LIU, L., & ÖZSU, M. T (Ed). **Encyclopedia of database systems**. New York, NY, USA:: Springer, 2009.

LIU, Y., SONG, M., LIU, X., ZHANG, Y., HUI, L., KONG, L., & FENG, M. Characterization and sources of volatile organic compounds (VOCs) and their related changes during ozone pollution days in 2016 in Beijing, China. **Environmental Pollution**, v. 257, p. 113599, 2020.

LIU, Y., YIN, T., FENG, Y., CONA, M.M., HUANG, G., Liu, J., & BORNMANS, G. Mammalian models of chemically induced primary malignancies exploitable for imaging-based preclinical theragnostic research. **Quantitative imaging in medicine and surgery**, v. 5, n. 5, p. 708, 2015.

LOMONACO, T., MANCO, E., CORTI, A., LA NASA, J., GHIMENTI, S., BIAGINI, D., & CASTELVETRO, V. Release of harmful volatile organic compounds (VOCs) from photo-degraded plastic debris: A neglected source of environmental pollution. **Journal of hazardous materials**, v. 394, p. 122596, 2020.

LOUREIRO, A. P. M., DI MASCIO, P., & MEDEIROS, M.H. Formação de adutos exocíclicos com bases de DNA: implicações em mutagênese e carcinogênese. **Química Nova**, v. 25, n. 5, p. 777-793, 2002.

MAHUL, B. A., STEPHEN, E., FREDERICK, L. G., DAVID, R. B., ROBERT, K., MARY, K. W., & DANIEL, C. S. **AJCC cancer staging manual**. Eight Edition: Springer Nature. 2017.

MANSUROVA, M., EBERT, B. E, BLANK, L. M., I., & BÁÑEZ, A. J. A. breath of information: the volatilome. **Current genetics**, v. 64, n. 4, p. 959-964, 2018.

MARGARITIS, A., SOENEN, H., FRANSEN, E., PIPINTAKOS, G., JACOBS, G., & BLOM, J. Identification of ageing state clusters of reclaimed asphalt binders using principal component analysis (PCA) and hierarchical cluster analysis (HCA) based on chemo-rheological parameters. **Construction and Building Materials**, 244, 118276, 2020.

MARKAR, S. R., CHIN, S. T., ROMANO, A., WIGGINS, T., ANTONOWICZ, S., PARASKEVA, P., & HANNA, G. B. Breath volatile organic compound profiling of colorectal cancer using selected ion flow-tube mass spectrometry. **Annals of surgery**, v. 269, n. 5, p. 903-910, 2019.

MARRUBINI, G., DUGHERI, S., CAPPELLI, G., ARCANGELI, G., MUCCI, N., APPELBLAD, P., & SPELTINI, A. (2020). Experimental designs for solid-phase

microextraction method development in bioanalysis: A review. **Analytica Chimica Acta**, v. 1119, p. 77-100, 2020.

MARTÍ-ÁLAMO, S.; MANCHEÑO-FRANCH, A.; MARZAL-GAMARRA, C.; & CARLOS-FABUEL, L. Saliva as a diagnostic fluid. Literature review. **Journal of clinical and experimental dentistry**, v. 4, n. 4, p. e237, 2012.

MARTÍNEZ-CAMBLOR, P., BAYÓN, G. F., & PÉREZ-FERNÁNDEZ, S. Cumulative/dynamic ROC curve estimation. **Journal of Statistical Computation and Simulation**, v. 86, n. 17, p. 3582-3594, 2016.

MCMINN, D. Detectors: General (flame ionization detectors and thermal conductivity detectors). *Encycl Sep Sci*, 2000.

MIEKISCH, W., SCHUBERT, J. K., & NOELDGE-SCHOMBURG, G. F. Diagnostic potential of breath analysis—focus on volatile organic compounds. **Clinica chimica acta**, v. 347, n. 1-2, p. 25-39, 2004.

MINA, A. Comparison of several immunoassays used in drugs of abuse screening: Assessment against gold standard methods and calculation of measurement uncertainty. **Journal of pharmacological and toxicological methods**, v. 101, p. 106649, 2020.

