TD

Non-invasive, Innovative and Promising Strategy for Breast Cancer Diagnosis Based on Metabolomic Profile of Urine, Cancer Cell Lines and Tissue

DOCTORAL THESIS

Catarina Grace Sousa Luís Silva
DOCTORATE IN CHEMISTRY
SPECIALITY IN BIOCHEMISTRY

UNIVERSIDADE da MADEIRA
A Nossa Universidade
www.uma.pt

March | 2019

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

Non-invasive, Innovative and Promising Strategy for Breast Cancer Diagnosis Based on Metabolomic Profile of Urine, Cancer Cell Lines and Tissue

DOCTORAL THESIS

Catarina Grace Sousa Luís Silva
DOCTORATE IN CHEMISTRY
SPECIALITY IN BIOCHEMISTRY

SUPERVISOR
José de Sousa Câmara

CO-SUPERVISOR
Helena Maria Pires Gaspar Tomás

Acknowledgments

First, I would like to express my sincere gratitude to **Professor José Câmara** and **Professor Helena Tomás**, my supervisors, for their guidance, enthusiastic encouragement and very useful constructive suggestions during the planning and development of this thesis.

I would like to thank Centro de Química da Madeira (CQM) and Universidade da Madeira (UMa) for the excellent facilites to perform this work.

To the *Fundação para a Ciência e Tecnologia (FCT), I.P.*, for the Ph.D. scholarship (ref. SFRH/BD/97039/2013), the CQM Strategic Project PEstOE/QUI/UI0674/2019 (Portuguese Government funds), the NMR Network PTNMR-2015-2016 and the research project INN-INDIGO/0001/2015. The Project PROEQUIPRAM - *Reforço do Investimento em Equipamentos e Infraestruturas Científicas na RAM* (M1420-01-0145-FEDER-000008) and *Centro de Química da Madeira - CQM*$^{+}$ (M1420-01-0145-FEDER-000005, ARDITI) funded by Madeira Regional Operational Programme (Madeira 1420), are also acknowledged.

I acknowlegde the technicians from UMa, Ana Paula Andrade and Paula Vieira for their support and assistance through the development of this thesis.

I also acknowlegde the Hemato-Oncology Unit and Blood Donors Bank from Hospital Dr. Nélio Mendonça for the urine samples supply.

I acknowlegde Dr. Joaquim Vieira, Director of Gynecology/Obstetrics Service (Breast Disease Unit) from Hospital Dr. Nélio Mendonça for the support and guidance through this work. Also, I would like to thank the administrative staff, Carmo Silva and Rosalina Pimentel for their availability. I extend my thanks to the Pathologic Anatomy Service from Hospital Dr. Nélio Mendonça for the tissue samples supply, and also to Dr. José Camacho (Director of Pathologic Anatomy Service), Dr. Filipa Capelinha and the technicians of the Macroscopy Laboratory, namely Sara Silva and Filipa Andrade for their availability in clarify my doubts and help in the collection of tissue samples.

To all my colleagues and friends, that helped with my work and also giving many happy moments inside and outside of the laboratory. A special thanks to Ivo Ferreira, Ana Olival, Joana Pinto, Carla Alves, Mara Gonçalves, Rita Castro, Pedro Silva, Carla Miguel, and João Gonçalves, for their friendship, happy moments and motivation to continue my work.

Many thanks to my friend Rosa Perestrelo for your sincere friendship all these years and for the encouragement that was crucial to achieve this goal.

Thank you Pedro Ideia for your friendship and all the happy moments, always funny, you provided through this time.

Since life is not made only from work, I wisk to thank Elmano Freitas, my swim coach for teaching that is never too late to learn. My swim partners that became good friends, Pedro Ideia, Susana Sousa Gomes, João Pedro Sousa and Mayra Santos, with whom I began my day swimming at 6 a.m., getting ready for the daily challenges.

Last but not the least, I would like to thank my family, namely parents, husband Miguel and daughters, Maria & Sofia, for supporting me throughout this journey. Thanks for being there always.

Abstract

The work presented in this thesis aimed to establish the metabolomic profile of urine and breast cancer (BC) tissue from BC patients (samples cordially provided by Funchal Hospital), in addition to BC cell lines (MCF-7, MDA-MB-231, T-47D) as a powerful strategy to identify metabolites as potential BC biomarkers, helping on the development of non-invasive approaches for BC diagnosis and management. To achieve the main goal and obtain a deeper and comprehensive knowledge on BC metabolome, different analytical platforms, namely headspace solid-phase microextraction (HS-SPME) combined with gas chromatography-quadrupole mass spectrometry (GC-qMS) and nuclear magnetic ressonance (^{1}H NMR) spectroscopy were used.

The application of multivariate statistical methods - principal component analysis (PCA) and orthogonal partial least square – discriminant analysis (OPLS-DA), to data matrix obtained from the different target samples allowed to find a set of highly sensitive and specific metabolites metabolites, namely, 4-heptanone, acetic acid and glutamine, able to be used as potential biomarkers in BC diagnosis. Significant group separation was observed in OPLS-DA score plot between BC and CTL indicating intrinsic metabolic alterations in each group. To attest the robustness of the model, a random permutation test with 1000 permutations was performed with OPLS-DA. The permutation test yielded R^2 (represents goodness of fit) and Q^2 values (represents predictive ability) with values higher than 0.717 and 0.691, respectively. Several metabolic pathways were dysregulated in BC considering the analytical approaches used. The main pathways included pyruvate, glutamine and sulfur metabolisms, indicating that there might be an association between the metabolites arising from the type of biological sample of the same donor used to perform the investigation.

The integration of data obtained from different analytical platforms (GC-qMS and ^{1}H NMR) for urinary and tissue samples revealed that five metabolites (e.g., acetone, 3-hexanone, 4-heptanone, 2-methyl-5-(methylthio)-furan and acetate), were found significant using a dual analytical approach.

Keywords: Breast Cancer; Cell lines, urine and BC tissue; Metabolomics; NMR; GC-qMS; Chemometric tools.

Resumo

O trabalho apresentado nesta tese teve como objetivo estabelecer o perfil metabolómico da urina e do tecido da mama de doentes com cancro de mama (BC) (amostras cordialmente fornecidas pelo Hospital do Funchal), além das linhas celulares de BC (MCF-7, MDA-MB-231, T -47D) como uma poderosa estratégia para identificar metabolitos como potenciais biomarcadores de BC, auxiliando no desenvolvimento de abordagens não invasivas para o diagnóstico e a gestão da patologia. Para obter um conhecimento mais profundo e abrangente do metaboloma de BC, diferentes plataformas analíticas, nomeadamente a microextração em fase sólida em modo *headspace* (HS-SPME) combinada com a cromatografia em fase gasosa acoplada à espectrometria de massa (GC-qMS) e espectroscopia de ressonância magnética nuclear (^{1}H RMN), foram usadas para atingir o objetivo principal.

A aplicação de métodos estatísticos multivariados - análise de componentes principais (PCA) e análise discriminante de mínimos quadrados parciais ortogonais (OPLS-DA) à matriz de dados obtida a partir das diferentes amostras alvo, permitiu estabelecer um grupo de metabolitos sensíveis e específicos, nomeadamente a 4-heptanona, o ácido acético e a glutamina, possíveis de serem utilizados como potenciais biomarcadores no diagnóstico de BC. Uma separação significativa entre os grupos BC e CTL foi observada pelo OPLS-DA, indicando alterações metabólicas em cada grupo. Para verificar a robustez do modelo, foi realizado um teste de permutação aleatória com 1000 permutações com o sistema OPLS-DA. Valores de R^2 (representa o ajuste) e Q^2 (representa a capacidade preditiva) superiores a 0,717 e 0,691, foram obtidos utilizando o teste da permutação. Diversas vias metabólicas estavam desreguladas no BC considerando as abordagens analíticas utilizadas. As principais vias incluíram os metabolismos do piruvato e glutamina, indicando que poderá haver uma associação entre os metabolitos derivados do tipo de amostra biológica do mesmo doador utilizado para realizar a investigação.

A integração de dados obtidos pelas diferentes plataformas analíticas (GC-qMS e ^{1}H RMN) para amostras urinárias e de tecido revelou cinco metabolitos significativos usando a dupla abordagem analítica. (i.e., acetona, 3-hexanona, 4-heptanona, 2-metil-5- (metiltio) - furano e acetato).

Palavras-chave: Cancro da Mama; Linhas celulares, urina e tecido; Metabolómica; NMR; GC-qMS; Ferramentas estatísticas.

AA – Acetaldehyde

ACS – American cancer society

ADH – Alcohol dehydrogenase

ANOVA – Analysis of variance

ASCO – American society of clinical oncology

ATM – Ataxia-telangiectasia mutated

AUC – Area under the curve

BC – Breast cancer

BRCA1 – Mutation on chromosome 17

BRCA2 – Mutation on chromosome 13

CA 15.3 – Cancer antigen 15.3

CA 27.29 – Cancer antigen 27.29

CAE – Carcinoembryonic antigen

CC – Colon cancer

CCD – Central composite design

CE-MS – Capillary electrophoresis mass spectrometry

CI – Chemical ionization

CTLs –-Controls

DART-MS – Direct analysis in real time mass spectrometry

DI-ESI-MS – Direct infusion electrospray ionization mass spectrometry

DCIS – Ductal carcinoma in situ

DNA – Deoxyribonucleic acid

EBC – Early breast cancer

EI – Electron ionization

EMs – Endogenous metabolites

ER – Estrogen receptors

FDA – Food and drug administration

FO – Frequency of occurence

GC × GC-TOFMS – Comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry

GPC – Glycerophosphocholine

GSH – Glutathione sulfide

GSSG – Glutathione disulfide

HCA – Hierarchical cluster analysis

HER2 – Human epidermal growth factor

HMEC – Human mammary epitelial cells

HR-MAS – High-resolution magic angle spinning

HR-MAS-NMR – High-resolution magic angle spinning nuclear magnetic ressonance

HS-SPME – Headspace solid-phase microextration

IARC – International agency for research on cancer

IDC – Infiltrating ductal carcinoma

IGF – Insulin-like growth factor

IP3R – Inositol 1, 4, 5-trisphosphate receptor

KS-test – Kolmogorov-Smirnov test

KEGG – Kyoto encyclopedia of genes and genomes

LC – Liquid chromatography

LC-FT-ICR-MS – Liquid chromatography/Fourier transform ion cyclotron resonance mass spectrometry

LCIS – Lobular carcinoma in situ

LC-MS-NMR – Liquid chromatography-mass spectrometry-nuclear magnetic ressonance spectroscopy

LDA – Linear discriminant analysis

MALDI-FT-ICR-MS – Matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry

MANOVA – Multivariate analysis of variance

MBC – Metastatic breast cancer

MCF-7 – Human breast cancer cell line

MDA-MB-231 – Human breast cancer cell line

MLR – Multiple linear regression

MRI – Magnetic ressonance imaging

MS – Mass spectrometry

MUC1 – Gene mucin 1

NMR – Nuclear magnetic ressonance

OLS – Ordinary least squares

OPLS-DA – Orthogonal projections to latent structures discriminant analysis

PAI-1 – Plasminogen activator inhibitor 1

PCA – Principal component analysis

PCho – Phosphocholine

pCR – Pathological complete response

PLS-DA – Partial least square discriminant analysis

PR – Progesterone receptors

PTEN – Phosphatase and tensin homolog deleted on chromosome ten

RF – Random forest

RNA – Ribonucleic acid

RSM – Response surface methodology

SVM – Support vector machines

T-47D – Human breast cancer cell line

TOF – Time-of-flight

TP53 – tumor protein p53

TSP – 3-(trimethylsilyl)propionic-2,2,3,3-*d4* acid sodium salt

uPa – Urokinase plasminogen activator

VOMs – Volatile organic metabolites

Table of Contents

List of Figures

SECTION 1

SECTION 2

SECTION 3

SUB-SECTION 3.1

SUB-SECTION 3.2

SUB-SECTION 3.3

SUB-SECTION 3.4

SUB-SECTION 3.5

List of Tables

SECTION 1

SECTION 3

SUB-SECTION 3.1

SUB-SECTION 3.2

SUB-SECTION 3.3

SUB-SECTION 3.4

SUB-SECTION 3.5

Catarina Luís Silva is a Ph.D. student at Universidade da Madeira (UMa) and Centro de Química da Madeira (CQM).

After her graduation in Biochemistry in 2004 (Faculty of Sciences, University of Porto – UP), she continued her studies and got a Master degree in Applied Biochemistry (UMa) in 2009, where she established the volatile metabolic profile of potential biomarkers in individuals with oncologic pathologies, under the supervision of Professor José Câmara.

In 2014, she was awarded a Ph.D. research grant by *Fundação para a Ciência e Tecnologia* (FCT), and since then she have been developing her experimental work at CQM, under the supervision of Professor José Câmara and co-supervision of Professor Helena Tomás.

During her Ph.D., she participated in several courses and scientific meetings. She also contributed for many research projects and activities developed in CQM.

As a result of her scientifc activity, she is author/co-author of 28 SCI papers (H-index: 12; Number of scopus citations: 458) and 4 book chapters in international peer-reviewed journals. Also, her work resulted in 4 oral communications and more than 40 poster presentations in International and National conferences.

Participation In Research Projects

HCV project (Inn-Indigo/0001/2015) – "Exploring the volatome of noncommunicable diseases as a promising innovative and integrating approach for its rapid diagnostics. The case study of cancer and neurodegenerative diseases". From June 2016 to the present.

New Indigo project - "An attractive and promising strategy for early cancer diagnosis through the assembly of the Human Cancer Volatome" (New-Indigo/0003/2012). Research Grant (BI-Master) from October 2013 to September 2014.

Project Interreg VinSaudeMAC - "Evaluation of the biological effects of red wines in the prevention of arteriosclerotic, oncologic and neurologic pathologies – establishing of winemaking processes to improve wine antioxidant composition" code MAC/1/M105, co-financed by Program MAC 2007-2013 Transnational Cooperation. Research Grant from October 2009 to May 2013.

Project Interreg ANTIVINMAC - "Determination of the antioxidant potential, aminoacids and biogenic amines in wines from Canary, Azores and Madeira Islands". Research Grant from October 2009 to May 2013. Research Grant from October 2007 to October 2009.

Participation In Advance Courses

- 10th CERMAX practical course on basic NMR, ITQB, Lisbon, Portugal: July **2017**.
- III Workshop on Cancer Research, IPATIMUP, Porto, Portugal: May **2013**.
- Intensive course on "Empower software – Level I". Via Athena, Gestão de Laboratórios, Lda., May **2010,** Sacavém, Portugal

Publications

Papers in international scientific periodicals with referees

1. Silva C., Perestrelo R., Silva P., Capelinha F., Tomás H., Câmara J.S.; Integrated metabolomics based on GC-MS and NMR data as powerful strategy to search potential breast cancer biomarkers (*submitted to European Journal of Cancer, EJC-D-19-00135*).

2. Silva C., Olival A., Perestrelo R., Silva P., Tomás H., Câmara J.S.; ^{1}H NMR metabolomic urinary profile of breast cancer: a feasibility study (*submitted to Metabolites, Special issue Cancer Metabolomics 2018, 451707*).

3. Silva C., Perestrelo R., Silva P., Tomás H., Câmara J.S.; Breast cancer metabolomics: from analytical platforms to multivariate data analysis. A Review. *Metabolites* **2019**, Special issue *Cancer Metabolomics 2018*, 9, 102-133).

4. Silva C., 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 volatomic expression. A first approach for breast cancer. *Metabolomics* **2019***;*15: 1-13.

5. Silva C., Perestrelo R., Silva P., Capelinha F., Tomás H., Câmara J.S.; Volatomic pattern of breast cancer tissue as a powerful strategy to identify potential biomarkers. *Analyst* **2019***;* 144: 4153-4161.

6. Perestrelo R., Silva P., Porto-Figuerira P., Pereira J., Silva C., Medina S., Câmara J.S.; QuEChERS - fundamentals, relevant improvements, applications and future trends. *Analytica Chimica Acta* **2019**; 1070: 1-28.

7. Perestrelo R., Silva C., Silva P., Medina S., Câmara J.S.; Differentiation of Fresh and Processed Fruit Juices using Volatile Composition. *Molecules* **2019**; 24, 974.

8. Perestrelo R., Silva C., Silva P., Medina S., Câmara J.S.; Untargeted fingerprinting of cider volatiles from different geographical regions by HS-SPME/GC-MS. *Michrochemical Journal* **2019**; 148: 643-651.

9. Perestrelo R., Silva C., Silva P., Câmara J.S.; Unraveling Vitis vinifera L. grape maturity markers based on integration of terpenic pattern and chemometric methods. *Microchemical Journal* **2018**; 142:367-376.

10. Silva P., Silva C., Perestrelo R., Nunes F., Câmara J.S.; Fingerprint targeted compounds in authenticity of sugarcane honey - An approach based on chromatographic and statistical data. *LWT* **2018**; 96:82-89.

11. Perestrelo R., Silva C., Silva P., Câmara J.S.; Rapid spectrophotometric methods as a tool to assess the total phenolics and antioxidant potential over grape ripening: a case study of Madeira grapes. *Journal of Food Measurement and Characterization* **2018;** 12:1754-1762.

12. Perestrelo R., Silva C., Silva P., Câmara J.S.; Impact of storage time and temperature on volatomic signature of Tinta Negra wines by LLME/GC-ITMS. *Food Research International* **2018**; 109:99-111.

13. Casado N., Perestrelo R., Silva C.L., Sierra I., Câmara, J.S.; An improved and miniaturized analytical strategy based on μ-QuEChERS for isolation of polyphenols. A powerful approach for quality control of baby foods. *Microchemical Journal* **2018**; 139:110-118.

14. Perestrelo R., Silva C., Silva P., Câmara J.S.; Establishment of the volatile signature of wine-based aromatic vinegars subjected to maceration. *Molecules* **2018**; 23:499.

15. Silva C., Perestrelo R., Silva P., Tomás H., Câmara J.; Volatile metabolomic signature of human breast cancer cell lines. *Scientific Reports* **2017**; 7:43969.

16. Silva P., Silva C., Perestrelo R., Nunes F., Câmara J.S.; A useful strategy based on chromatographic data combined with quality-by-design approach for food analysis applications. The case study of furanic derivatives in sugarcane honey. *Journal of Chromatography A* **2017**; 150:117-126.

17. Perestrelo R., Silva C., Silva P., Câmara J.S.; Global volatile profile of virgin olive oils flavoured by aromatic/medicinal plants. *Food Chemistry* **2017**; 227:111-121.

18. Silva P., Freitas J., Silva C., Perestrelo R., Nunes F., Câmara J.S.; Establishment of authenticity and typicality of sugarcane honey based on volatile profile and multivariate analysis. *Food Control* **2017**; 73:1176-1188.

19. Ruiz-García Y., Silva C.L., Gómez-Plaza E., Câmara J.S.; A Powerful Analytical Strategy Based on QuEChERS-Dispersive Solid-Phase Extraction Combined with Ultrahigh Pressure Liquid Chromatography for Evaluating the Effect of Elicitors on Biosynthesis of trans-Resveratrol in Grapes. *Food Analytical Methods* **2016**; 9:670-679.

20. Perestrelo R., Silva C., Rodrigues F., Caldeira M., Câmara J.S.; A powerful approach to explore the potential of medicinal plants as a natural source of odor and antioxidant compounds. *Journal of Food Science and Technology* **2016**; 53:132-144.