MONEDEIRO, F., MONEDEIRO-MILANOWSKI, M., LIGOR, T., & BUSZEWSKI, B. A review of GC-based analysis of non-invasive biomarkers of colorectal cancer and related pathways. **Journal of clinical medicine**, v. 9, n. 10, p. 3191, 2020.

MONEDEIRO, F., MONEDEIRO-MILANOWSKI, M., ZMYŚŁOWSKI, H., DE MARTINIS, B. S., & BUSZEWSKI, B. (2021). Evaluation of salivary VOC profile composition directed towards oral cancer and oral lesion assessment. **Clinical Oral Investigations**, v. 25, n. 7, p. 4415-4430, 2021.

MONEDEIRO, F.; DOS REIS, R.B.; PERIA, F.M.; SARES, C.T.G.; DE MARTINIS, B. S. Investigation of sweat VOC profiles in assessment of cancer biomarkers using HS-GC-MS. **Journal of breath research**, v. 14, n. 2, p. 026009, 2020.

MONEDEIRO, Fernanda Ferreira da Silva Souza. **Investigação de compostos orgânicos voláteis em amostras de suor como biomarcadores no diagnóstico de câncer**. Tese de Doutorado. Universidade de São Paulo. Ribeirão Preto. 2018.

MONTEIRO, M., CARVALHO, M., HENRIQUE, R., JERONIMO, C., MOREIRA, N., DE LOURDES BASTOS, M., & DE PINHO, P. G. Analysis of volatile human urinary metabolome by solid-phase microextraction in combination with gas chromatography–mass spectrometry for biomarker discovery: Application in a pilot study to discriminate patients with renal cell carcinoma. **European Journal of Cancer**, 50(11), 1993-2002, 2014.

MONTEIRO, M., CARVALHO, M., HENRIQUE, R., JERONIMO, C., MOREIRA, N., DE LOURDES BASTOS, M., & DE PINHO, P. G. Analysis of volatile human urinary metabolome by solid-phase microextraction in combination with gas chromatography–mass spectrometry for biomarker discovery: Application in a pilot study to discriminate patients with renal cell carcinoma. **European Journal of Cancer**, 50(11), 1993-2002, 2014.

MONTEIRO, M., MOREIRA, N., PINTO, J., PIRES-LUÍS, A. S., HENRIQUE, R., JERÓNIMO, C., & GUEDES DE PINHO, P. GC-MS metabolomics-based approach for the identification of a potential VOC-biomarker panel in the urine of renal cell carcinoma patients. **Journal of cellular and molecular medicine**, v. 21, n. 9, p. 2092-2105, 2017.

MORAES, Rafaela dos Santos. **Complexos homo e heterobinucleares de ligantes derivados da isoniazida como potenciais agentes antitumorais**. Tese de Doutorado. Pontificia Universidade Católica do Rio de Janeiro. Rio de Janeiro. 2016.

MOTAMAYEL, F.A., DAVOODI, P., DALBAND, M., & HENDI, S. S. Saliva as a mirror of the body health. **Avicenna Journal of Dental Research**, v. 1, n. 2, p. 41-55, 2018.

MURPHY, P. M. Chemokines and the molecular basis of cancer metastasis. **New England Journal of Medicine**, v. 345, n. 11, p. 833-835, 2001.

NASCIMENTO, José Orete do Nascimento. **Análise da produção científica brasileira em Oncologia: um estudo comparativo entre artigos sobre câncer na base Lattes e os registros hospitalares de câncer**. Tese de Doutorado. Universidade Federal do Rio de Janeiro. Rio de Janeiro. 2020.

NCI. National Cancer Institute. 2021. **Head and Neck Cancers**. 2020. Disponível em <<https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet#r1>>. Acesso em 05 Mai. 2021.

NIELSEN, N. J., BALLABIO, D., TOMASI, G., TODESCHINI, R., & CHRISTENSEN, J. H. Chemometric analysis of gas chromatography with flame ionisation detection chromatograms: A novel method for classification of petroleum products. **Journal of Chromatography A**, v. 1238, p. 121-127, 2012.

OENNING, A. L., MORÉS, L., DIAS, A. N., & CARASEK, E. A new configuration for bar adsorptive microextraction (BA μ E) for the quantification of biomarkers (hexanal and heptanal) in human urine by HPLC providing an alternative for early lung cancer diagnosis. **Analytica chimica acta**, v. 965, p. 54-62, 2017.

OKSANEN, J., BLANCHET, F. G., KINDT, R., LEGENDRE, P., MINCHIN, P. R., O'HARA, R. B., & OKSANEN, M. J. Package 'vegan'. Community ecology package, version, v. 2, n. 9, p. 1-295, 2013.

OPITZ, P., & HERBARTH, O. The volatilome—investigation of volatile organic metabolites (VOM) as potential tumor markers in patients with head and neck squamous cell carcinoma (HNSCC). **Journal of Otolaryngology-Head & Neck Surgery**, v. 47, n. 1, p. 1-13, 2018.

PAULING, L., ROBINSON, A. B., TERANISHI, R., & CARY, P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. **Proceedings of the National Academy of Sciences**, v. 68, n. 10, p. 2374-2376, 1971.

PERROTTI, F., ROSA, C., CICALINI, I., SACCHETTA, P., DEL BOCCIO, P., GENOVESI, D., & PIERAGOSTINO, D. **International journal of molecular sciences**, v. 17, n. 12, p. 1992, 2016.

PINTO, J., AMARO, F., LIMA, A. R., CARVALHO-MAIA, C., JERÓNIMO, C., HENRIQUE, R., & GUEDES DE PINHO, P. Urinary volatilomics unveils a candidate biomarker panel for noninvasive detection of clear cell renal cell carcinoma. **Journal of Proteome Research**, v. 20, n. 6, p. 3068-3077, 2021.

POTTER, M., NEWPORT, E., & MORTEN, K. J. The Warburg effect: 80 years on. **Biochemical Society Transactions**, v. 44, n. 5, p. 1499-1505, 2016.

R Core Team, 2020. RStudio: Integrated Development Environment for R, Boston, MA. Disponível em <<http://www.rstudio.com/>>.

RAMOS, Á. G., ANTÓN, A. P., DEL NOGAL SANCHEZ, M., PAVÓN, J. L. P., & CORDERO, B. M. Urinary volatile fingerprint based on mass spectrometry for the discrimination of patients with lung cancer and controls. **Talanta**, v. 174, p. 158-164, 2017.

REN, S., HINZMAN, A. A., KANG, E. L., SZCZESNIAK, R. D., & LU, L. J. Computational and statistical analysis of metabolomics data. **Metabolomics**, 11(6), 1492-1513, 2015.

RESTEK, A. Technical guide for static headspace analysis using GC. **Restek Corp**, p. 11-12, 2000. Disponível em < <https://d1lqgfmy9cwjff.cloudfront.net/csi/pdf/e/rk67.pdf>> Acesso em 25 Fev. 2022.

ROBIN, X., TURCK, N., HAINARD, A., TIBERTI, N., LISACEK, F., SANCHEZ, J. C., & MÜLLER, M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. **BMC bioinformatics**, v. 12, n. 1, p. 1-8, 2011.

RODRIGUES, D., PINTO, J., ARAÚJO, A. M., MONTEIRO-REIS, S., JERÓNIMO, C., HENRIQUE, R., & CARVALHO, M.. Volatile metabolomic signature of bladder cancer cell lines based on gas chromatography–mass spectrometry. **Metabolomics**, v. 14, n. 5, p. 1-15, 2018.

ROSER, M., & RITCHIE, H. **Cancer**. 2019. Disponível em <<https://ourworldindata.org/cancer>>. Acesso em 27 Jul. 2021.

RUDNICKA, J., KOWALKOWSKI, T., & BUSZEWSKI, B. Searching for selected VOCs in human breath samples as potential markers of lung cancer. **Lung Cancer**, v. 135, p. 123-129, 2019.

RUZSÁNYI, V., & KALAIPOS, M. P. Breath acetone as a potential marker in clinical practice. **Journal of breath research**, v. 11, n. 2, p. 024002, 2017.

SANTANA, F. B. D., SOUZA, A. M. D., ALMEIDA, M. R., BREITKREITZ, M. C., FILGUEIRAS, P. R., SENA, M. M., & POPPI, R. J. Experimento Didático de Quimiometria Para Classificação de Óleos Vegetais Comestíveis Por Espectroscopia No Infravermelho Médio Combinado Com Análise Discriminante Por Mínimos Quadrados Parciais: Um Tutorial, Parte V. **Química Nova**, v. 43, p. 371-381, 2020.