21. Perestrelo R., Silva C.L., Câmara J.S.; Determination of urinary levels of leukotriene B4 using ad highly specific and sensitive methodology based on automatic MEPS combined with UHPLC-PDA analysis. *Talanta* **2015**; 144:382-389.

22. Perestrelo R., Silva C.L., Câmara J.S.; Quantification of furanic derivatives in fortified wines by a highly sensitive and ultrafast analytical strategy based on digitally controlled microextraction by packed sorbent combined with ultrahigh pressure liquid chromatography. *Journal of Chromatography A* **2015**; 1381:54-53.

23. Gonçalves J.L., Figueira J.A., Rodrigues F.P., Ornelas L.P., Branco R.N., Silva C.L., Câmara J.S.; A powerful methodological approach combining headspace solid phase microextraction, mass spectrometry and multivariate analysis for profiling the volatile metabolomic pattern of beer starting raw materials. *Food Chemistry* **2014**; 160:266-280.

24. Silva C., Cavaco C., Perestrelo R., Pereira J., Câmara J.S.; Microextraction by packed Sorbent (MEPS) and solid-phase microextraction (SPME) as sample preparation procedures for the metabolomic profiling of urine. *Metabolites* **2014**; 4:71-97.

25. Perestrelo R., Silva C., Câmara J.S.; A useful approach for the differentiation of wines according to geographical origin based on global volatile patterns. *Journal of Separation Science* **2014**; 37:1974-1981.

26. Pereira J., Silva C.L., Perestrelo R., Gonçalves J., Alves V., Câmara J.S.; Re-exploring the high-throughput potential of microextraction techniques, SPME and MEPS, as powerful strategies for medical diagnostic purposes. Innovative approaches, recent applications and future trends Microextraction Techniques. *Analytical and Bioanalytical Chemistry* **2014**; 406:2101-2122.

27. Silva C.L., Câmara J.S.; Profiling of volatiles in the leaves of Lamiaceae species based on headspace solid phase microextraction and mass spectrometry. *Food Research International* **2013**; 51:378-387.

28. Gonçalves J., Silva C.L., Castilho P.C., Câmara J.S.; An attractive, sensitive and high-throughput strategy based on microextraction by packed sorbent followed by UHPLC-PDA analysis for quantification of hydroxybenzoic and hydroxycinnamic acids in wines. *Microchemical Journal* **2013**; 106:129-138.

29. Silva C.L., Haesen N., Câmara J.S.; A new and improved strategy combining a dispersive-solid phase extraction-based multiclass method with ultra high pressure liquid chromatography for analysis of low molecular weight polyphenols in vegetables. *Journal of Chromatography A* **2012**; 1260:154-163.

30. Silva C.L., Gonçalves J.L., Câmara J.S.; A sensitive microextraction by packed sorbent-based methodology combined with ultra-high pressure liquid chromatography as a powerful technique for analysis of biologically active flavonols in wines. *Analytica Chimica Acta* **2012**; 739: 89-98.

31. Gonçalves J., Mendes B., Silva C.L., Câmara J.S.; Development of a novel microextraction by packed sorbent-based approach followed by ultrahigh pressure liquid chromatography as a powerful technique for quantification phenolic constituents of biological interest in wines. *Journal of Chromatography A* **2012**; 1229:13-23.

32. 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* **2012**; 89:360-368.

33. Silva C.L., Passos M., Câmara 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* **2011**; 105:1894.

34. Silva C.L., Pereira J., Wouter V.G., Giró C., Câmara J.S.; A fast method using a new hydrophilic-lipophilic balanced sorbent in combination with ultra-high performance liquid chromatography for quantification of significant bioactive metabolites in wines. *Talanta* **2011**; 86:82-90.

35. Mateus N., Carvalho E., Luís C., Freitas V.; Influence of the tannin structure on the disruption effect of carbohydrates on protein–tannin aggregates. *Analytica Chimica Acta* **2004**; 513:135-140.

Book Chapters

1. Perestrelo R., Silva C.L., Pereira J., Câmara J.S., "Wines; Madeira, Port and Sherry Fortified Wines - The Sui Generis and Notable Peculiarities. Major Differences and Chemical Patterns; Encyclopedia of Food and Health, **2016**, 534-555; ISBN: 978-0-12-384953-3; p534-555.

2. Perestrelo R., Silva C.L., Pereira J., Câmara J.S., "Healthy effects of bioactive metabolites from Vitis Vinifera L. grapes: a review", Nova Science Publishers, New York, Chapter 13, **2014**; ISBN 978-1-63321-410-1.

3. Pereira J., Silva C.L., Gonçalves J., Câmara J.S., The Potentialities of Ultra-Performance Liquid Chromatography Combined with Photodiode Array Detection in the Analysis of Wine Metabolites with Bioactive Effects: The Case Study of Polyphenols, in Wine: Types, Production and Health; Series: Food Science and Technology, Food and Beverage Consumption and Health, Editor Arthur S. Peeters, Nova Science Publishers; ISBNs: 978-1-61470-635-9 (hardcover), 978-1-61470-804-9 (ebook).

4. Pereira J., Gonçalves J., Silva C.L., Mendes B., Silva P., Alves V., Câmara J.S., Metabolomic applications of liquid chromatography - from food bioactive metabolites to disease biomarkers

research in Liquid Chromatography: Principles, Technology and Applications, N. Publishers, Editor. **2012**, Nova Science Publishers; ISBNs: 978-1-62618-739-9

Oral communications in International meetings

1. Silva C.L., Olival A., Perestrelo R., Silva P., Tomás H., Câmara J.S., ^{1}H NMR metabolomic urinary profile of breast cancer, 3rd International Symposium in profiling, September **2017**, Costa da Caparica, Portugal.

2. Silva P., Freitas J., Perestrelo R., Silva C.L., Nunes F., Câmara J.S., Development of an effective tool based on furanic profile to evaluate the quality and prevent frauds of sugarcane honey, 31st International Symposium on Chromatography, August-September **2016**, Cork, Ireland.

3. Silva C.L., Tomás H., Câmara J.S., Metabolomic approach for early breast cancer diagnosis based on urinary volatile profile, Urinomics & Nephromics 2015, September **2015**, Costa da Caparica, Portugal.

4. Silva C.L., Tomás H., Câmara J.S., The potential of urinary volatile metabolites as a non-invasive, innovative and promising strategy for early diagnosis of cancer, 2nd HCV Project Meeting, November **2014**, Funchal, Portugal.

Oral communications in CQM meetings

1. Silva C.L., Olival A., Perestrelo R., Silva P., Tomás H., Câmara J.S., ^{1}H NMR Metabolomic Urinary Profile of Breast Cancer, CQM Annual Meeting, February **2017**, Funchal, Portugal.

2. Silva C.L., Tomás H., Câmara J.S., Volatomic fingerprint of breast cell lines, CQM Annual Meeting, April **2016**, Funchal, Portugal.

3. Silva C.L., Perestrelo R., Câmara J.S., An improved and highly sensitive procedure based on microextraction by packed sorbent (MEPS) combined with ultra high pressure liquid chromatography for quantification of urinary levels of leukotriene B4, CQM Annual Meeting, February **2014**, Funchal, Portugal.

Poster communications

1. Silva P., Freitas J., Perestrelo R., Silva C.L., Nunes F., Câmara J.S., Volatile profile of sugarcane honey at different stages of processing and storage. The influence of browning reactions and microbial activity on the typicality of Sugarcane Honey, 32st EFFoST International Conference, November 2018, Nantes, France.

2. Silva P., Freitas J., Perestrelo R., Silva C.L., Nunes F., Câmara J.S., An useful strategy based on MEPS/UHPLC combined with Quality-by-Design approach for Food Safety applications. The case study of phenolic acids in sugarcane honey, 32st EFFoST International Conference, November 2018, Nantes, France.

3. Silva P., Freitas J., Perestrelo R., Silva C.L., Nunes F., Câmara J.S., Fingerprint targeted compounds for use in authenticity of sugarcane honey – na approach based on chromatographic and statistical data, 10º Encontro Nacional de Cromatografia, December 2017, Bragança, Portugal.

4. Silva P., Freitas J., Perestrelo R., Silva C.L., Nunes F., Câmara J.S., An useful strategy based on MEPS/UHPLC combined with Qaulity-by-design approach for Foodomocs applications. The case study of furanic derivatives in sugarcane honey, 45th International Symposium oh High Performance Luiquid Phase Separations, HPLC 2017, June 2017, Prague, Czech Republic.

5. Silva C.L., Olival A., Perestrelo R., Silva P., Tomás H., Câmara J.S., 1H NMR metabolomic urinary profile of breast cancer, 3rd International Symposium in profiling, September 2017, Costa da Caparica, Portugal.

6. Silva P., Freitas J., Perestrelo R., Silva C.L., Nunes F., Câmara J.S., Development of an effective tool based on furanic profile to evaluate the quality and prevent frauds of sugarcane honey, 31st International Symposium on Chromatography, August-September 2016, Cork, Ireland.

7. Silva C.L., Tomás H., Câmara J.S., Metabolomic approach for early breast cancer diagnosis based on urinary volatile profile, Urinomics & Nephromics 2015, September 2015, Costa da Caparica, Portugal.

8. Silva C.L., Tomás H., Câmara J.S., The potential of urinary volatile metabolites as a non-invasive, innovative and promising strategy for early diagnosis of cancer, 2nd HCV Project Meeting, November, 2014, Funchal, Portugal.

9. Silva P., Freitas J., Perestrelo R., Silva C.L., Nunes F., Câmara J.S., An useful strategy based on MEPS/UHPLC combined with quality by-design approach for foodomics applications. The case study of furanic derivatives in sugarcane honey, 45th International Symposium on High Performance Liquid Phase Separations, June 2017, Prague, Czech Republic.

10. Silva P., Freitas J., Perestrelo R., Silva C.L., Câmara J.S., Furanic profile of sugarcane honey as an effective tool to evaluate quality and prevent adulteration, 9º Encontro Nacional de Cromatografia/ XVI Congresso Latino-Americano de Cromatografia, January 2016, Portugal.

11. Silva P., Freitas J., Perestrelo R., Silva C.L., Câmara J.S., The highthrouput potential of chromatography on food authenticity establishment. The case study of sugarcane honey, 9º Encontro Nacional de Cromatografia/ XVI Congresso Latino-Americano de Cromatografia, January 2016, Portugal.

12. Silva C.L., Tomás H., Câmara J.S., Volatomic signature of human breast cancer cell lines by solid-phase microextraction combined with gas chromatography-mass spectrometry, 31st International Symposium on Chromatography 2016, Cork, Ireland.

13. Silva P., Freitas J., Perestrelo R., Silva C.L., Nunes F., Câmara J.S., Volatomic pattern and multivariate analysis as useful approach to establish the authenticity and typicity of sugarcane honey, 1st Food Chemistry Conference, October 2016, Amsterdam, The Netherlands.

14. Silva P., Freitas J., Perestrelo R., Silva C.L., Nunes F., Câmara J.S., Furanic profiling as an effective tool to establishment of authenticity and quality of sugarcane honey, 1st Food Chemistry Conference, October 2016, Amsterdam, The Netherlands.

15. Silva P., Silva C.L., Caldeira M., Aveiro F., Câmara J.S., Profiling urinary volatile metabolites as a non-invasive and powerful strategy to detect potential lung cancer-specific biomarkers, 31st International Symposium on Chromatography 2016, Cork, Ireland.

16. Silva C.L., Tomás H., Câmara J.S., A combinatory approach based on MS and metabolomic data as a powerful strategy for identification of urinary volatile metabolites as potential breast cancer biomarkers, 31st International Symposium on Chromatography 2016, Cork, Ireland.
17. Silva C.L., Tomás H., Câmara J.S., Metabolomic approach for early breast cancer diagnosis based on urinary volatile profile, Urinomics & Nephromics, September 2015, Costa da Caparica, Portugal.

18. Cavaco C., Aveiro F., Silva C.L, Pereira J., Câmara J.S., "Establishment of the saliva volatomic profile of breast cancer patients", ISC 2014, 30th International Symposium on Chromatography, September 2014, Salzburg, Austria.

19. Silva C.L., Aveiro F., Câmara J.S., "Non-invasive, innovative and promising strategy for early diagnosis of breast cancer based on urinary volatile metabolites", ISC 2014, 30th International Symposium on Chromatography, September 2014, Salzburg, Austria.

20. Silva C.L., Câmara J.S., "Integration of solid phase microextraction, mass spectrometry and metabolomic data as a powerful strategy for identification of urinary volatile metabolites as potential cancer biomarkers", 38th International Symposium on Capillary Chromatography and 9th GCxGC Symposium, May 2014, Riva del Garda, Italy.

21. Perestrelo R., Silva C.L., Câmara J.S., "A new, reliable and high throughput analytical approach for the simultaneous determination of furans in wine matrices", ExTech 2014, 16th International Symposium on Advances in Extraction Technologies, May, 2014, Chania, Crete, Greece.

22. Câmara J.S., Perestrelo R., Silva C.L., "An unexplored strategy based on semi-automatic MEPS procedure followed by UHPLC-PDA as an highly sensitive and specific methodology to quantify the urinary levels of leukotriene B4", 38th International Symposium on Capillary Chromatography and 9th GCxGC Symposium, May, 2014, Riva del Garda, Italy.

23. Perestrelo R., Silva C.L., Câmara J.S., An improved and highly sensitive procedure based on microextraction by packed sorbent (MEPS) combined with ultrahigh pressure liquid chromatography for quantification of urinary levels of leukotriene B4", 8º Encontro Nacional de Cromatografia, December, 2013, Portugal.

24. Perestrelo R., Silva C.L., Pereira J., Câmara J.S., "Exploring the potential of microextraction by packed sorbent (MEPS) combined with ultrahigh pressure liquid chromatography to quantify the urinary leukotriene B4 levels in asthmatic patients", 39th International Symposium on High-Performance-Liquid-Phase Separations and Related Techniques (HPLC 2013), June, 2013, Amsterdam.

25. Silva C.L., Castro R., Tomás H., Câmara J.S., "Effect of red wine bioactive polyphenols in Caco-2 cell line viability", XXII Porto Cancer Meeting & Porto-Bordeaux Joint Meeting, 11-12 April 2013, IPATIMUP, Porto, Portugal.

26. Silva C.L., Aveiro F., Câmara J.S., "The potential of urinary volatile metabolites as a non-invasive, innovative and promising strategy for early diagnosis of cancer", XXII Porto Cancer Meeting & Porto-Bordeaux Joint Meeting, 11-12 April 2013, IPATIMUP, Porto, Portugal.

27. Silva C.L., Silva P., Aveiro F., Câmara J.S., "Urinary metabolomic pattern as a powerful mean to finding potential lung cancer biomarkers", XXII Porto Cancer Meeting & Porto-Bordeaux Joint Meeting, 11-12 April 2013, IPATIMUP, Porto, Portugal.

28. Silva C.L., Haesen N., Câmara J.S.," A new and improved strategy combining a modified QuEChERS-based multi-residue method with dispersive-solid phase extraction followed by ultrahigh pressure liquid chromatography for analysis of polyphenols in vegetables", 36th International Symposium on Capillary Chromatography, 27th May-June 1st, 2012, Riva del Garda, Italy.

29. Pereira J., Mendes B., Silva P., Silva C.L. and Câmara J.S., "Semi-automatic micro-extraction in packed syringe combined with ultrahigh pressure liquid chromatography as a fast and simplified methodology to assess oxidative damage through quantification of 8-iso-prostaglandin F2α biomarker in urine", 36th International Symposium on Capillary Chromatography, 27th May-June 1st, 2012, Riva del Garda, Italy.

30. Mendes B., Silva P., Pereira J., Silva C.L., Câmara J.S., "A powerful and sensitive MEPS/UHPLC-PDA-based methodology for assessment of urinary biomarkers of oxidative DNA damage", 36th International Symposium on Capillary Chromatography, 27th May-June 1st, 2012, Riva del Garda, Italy.

31. Figueira J.A., Gonçalves J., Rodrigues F., Ornelas L., Branco N., Silva C.L., Câmara, J.S. "Profiling the volatile metabolomics pattern of beer starting raw-materials based on headspace solid phase microextraction, mass spectrometry and multivariate analysis", 36th International Symposium on Capillary Chromatography, 27th May-June 1st, 2012, Riva del Garda, Italy.

32. Figueira J., Câmara H., Silva C.L., Câmara J.S., "A powerful strategy coupling headspace solid phase microextraction and mass spectrometry to profile the volatile metabolome pattern of different Lycopersicon esculentum L. Cultivars", 36th International Symposium on Capillary Chromatography, 27th May-June 1st, 2012, Riva del Garda, Italy.

33. Rodrigues F., Gonçalves J., Figueira J.A., Câmara H., Figueira P., Silva C.L., Ornelas L., Branco N., Câmara J.S. "Optimization and validation of a new strategy coupling a semi-automatic micro-extraction by packed syringe and ultrahigh pressure liquid chromatography for determination of xanthohumol and Isoxanthohumol in beer", 36th International Symposium on Capillary Chromatography, 27th May-June 1st, 2012, Riva del Garda, Italy.

34. Silva P., Nunes A., Silva C.L., Aveiro F., Câmara J.S. "An exploratory study based on mass spectrometry metabolomics to establish the urine molecular pattern in lung cancer patients", 36th International Symposium on Capillary Chromatography, 27th May-June 1st, 2012, Riva del Garda, Italy.

35. Silva P., Nunes A., Silva C.L., Aveiro F., Câmara J.S. "Comparison on volatile urinary metabolomic profile of lung cancer patients and normal individuals as a powerful tool to finding potential biomarkers", 36th International Symposium on Capillary Chromatography, 27th May-June 1st, 2012, Riva del Garda, Italy.

36. Silva C.L., Passos M., Câmara J.S. "Solid phase microextraction, mass spectrometry and metabolomic approaches as powerful strategies for detection of urinary volatile metabolites as potential cancer biomarkers", 7th National Meeting on Chromatography, 9-11th January 2012, Porto, Portugal.

37. Silva C.L., Gonçalves J., Câmara J.S. "Efficient and high throughput Microextraction by Packed Sorbent-Based Methodology Combined with UHPLC-PDA as a Powerful Technique for Analysis of Biologically Active Flavonols in Wines", 7th National Meeting on Chromatography, 9-11th January 2012, Porto, Portugal.

38. Haesen N., Silva C.L., Câmara J.S. "New strategy to enhance the extraction efficiency of polyphenols from dietary vegetables using a modified QuEChERS combined with UHPLC-PDA system", 7th National Meeting on Chromatography, 9-11th January 2012, Porto, Portugal;

39. Silva C.L., Pereira J., Wouter V.G., Giró C., Câmara J.S. "A new hydrophilic-lipophilic balanced sorbent combined with ultra-high performance liquid chromatography for quantification of bioactive phenolics in wines". 5th International Conference on Polyphenols and Health (ICPH2011), 17th-20th October 2011, Barcelona, Spain.

40. Gonçalves J., Mendes B., Silva C.L., Câmara J.S. "A novel microextraction by packed sorbent-based methodology combined with ultra-high performance liquid chromatography for the determination of significant bioactive metabolites in wines", Euroanalysis, 11-15th September 2011, Belgrade, Serbia.

41. Gonçalves J., Mendes B., Silva C.L., Câmara J.S. "Microextraction in packed sorbents (MEPS) combined with ultra-performance liquid chromatography for the determination of significant bioactive metabolites in wines", XXXIV World Congress of Vine and Wine (The wine construction) 20-27th June 2011, Porto, Portugal.