SCHMIDT, K., & PODMORE, I. Current challenges in volatile organic compounds analysis as potential biomarkers of cancer. **Journal of biomarkers**, v. 2015, 2015.

SCHUTTE, H. W., HEUTINK, F., WELLENSTEIN, D. J., VAN DEN BROEK, G. B., VAN DEN HOOGEN, F. J., MARRES, H. A., & TAKES, R. P. Impact of time to diagnosis and treatment in head and neck cancer: a systematic review. **Otolaryngology–Head and Neck Surgery**, v. 162, n. 4, p. 446-457, 2020.

SEIDI, S., REZAZADEH, M., & ALIZADEH, R. Miniaturized sample preparation methods for saliva analysis. **Bioanalysis**, v. 11, n. 02, p. 119-148, 2019.

SHIGEYAMA, H., WANG, T., ICHINOSE, M., ANSAI, T., & LEE, S. W. Identification of volatile metabolites in human saliva from patients with oral squamous cell carcinoma via zeolite-based thin-film microextraction coupled with GC–MS. **Journal of Chromatography B**, v. 1104, p. 49-58, 2019.

SILVA, A. E., SERAKIDES, R., & DANTAS, G. Carcinogênese hormonal e neoplasias hormônio-dependentes. **Ciência Rural**, v. 34, n. 2, 2004.

SILVA, C. L., PASSOS, M., & CAMARA, J. S. Investigation of urinary volatile organic metabolites as potential cancer biomarkers by solid-phase microextraction in combination with gas chromatography-mass spectrometry. **British journal of cancer**, v. 105, n. 12, p. 1894-1904, 2011.

SILVA, C. L., PASSOS, M., & CÂMARA, J. S. Solid phase microextraction, mass spectrometry and metabolomic approaches for detection of potential urinary cancer biomarkers—a powerful strategy for breast cancer diagnosis. **Talanta**, v. 89, p. 360-368, 2012.

SILVA, C. L., PERESTRELO, R., SILVA, P., TOMÁS, H., & CÂMARA, J. S. Implementing a central composite design for the optimization of solid phase microextraction to establish the urinary volatome expression: A first approach for breast cancer. **Metabolomics**, v. 15, n. 4, p. 1-13, 2019.

SITHERSINGH, M. J., & SNOW, N. H. Headspace gas chromatography. In: **Gas chromatography**. Elsevier, 2021. p. 251-265.

SMOLINSKA, A., BODELIER, A. G. L., DALLINGA, J. W., MASCLÉE, A. A. M.; JONKERS, D.M.; VAN SCHOOTEN, F.J.; PIERIK, M.J. The potential of volatile organic compounds for the detection of active disease in patients with ulcerative colitis. **Alimentary pharmacology & therapeutics**, v. 45, n. 9, p. 1244-1254, 2017.

SMOLINSKA, A., HAUSCHILD, A. C., FIJTEN, R. R. R., DALLINGA, J. W., BAUMBACH, J., & VAN SCHOOTEN, F. J. Current breathomics—a review on data pre-processing techniques and machine learning in metabolomics breath analysis. **Journal of breath research**, v. 8, n. 2, p. 027105, 2014.

SNOW, N. H., & SLACK, G. C. Head-space analysis in modern gas chromatography. **TrAC Trends in Analytical Chemistry**, v. 21, n. 9-10, p. 608-617, 2002.

STIPP, M. C., & ACCO, A. Involvement of cytochrome P450 enzymes in inflammation and cancer: a review. **Cancer Chemotherapy and Pharmacology**, v. 87, n. 3, p. 295-309, 2021.

STOYANOV, K., & WALMSLEY, A. D. Response-Surface Modeling and Experimental Design. **Practical guide to chemometrics**, p. 263, 2006.

STRIMBU, K.; TAVEL, J. A. What are biomarkers?. **Current Opinion in HIV and AIDS**, v. 5, n. 6, p. 463, 2010.

SUNG, H., FERLAY, J., SIEGEL, R. L., LAVERSANNE, M., SOERJOMATARAM, I., JEMAL, A., & BRAY, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: a cancer journal for clinicians**, v. 71, n. 3, p. 209-249, 2021.

TAUNK, K., TAWARE, R., MORE, T. H., PORTO-FIGUEIRA, P., PEREIRA, J. A., MOHAPATRA, R., & RAPOLE, S. A non-invasive approach to explore the discriminatory potential of the urinary volatilome of invasive ductal carcinoma of the breast. **RSC advances**, v. 8, n. 44, p. 25040-25050, 2018.

TAWARE, R., TAUNK, K., PEREIRA, J. A., DHAKNE, R., KANNAN, N., SONEJI, D., & RAPOLE, S. Investigation of urinary volatome alterations in head and neck cancer: a non-invasive approach towards diagnosis and prognosis. **Metabolomics**, v. 13, n. 10, p. 1-11, 2017.

TAWARE, R., TAUNK, K., PEREIRA, J. A., SHIROLKAR, A., SONEJI, D., CÂMARA, J. S., & RAPOLE, S. Volatilomic insight of head and neck cancer via the effects observed on saliva metabolites. **Scientific reports**, v. 8, n. 1, p. 1-10, 2018.

TIPLER, A. An introduction to headspace sampling in gas chromatography fundamentals and theory. 2013. Disponível em https://resources.perkinelmer.com/corporate/pdfs/downloads/gde_intro_to_headspace.pdf> Acesso em 25 Fev. 2022.

TOMASETTI, C., & VOGELSTEIN, B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. **Science**, v. 347, n. 6217, p. 78-81, 2015.

TYAGI, H., DAULTON, E., BANNAGA, A. S., ARASARADNAM, R. P., & COVINGTON, J. A. Non-invasive detection and staging of colorectal cancer using a portable electronic nose. **Sensors**, v. 21, n. 16, p. 5440, 2021.

UBHI, B. K., CHENG, K. K., DONG, J., JANOWITZ, T., JODRELL, D.; TAL-SINGER, R., & CONNOR, S. C. Targeted metabolomics identifies perturbations in amino acid metabolism that sub-classify patients with COPD. **Molecular BioSystems**, v. 8, n. 12, p. 3125-3133, 2012.

ULANOWSKA, A., TRAWIŃSKA, E., SAWRYCKI, P., & BUSZEWSKI, B. Chemotherapy control by breath profile with application of SPME-GC/MS method. **Journal of Separation Science**, v. 35, n. 21, p. 2908-2913, 2012.

VAN DE GOOR, R. M.; HARDY, J.C.; VAN HOOREN, M.R.; KREMER, B.; KROSS, K.W. 2019. Detecting recurrent head and neck cancer using electronic nose technology: a feasibility study. **Head & neck**, v. 41, n. 9, p. 2983-2990, 2019.

VAN DE GOOR, R., VAN HOOREN, M., DINGEMANS, A. M., KREMER, B., & KROSS, K. Training and validating a portable electronic nose for lung cancer screening. **Journal of Thoracic Oncology**, v. 13, n. 5, p. 676-681, 2018.

VAN DE KANT, K.D.G.; VAN DER SANDE, L.J.T.M.; JÖBSIS, Q.; VAN SCHAYCK, O.C.P.; DOMPELING, E. Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. **Respiratory Research**. v. 117, n. 13, 2012.

VAN DEN BERG, R. A., HOEFSLOOT, H. C., WESTERHUIS, J. A., SMILDE, A. K., & VAN DER WERF, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. **BMC genomics**, v. 7, n. 1, p. 1-15, 2006.

VAN VLIET, D., SMOLINSKA, A., JÖBSIS, Q., ROSIAS, P., MURIS, J., DALLINGA, J., & VAN SCHOOTEN, F. J. **Journal of breath research**, v. 11, n. 1, p. 016016, 2017.

VANDER HEIDEN, M. G., CANTLEY, L. C., & THOMPSON, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. **Science**, v. 324, n. 5930, p. 1029-1033, 2009.

VETTORE, L., WESTBROOK, R. L., & TENNANT, D. A. New aspects of amino acid metabolism in cancer. **British journal of cancer**, v. 122, n. 2, p. 150-156, 2020.

VINCENT, T. L., & GATENBY, R. A. (2008). An evolutionary model for initiation, promotion, and progression in carcinogenesis. **International journal of oncology**, 32(4), 729-737.