42. Silva C.L., Pereira J., Wouter V.G., Giró C., Câmara J.S. "Characterization of wine polyphenols using reverse solid-phase extraction and ultra performance chromatography", XXXIV World Congress of Vine and Wine (The wine construction) 20-27th June 2011, Porto, Portugal.

43. Miguel C., Camacho J., Craveiro P., Silva C.L., Pereira J., Câmara J.S. "High throughput SPE/UPLC-PDA-based methodology for the simultaneous determination of bioactive phenolic metabolites in food dietary products." HPLC2011, 19th-23rd June 2011, Budapest, Hungary;

44. Pereira J., Silva C.L., Câmara J.S. "High throughput methodology to characterize red wine polyphenols using solid phase extraction combined with ultra-performance liquid chromatography." HPLC 2011, 19th-23rd June 2011, Budapest, Hungary;

45. Silva C.L., Pereira J., Wouter V.G., Giró C., Câmara J.S. "A reverse-phase SPE using a new hydrophilic-lipophilic balanced sorbent combined with an ultra-fast technique for the determination of significant bioactive phenolic compounds in wines", ExTech 2010, 20-22th September 2010, Poznań, Poland;

46. Silva C.L., Câmara J.S. "One-dimensional gas chromatography followed by mass spectrometry as a powerful approach to detect cancer volatile biomarkers", 34th International Symposium on Capillary Chromatography, 1st-4th June 2010, Riva del Garda, Italy;

47. Silva C.L., Câmara J.S. "Detection of urinary volatile metabolites in breast cancer patients by the use of solid phase microextraction and gas chromatography-mass spectrometry", 34th International Symposium on Capillary Chromatography, 1st-4th June 2010, Riva del Garda, Italy;

48. Silva C.L., Perestrelo R., Câmara J.S. "Analysis of urinary volatile metabolites as potential breast cancer biomarkers using HS-SPME/GC-qMS", 2nd Portuguese Young Chemists Meeting, 21-23rd April 2010, Aveiro;

49. Luís C., Câmara J.S. "Headspace solid phase microextraction-gas chromatography-mass spectrometry based metabolomic approaches in urinary biomarker study of cancers", 6th National Meeting on Chromatography, 14-16th December 2009, Funchal;

50. Luís C., Câmara J.S. "Characterization of volatiles in individuals with breast cancer by HS-SPME-GC-qMSD", 6th National Meeting on Chromatography, 14-16th December 2009, Funchal;

51. Perestrelo R., Luís C., Câmara J.S. "Differentiation of wines from different geographic regions based on volatile profile", 6th National Meeting on Chromatography, 14-16th December 2009, Funchal;

52. Luís C., Câmara J.S. "Monitorization of volatile compounds as potential biomarkers in Individuals with Oncologic Pathologies by HS-SPME-GC-qMSD", 6th National Meeting on Chromatography, 14-16th December 2009, Funchal;

53. Luís C., Perestrelo R., Câmara J.S. "Characterization of volatile biomarkers in individuals with oncologic pathologies by HS-SPME-qMSD", EUROANALYSIS 2009, 6-10th September 2009, Innsbruck, Austria;

54. Luís C., Perestrelo R., Pérez-Pont M., Trijullo J.P., Marques J.C., Álvares M., Conde J.E., Câmara J.S. "Solid-Phase Extraction Methodology for Wines Differentiation from Different Geographic Regions", In Vino Analytica Scientia, 2-4th July 2009, Angers, France;

55. Luis C., Perestrelo R.., Pérez-Pont M., Pedro J., Marques J.C., González J.E., Câmara J.S. "Caracterização do perfil volátil de vinhos produzidos nas ilhas atlânticas usando a extracção em fase sólida seguida de GC-qMS", 9th meeting on Food Chemistry, 29 April-2nd May 2009, Angra do Heroísmo, Açores;

56. Luís C., Câmara J.S. "Characterization of Volatile Biomarkers in Individuals with Oncologic Pathologies by HS-SPME-GC-qMSD", 1st Portuguese Young Chemists Meeting, 15-17th October 2008, Lisbon;

57. Luís C., Perestrelo R., Pérez-Pont M., Pedro J., Marques J.C., González J.E., Câmara J.S. "Characterization of volatile profile in different wines from Madeira, Azores and Canary by solid phase extraction combined with GC-qMSD", ISC 2008 - 27th International Symposium on Chromatography, 21-25th September 2008, Münster, Germany;

58. Perestrelo R., Luís C., Caldeira M., Câmara J.S. "Analysis of volatiles in commercial olive oil using LLE/GC-MS and HS-SPME/GC-MS", XXI National Meeting SPQ, Chemistry and Inovation, June 2008, Engineering Faculty, University of Porto;

59. Luís C., Rodrigues F., Caldeira M., Paixão N., Câmara J.S. "Screening of the volatile composition of four medicinal plants by HS-SPME-GC-qMS: Thymus vulgaris, Mentha piperita, Mentha spicata and Rosmarinus officinalis", 31st Symposium on Capillary Chromatography & Electrophoresis, November 2007; Albuquerque, New Mexico, USA;

60. Luís C., Rodrigues F., Caldeira M., Marques J.C., Câmara J.S. "Screening of the volatile profile in medicinal plants: Mentha piperita and Mentha spicata", 5th National Meeting on Chromatography, December 2007; Aveiro;

61. Perestrelo R., Luís C., Rodrigues F., Caldeira M., Câmara J.S. "Estimation of volatile composition in Ruta chalepensis plant", 5th National Meeting on Chromatography, December 2007; Aveiro;

62. Caldeira M., Rodrigues F., Perestrelo R., Luís C., Câmara J.S. "Identification of compounds with antioxidant potential in several plants", 5th National Meeting on Chromatography, December 2007; Aveiro;

63. Perestrelo R., Luís C., Rodrigues F., Caldeira M., Câmara J.S. "Characterization of madeira endemic plant Ruta chalepensis by HS-SPME-GC-qMS and UV-Vis", 31st Symposium on Capillary Chromatography & Electrophoresis, November 2007, USA;

64. Mateus N., Carvalho E., Luís C., Freitas V. "Nephelometric study of the influence of carbohydrates on the interaction between BSA and procyanidin oligomers", In vino Analytica Scientia, July 2003, Aveiro.

Other activities

- ✓ Participation in several science dissemination activities carried out in the scope of CQM, such as, *A Química é Divertida* ("Chemistry is fun"). Collaborator in the period 2007-2018.
- ✓ Member of the Organizing Committee of Mad-Nano18: Madeira International conference on Emerging Trends and Future of Nanomaterials for Human Health, December 2018, Funchal, Portugal.

- ✓ Member of the Organizing Committee of Mad-Nano16: Madeira International conference on Emerging Trends and Future of Nanomaterials for Human Health, November 2016, Funchal, Portugal.

Outline of the Thesis

Regarding to the structure of this thesis, the publications of original research that resulted from the experimental data were used to organize the results obtained presented in section 3. In order to organize the information and present the developed research, a general introduction, integrated discussion and conclusions were made. In summary, this thesis is divided in the following sections:

1. Introduction

In this section, a brief introduction will be presented about breast cancer (BC) comprising a review with the most current analytical thecniques used in BC metabolomic studies and their applications to identify metabolites as potential BC biomarkers based on the main advantages and advances in metabolomics research. In addition, chemometric methods used in metabolomics will be also focused.

2. Aims and Scope

The main objectives of this thesis will be presented according to the proposed research.

3. Metabolomic Pattern in Breast Cancer

In this section, the results obtained as well as their discussion will be presented by manuscripts.

4. Integrated Discussion

This section provides na integrated discussion of the results obtained through this thesis and compared with the results obtained by other authors.

5. Conclusions and Future Perspectives

A general conclusion is presented regarding to the ability of the analytical techniques to establish potential BC biomakers. Also, the improvements that should be taken into account to validate the results obtained and to go deeper in knowledge about BC metabolomics.

SECTION 1| Breast cancer metabolomics: from analytical platforms to multivariate data analysis

(Silva et al.; Metabolites **2019***; 9:102)*

Introduction

Cancer is a public health problem and causes a tremendous burden on patients, families and society creating a significant problem on global economy. Although has been extensively investigated, cancer still remains one of the leading causes of death in the world after coronary diseases [1]. Globally, breast cancer (BC) remains at the top of women´s cancers worldwide followed by colorectal, lung, cervix, and stomach cancers according to GLOBOCAN series of the International Agency for Research on Cancer (IARC), contributing with more than 11.6 % of all cancer types. (Figure 1. 1 A).

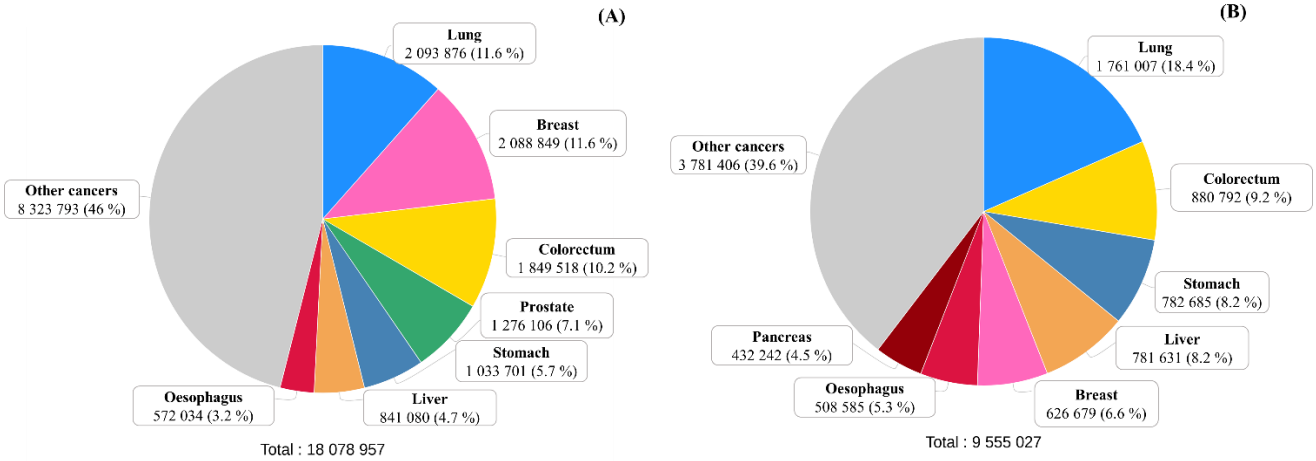

Figure 1. 1 - Estimated cancer incidence rates (A) and (B) number of deaths worldwide for 2018. Adapted from GLOBOCAN [1].

In addition, around 2.1 million BC new cases were diagnosed in 2018 and occured 630 thousands deaths (6.6 % of all cancers) (Figure 1. 1 B).

The incidence rates are highest in North America, Australia and Europe and lowest in Asia. These differences might be related to societal changes, as result of industrialization, such as, unhealthy lifestyle, expressed by overweight and other symptoms, alcohol consumption, tobacco smoking, physical inactivity, early menarche, among others [2,3]. Although its incidence is high in some developed countries, mortality is higher in low and middle income countries [4]. The incidence of breast cancer increases with age and is generally in the 50–60 age group. Moreover, mortality rates affect the group below 35 years and above 75 years, due to the younger group have the most aggressive type of the disease and in the older group the treatment cannot be so aggressive [5]. Concerning the incidence rates and mortality for breast cancer in Europe, it was observed that in 2012, the incidence of breast cancer was around 361,608 cases with 91,585 deaths. For 2020, around 400 thousand new cases will be diagnosed resulting in 100 thousand deaths according to International Agency for Research on Cancer (IARC). For Portugal and USA, the expected number of breast cancer cases in 2020 will be nearly 6000 and 270 thousand resulting in around 1700 and 51,000 deaths, respectively as shown in Figure 1. 2**.**

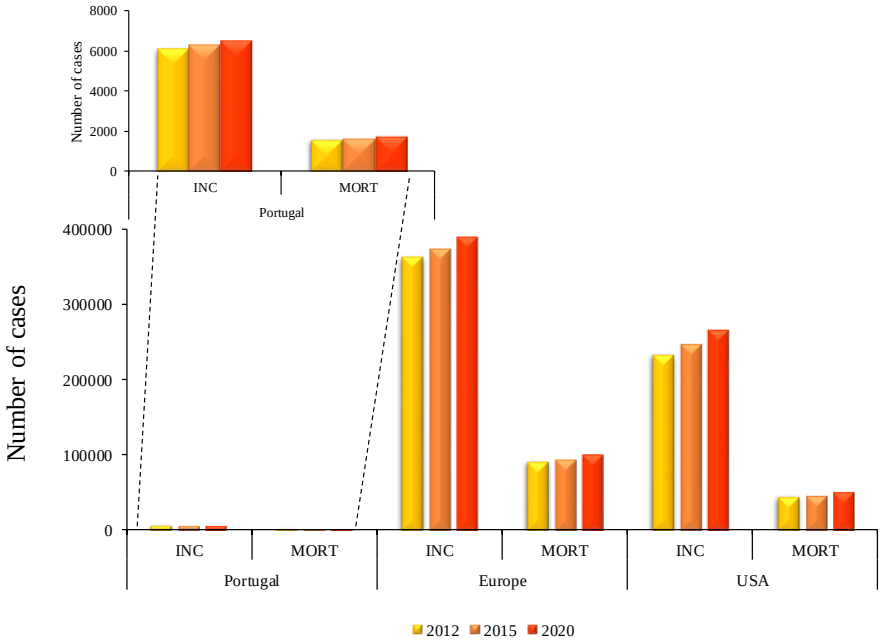

Figure 1. 2 - BC incidence and mortality rates in Portugal, Europe and USA from 2012, 2015 and expected rates for 2020. Data available at IARC. Legend: INC: incidence; MORT: mortality.

This trend might be as consequence by the availability of better screening procedures resulting in an early detection and also in the development of new treatments [2,6,7], which lead to an improved survival. Several risk factors associated with BC have been already recognized, namely epidemiological factors (e.g., age, reproductive factors, socioeconomic status, ethnicity), often using standard analysis approaches (e.g., logistic regression) with adjustment for multiple comparisons. Other factors as lifestyle (e.g., alcohol, tobacco, obesity, physical activity), and exposure to radiation [8] are also associated. The risk of developing BC increases with age being rare in women younger than 25 years, but tending to be more aggressive in younger people. The most common BC that occurs is the invasive type independently of age [9]. The highest risk of family history is associated with increasing number of first-degree relatives diagnosed with BC (age under 50 years). The risk is further increased when the affected relative is diagnosed in both breasts [10]. Particularly, the mutations in genes BRCA1, BRCA2 and TP53 are strongly associated with the development of BC [9], even if these mutations are low, accounting for a small portion of the total BC incidence [3]. Consistent physical activity has many benefits and greater activity has been related to lower BC risk by decreasing the circulating estrogen levels in postmenopausal women [11,12]. Extensive literature has linked alcohol consumption to BC risk and reveal the role of ethanol in carcinogenesis altering estrogen levels through acetaldehyde. Briefly ethanol is converted to acetaldehyde (AA) through alcohol dehydrogenase (ADH), that then binds to DNA interfering with the DNA synthesis and repair [13]. Obesity is another BC risk factor to take into account as it is involved in insulin resistance and

hyperinsulinemia [14]. Insulin has anabolic effects on cellular metabolism and an overexpression of insulin receptor has been demonstrated in human cancer cells [9,15]. The involvement of insulin-like growth factor (IGFs) in carcinogenesis is attributed to their role in linking high energy intake, increased cell proliferation, and suppression of apoptosis to cancer risks [15,16]. With regard to obesity and BC risk, some studies indicate that is strongly associated with increased invasive BC risk in postmenopausal women particularly for estrogen receptor–positive cancers (ER^+) [17–19]. In clinical practice, there are nowadays several biomarkers routinely used for prognosis and identification of tumors, including the estrogen receptor (ER), progesterone receptor (PR) and the human epidermal growth factor receptor-2 (HER2) [20,21]. Another promising prognostic and predictive biomarker of BC is Ki-67 (present in dividing cells) as indicator of cell proliferation and also as an endpoint for neoadjuvant systemic therapy [20]. However there are other proposed markers of proliferation measured by immunohistochemistry (IHC), such as, cyclin D, cyclin E, p27, p21, among others that are used to determine the predictive and prognostic levels [22].

In the last years, metabolomics emerged as a powerful approach in the advanced disease biomarker discovery which includes the comprehensive study of metabolites that are present in biological samples [23]. The study of metabolome to search biomarkers for any disease involves the identification of endogenous metabolites that have the potential to discriminate between samples obtained from healthy subjects and diseased patients. Plasma, serum, urine, tissue and cerebrospinal fluid (CSF), are the most commonly used biological samples in metabolomic studies. These biological samples contain hundreds of metabolites that vary in chemical and physical properties and concentration levels. Metabolomic studies includes two main approaches – targeted and untargeted. The targeted analysis is focused in specific groups of chemical characterized and annotated metabolites and their related pathways, whereas in the untargeted analysis the study includes a comprehensive measurement of all metabolites present in samples [24,25].

The type of approach chosen will determine the experimental design, sample preparation, and which analytical techniques can be used to obtain the results. Both targeted and untargeted follow the similar pipeline. Briefly, the study design includes the population that will be part in the study and also the determination of the conditions that are relevant for the hypothesis in investigation, namely the sample size, randomization (as a study design consideration), storage (as a sample handling issue), freeze/thaw cycles and timing during sample preparation are the most common factors that should be taken into account to guarantee reproducible and successful experiments minimizing variability. There are three main analytical platforms frequently used in metabolomic studies, which include mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy [26]. Moreover, after data acquisition, the obtained dataset, normally is subjected to statistical analysis (univariate and

multivariate methods) to find significant variations that allow the discrimination of patients with a specific disease (in this case, BC) from a control group [27]. The most common approaches for the identification of important metabolites comprise the application of unsupervised methods, such as, principal component analysis (PCA), hierarchical cluster analysis (HCA), as well as supervised methods, like partial least squares discriminant analysis (PLS-DA), random forest (RF) and support vector machines (SVM) [26,28]. A training set is used to construct the multivariate analysis models (e.g., PCA or PLS-DA), followed by an external validation set to predict the new cohort of samples using the model constructed with the training model. Finally, the putative biomarkers can be placed in metabolic networks to allow the biological interpretation or which pathways are up- or down-regulated.

OMICS science

The OMICs is a neologism broadly adopted in biomedical research, that comprises the dataset of genomics (DNA), transcriptomics (RNA), proteomics (proteins) and metabolomics (metabolites) based on the central dogma of molecular biology [29]. The purpose of OMICs science in cancer research is to discover cancer-specific biomarkers (diagnostic, prognostic and/or putative). The Food and Drug Administration (FDA) defined biomarkers as a "characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or biological responses to a therapeutic intervention" [30]. Biomarkers are powerful tools, when used for the early cancer detection and selection of therapeutic strategy, thus improving the outcome of cancer treatment and reduce cancer-related mortalities.