VISENTAINER, J. V., CLAUS, T., SANTOS JR, O. O., CHIAVELLI, L. U. R., & MARUYAMA, S. A. Analytical aspects of the flame ionization detection in comparison with mass spectrometry with emphasis on fatty acids and their esters. London: IntechOpen; 2014.

VISWANATH, B., CHOI, C. S., LEE, K., & KIM, S. Recent trends in the development of diagnostic tools for diabetes mellitus using patient saliva. **TrAC Trends in Analytical Chemistry**, v. 89, p. 60-67, 2017.

WANG, C., LI, P., LIAN, A., SUN, B., WANG, X., GUO, L., & LI, E. Blood volatile compounds as biomarkers for colorectal cancer. **Cancer biology & therapy**, v. 15, n. 2, p. 200-206, 2014.

WANG, P., RASHID, M., LIU, J., HU, M., & ZHONG, G. Identification of multi-insecticide residues using GC-NPD and the degradation kinetics of chlorpyrifos in sweet corn and soils. **Food chemistry**, v. 212, p. 420-426, 2016.

WEI, Z., LIU, X., CHENG, C., YU, W., & YI, P. Metabolism of amino acids in cancer. **Frontiers in cell and developmental biology**, p. 1628, 2021.

WEN, Q., BOSHIER, P., MYRIDAKIS, A., BELLUOMO, I., & HANNA, G. B. Urinary volatile organic compound analysis for the diagnosis of cancer: A systematic literature review and quality assessment. **Metabolites**, v. 11, n. 1, p. 17, 2020.

WHO, World health organization. **Cancer**. 2022. Disponível em: <http://www.who.int/news-room/fact-sheets/detail/cancer>. Acesso em 27 Jul. 2021.

WHO, World health organization. **Mechanisms of carcinogenesis**. 2005. Disponível em: https://www.iarc.fr/en/publications/pdfs-online/wcr/2008/wcr_2008_5.pdf. Acesso em 27 Jul. 2021.

WHO, World health organization. **Population Fact Sheets – Brazil**. 2021. Disponível em <https://gco.iarc.fr/today/data/factsheets/populations/76-brazil-fact-sheets.pdf> Acesso em 27 Jul 2021.

WHO, World health organization. **World Cancer Report: Cancer research for cancer prevention**. 2020. Disponível em: https://www.iarc.who.int/cards_page/world-cancer-report/ Acesso em 27 Jul. 2021.

WORLEY, B., & POWERS, R. Multivariate analysis in metabolomics. **Current metabolomics**, v. 1, n. 1, p. 92-107, 2013.

XU, Z. Q., BROZA, Y. Y., IONSECU, R., TISCH, U., DING, L., LIU, H., & HAICK, H. A nanomaterial-based breath test for distinguishing gastric cancer from benign gastric conditions. **British journal of cancer**, v. 108, n. 4, p. 941-950, 2013.

YOSHIZAWA, J. M., SCHAFER, C. A., SCHAFER, J. J., FARRELL, J. J., PASTER, B. J., & WONG, D. T. Salivary biomarkers: toward future clinical and diagnostic utilities. **Clinical microbiology reviews**, v. 26, n. 4, p. 781-791, 2013.

ZAIDI, N., LUPIEN, L., KUEMMERLE, N. B., KINLAW, W. B., SWINNEN, J. V., & SMANS, K. Lipogenesis and lipolysis: the pathways exploited by the cancer cells to acquire fatty acids. **Progress in lipid research**, v. 52, n. 4, p. 585-589, 2013.

ZHANG, J., WEN, X., LI, Y., LI, X., QIAN, C., TIAN, Y., & DUAN, Y. Diagnostic approach to thyroid cancer based on amino acid metabolomics in saliva by ultra-performance liquid chromatography with high resolution mass spectrometry. **Talanta**, v. 235, p. 122729, 2021.

ZHOU, B., XU, J. W., CHENG, Y. G., GAO, J. Y., HU, S. Y.; WANG, L., & ZHAN, H. X.
Early detection of pancreatic cancer: Where are we now and where are we
going?. **International journal of cancer**, v. 141, n. 2, p. 231-241, 2017.

ANEXO 1 – Aprovação do Comitê de Ética em Pesquisa da FCFRP**FACULDADE DE CIÊNCIAS FARMACÊUTICAS
DE RIBEIRÃO PRETO****Comitê de Ética em Pesquisa**

Of. CEP/FCFRP nº. 001/2023
kms

Ribeirão Preto, 03 de janeiro de 2023.


Ao pós-graduando
Bruno Ruiz Brandão da Costa
Orientador: Prof. Dr. Bruno Spinosa De Martinis
FCFRP/USP

Prezado Pesquisador,

Informamos que a Emenda 2 do protocolo de pesquisa intitulado **“Análise híbrida do perfil de compostos orgânicos voláteis em amostras de fluido oral e urina como possível método de diagnóstico do câncer”** (CAAE: 02519118.8.0000.5403 - Protocolo CEP/FCFRP nº 493), foi aprovada *ad referendum* do Comitê de Ética em Pesquisa da FCFRP em 01/11/2022, conforme Parecer Consubstanciado nº 5.733.512.

Informamos que deverá ser encaminhado ao CEP, através de **notificação**, o relatório parcial (após 6 meses) e o relatório final da pesquisa em formulário próprio deste Comitê, bem como comunicada qualquer alteração, intercorrência ou interrupção do mesmo, tais como eventos adversos e eventuais modificações no protocolo ou nos membros da equipe, através da interposição de **emenda** na Plataforma Brasil.

Atenciosamente,


Profa. Dra. Fabíola Attí de Castro
Coordenadora do CEP/FCFRP

ANEXO 2 – Aprovação do Comitê de Ética em Pesquisa da FFCLRP

Universidade de São Paulo
Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto
Comitê de Ética em Pesquisa

OF.138/CEP/FFCLRP/USP/22.11.2022

Prezado Pesquisador,

Comunicamos a V. Sa. que a emenda ao projeto de pesquisa intitulado **"Análise híbrida do perfil de compostos orgânicos voláteis em amostras de fluido oral e urina como possível método de diagnóstico do câncer"** foi analisada *ad referendum* do Comitê de Ética em Pesquisa da FFCLRP-USP e enquadrada na categoria: **APROVADA** (CAAE nº 02519118.8.3001.5407).

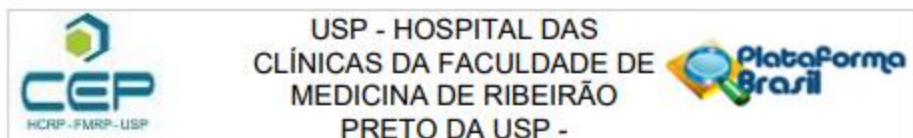
De acordo com a Resolução nº466 de 12.12.2012, devem ser entregues relatórios semestrais e, ao término do estudo, um relatório final, sempre via Plataforma Brasil.

Atenciosamente,

Prof. Dra. Sylvia Domingos Barrera
Coordenadora

Ilmo. Sr.
Bruno BRBC Costa
Faculdade de Ciências Farmacêuticas de Ribeirão Preto - USP

ANEXO 3 – Aprovação do Comitê de Ética em Pesquisa da FMRP



Continuação do Parecer: 5.739.695

TCLE / Termos de Assentimento / Justificativa de Ausência	consentimento.pdf	11/09/2018 13:27:05	Bruno BRBC Costa	Aceito
---	-------------------	------------------------	------------------	--------

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

RIBEIRAO PRETO, 04 de Novembro de 2022

Assinado por:
MARCIA GUIMARÃES VILLANOVA
 (Coordenador(a))

Endereço: CAMPUS UNIVERSITÁRIO
 Bairro: MONTE ALEGRE CEP: 14.048-900
 UF: SP Município: RIBEIRAO PRETO
 Telefone: (16)3602-2228 Fax: (16)3633-1144 E-mail: cep@hcrp.usp.br

ANEXO 4 - Termo de consentimento livre e esclarecido

Você está sendo convidado para participar de uma pesquisa como voluntário. Esta pesquisa está sendo realizada pelos pesquisadores do Programa de Pós-Graduação em Toxicologia da Faculdade de Ciências Farmacêuticas de Ribeirão Preto - USP. Os pesquisadores responsáveis são: Bruno Ruiz Brandão da Costa e Prof. Dr. Bruno Spinosa De Martinis.

O Título da pesquisa é “Análise híbrida do perfil de compostos orgânicos voláteis em amostras de fluido oral e urina como possível método de diagnóstico do câncer”.