One of the newest promising OMICs sciences is metabolomics being a suitable tool that provides state of the art of analytical instrumentation tandem with pattern recognition procedures and chemometric tools to discover new disease- biomarkers providing novel insights into disease etiology, and more robust assessment of etiological pathways [30,31]. Metabolomics studies the complex interaction in biological systems providing a comprehensive and detailed information of the phenotype and molecular physiology as result of environmental factors, genetic as well as exogenous and endogenous factors (e.g. age, gender, race, diet, drugs, exercise, gut microbiota) [31]. In addition, metabolomics can be used in early detection and diagnosis of cancer, in the assessment of therapies and medical interventions, since cancer is a disease that promotes changes in cellular metabolism [32–35] . This OMICs tool has been extensively applied in clinical health practice due to its ability to quickly analyze biological samples (e.g. blood, tissue, saliva and urine) with relatively simple sample preparation (10 – 30 min), cost-effective and high-throughput [30,36,37]. Nevertheless, metabolomics present several drawbacks resulting from biological and experimental features, such as

sampling variability, inter- and intra-individual differences and a lack of validated protocols for biological samples handling which have a significant impact on the OMICs data approach [38,39].

It will be further discussed the the metabolic profile of several biological samples, including lipidomics (lipids), labeled substrates (e.g., ^{13}C labeled glucose), volatomic (volatile organic metabolites), and metabolites resulting from Krebs cycle [30,39] with the purpose of an early diagnosis, metabolic reprogramming, cancer typing, staging and therapeutic intervention response [29,37]. Regarding the Krebs/TCA cycle, there is evidence that the role of TCA for energy production and macromolecule synthesis by cancer cells, especially those with dysregulated oncogene and tumor suppressor expression [40–42]. Over the last years, there has been a rapidly growing number of metabolomic studies intended to discover new biomarkers or make disease diagnosis using different biological matrices , such as cell lines [43–46], blood [47], exhaled breath [48], plasma [33,49,50], saliva [51–54], tissues [55–57], serum [58] and urine [59]. In Table 1 are resumed the most common analytical approaches used in metabolomic studies grouped by type of biological sample and objective of the study. Interestingly, the main studies involve a diagnostic purpose using BC cell lines with the aim of search biomarkers, inspect the metabolome (endo- and exo-). Moreover, lipids as building blocks of cell membranes have their levels changed during the malignant transformation. Lipid metabolism plays a vital role in oxidative stress and is correlated with other parameters linked to BC risk (e.g. hormonal balance, body mass index, breast density, drug metabolism and growth of insulin levels) [43,60] (Table 1.1).

Table 1. 1 - Summary of metabolomics studies performed in breast cancer biomarker discovery in different biological matrices.

Biological sample	Sample groups	Aim	Analytical approaches	Main conclusions	References
Human cell lines					
Diagnostic biomarkers	BC (ZR-75-1, T-74D, MCF7, MDA-MB-231, MDA-MB-453, MDA-MB-468, SK-BR-3, BT-474, BT-549), Control (MCF10A)	To compare the differences in the lipidomic compositions of human cell lines derived from normal and BC tissues, and tumor *vs*. normal tissues obtained after the surgery of BC patients.	LC-MS/MS, GC-MS	* 123 lipids were identified, and a differentiation was observed for MDA cells	[29,43]
Diagnostic biomarkers	BC (MDA-MB-231, -453, BT-474), Control (MCF-10A)	To determine endo- and exo-metabolite analysis of the BC cell lines	UPLC-MS/MS, LC-MS/MS	* Statistical analysis allowed a discrimination of the breast epithelial cells from the BC cell lines * MDA-MB-231 showed an increase in nicotinamide levels, namely in 1-ribosyl-nicotinamide and NADþ	[46]
Diagnostic biomarkers	BC (T-47D, MDA-MB-231, MCF-7), Control (HMEC)	To establish the BC cell lines volatile metabolomic signature	GC–MS	* 60 VOMs were identified and six of them were detected only in the headspace of cancer cell lines	[44]
Diagnostic biomarkers	BC (MDA-MB-468, SKBR3, MCF-7)	To quantify specific metabolites in BC cell extracts	NMR	* Significantly differences were observed between cell lines, namely in the concentrations of 15 metabolites* The current method represented a useful tool for the establishment of potential biomarkers	[61]
Diagnostic biomarkers	BC (Cal 51, SKBR3, MCF-7)	To measure the absolute metabolite concentrations in complex mixtures with a high precision in a reasonable time	NMR	* The proposed approach represented a powerful tool to quantify 14 metabolites (alanine, lactate, leucine, threonine, taurine, glutathione, glutamate, glutamine, choline, valine, isoleucine, myo-inositol, proline, and glucose) in cell extracts within 20 min	[45]

Diagnostic biomarkers	BC cell lines (MCF-7, HCC70, MDA-MB-231, MDA-MB-436, MDA-MB-468), BC patients (n= 35)	To investigate the metabolic profiles of human BC cell lines carrying BRCA1 pathogenic mutations	LC-MS/MS	* It was possible to collect differential metabolic signature for BC cells based on the BRCA1 functionality	[50]
Therapy response	BC cell line (MCF-7)	To develop a robust and highly sensitive platform to identify endogenous estrones in clinical specimens	MALDI-MS, LC-MS/MS	* The results suggested that MALDI-MS-based quantitative approach can be a broad method for the ketone-containing metabolites target analysis thus replicating the clinical stage.	[62]
Therapy response	BC tissue (n = 40), Blood (n = 27), BC cell lines (n = 3)	To detect alterations in metabolites and their linkage to metabolic processes in several pathological conditions including BC	NMR	* Functional of IP3Rs in causing metabolic disruption was observed in MCF-7 and MDA MB-231 cells * The results offered new insights regarding the relationship of BC metabolites with IP3R.	[63]
Metabolic reprogramming	MDA-MB-231, BC xenografts	To study toxic effects of bisphenol and the underlying mechanisms on tumor metastasis-related tissues	LC-MS/MS, MALDI-MS	* Metabolites-based studies might be suitable for BC diagnosis * The data provided good indication for BPA screening secure option	[64]
Human Blood, plasma, serum					
Diagnostic biomarkers	BC patients (n = 258), Benign mammary gland (n = 159), Control (n = 78)	To screen metabolite markers with BC diagnosis potentials	MS	* The method developed allowed the discrimination of BC from non-BC using six blood metabolites	[47]

Diagnostic biomarkers	Metastatic BC patients (n = 95), Early-stage BC patients (n = 80)	To explore whether serum metabolomic spectra could distinguish between early and metastatic BC patients and predict disease relapse	NMR	* Disease relapse was linked with lower and higher levels of histidine and glucose, respectively	[58]
Diagnostic biomarkers	BC patients (n= 132), Control (n= 76)	To develop a new computational method using personalized pathway dysregulation scores for disease diagnosis	LC-TOF-MS, GC-TOF-MS	* The method allowed to determine important metabolic pathways signature for BC diagnosis, representing a suitable tool for diagnostic and therapeutic interventions.	[65]
Diagnostic biomarkers	BC patients (n = 45), Control (n = 45)	To detect differences between BC and healthy individuals	UHPLC-MS, GC-MS	* 661 metabolites were detected, but only 338 metabolites were found in all samples, and 490 in more than 80% of samples.	[66]
Diagnostic biomarkers	BC patients (n = 29), Control (n = 29)	To establish a plasma metabolic fingerprint of Colombian Hispanic women with BC	LC-MS, GC-MS, NMR	* The current report showed the effectiveness of multiplatform strategies in metabolic/lipid fingerprinting works	[49]
Diagnostic biomarkers	BC patients (n = 91), Control (n = 20)	To explore whether serum metabolomic profile can discriminate the presence of human BC irrespective of the cancer subtype	LC-MS/MS	* From the 1269 metabolites identified in plasma from controls and patients; only 35 metabolites were related to BC.	[33]
Diagnostic biomarkers	BC patients (n = 27), control (n = 30)	To apply 1H NMR and DART-MS for the metabolomics analysis of serum samples from BC patients and healthy controls.	NMR, DART-MS	* The approach allowed the disease classification and the biochemical validation useful to identify the mechanisms associated to BC development.	[67]
Diagnostic biomarkers	Metastatic BC patients (n = 39 + 51 for validation), Early-stage BC patients (n = 85 + 112 for validation)	To distinguish between early and metastatic BC	NMR	* Metabolic phenotyping by NMR showed a robust potential for the diagnosis, prognosis, and management of BC cancer patients	[68]
Diagnostic biomarkers	BC patients (n = 40) BE patients (n = 40) and healthy controls (n = 34). BE patients with	To investigate the free fatty acid (FFA) metabolic profiles to identify biomarkers that can be used to distinguish patients with BC (BC) from benign (BE) patients or healthy controls.	GC-MS	The FFA biomarkers proved to be helpful for the prevention and characterization of BC patients.	[69]

	fibroma (n = 25) and chronic fibroadenosis of breast (n = 15)				
Therapy response	BC patients (n = 19)	To compare metabolite concentrations and Pearson's correlation coefficients to examine concomitant changes in metabolite concentrations and psychoneurologic symptoms before and after chemotherapy.	UPLC-MS/MS	* The post-chemotherapy global metabolites were characterized by higher and lower amounts of acetyl-L-alanine and indoxyl sulfate and 5-oxo-L-proline, respectively. * Metabolomics was useful for further understanding of biological mechanisms associated with psychoneurologic symptoms.	[70]
Therapy response	BC patients (n = 28)	To identify potential biomarker candidates that can predict response to neoadjuvant chemotherapy for BC	LC-MS, NMR	* The concentrations of threonine, isoleucine, glutamine, linolenic acid had significantly different responses to chemotherapy * The purposed approach clearly discriminates patients regarding the response to drugs providing a valuable tool for a non-invasive prognosis of the treatment strategy.	[71]
Endogenous factors	BC patients (n = 206), Control (n = 396)	To investigate whether plasma untargeted metabolomic profiles could contribute to predict the risk of developing BC	NMR	* The study contributed to the development of screening approaches for the identification of BC at-risk women.	[31]
Endogenous factors	BC patients (n = 621), Control (n= 621)	To evaluate associations of diet-related metabolites with the risk of BC in the prostate, lung, colorectal and ovarian cancer screening trial	GC-MS, LC-MS/MS	* The data obtained showed how nutritional metabolomics might identify diet-related exposures associated to cancer risk.	[72]
Human urine					
Diagnostic biomarkers	BC patients (n= 30), CC (n = 30), Control (n = 30)	To discriminate different types of cancer based on urinary volatomic biosignature	GC-MS	* The butanoate metabolism was highly activated in studied cancers, as well as tyrosine metabolism, but in a reduced proportion * Different clusters allowed to establish sets of VOMs fingerprints resulted in the discrimination of the studied cancers	[59]
Therapy response	BC patients (n = 31), Control (n = 29)	To identify metabolites which can be helpful in the understanding of metabolic alterations driven by BC as well as their potential usage as biomarkers	LC-MS, GC-MS	* The analytical multiplatform approach enabled a wide coverage of urine metabolites revealing significant alterations in BC samples	[73]

Human Saliva					
Diagnostic biomarkers	BC patients (primary, n = 8; relapse, n = 22), Control (n = 14)	To determine polyamines including N-acetylated forms in human saliva and the diagnostic approach to BC Patients	UPLC−MS/MS	* The increase on polyamines level in BC patients Ac-SPM, DAc-SPD, and DAc-SPM levels were significantly higher only in the relapsed patients	[52]
Diagnostic biomarkers	BC patients (n = 30), Control (n = 25)	To screen the potential salivary biomarkers for BC diagnosis, staging, and biomarker discovery.	UPLC-MS	* Saliva metabonomics approach may provide new insights into the discovery of BC diagnostic biomarkers.	[53]
Diagnostic biomarkers	BC patients (n = 111), Control (n = 61)	To determine of polyamines including their acetylated structures for the diagnosis of BC patients.	UPLC-MS/MS	* The ratio of N8-Ac-SPD/ (N1-Ac-SPD + N8-Ac-SPD) can be used as a health status index after the surgical treatment.	[54]
Diagnostic biomarkers	BC patients (n = 66), Control (n = 40)	To explore the potential of the volatile composition of saliva samples as biosignatures for BC non-invasive diagnosis	GC-MS	* This study defined an experimental layout appropriate for the characterization of volatile fingerprints from saliva as potential biosignatures for BC non-invasive diagnosis.	[51]
Human Exhaled breath					
Diagnostic tool	BC patients (n = 14), Control (n = 11)	To detect and identify human exhaled BC–related volatile profile	MS	* Eight metabolites enabled a clear discrimination of exhaled breath of BC patients from controls. * The analytical technique provided a non-invasive strategy to detect VOMs for the BC diagnosis.	[48]

Human Tissues					
Diagnostic biomarkers	BC patients (n = 10)	To establish a detailed lipidomic characterization with the goal to find the statistically differences between BC and normal tissues.	HPLC-MS	* Total concentrations for phosphatidylinositols, phosphatidylcholines, phosphatidylethanolamines and lysophosphatidylcholines were increased leading to a clear differentiation by PCA and OPLS-DA.	[55]
Diagnostic biomarkers	Paired tumor and non-tumor liver (n = 60), breast (n = 130) and pancreatic (n = 76)	To assess the metabolomic profiling as a novel tool for multiclass cancer characterization	GC-MS, LC-MS	* The findings provided a framework to validate cancer-type specific metabolite levels in tumor tissues.	[56]
Diagnostic biomarkers	BC patients (n = 37), Control (n = 35)	To identify potential biomarkers that differs TNBC from ER^+ BC	GC-MS, LC-MS/MS	* 133 metabolites presented significant differences between ER+ and TNBC tumors * The metabolic pathway of tumors can provide new treatment targets.	[57]
Diagnostic biomarkers	BC patients (n = 47), Control (n = 35)	To identify how TNBC differs from LABC subtypes within the African-American and Caucasian BC patients	HR-MAS-NMR	* Increased pyrimidine synthesis was related to TNBC in Caucasian women* Novel treatment targets for TNBC could be explored through the metabolic changes	[74]
Diagnostic biomarkers	BC patients (n = 228)	To distinguish between tumor and non-involved adjacent tissue	HR-MAS-NMR	* Metabolic profiling of tumor tissues by NMR can be a suitable method for the analysis of the resection margins during BC surgery	[75]
Diagnostic biomarkers	BC patients (n = 25), Control (n = 5)	To establish metabolic profiles of ER+ vs. ER– and of ER– subtypes linked to genetics	GC-MS, LC-MS	* Changes in the metabolic profile of ER^- vs. ER^+ breast tumors were observed * The data represents a potential tool for the hypothesis testing of tumor metabolism	[76]
Diagnostic biomarkers	BC patients (n = 270), Control (n = 97)	To quantify the dysregulation of the glutamate-glutamine equilibrium in BC	GC-TOFMS	* A positive correlation between glutamate and glutamine in normal breast tissues was observed, whereas a negative correlation was obtained for normal tissues	[77]

Diagnostic biomarkers	95 OC (84 peritoneal, 11 pleural), 10 BC (7 pleural, 2 peritoneal, 1 pericardial), and 10 malignant mesotheliomas (6 peritoneal, 4 pleural)	To identify the metabolic differences between ovarian serous carcinoma effusions obtained pre- and post-chemotherapy and compare ovarian carcinoma (OC) effusions with breast carcinoma and malignant mesothelioma specimens.	1H-NMR	* Differences in metabolic profiles of different malignant effusions were detected * Metabolic characterization by NMR can be a technique to additional knowledge the mechanisms of effusion development	[78]
Therapy response	BC patients (n = 122)	To explore the effect of neoadjuvant therapy on metabolic profiles of BC tissues	HR-MAS-NMR	* Non-metastatic breast tumor tissue reflected different alterations in all patient groups after treatment. * Metabolic profiles discriminated pNRs from pMRD patients thus complementing other molecular assays allowing the knowledge of the underlying mechanisms affecting the response.	[79]
Therapy response	BC patients (n = 18)	To study metabolite levels in human BC tissue, assessing, for instance, correlations with prognostic factors, survival outcome or therapeutic response	HR-MAS-NMR	* Significant changes between the tumors were identified, indicating that the intertumoral changes for numerous metabolites were greater than the intratumoral changes for these three tumors.	[80]
Therapy response	BC patients (n = 37)	To determine whether metabolic profiling of core needle biopsy (CNB) samples using HR-MAS-NMR could be used for predicting pathologic response to neoadjuvant chemotherapy (NAC) in patients with locally advanced BC	HR-MAS-NMR	* The purposed method can be applied to predict the pathologic response before neoadjuvant chemotherapy	[81]
Therapy response	BC patients (n = 271)	To establish metabolic signatures for ER^+ vs. ER^- BC	GC-TOFMS	Some metabolites levels were increased in ER^- subtype, such as, beta-alanine, glutamate and xanthine The down-regulation of the ABAT protein in ER^- BC was confirmed by immunohistological analysis.	[82]

Mouse BC tissue					
Metabolic reprogramming	MMTVPyMT, MMTV-PyMT-DB, MMTV-Wnt1, MMTV-Her2/neu, and C3(1)-SV40 T-antigen (C3-TAg)	To identify global metabolic profiles of breast tumors isolated from multiple transgenic mouse models and to identify unique metabolic signatures driven by these oncogenes	GC-MS, LC-MS/MS, CE-MS	* C3-TAg was the only cohort with a tumor metabolic signature composed of ten metabolites with significance prognostic value in BC patients	[83]

ANOVA – Analysis of variance; AUC – Area under the curve; BC – BC; BFS– Bootstrap feature selection; CE-MS – Capillary electrophorese-mass spectrometer; DART-MS – Direct analysis in real time mass spectrometry; GC-MS – gas chromatography – mass spectrometry; GC-TOF-MS – Gas chromatography time-of-flight mass spectrometry; GGM – Gaussian graphical modelling; HCA – Hierarchical cluster analysis; HR-MAS-NMR - High resolution magic angle spinning nuclear magnetic resonance spectroscopy; LC-MS/MS – Liquid chromatography tandem with mass spectrometer; LC-TOF-MS – Liquid chromatography time-of-flight mass spectrometry; LDA – Linear discriminant analysis; MALDI-MS – Matrix-assisted laser desorption/ionization mass spectrometry; MCCV – Monte Carlo cross validation; MS – Mass spectrometry; MWT – Mann Whitney U test; NMR – Nuclear resonance magnetic; NRI – Net reclassification improvement; OPLS-DA – Orthogonal projections to latent structures discriminant analysis; OSC-PLS – Orthogonal signal correction partial least squares; PC – Pearson correlation; PCA – Principal component analysis; PEA – Pathway enrichment analysis; PLS-DA – Partial least squares discriminant analysis; RF – Random Florest classifier; ROC – Receiver operating characteristic; SCC – Spearman correlation coefficient; SVM – Support vector machine ; TNBC – Triple negative BC; UPLC-MS/MS – Ultra performance liquid chromatography tandem mass spectrometer; VIP – variable importance in projection.

In literature, the reports performed involving human cell lines focus mainly in diagnostic purpose. As for example in the volatile composition (VOMs) as described by Silva et al. [44] where the volatomic signature of BC cell lines was established, and based on the results, 2-pentanone, 2-heptanone, 3-methyl-3-buten-1-ol, ethyl acetate, ethyl propanoate and 2-methyl butanoate were detected only in cultured BC cell lines. These VOMs are formed endogenously or obtained from exogenous sources (e.g., environmental, lifestyle, biological agents) [51], and can be recognised as a useful tool to BC non-invasive diagnosis [44,51]. Other study by Willmann et al. [46] observed the changes of the exo- and endometabolite profiles in BC cell lines by LC-MS/MS and observed a clear discrimination of the breast epithelial from the BC cell lines through statistical tools. Moreover, a decrease on ratio of glutathione (GSH) and glutathione disulfide (GSSG) was observed in BC cell lines as a result of oxidative stress. The lipidomic profile of several BC cell lines was compared with normal cells obtained from non-cancerous tissues by LC-MS/MS and GC-MS that changes observed in breast tumor tissues were caused mainly by difference in lipidomic profiles of tumor cells and these alterations can be correlated with the lipidomic composition of the nine breast cancer cell lines. Furthermore, Martineau et al. [61], determined the absolute concentration of several metabolites (e.g., alanine, lactate, threonine, taurine, glutathione, glutamate, glutamine, choline, valine, isoleucine, myo inositol, serine, proline, aspartate and histidine), revealing the usefulness for the establishment of potential biomarkers. Also, BC cell lines with BRCA1 pathogenic mutations were investigated by LC-MS/MS in order to obtain their metabolic signature as possible diagnostic approach.

Regarding plasma, serum or blood, many studies have been conducted as observed in Table 1, with multiple aims as Cala et al. [49] that developed a pilot control case-study, where a metabolomic and lipidomic approach was performed in order to establish a plasma metabolic fingerprint of Colombian Hispanic women with BC. According to these authors, the plasma metabolites could contribute to an enhanced knowledge of the underlying metabolic shifts driven by BC in women of Colombian Hispanic origin. Moreover, despite racial differences, the mapped metabolic signatures in BC were comparable but not identical to those described for non-Hispanic women. Wang et al. [47] used a dried blood spot approach for rapid BC detection. In the first study, the target analytes were 23 amino acids and 26 acylcarnitines, and based on the results piperamide, asparagine, proline, tetradecenoylcarnitine/palmitoylcarnitine, phenylalanine/tyrosine, and glycine/alanine could be used as potential biomarkers to diagnose BC. Lyon et al. [70] established a serum metabolome analysis from the tryptophan pathway of 19 women with early-stage BC. The targeted analysis indicated higher kynurenine levels and kynurenine/tryptophan ratios post-chemotherapy. Also, the symptoms of pain and fatigue had association with several targeted metabolites. An improved metabolic profile of human serum samples was obtained using complementary thecniques, namely MS and NMR and

this approach may be useful to achieve more accurate disease detection and gain more insights regarding disease mechanisms and biology [67].

Another study conducted by Lécuyer et al. [31] combined metabolomic and epidemiological approaches by NMR to investigate whether plasma untargeted metabolomic profiles could contribute to the identification of BC at-risk women, whereas Playdon et al. [72] focused on the evaluation of the associations of diet-related metabolites with the risk of breast cancer. It was possible to verify that the prediagnostic serum concentrations of metabolites related to alcohol, vitamin E, and animal fats were associated with ER+ breast cancer risk.

Urine became a very interesting biological sample to investigate as diagnostic tool or as result of a treatment, as it is easy to collect, and also as ending point of all reactions that occur in the body. Furthermore, Porto-Figueira [59] established the urinary volatomic biosignature from breast (BC), and colon (CC) cancer patients as well as healthy individuals. This last work observed that several pathways are over activated in cancer patients, being phenylalanine pathway in BC and limonene and pinene degradation pathway in CC the most relevant. Yu et al. [84] explored the relationship between urinary metabolites and clinical chemotherapy response in BC. As results, chemotherapy-sensitive patients exhibited 30 % of change in metabolite levels when compared to healthy individuals, while chemotherapy-insensitive patients showed only 9 % of change in metabolite levels when compared to healthy people that presented recurrence.

Another explored biological fluid is saliva as described by Zhong et al. [53] that screened the putative salivary biomarkers for BC diagnosis, staging, and biomarker discovery. As a result, 18 biomarkers were identified, but only three up-regulated metabolites, displayed the area under the curve (AUC) values higher than 0.920, indicating the high accuracy to predict BC. Also, Cavaco et al. [51] screened salivary volatiles for a putative BC discrimination, and from metabolites identified, only 3-methyl-pentanoic acid, 4-methyl-pentanoic acid, phenol, p-tert-butyl-phenol, acetic, propanoic, benzoic acids, 1,2-decanediol, 2-decanone, and decanal were statistically relevant for the discrimination of BC patients in the populations analyzed. Another type of molecules, the polyamines were associated with tumor growth due to their biosynthesis and accumulation [54]. In this context, Tsutsui et al. [52] and Takayama et al. [54] determined polyamines including N-acetylated forms in saliva to diagnose BC. According to Tsutsui et al. [52], the level of polyamines increased in BC patients, and the levels of N^1-acetyl-spermine, N^1N^8-diacetyl-spermidine and *N*,*N*-diacetyl-spermine were significantly higher only in the relapsed patients. Takayama et al. [54] demonstrated that eight polyamines are strongly correlated with the BC patients. Furthermore, the ratio of N^8-acetyl-spermidine/ (N^1-acetylspermidine + N^8-acetyl-spermidine) may be adopted as an index of the health status after the surgical treatment.

In-vitro analysis of BC tissues can be a valuable tool to inspect the metabolic differences between tissue classes, either using the hydrophilic or the lipophilic part. As a result, one might use the metabolomic profile as a novel tool for cancer characterization. Breast tissue is also an interesting biological sample used for diagnostic purposes and /or response to a treatment as demonstrated by Euceda et al. [79] that explored the effect of the antiangiogenic drug bevacizumab on metabolic profile from BC tissue. On the other hand, Budczies et al. [77] studied the glutamate enrichment as a new diagnostic opportunity in BC, and a positive correlation between glutamate and glutamine in normal breast tissues switched to negative correlation between glutamate and glutamine in BC tissues. Euceda et al. [79] observed a metabolic alteration indicating a decline in glucose consumption as an effect of chemotherapy. In addition, a lower glucose and higher lactate level was observed in patients ($\geq$ 90% of tumor reduction) when compared to those with no response ($\leq$ 10% of tumor reduction). In turn, Choi et al. [81] determined the metabolic profiling of core needle biopsy samples in order to predict pathologic response to neoadjuvant chemotherapy in patients with locally advanced BC. These authors observed that there was a trend of lower levels of phosphocholine/creatine ratio and choline-containing metabolite concentrations in the pathologic complete response group when compared to the non-pathologic complete response group. Most of the BC patients undergo a cycle or more of chemo being the general treatment that uses cancer-killing drugs before (neoadjuvant or preoperative therapy) and after (adjuvant therapy) surgery [31,36], Then, the therapeutic chemo effect may shift significantly between patients, as a result of BC phenotypes [37] of and intra- and inter-individual differences. For this reason, it is necessary to punctually and accurately evaluate the therapeutic effects of chemotherapy, which could help to adjust the chemotherapy regimen [71,84]. whereas the advances in treatment increased significantly the survival rates for women with BC, as women often report psychoneurologic symptoms (e.g. pain, fatigue, depression) during and after chemotherapy cycles.

Regarding exhaled breath a less explored biological sample in terms of BC diagnostic purpose. In a study performed by Martinez-Lozano Sinues et al. [48] who developed a pilot study to identify cancer–related volatile profile in exhaled breath of BC patients. Concerning exhaled breath and the possible mechanisms involved in the production of endogenous VOMs, in Figure 1. 3 is represented a schematic illustration about the possible pathways.

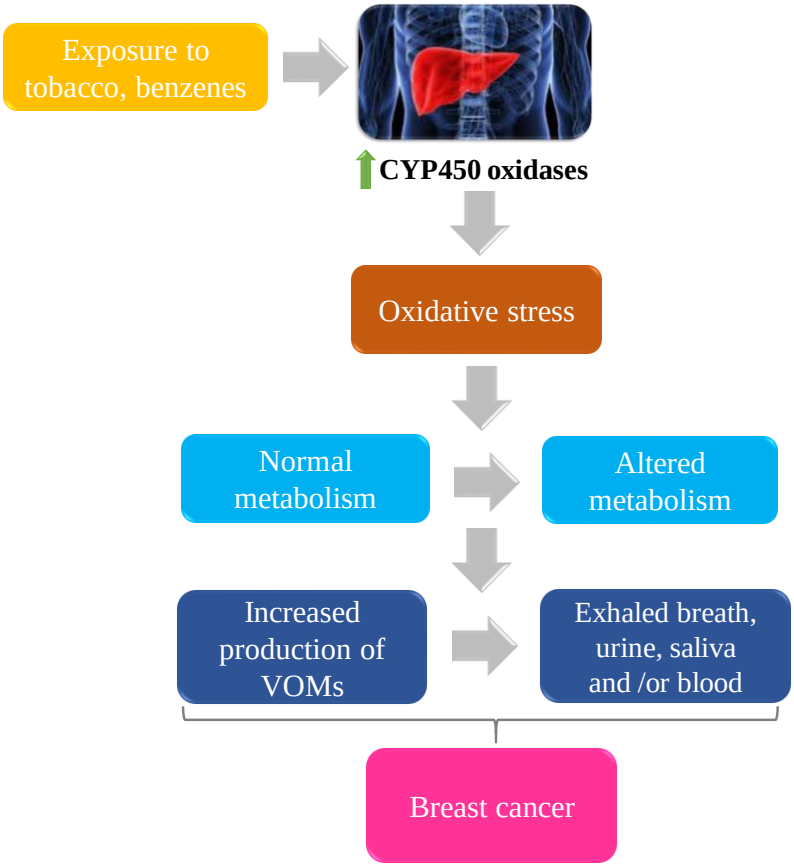

Figure 1. 3 - Scheme to illustrate the possible origin of some VOMs.

The principle behind this is based on the fact that cancer growth is promoted by the progressive accumulation of genetic and epigenetic changes leading to cellular oxidative stress, which in turn increases the liver's production of cytochrome P-450 (CYP450) oxidase enzymes to take into account with stress. Both processes affect the abundance of VOMs in breath once oxidative stress causes lipid peroxidation of polyunsaturated fatty acids (PUFA) in membranes, producing alkanes and methylalkanes which are catabolized by CYP450 [85].

Analytical approaches

Metabolomics encompasses targeted and non-targeted analysis of endogenous and exogenous metabolites (<1500 Da), such as lipids, amino acids, hormonal steroids, peptides, nucleic acids, organic acids, vitamins, thiols and carbohydrates, which represent a promising tool for biomarker discovery [86,87]. The complexity of the metabolome, the metabolites properties and their concentration levels in biological samples complicates the separation and detection on a single analytical platform. For this fact, the integration of high resolution analytical frameworks, mass spectrometry (MS) and nuclear magnetic resonance (NMR), appear as an outcome in metabolomics studies, providing sensitive, reliable detection and quantification of thousands of metabolites in a

biological sample and related metabolic pathways within a few minutes [27,86,87] as shown in Figure 1.4.

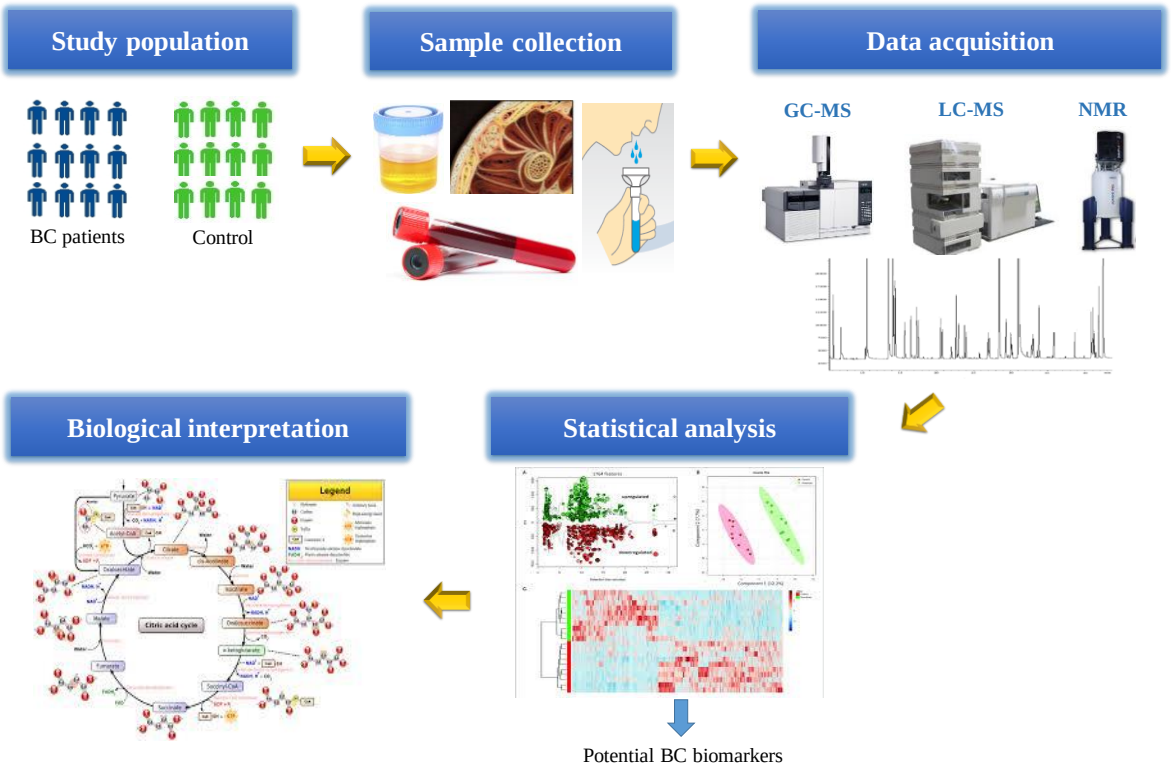

Figure 1. 4 - General flowchart in targeted/untargatet metabolomic approaches.

MS - based metabolomics

MS is an analytical tool extensively used in metabolomics applications, ranging from understanding the structural characterization of important metabolites to biomarker discovery [86]. Metabolic fingerprinting is general obtained by MS direct-injection, but this approach presents several drawbacks namely co-suppression and low ionization efficiencies. Thus, generally MS based metabolomics includes a separation step, based on gas chromatography (GC–MS) [43,44,51,59,65,66,77,82], liquid chromatography (LC–MS) [33,43,46,50,52–55,70] or capillary electrophoresis (CE-MS) [83,84], to solve co-suppression and to decrease the complexity of the biological sample. The integration of MS with a chromatographic technique (GC, LC) and capillary electrophoresis showed high sensitivity, speed, selectivity and improves the accuracy of compound identification, detection and quantification. In addition, GC-, LC- and CE-MS are destructive methods, requires sample preparation and are expensive, being these facts the main drawbacks of these hyphenated frameworks [86,88,89].

Gas Chromatography-Mass Spectrometry (GC-MS) - based metabolomics

In the last decades, MS and chromatography have been broadly developed, and GC-MS becomes a core and reliable separation, detection and identification analytical framework on metabolomics analysis [43,44,51,59,65,66,77,82]. After sample collection and metabolites extraction, a small volume of sample is common injected in splitess mode, once the metabolites are in trace levels, to improve the sensitivity and carrier gas propels the sample through the high resolution capillary columns (30 or 60 m columns with 5–50% phenyl stationary phases). The separation in GC occurs in an oven at high temperatures, and the metabolites need be thermally stable and volatile (e.g., aldehydes, ketones, alkanes, organic acids) or non-volatile metabolites requiring derivatization (e.g., amino acids, sugars, phosphorylated metabolites, amines, lipids) [86,88,89]. The samples are ionized by electron-impact (EI) or chemical ionization (CI) for MS detection, being EI the most used since it provides molecular ion fragmentation to obtain a mass spectrum revealing of the metabolite's structure [88]. The MS employed influences the sensitivity of detection, being the quadrupole (q), time-of-flight (TOF) and ion trap the most usually applied in metabolomics. GC-qMS was used to screen salivary volatiles for putative BC as an exploratory study involving geographically distant populations [51], also to establish the metabolomic signature of human BC cell lines [44] and to discriminate different types of cancer based on urinary volatomic biosignature [59], among other examples reported in Table 1. In the first study, up to 120 volatiles from distinct chemical classes, with significant variations among the groups, were identified [51], whereas Silva et al. [44] and Porto-Figueira et al. [59] identified 60 and 130 volatiles in BC cell lines and urine, respectively. On the other hand, Budczies et al. [77,82] used GC × GC-TOFMS framework to evaluate the glutamate enrichment as new diagnostic opportunity in BC and to accomplish a comparative metabolomics of estrogen receptor positive (ER^+) and estrogen receptor negative (ER^-) in BC. Budczies et al. [82] identified 19 metabolites and the GC–TOFMS based analysis of metabolites present in BC tissues revealed significantly differences in central metabolism in the more aggressive ER^- compared to the ER^+ type. The detected changes included the metabolism of glutamine with a decrease in concentration of glutamine and an increase in concentration of glutamate and 2-hydroxyglutaric [82]. In turn, Dougan et al. [66] used GC-MS to evaluate the detectability, reliability, and distribution of metabolites measured in pre-diagnostic plasma samples in a pilot study of women listed in the Northern California site of the BC Family Registry. In this study. 661 metabolites were detected, 338 (51 %) of them were found in all samples, and 490 (74 %) in more than 80 % of samples.

The main advantages of GC-MS-based metabolomics are sensitivity, specificity, high-throughput technology to handle a large volume of samples and reproducible. Nevertheless, this hyphenated technique has limited in mass range (m/z 30 – 550), the molecular ion is often not detected owing to

fragmentation, which makes more difficult the identification of unknown metabolites and the metabolites need be volatile and thermally stable [89,90].

Liquid Chromatography-Mass Sectrometry (LC-MS) - based metabolomics

Currently, liquid chromatography (LC) in particular high performance liquid chromatography-mass spectrometry (HPLC-MS, LC-MS) represents an easy-going tool on separation and characterization of a metabolites pool, namely salts, acids, bases, hydrophilic and hydrophobic metabolites. The versatility of LC-MS is due to the several separation procedures and wide-ranging mass analyzers [90]. Contrarily to GC-MS, HPLC-MS is not limited to volatile and thermo stable metabolites and it is a promising tool for global metabolomics and the establishment of disease biomarkers.

Basically, the metabolites are eluted through a column based on their selective partition between a stationary phase (column material) and a mobile liquid phase. The metabolites according to the type of stationary phase can be eluted based on their charge, size, hydrophobicity and molecular weight [91]. Nowadays, the evolution of the HPLC is focused in miniaturization, smaller columns and low solvent volumes to attain a faster separation of metabolites. Ultra-high performance chromatography (UHPLC) appears as solution, since compared to HPLC promotes the resolution within a low analysis time and requires low volumes of solvent [92,93]. UHPLC columns are packed with 2 μm particles and the system operates at higher pressures (1000 bar) and tandem with MS, results in higher peak capacity, resolution, specificity and high-throughput abilities (reduced run time per sample) compared with HPLC [86,90,92–94].

Furthermore, Willmann et al. [46] analyzed the endo- and exometabolite of the BC cell lines MDA-MB-231, -453 and BT-474 as well as the breast epithelial cell line MCF-10A through two different analytical platforms: UHPLC-ESI-QTOF and HPLC-ESI-QqQ, which resulted in the identification of 92 annotated exometabolites and 58 endometabolites. In turn, Jové [33] used LC-ESI-qTOFMS/MS to establish the metabolomic profile of BC, whereas HPCL-ESI-MS was used to determine the determine the lipidomic differences between human BC and the surrounding normal tissues [55]. UHPLC tandem with MS was applied to explore novel blood plasma biomarkers associated to the BRCA1-mutated phenotype of BC [50], to determine polyamines including N-acetylated forms in saliva [52,54], and to screen the potential salivary biomarkers for BC diagnosis, staging, and biomarker discovery [53].