Após ser esclarecido sobre as informações a seguir, no caso de aceitar fazer parte da pesquisa, por favor, assine ao final deste documento, o qual está em duas vias. Uma das vias deste termo ficará com você e a outra é dos pesquisadores responsáveis.

Breve explicação do projeto: partindo-se da ideia de que o indivíduo, quando doente, passa a produzir e eliminar substâncias químicas diferentes das comuns, o objetivo desta pesquisa é investigar essas substâncias e verificar como elas podem estar relacionadas ao câncer. Foi escolhido analisar essas substâncias em saliva e urina, uma vez que são métodos baratos, rápidos e não invasivos, o que pode aumentar o número de pessoas interessadas em participar do processo. Após a avaliação de quais substâncias presentes na saliva e urina são específicas para indivíduos com câncer, é esperado que a identificação dessas substâncias auxilie no desenvolvimento de um método para diagnosticar o câncer precocemente e de maneira simples.

Se você concordar em participar da pesquisa, será fornecido um chumaço de algodão, mais conhecido como Salivette®, para a coleta da saliva. O algodão serve **exclusivamente** para recolher a saliva do participante, não contendo nenhuma substância que possa ser transferida para seu corpo. O Salivette® deve ficar em contato com a boca do participante por 1 minuto. Além disso, será fornecido um pote de coleta universal para que seja coletada a urina. As duas amostras serão coletadas no momento da consulta, com a devida orientação do pesquisador responsável.

Um possível desconforto será alguma irritação no contato do algodão com a boca, mas o material é feito para não causar nenhum tipo de problema ou alergias, ainda não existindo relatos de qualquer tipo de dano à boca. Caso você sinta algum desconforto, poderá retirar o Salivette® a qualquer momento, sem que ocorra qualquer tipo de penalização por isso.

Os seus dados são confidenciais e você será identificado por um número. As informações obtidas no resultado da análise sobre a presença ou ausência das substâncias nas amostras coletadas serão mantidas em segredo. Além disso, gostaríamos de pedir

autorização para verificar seu prontuário, com o objetivo de confirmar quais medicamentos você faz uso e também verificar outros diagnósticos relatados pelos médicos.

Você é livre para não querer participar ou desistir da pesquisa a qualquer momento que desejar. Caso mude de ideia você não será prejudicado. **Se você recebe atendimento no Hospital das Clínicas de Ribeirão Preto, tal atendimento continuará sendo dado da mesma forma.** Você tem o direito de perguntar o que desejar em qualquer etapa da pesquisa. Você não receberá nenhuma gratificação por participar desta pesquisa e este projeto não lhe trará benefícios diretos. Caso você tenha despesas adicionais de transporte decorrentes de sua participação na pesquisa vocês receberá um ressarcimento integral.

Se você tiver alguma pergunta ou qualquer dúvida sobre esta pesquisa, entre em contato com:

Bruno Ruiz Brandão da Costa: e-mail: bruno.ruiz.costa@usp.br

Prof. Dr. Bruno Spinosa De Martinis: (16) 33150471, e-mail: martinis@usp.br

Endereço do laboratório: Avenida Bandeirantes, 3900 – DQ-FFCLRP, bloco 8 superior, sala 47. (O contato por telefone ou e-mail pode ser feito tanto em horário comercial ou fora do comercial).

Em caso de dúvida sobre os aspectos éticos da pesquisa, você pode procurar o Comitê de Ética em Pesquisa da Faculdade de Ciências Farmacêuticas de Ribeirão Preto – FCFRP. Endereço do Comitê de Ética: Avenida do Café, sem número, bairro Monte Alegre, Campus USP Ribeirão Preto - Bloco H - Sala 116, CEP 14040-903 - Ribeirão Preto – SP. Horário de funcionamento: dias úteis, das 8 às 12 horas. tel. (16) 3315-4213.

Eu expliquei o propósito deste estudo ao participante. Estou certo de que ele entendeu o propósito, os procedimentos, riscos e benefícios deste estudo.

Nome do pesquisador: _____ Data: ____/____/____

Assinatura do pesquisador: _____

Li e concordo em participar da pesquisa:

Nome do participante: _____ Data: ____/____/____

Assinatura do participante: _____ RG _____