NMR – based metabolomics

NMR spectroscopy has been announced as a promising tool of metabolomics, providing a comprehensive view of metabolite fingerprinting, profiling and metabolic flux analysis under specific conditions, despite its inherent lower sensitivity compared to MS, limiting its skill with trace level metabolites. The main advantages of NMR are automation, requires low or no sample preparation, high reproducibility, non-destructive, non-selectivity in metabolite detection and the ability to simultaneously quantify multiple classes of metabolites [29,87].

The principle of NMR spectroscopy is based on the fact that the nucleic of many isotopes (e.g., ^{1}H, ^{13}C, ^{14}N, ^{15}N, ^{17}O), when placed in a magnetic field, absorb radiation at a specific frequency [90]. The result is a NMR spectrum which corresponds to a unique metabolite pattern and provides structural information that can simplify the identification of unknown metabolites [86,89]. A fast identification of metabolite results from a combination of chemical shifts, spin–spin coupling, and relaxation or diffusion information [86,89]. Jobard et al. [68] reported a ^{1}H NMR-based metabolic phenotyping study aiming the identification of metabolic serum changes associated with advanced metastatic BC (MBC) in comparison to the localized early disease (EBC). Histidine, acetoacetate, glycerol, pyruvate, glycoproteins (N-acetyl), mannose, glutamate and phenylalanine were the metabolites that allowed the discrimination between MBC and EBC groups. NMR was also used by Tenori et al. [58] to explore whether serum metabolomic spectra could distinguish between early and metastatic BC patients and predict disease relapse, whereas Singh et al. [63] used NMR to detect alterations in metabolites and their linkage to metabolic processes in a number of pathological conditions including BC. In the last study, the authors observed an increase in lipoprotein, lactate, lysine and alanine level and a decrease in the levels of pyruvate and glucose in serum of inositol 1, 4, 5-trisphosphate (IP3R) receptor group patients when compared to control. In addition, NMR offers the possibility to study tissue through high-resolution magic angle spinning (HR-MAS) to reduce line widths in NMR spectra of tissue samples [74,75,79–81]. Tayyari et al. [74] performed the metabolomic analysis of triple-negative and luminal A BC subtypes in African-American using HR-MAS-NMR. A total of 27 metabolites were assigned and the metabolic profiles of these subtypes were also distinct from those revealed in Caucasian women. In turn, the feasibility of HR-MAS-NMR of small tissue biopsies to distinguish between tumor and non-involved adjacent tissue was investigated by Bathen et al. [75]. The results showed that the samples with a low tumor content have higher levels of glucose, while samples with a high tumor content have higher levels of ascorbate, lactate, creatine, glycine, taurine and the choline-containing metabolites (glycerophosphocholine (GPC), phosphocholine (PCho), and free choline). Euceda et al. [79] evaluate the metabolomic changes during neoadjuvant chemotherapy combined with bevacizumab in BC using HR-MAS-

NMR. According to these authors, despite metabolic profiles not being able to predict the pathological complete response (pCR) prior to treatment, a significant metabolic difference in pCR^+ patients compared to pCR^- was detected after neoadjuvant chemotherapy.

Comprehensive analytical frameworks on metabolomics approach

Comprehensive analytical frameworks are gained popularity on metabolomics fields [86], being hundreds of metabolites detected simultaneously through analytical frameworks such as GC×GC-TOFMS, HPLC-CE-MS, LC×LC-MS, LC-MS-NMR, MALDI-FT-ICR-MS, LC-FT-ICR-MS, among others.

On the last decade, two dimension (2D) liquid-liquid chromatography (LC×LC) as well as gas-gas chromatography (GC×GC) have been gained increasing attention since overcome overlapping of metabolites by diverting each peak from a GC or LC column to a second GC or LC column, improve sensitivity and complementary selectivity being a promising tool in metabolomics field [95]. Nevertheless, other comprehensive analytical framework has been purposed in metabolomic field, in this context LC-MS-NMR platform is used in the identification of unknown metabolites in biological samples at trace levels, providing sample efficiency higher than the conventional flow injection methods [86]. In this sense, Reichenbach and co-workers [96] developed a suitable approach based on GC×GC-HRMS to analyze a cohort of 18 samples from BC tumors. This approach avoided the intractable problem of comprehensive peak matching, through a few reliable peaks for alignment and peak-based retention-plane windows to define comprehensive features that can be consistently matched for cross-sample analysis. In addition, a clear discrimination was achieved between sample of different grades and establish potential BC biomarkers. On the other hand, Yu et al. [97] optimized GC×GC-MS for robust BC cells, tissue, serum and urine metabolite profiling. GC×GC-MS analysis revealed detection around 600 molecular features from which 165 were characterized representing different chemical groups, such as amino acids, fatty acids, lipids, carbohydrates, nucleosides and small polar components of glycolysis and the Krebs cycle using EI spectrum matching. NanoLC-FT-ICR MS was used to analyze protein digests of ~3000 laser capture microdissection (LCM)-derived tumor cells from breast carcinoma tissue, corresponding to ~300 ng of total protein [98].

Data analysis

Data analysis is crucial in the metabolomics, being indispensable in every step of research, namely in sampling and experiment designs, data pre-processing and metabolite identification, as well in variables selection, classification modeling and validation procedures. The great challenge of data analysis in metabolomics is high dimensionality and complexity of datasets under analysis. Several

chemometric tools and statistical software's are used in order to attribute value for high-dimensional metabolomic information obtained previously by the analytical tools [99,100]. Normally, a complete data analysis procedure in metabolomics is based on the following steps: dataset pre-treatment (centering, scaling, normalization), pre-processing (exploratory projection, variables selection), processing (predictive models), validation (model verification) and post-processing (pathway analysis) [101]. However, data analysis is adaptable and dependent on the objective of the study, and may be simple exploratory research or complex discovery of biomarkers and metabolic pathways, for this reason not all steps are always present or are not followed in this order. The data analysis procedures of recent metabolomics studies in BC are described in Table 2.

Dataset pre-treatment

Dataset pre-treatment is the initial step in data analysis, being extensively used in metabolomics to resolve the heteroscedasticity of high-dimensional datasets. Commonly, pre-treatment in BC metabolomics is done through normalization of dataset based on the centering, scaling, transformation and/or experimental corrections of variables values [102–104]. Centering is performed when the data analysis is focused on the differences between variables, where all measurements (e.g., concentrations, areas) are converted to values around zero based on variation measures. Mean [46,67,68,79] is the measure normally used in centering. Scaling is used to adjust the variables measurements based on a scaling factor, converting the measurements of all variables into values relative to the scaling factor. The scaling factor selected can be a dispersive measure (e.g., standard deviation) or size measure (e.g., mean). The main scaling approaches based on dispersive measures are autoscaling (standard deviation) [46,51,59] and pareto scaling (square root of the standard deviation) [43,53,55]. On the other hand, the most of size measure approaches uses scaling factors based on the mean [80], median [51,57,59,66,75,78,83] or total intensity value [53,58,67,68,71,73,74,81]. Transformations are mathematical approaches used to decrease the heteroscedasticity of dataset, which the variability between variables is dramatically reduce. Log [43,57,66,69,70,72,76,79] is the main transformation in BC metabolomics. However, cubic root [51,59] and quantile [48] transformations are also used.

Table 1. 2 - Chemometric methods applied to metabolomic studies.

Biological sample	Data Pre-treatment	Pre-processing	Processing	Validation	Post-processing	References
Diagnostic tool						
Human BC cell lines	Scaling (Pareto scaled), Transformation (log transformed)	PCA, HCA	OPLS-DA	LOOCV, ROC	none	[43]
	Centering (mean centered), Scaling (autoscaled)	ANOVA, PCA, HCA, Pearson correlation	PLS-DA	LOOCV	none	[46]
	Experimental correction (sample weight corrected)	PCA	none	none	none	[61]
	none	none	none	none	none	[45]
	none	ANOVA, PCA	PLS, LDA	K-CV	none	[44]
Human blood	none	T-test	PLS-DA, LRA	ROC, Permutation test	none	[47]
	Scaling (total intensity value scaled)	Wilcoxon test	RF	ROC, Bootstrapping	none	[58]
Human Exhaled breath	Transformation (quantile transformed)	T-test	RF, SVM	LOOCV, ROC, Bootstrapping	none	[48]
Human plasma	none	Correlation feature selection (CFS)	LRA, SVM, RF	K-CV, ROC	Pathway-based metabolite sets analysis (pathifier)	[65]
	Scaling (median value scaled), Transformation (log transformed)	ANOVA, PCA	none	none	none	[66]
	none	T-test, PCA, HCA	PLS-DA, RF	K-CV, ROC	Pathway enrichment analysis (metaboanalyst)	[33]

Biological sample	Data Pre-treatment	Pre-processing	Processing	Validation	Post-processing	References
Human BC cell lines, plasma	none	KS-test, T-test, PCA	none	none	none	[50]
Human saliva	none	none	none	none	none	[52]
	Scaling (Pareto and total intensity value scaled)	T-test, PCA	PLS-DA	ROC, Permutation test	none	[53]
	none	none	LDA	K-CV, ROC	none	[54]
	Scaling (autoscaled and median value scaled), Transformation (cubic root transformed)	MW-test, HCA	PLS-DA, OPLS-DA	MCCV, Permutation test	none	[51]
	Experimental correction (internal standard corrected)	MW-test, PCA	PLS-DA, SVM, LRA	K-CV, ROC	none	[105]
Human tissues	Scaling (Pareto scaled)	PCA	OPLS	K-CV	none	[55]
	none	PCA, HCA	none	none	none	[56]
	Scaling (median scaled), Transformation (log transformed)	T-test	none	none	none	[57]
	Scaling (total intensity value scaled)	T-test	PLS-DA	LOOCV, ROC	Pathway enrichment analysis (metaboanalyst)	[74]
	Scaling (median scaled)	PCA	PLS-DA	LOOCV	none	[75]
	Scaling (median scaled)	T-test, PCA	PLS-DA	LOOCV	none	[78]
	Transformation (log transformed)	T-test, Pearson correlation, HCA	none	none	none	[76]
	none	T-test, Pearson correlation	PLS-DA	K-CV, ROC	none	[77]

Biological sample	Data Pre-treatment	Pre-processing	Processing	Validation	Post-processing	References
Human serum	Centering (mean centered), Scaling (total intensity value scaled)	PCA	PLS-DA, OPLS-DA	K-CV, ROC	none	[67]
	Centering (mean centered), Scaling (total intensity value scaled)	T-test, PCA, ANOVA	OPLS	K-CV, ROC, Bootstrapping	none	[68]
	Transformation (log transformed), Experimental correction (internal standard corrected)	ANOVA, PCA	PLS-DA, LRA	K-CV, ROC	none	[69]
Human urine	Scaling (autoscaled and median value scaled), Transformation (cubic root transformed)	T-test, HCA	PLS-DA, SVM, RF	MCCV, ROC	Pathway enrichment analysis (metaboanalyst)	[59]
Drug therapy						
BC cell line	none	T-test	none	none	none	[62]
Human blood	Transformation (log transformed)	T-test, Pearson correlation	none	none	none	[70]
BC tissues	Centering (mean centered), Transformation (log transformed - only in univariate analysis)	T-test, Pearson correlation, PCA	PLS-DA	K-CV, Permutation test	none	[79]
	Scaling (mean scaled - only in PCA)	ANOVA, Spearman correlation, PCA	RF	K-CV, Bootstrapping, Permutation test	none	[80]
	Scaling (total intensity value scaled)	MW-test	OPLS-DA	LOOCV	none	[81]
	none	Spearman correlation	none	none	none	[82]
Serum	Scaling (total intensity value scaled)	T-test	PLS, PLS-DA	LOOCV, ROC	none	[71]

Biological sample	Data Pre-treatment	Pre-processing	Processing	Validation	Post-processing	References
Serum, tissues, cell lines	none	T-test, ANOVA, PCA	PLS-DA	K-CV, ROC	Pathway enrichment analysis (metaboanalyst)	[63]
Urine	Scaling (total intensity value scaled)	KS-test, L-test, SW-test, T-test, PCA	OPLS-DA	K-CV, ROC	Pathway enrichment analysis (metaboanalyst)	[73]
	none	T-test, PCA	PLS-DA	K-CV	none	[84]
Metabolic reprogramming						
Human BC cell lines, BC xenografts	none	ANOVA, PCA	PLS-DA	K-CV	none	[64]
Mouse BC tissue	Scaling (median scaled)	ANOVA, PCA	none	none	none	[83]
Endogenous factors						
Human plasma	none	T-test, Spearman correlation, PCA	LRA	ROC	none	[31]
Human serum	Transformation (log transformed)	Pearson correlation, PCA	LRA	none	none	[72]

ANOVA – Analysis of variance; ROC – Receiver operating characteristic; LOOCV – leave-one-out-cross validation; AUC – Area under the curve; BFS– Bootstrap feature selection; GGM – Gaussian graphical modelling; HCA – Hierarchical cluster analysis; LDA – Linear discriminant analysis; MCCV – Monte Carlo cross validation; MWT – Mann Whitney U test; NRI – Net reclassification improvement; OPLS-DA – Orthogonal projections to latent structures discriminant analysis; LRA - logistic regression analysis; OSC-PLS – Orthogonal signal correction partial least squares; PC – Pearson correlation; PCA – Principal component analysis; PEA – Pathway enrichment analysis; PLS-DA – Partial least squares discriminant analysis; RF – Random Florest classifier; SCC – Spearman correlation coefficient; SVM – Support vector machine; VIP – variable importance in projection.

Other normalization approaches based on experimental corrections are also used in metabolomics, such as internal standards [105,106] and sample weight [61]. Internal standards normalization assumes that the heteroscedasticity of all variables is systematic and can be corrected by variance of internal standards. Sample weight normalization is the direct correction of variables values by experimental sample measures (e.g., volume and weight).

Pre-processing

Pre-processing methods are performed to obtain an exploratory projection of dataset or an overview of variables importance prior to prediction models processing. Primarily, normality tests are used to determine if the data distribution is normal (parametric) or not normal (non-parametric). The most commonly used are Kolmogorov-Smirnov test (KS-test) [50,73], Shapiro-Wilk test (SW-test) [73] and Lilliefors test (L-test) [73]. Two types of approaches are normally used in exploratory projections/variables importance ranking of BC metabolomics datasets: univariate and multivariate analysis. Univariate statistical methods are used to analyzed only one variable at a time, being useful to easily discover significant differences or measure correlations between samples groups. The differentiation is based on variance between groups by rejection of the null hypothesis or acceptation the alternate hypothesis [101,107,108]. The most common methods used when the data is parametric T-tests [31,47,68,70,71,74,76–79,84,48–50,53,57,59,62,63] and ANOVA [33,44,46,63,64,66,68,69,72,83]. T-tests, such as Student and Welch's tests, are recommended to analyze differences between two groups, and ANOVA-based methods, such as one-way ANOVA, two-way ANOVA, factorial ANOVA and MANOVA are used to evaluate more than two groups. Alternative univariate methods are implemented when the assumption of the normal distribution is non-parametric, such as Mann-Whitney test (MW-test) [51,81,105] and Wilcoxon test (W-test) [58]. In addition, univariate methods are also widely used to measure the correlations between continuous variables and response. The Pearson correlation [46,70,72,76,77,79] is the preferred option for linear relationships in populations with normal distribution. On the other hand, the Spearman correlation [31,80,82] is usually used in non-parametric datasets [109]. More complex correlation methods are also used in data analysis, such as Correlation Feature Selection (CFS) [65], where the appropriate correlation measure and a heuristic search strategy are performed by experiments on artificial and natural datasets based on algorithms.

Similarly, the multivariate methods are also widely used for exploratory studies to obtain dataset patterns based on relationships between groups, being divided into two sub-groups, unsupervised and supervised methods. Unsupervised methods are the preferential option for exploratory studies, where the modeling process is based only on the explanatory variables, without external intervention of user

[100]. The most commons are principal component analysis (PCA) [31,33,63,64,66–69,72,75,78,79,43,80,83,84,105,44,46,49,53,55,56,61] and hierarchical cluster analysis (HCA) [33,43,51,56,59,76]. PCA provides the projection of dataset into low dimensional based on orthogonal transformation, converting the variables variability from a set of observations into score vectors and loadings, called principal components [100,110]. HCA methods are used to form subsets of samples at ordered levels based on variables similarities/dissimilarities (such as distances or correlations), and can be performed in agglomerative mode (samples are aggregate into clusters) or divisive mode (complete dataset is divide into clusters). In both modes, the linkage criterion need to be selected, being that the most commonly used are single-linkage clustering (the minimum of distances) and complete linkage clustering (the maximum of distances) [111,112].

Processing methods

After the explorative studies and variable selection, the next step is the processing of dataset in order to create a predictive response model to classification of new samples (ex. diagnostic tools), identification of valuable variables (ex. biomarkers) or exploring the mechanisms of metabolomic studies (ex. metabolic pathways). In this stage, the supervised methods are the preferential choice, where the response models are mainly based on two types, continuous (regression) and discrete (classification) [100,102]. The main methods for continuous response are based on multiple linear regression (MLR), sometimes called ordinary least squares (OLS). MLR is performed to predict the values of a dependent variable (response) based on a set of continuous explanatory variables, assuming a linear combination of the explanatory variables [110]. The most applied MLR-based method in metabolomics is partial least squares (PLS) [44,55,68,71]. Unlike PCA, which uses only the variables variation, PLS is a predictive and supervised method that use an informative response to maximize the covariance between the explanatory variables and the response, producing score vectors and loading vectors. The prediction model is based on interaction between the variables and response, ignoring the variables with irrelevant importance. The importance of each variable is defined according the PLS-based criteria, such as loading weights, variable importance on projection scores, regression coefficient, target projection and selectivity ratio [100,101,110]. However, when categorical variables are introduced, the discrete models should be used. Discrete models provide a predictive classification of response based on continuous and categorical variables, being classified into linear or non-linear. In linear methods, the classification is performed by highest probability based on linear relationships between explanatory variables, where exist a grouping variable (categorical). Linear discriminant analysis (LDA) [44,54] is the preferential method to classification models of discrete responses. LDA perform linear transformations of explanatory variables to create

discriminant functions that will maximize the separation between multiple classes of samples (groups) based on the information of the categorical variables (Liland, 2011). Among the various LDA-based methods, PLS-DA [33,46,74,75,77–79,84,105,113,47,51,53,59,63,64,67,69] is most widely used in metabolomics studies. PLS-DA is a successful combination of PLS and LDA that provides a visual low-dimensional pattern of samples discrimination based on the analysis of relationships between continuous and categorical variables [101,110]. Recently, some extensions of PLS-DA were used in BC metabolomics, namely the OPLS-DA [43,51,67,73,81]. OPLS-DA separates out response orthogonal variations in rotations of the original component [110].

On the other hand, non-linear methods are used when metabolomics dataset follow a non-linear response. The most applied non-linear methods are support vector machines (SVM) [48,59,65,105], random forests (RF) [33,48,58,59,65,80] and logistic regression analysis (LRA) [31,47,65,69,72,105]. SVM is a kernel-based model used for regression and classification of non-linear datasets, transforming the non-linear data into more general spaces (linear) by algorithm based on kernels functions. SVM perform the mapping of dataset into a high-dimensional space through kernels functions for the separation of two groups of samples into distinctive regions. The separation is based on support vectors, which are points (samples) on the boundary or on the incorrect side of the margin supporting the separation. SVM is a versatile method that transforms non-linear complex datasets into a high-dimensional space where classes are linearly separable [100,101,110]. RF is a non-linear method for regression and classification of high-dimensional datasets, where a large number of classification and regression trees are created by bootstrapping (replacement) based on random selection of a training samples from the original dataset. Afterwards, bootstrapping is performed systematically to build a large group of simple trees that are used to estimate classification accuracy of the model [100,101]. Another non-linear predictive method widely used is LRA, which is similar to linear regression, but with a binomial response variable. LRA is used to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables [114].

Model validation

The validation of predictive models is a key step in data analysis of metabolomics studies. Validation process analyzes the performance/ability of model to predict correctly the hypothesized relationships between variables and responses [101]. Several validation methods have been used in BC metabolomics. The coefficient of determination (R^2) is the simplest method to evaluate the ability of predictive model, being used for continuous responses. The R^2 is expressed as the ratio between 0 and 1, where a value of 1 indicates the perfect prediction. However, this validation is recommended

for small datasets, due to fact that the R^2 value tends to be increased when a predictor variable is added to the model [115]. However, in validation of predictive models used to high-dimensional and complex datasets, as the case of metabolomics studies, the cross validation (CV) methods are the preferential option. CV provides qualitative and quantitative analysis of the model ability to model's ability to predict new independent samples without collecting additional data. During the CV, the available data are split into two sets, where one set is used to create a predictive model using the values of continuous and predictor variables (training set). The second set is used to test the performance of predictive model (validation set) [100]. The most applied CV procedure is k-fold (K-CV) [33,44,73,77,79,80,84,105,54,55,63–65,67–69]. K-CV processing is based on random partition of original dataset into equal sized subsamples (k). A single k subsample is used as the validation set for testing the model, and the remaining k -1 subsamples are used as a training set. This process is then repeated k times (folds), with each of the k subsamples being used exactly one time as the validation set [106]. One special type of K-CV is the leave-one-out cross validation (LOOCV) [43,46,48,74,75,78,81], where the number of folds equals the number of k subsamples. LOOCV is considered an exhaustive CV, being recommend for small datasets [106,115]. Another type of CV is the Monte Carlo cross validation (MCCV) [51,59]. Although less used in metabolomics than LOOCV, MCCV is asymptotically consistent and showed better prediction ability. In MCCV proceeding, significant part of dataset is leaved out at a time during model validation, repeating systematically this procedure several times [116,117]. The Q^2 value, which is the equivalent R^2 value, is the preferential coefficient of determination for CV procedures.

A visual and easy model validation method is the receiver operating characteristic (ROC) curve [31,33,65,67–69,71,73,74,77,105,43,47,48,53,54,58,59,63] which the prediction ability of a model is validated considering the specificity (ratio of the correctly predicted negatives) and sensitivity (ratio of correctly predicted positives). The ROC curve is given by plotting the sensitivity versus (1 - specificity) across a series of cutoff points. The area under curve (AUC) is a quantitative measure (between 0 and 1) of the ability of predictive model, where a AUC value close to 1 indicates a nearly perfect prediction response [100,115].

Random resampling-based methods are a robust alternative for model validation. The most used in BC metabolomics are bootstrapping [48,58,68,80] and permutation tests [47,51,53,79,80]. Bootstrapping is a model validation based on replacement of samples, which can be considered non-parametric when the replacement is from the original dataset, or parametric when random noise is added from a recognized distribution to the dataset to estimate underlying sampling distribution or establish robust confidence intervals. Normally, in metabolomics studies the common approach is non-parametric bootstrapping [115,118]. Permutation tests provide the exact control of false positives

from a predictive model (linear or non-linear), under minimal assumptions, based on differences between the randomly permuted response variables model and the original model. Permutation tests are based on a repeatedly permuting (repetitive reordering) of the N entries in the response variable. Permuted vectors containing integers between 1 and N are produced in a random number generator, creating new scrambled response variables only by switching their internal positions. The scrambled vectors are modelled one by one, where for every test, the R^2 and Q^2 values are calculated and saved. After, these values are compared with the values calculated from the original data. The results of permutation tests are displayed as a percentage overlap between the real and permuted R^2 and Q^2 values, where a 0% of overlap is the optimal result [110,119].

Post-processing

The post-processing step consists in interpretation of metabolomic responses from original dataset. Normally, pathway analysis is the most used strategy to provide an overview of association/relationship between identified metabolites and metabolic pathways and other general biological networks. Pathifier [65] and metaboanalyst [33,59,63,73,74] are the most used software for this propose in metabolomics.

Future directions

The advances in analytical techniques and chemometric methods in metabolomics have been growing rapidly becoming possible the identification of potential biomarkers. Furthermore, the integration of analytical platforms increases the comprehensive analysis of metabolites in biological samples. In this context, metabolites became valuable identifications, regardless their hierarchical source, enabling the phenotypic properties in a biological system. Additionally, the identification of key metabolic pathways from which significant metabolites are linked, it is possible to reveal potential targets for cancer therapy.

Also, standard procedures for sample collection, data analysis and shared in repositories have potential to be adopted by both researchers and medical communities.

Since the metabolome instantly responds to environmental stimuli including therapeutic or surgical intervention, could be also used to monitor the metabolic status of the individual and indicate any possible toxic effects. Moreover, metabolomics may help in the detection of potential cancer biomarkers, being useful for example in the development of different devices, including biosensors, that can significantly improve the cancer diagnosis. These devices include a biorecognition element within a biosensor system. The biorecognition molecules interact with the target, which is then converted into a measurable signal by a transducer. Basically, these molecules, usually enzymes or

antibodies, can be immobilized on the transducer surface and interact with the target (biomarker) to produce a signal is interpreted, providing information about the disease and their possible recurrence after therapy.

SECTION 2| Thesis Aims and Scope

Although breast cancer (BC) research is directed towards the improvement of diagnostic/screening techniques in the last decades, new tools are needed for BC diagnosis due to the limited sensitivity and specificity of current screening methods, as well as the perceived discomfort of mammography and potentially hazardous exposure to radiation. In recent years a promising and attractive hypothesis is based on small molecular-weight volatile organic metabolites (VOMs), consumed or released by cancerous cells, is emerging as an attractive and non-invasive approach for cancer diagnostics. These molecules, which can be perceived as odors, have been shown to function as cancer ‘‘biosignatures”. In this context, the present thesis incompassed the study of the metabolomic profile of urine, cell lines and tissues from BC patients and CTLs with the aim of discover a set of metabolites able to be used as potential BC biomarkers.

To achieve the main goal, the experimental design was followed being summarized in the flowchart:

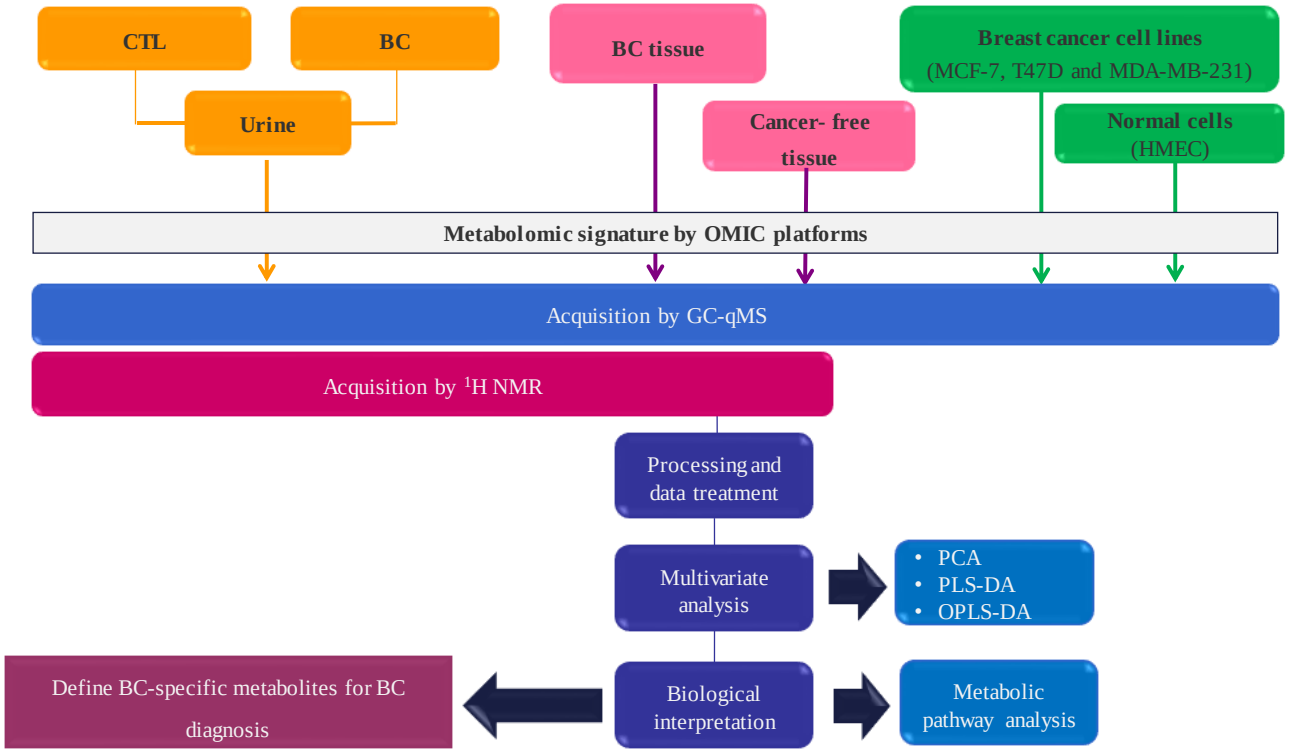

Figure 2. 1 - General diagram of the study performed in the thesis.

The specific objectives of the thesis included:

I. The application of a multivariate experimental design (central composite design) to optimize the SPME-influencing parameters as a mean to improve VOMs extraction efficiency and their application to urine samples from BC patients and CTLs;

II. The establishment of the urinary metabolomic profile by ^{1}H NMR from BC patients as well as from CTLs as a powerful tool to detect potential BC-specific metabolites;

III. Study of VOMs released from human BC cell lines: MCF7, T47D and MDA-MB-231, and from normal breast cell lines, as primary mammary epithelial cells (HMEC);

IV. To characterize the metabolomic fingerprint of BC tissues and compare to breast normal tissue;

V. Combination of NMR- and GC-qMS-based method, multivariate statistics and metabolic correlation networks, in order to find a panel of biologically significant metabolites as potential BC biomarkers.

SECTION 3| Metabolomic Pattern in Breast Cancer

3.1| Implementing a Central Composite Design for the optimization of solid phase microextraction to establish the urinary volatomic expression. A first approach for breast cancer

(Silva et al.; Metabolomics **2019***; 15:64)*

Abstract

BC is positioned as the second among all cancers remaining at the top of women´s diseases worldwide followed by colorectum, lung, cervix, and thyroid cancers. The main drawback of most the screening/diagnostic methods is their low sensitivity/specificity and in some cases the invasive procedure required to obtain the samples. On the present investigation, we report a statistical design was to evaluate by central composite design the influence towards the optimization of the most significant variables of solid-phase microextraction (SPME) procedure for the isolation of volatile organic metabolites (VOMs) from urine of BC patients (*n*=31) and healthy individuals (CTL; *n*=40). The establishment of the urinary volatomic composition, through gas chromatography-mass spectrometry (GC-MS) analysis, can boost the identification of volatile organic metabolites (VOMs) potential BC biomarkers useful to be used together or to complement the current BC diagnostics tools. Better early detection methods are needed to improve the outcomes of patients with colorectal cancer (CRC). Several combinations of experiments were considered with a Central Composite Design (CCD) of Response Surface Methodology (RSM) for the urinary volatomic pattern. Three-level three-factor CCD was employed assessing the most important extraction-influencing variables - fiber coating, NaCl amount, extraction time and temperature. The optimal conditions were achieved using a carboxen/polydimethylsiloxane fiber with 15% (w/v) NaCl during 75 min at 50 ºC.

A total of ten VOMs belonging to sulfur compounds, terpenoids and carbonyl compounds presented the highest contribution towards discrimination of BC patients from CTL (variable importance in projection (VIP) >1, $p < 0.05$). The discrimination efficiency and accuracy of urinary metabolites was ascertained by receiver operating characteristic (ROC) curve analysis that allowed the identification of some metabolites with highest sensitivity and specificity to discriminate the groups. The results obtained with this approach suggest the possibility to identify endogenous metabolites as a platform to discovery potential BC biomarkers and paves a way to explore the related metabolomic pathways in order to improve BC diagnostic tools.

Keywords: Central composite design; Breast cancer; Urine; Metabolomics; Chemometric tools

Introduction

BC remains as the most common invasive cancer in women worldwide , accounting for 24.2% of all cancer cases, followed by colorectum, lung, cervix, and thyroid cancers according to the International Agency for Research on Cancer (IARC) [3]. Being a disease caused by a combination of genetic and environmental factors, BC is often characterized by an absence of early symptoms, which results in a late detection of the disease. Detection at advanced stages of BC implies that the treatment is harder and uncertain. Appropriate screening methods have been conducted within organized preventive examinations and have made significant contributions to early BC detection [39,120,121]. Moreover, if the disease is detected at an early stage of development, the healing percentage will increase. In this sense, it is necessary to detect disease as earlier as possible [121]. Furthermore, there is still the need for the development of new methodologies to support or monitor the disease together with current diagnostic tools, namely with mammography, ultrasound imaging or tumor marker analysis. Moreover, before the initiation of a BC therapy, complex and time-consuming analyzes are still required, being the most significant the determination of the histological type and grading, and the evaluation of ER, PR and HER-2, among others [122]. The main drawback of most of these screening/diagnostic methods is their low sensitivity and specificity and in some cases the invasive procedure required to obtain the samples [123]. Taking into account these aspects, the investigation is leading towards the research of new tools that can support the clinicians in BC treatment and monitoring of their recurrence [124]. In this sense, in recent years the -omic studies have emerged as a powerful tool to investigate the changes and/or metabolic responses of living systems to stimuli or genetic modifications [39]. The metabolome profile represents the quantitative and qualitative analysis of the complete set of metabolites present in cells, body fluids or tissues [125]. To date, based on the most used biological specimens (e.g., exhaled breath, urine, saliva, blood), the metabolome coverage in BC can be maximized by merging different technologies for metabolic profiling, namely with GC-MS. The results obtained can be useful to classify BC, helping to identify new prognostic and predictive markers and to discover new targets for future therapeutic interventions [126]. Many metabolomic studies have referenced the presence of glycolytic markers as an indication of the Warburg effect driven by malignancy, so that it remains a metabolic pathway of ongoing interest in cancer research [39]. Among them, the study of VOMs present in biological samples, namely in saliva, urine, exhaled breath and tissues can be useful for a cancer diagnosis, in particular for BC [107]. The principle is that disease and normal condition can be distinguished from each other by changes in their physiology and metabolic rates, which leads to the production of disease specific alterations in VOMs [30]. The most common procedure used in extraction of volatile compounds in biological samples is SPME, normally in headspace mode, being used in several

biological matrices [44,59,127,128]. In fact, SPME offers several advantages over conventional solvent extraction procedures: it is rapid, easy to use, solvent-free and does not require any concentration step before analysis [129]. Being a technique based on the equilibrium between sample matrix /headspace and between headspace/fiber coating, HS-SPME efficiency is dependent on several factors, namely the target compounds to be extracted, fiber coating, extraction temperature, extraction time and *salting-out* effect [127]. Taking into account that SPME is affected by several parameters and their optimization is a laborious and time-consuming process, the RSM can be a suitable tool for optimizing the process. This approach consists of a combination of statistical and mathematical techniques and has been highly used in optimization processes, since RSM reduces the number of experimental assays. Frequently, CCD is used as an experimental design to fit a second-order polynomial function by a least squares technique. An equation is used to define how test variables affect the response and determine the interrelationship among the variables [130,131]. Monteiro et al. [131] used CCD to optimize the SPME extraction parameters to discriminate patients with renal cell carcinoma, whereas Calejo et al. [132] used CCD to select the optimal conditions to extract carbonyl compounds to discriminate individuals with smoking habits using an in-solution derivatization with *O*-(2,3,4,5,6-pentafluorobenzyl)hydroxylamine hydrochloride (PFBHA) combined with SPME. As for example, GC-qMS was used to screen salivary volatiles for putative BC as an exploratory study involving geographically distant populations [51], also to establish the metabolomic signature of human BC cell lines [44] and to discriminate different types of cancer based on urinary volatomic biosignatures [59]. In the first study, up to 120 VOMs from distinct chemical families, with significant variations among the groups, were identified [51], whereas Silva et al. [44] and Porto-Figueira et al. [59] identified 60 and 130 VOMs in BC cell lines and urine, respectively. In turn, Dougan et al. [66] used GC-MS to evaluate the detectability, reliability, and distribution of metabolites measured in pre-diagnostic plasma samples of a pilot study from women listed in the Northern California site of the BC Family Registry. In this study, 661 VOMs were detected, 338 (51 %) of them were found in all samples, and 490 (74 %) in more than 80 % of samples.

In the current study, CCD was used for the optimization of experimental parameters, such as fiber coating, NaCl amount, extraction time and temperature as useful strategy to improve SPME extraction efficiency of VOMs from urine samples. The potential of SPME/GC-qMS data combined with chemometric tools was evaluated to discriminate the urinary volatomic biosignatures from BC patients and healthy controls (CTL), in order to identify a set of potential BC-specific biomarkers that could be used together with current BC techniques as a useful tool to improve its diagnosis.

Materials and Methods

Chemicals and materials

Sodium chloride (NaCl), hydrochloric acid (HCl) and 4-methyl-2-pentanol were supplied by Panreac (Barcelona, Spain) and Sigma Aldrich (St. Louis, MO, USA), respectively. The digital stirring plate (Cimarec™) was supplied by Thermo Scientific (Waltham, MA, USA) while SPME holder for manual sampling, together with 65 µm polydimethylsiloxane/divinylbenzene (PDMS/DVB), 50/30 µm (DVB/CAR/PDMS) and 75 µm carboxen/polydimethylsiloxane (CAR/PDMS) fibers, were purchased from Supelco (Bellefonte, PA, USA).

Urine samples collection and preparation

Urine samples (first urine morning) from BC patients (*n*=30) were taken at the Haemato-Oncology Unit from Dr. Nélio Mendonça Hospital, while the urine collection from healthy individuals (*n*=40) (Table 3.1.1) was carried out in Blood Transfusion Medicine Service in the same Hospital at the same time. Participants were instructed to collect the first urine morning (after the rejection of the first urine stream) into a sterile bottle using a disposable collector that was provided to them.

Table 3. 1. 1 - List of collected urine samples from breast cancer (BC) patients and healthy volunteers (CTL).

Urine Samples	**BC**	**CTL**
Number	30	40
Age (range, median)	(44-85, 65)	(43-80, 64)

The collected urine samples were aliquoted into 4 mL glass vials and stored at -80 °C until analysis. The analysis begun with the thawing of urine samples, after this they were centrifuged at 4000 rpm for 20 min at 4 °C. Then, the supernatant was collected and used for the extraction procedure. All the analyzes were performed in triplicate.

The research was approved by the Ethics Committee of Funchal Central Hospital Dr. Nélio Mendonça and have been performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. All the participants were fully informed of the objectives of the study and signed the informed consent.

HS-SPME optimization: CCD approach

Three types of fiber coatings (Supelco, Bellefonte, PA, USA) were used to evaluate the extraction performance of VOMs from urine samples: PDMS/DVB, DVB/CAR/PDMS and CAR/PDMS. They were daily conditioned according to the manufacturer´s recommendations, in order to avoid carryover between sets of analyzes. For the HS-SPME optimization, a CCD model was applied for each fiber, in order to build a second order (quadratic) model for the response. This model was defined by the equation $2^k + 2k + n$, where k is the number of factors and n is the number of centre runs. In this case, k and n were set at 3 and 5, respectively, meaning that 45 experiments (each experiment was performed in triplicate) were ran with 15 experiments for each fiber in the centre of the design. The factors were salt addition (NaCl, g), extraction time (t_{ext}, min) and extraction temperature (T_{ext}, ◦C). The CCD included five different levels (centre runs), k central points (mean value of each variable) and $2k$ star points (minimum and maximum level of the range). The axial distance (α) given by the equation, $\alpha = 2k/4$, was calculated to satisfy the rotatable of the design and was considered $\alpha = \pm 2$ [133]. The 45 experiments (combinations) were run in randomised order to minimise the effects of variability in the response due to irrelevant factors (Table 3.1.2).

Table 3. 1. 2 - Experimental conditions and values for the response (expressed in total areas) obtained for the CCD used for the optimization of the extraction conditions of urine samples by HS-SPME.

Run order	**Experimental conditions**					
	Fiber	**T_{ext}, ◦C**	**t_{ext}, min**	**NaCl (%)**	**Number VOMs**	**Response (total areas)**
1		40	30	10	35	1.89E+09
2		40	30	20	44	2.05E+09
3		40	60	10	49	4.98E+09
4		40	60	20	39	3.66E+09
5		60	30	10	56	5.75E+09
6		60	30	20	62	6.93E+09
7		60	60	10	72	1.18E+10
8	CAR/PDMS	60	60	20	79	1.11E+10
9		30	45	15	45	2.43E+09
10		70	45	15	108	1.27E+10
11		50	15	15	42	1.28E+09
12		50	75	15	108	1.96E+10
13		50	45	5	84	5.54E+09
14		50	45	25	83	4.97E+09
15		50	45	15	91	5.34E+09
16		40	30	10	54	2.47E+09
17		40	30	20	58	1.65E+09
18		40	60	10	63	2.81E+09
19	DVB/CAR/PDMS	40	60	20	66	3.18E+09
20		60	30	10	73	5.18E+09
21		60	30	20	82	6.41E+09
22		60	60	10	75	6.20E+09

Run order	Experimental conditions					
	Fiber	T_{ext}, °C	t_{ext}, min	NaCl (%)	Number VOMs	Response (total areas)
23		60	60	20	82	5.74E+09
24		30	45	15	56	1.52E+09
25		70	45	15	75	1.24E+10
26		50	15	15	57	1.53E+09
27		50	75	15	68	6.30E+09
28		50	45	5	49	2.67E+09
29		50	45	25	64	5.05E+09
30		50	45	15	57	3.45E+09
31		40	30	10	43	1.11E+09
32		40	30	20	45	2.21E+09
33		40	60	10	40	1.74E+09
34		40	60	20	48	2.13E+09
35		60	30	10	51	3.37E+09
36		60	30	20	61	4.68E+09
37		60	60	10	66	6.01E+09
38	PDMS/DVB	60	60	20	61	7.20E+09
39		30	45	15	50	1.69E+09
40		70	45	15	67	7.59E+09
41		50	15	15	57	1.38E+09
42		50	75	15	57	4.65E+09
43		50	45	5	49	3.82E+09
44		50	45	25	52	3.73E+09
45		50	45	15	49	4.65E+09

[a] T_{ext}, extraction temperature, [b] t_{ext}, extraction time

Application to urine samples

The optimal HS-SPME extraction conditions used to analyze urine samples from CTL and BC groups were obtained with CCD model as described in previous section. Briefly, urine samples were thawed and then 4 mL of urine was placed into 8 mL vials together with 17 % NaCl (w/v) and 100 µL of the internal standard (IS, 4-methyl-2-pentanol, 1.6 mg/L). The pH was adjusted to 2 with small amounts of HCl 5M. Then, the vial was capped with a Teflon (PTFE) septum using a screw cap and the CAR/PDMS fiber was introduced and exposed into the headspace during 75 min at 50 °C at 800 rpm (0.5 mm × 0.1 mm bar). After this period, the fiber was removed from the vial and inserted into the GC injection port and the extracted VOMs were desorbed for 10 min at 250 °C. Each sample was analyzed in triplicate and the blanks were performed before each analysis.

Gas chromatography quadrupole mass-spectrometry (GC-qMS) conditions

After the extraction procedure, the SPME fiber with the analytes was inserted into the injection port of an Agilent Technologies 6890N Network gas chromatograph system (Palo Alto, CA, USA) where the VOMs were desorbed at 250 °C for 10 min. The gas chromatograph was equipped with a 60 m × 0.25 mm I.D.× 0.25 µm film thickness, BP-20 (SGE, Dortmund, Germany) fused silica

capillary column and interfaced with an Agilent 5975 quadrupole inert mass selective detector. The following oven temperature profile was set: (a) 5 min at 45 °C; (b) increase temperature until 150 °C, at a rate of 2 °C min^{-1} (hold for 10 min); (c) 150 °C for 10 min; (d) increase temperature until 220 °C, at a rate of 7 °C min^{-1}; and (e) 220 °C for 10 min for a total GC run time of 87.5 min. The column flow was constant at 1.3 mL min^{-1} using Helium (He, N60, Air Liquide, Portugal) as the carrier gas. The injection port was operated in the splitless mode and held at 250 °C. For the 5975 MS system, the operating temperatures of the transfer line, quadrupole and ionization source were 270, 150 and 230 °C, respectively, while electron impact mass spectra were recorded at 70 eV ionization voltage and the ionization current was 10 µA. Data acquisition was performed in the scan mode (30–200 m/z). Metabolites identification was accomplished through manual interpretation through single ion monitorization (SIM) of spectra and matching against the Agilent MS ChemStation Software, equipped with a NIST05 mass spectral library with a similarity threshold higher than 80 % and comparison with commercially available standard samples when available. The analyzes were performed in triplicate and the results expressed by mean ± standard deviation.

Statistical Analysis

The statistical analysis for the CCD design was performed using Statistica, Version 10.0 (Statsoft Inc., Tulsa, OK, USA) while the web server Metaboanalyst 4.0 [134] was used for the multivariate statistical analysis - namely, principal component analysis (PCA), partial least squares-discriminant analysis (PLS-DA) and orthogonal projections to latent structures discriminant analysis (OPLS-DA) applied on urine metabolomic profile dataset in order to provide insights into the separations among the groups under study. For the evaluation of the applicability of the developed CCD method, data analysis was based on a non-targeted volatile approach. Moreover, the metabolites with Variable Importance in Projection (VIP) scores higher than 1.0 were selected by the PLS-DA analysis and used for the pathway analysis. Additionally, hierarchical cluster analysis by K-means of the two groups in study was carried out and Pearson´s correlation was used to build the heat map with the aim of identifying clustering patterns. The receiver operating characteristic curves (ROC) were attained to verify which metabolites had the highest sensitivity/specificity for a potential BC diagnosis. Also, Random forest (RF) classification was performed to determine the ability of VOMs to accurately classify the study subjects into their corresponding groups. Finally, the selected metabolites were used for the metabolic pathway analysis to identify the most relevant metabolic pathways involved in BC and CTL groups.

Results and Discussion

Optimization of HS-SPME methodology

Regarding the optimisation of the ideal SPME extraction-influencing parameters- for VOMs extraction from urine samples, three different fiber coatings were used. In this case, the authors choose to use the fibers with two or more coatings in order to cover the maximum amount of chemical classes of metabolites, which is the most appropriate procedure in an untargeted metabolomic approach. Overall, the performance of each fiber was evaluated based on the number of identified VOMs, relative GC peak area, and reproducibility expressed as the relative standard deviation (RSD%). The performance obtained in terms of relative GC peak area for each tested fiber is shown in Figure 3.1.1.

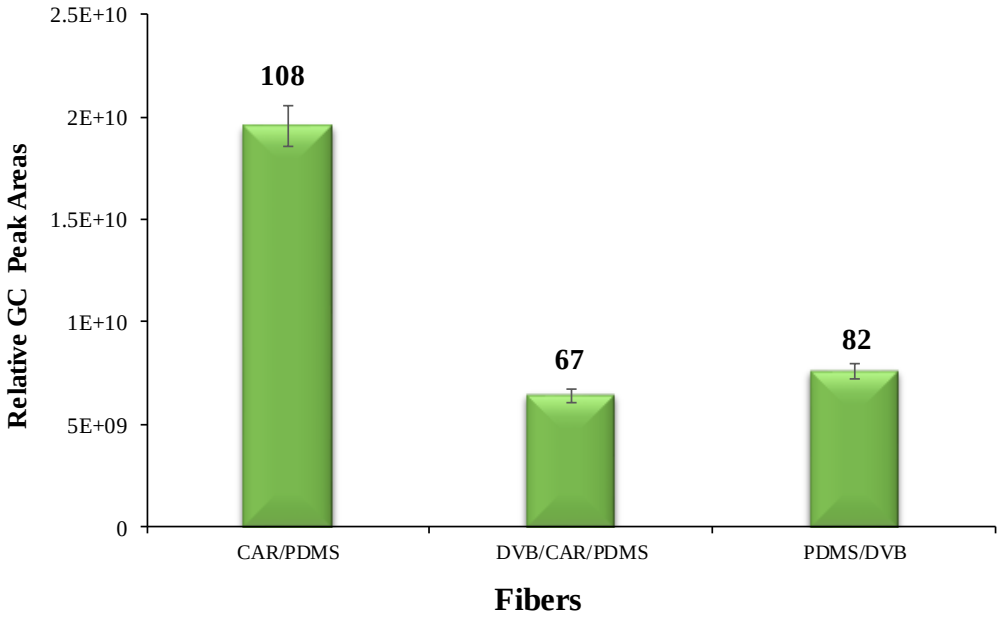

Figure 3. 1. 1 - Efficiency of fiber coatings in the extraction of VOMs from urine samples by HS-SPME. The numbers refer to the number of identified VOMs.

The CAR/PDMS fiber was the one that provided the uppermost extraction efficiency when compared with PDMS/DVB and DVB/CAR/PDMS fibers. In addition, the CAR/PDMS also presented the highest number of identified VOMs and the best reproducibility when compared with other tested fibers. Regarding the application of the CCD model, other factors that affected the SPME extraction efficiency were taken into account. This design was used as an experimental approach to assess the optimal value for each parameter with influence in SPME extraction efficiency and optimize the response of the process (Table 3.1.2). The particular value of a factor at which an experiment is run is called factor level. This approach allows the evaluation of the individual significance of each variable on the response, namely in the HS-SPME of VOMs from urine samples

and also the correlations among them. The other influencing parameters, such as the sample amount, pH and volume (4 mL of urine into a 8 mL glass vial), were fixed by the authors and based in previous studies with urine samples [127,128]. In the present case, 45 runs were performed randomly and the obtained response was based on the number of identified VOMs and sum of relative GC peak areas. The application of this type of statistical optimization procedure is useful over the univariate approach where the effect of possible interactions between important variables are not taken into account. Furthermore, this type of strategy enables the determination of the optimal conditions with fewer number of experiments when compared with full factorial design (FFD) [132,133]. The Pareto charts were built to illustrate the influence of each factor regarding the VOMs extraction performance as shown in Figure 3.1.2 (A and B).

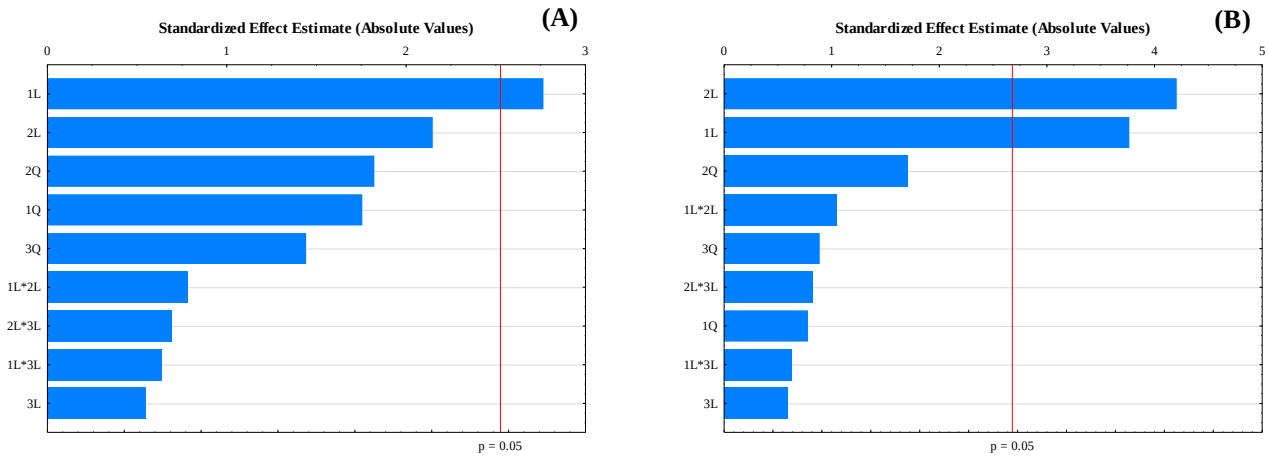

Figure 3. 1. 2 - Pareto charts of standardized effects of CCD factorial design for total GC peak area; (A) based on the extraction temperature and time, (B) extraction time and NaCl amount. The y axis of both graphs includes all variables considered in this study, their possible combinations and effects on the study. 1L-extraction temperature; 2L-extraction time; 3L-NaCl amount; Q-quadratic function; L-linear function.

This type of chart contains a bar for each evaluated factor as well as their combinations, which are ordered from the most to the least significant. Also, the vertical line indicates the critical value for Student´s t–test which was considered 0.05, meaning that any factor that extended that line was considered significant for the analysis. Regarding the charts obtained, in the case of the number of compounds identified, the variable with the greatest influence for the extraction of VOMs from urine samples was the extraction temperature (Figure 3.1.2 A) while, in terms of the total GC areas, was the extraction time and NaCl amount, as observed in Figure 3.1.2 B. Considering the combination of

variables, none of the evaluated combinations had a significant contribution for the analysis. In Figure 3.1.3 (A-F) the surface response graphs obtained for the CCD model are represented.

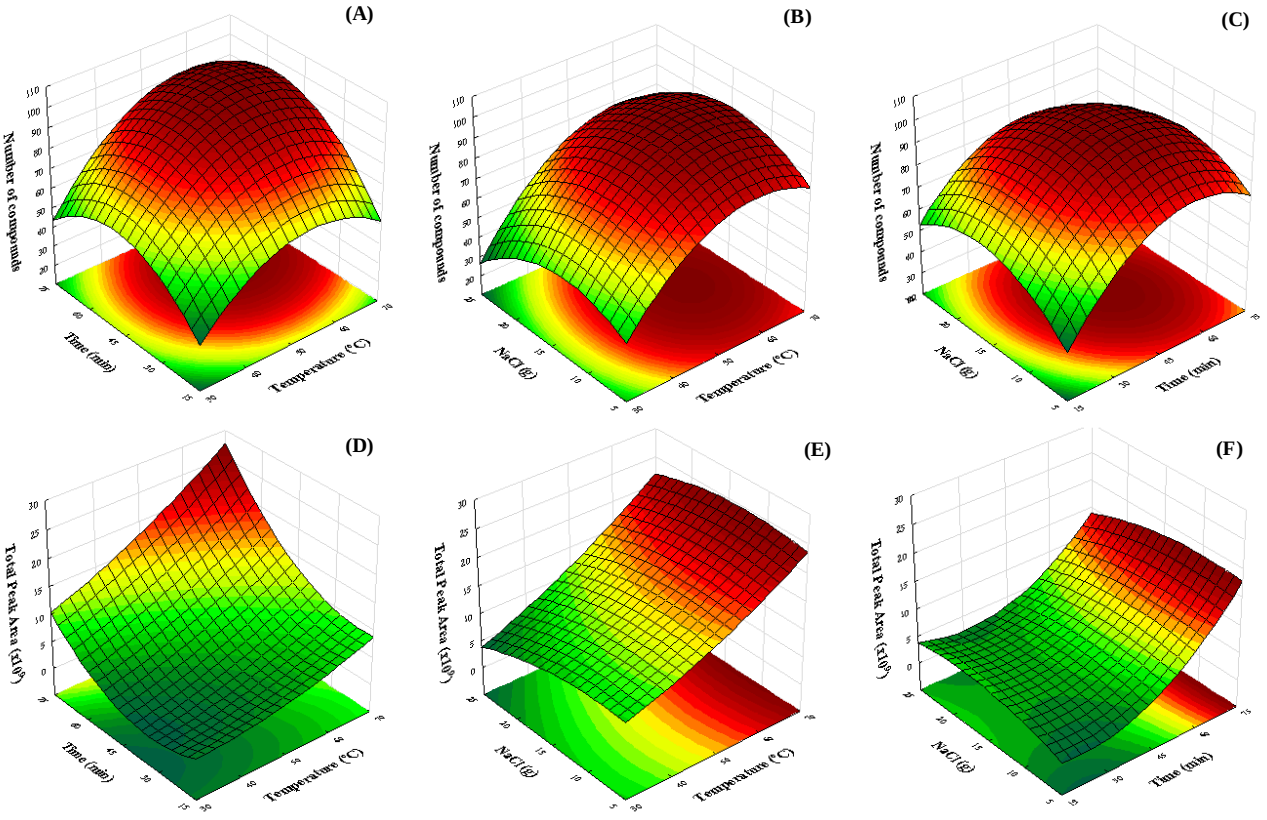

Figure 3. 1. 3 - Surface response obtained by central composite design (CCD) in the optimization of the SPME conditions and respective combinations. For number of identified compounds; (A) extraction time and temperature, (B) NaCl amount and extraction temperature and (C) NaCl amount and extraction time. Regarding the total peak areas, (D) extraction time and temperature, (E) NaCl amount and extraction temperature and (F) NaCl amount and extraction time.

Regarding the results, it was expected that for higher temperatures, the number of extracted VOMs increased, but also high temperatures are responsible for sample degradation and decrease the fiber lifetime. The first parameter that was taken into account was the number of identified VOMs (Figure 3.1.3 A-C) and as temperature, extraction time and NaCl amount increased (responsible for the "*salting-out*" effect), the extraction efficiency also increased, but when comparing with the total areas (Figure 3.1.3 D-F), the surface graphs indicated that the ideal extraction temperature was 70 °C. In this case, as mentioned before, this temperature was considered too high once there is the potential risk of metabolite degradation as reported in previous studies [127,128]. In this sense, 50 °C as extraction temperature and 75 min as extraction time were selected as optimal conditions. In sum, according to the results, the optimal conditions for the VOMs extraction from urine samples from BC

and CTL groups using 75 min as extraction time, 50 °C as extraction temperature and an addition of 15 % (w/v) of NaCl.

Urine volatomic pattern based on GC-qMS

A total of 116 VOMs were identified in BC and CTL urine samples which were classified in several chemical families, namely sulphur compounds, furanic compounds, phenols, terpenoids, carbonyl compounds, norisoprenoids, acids, alcohols and miscellaneous (Figure 3.1.4).

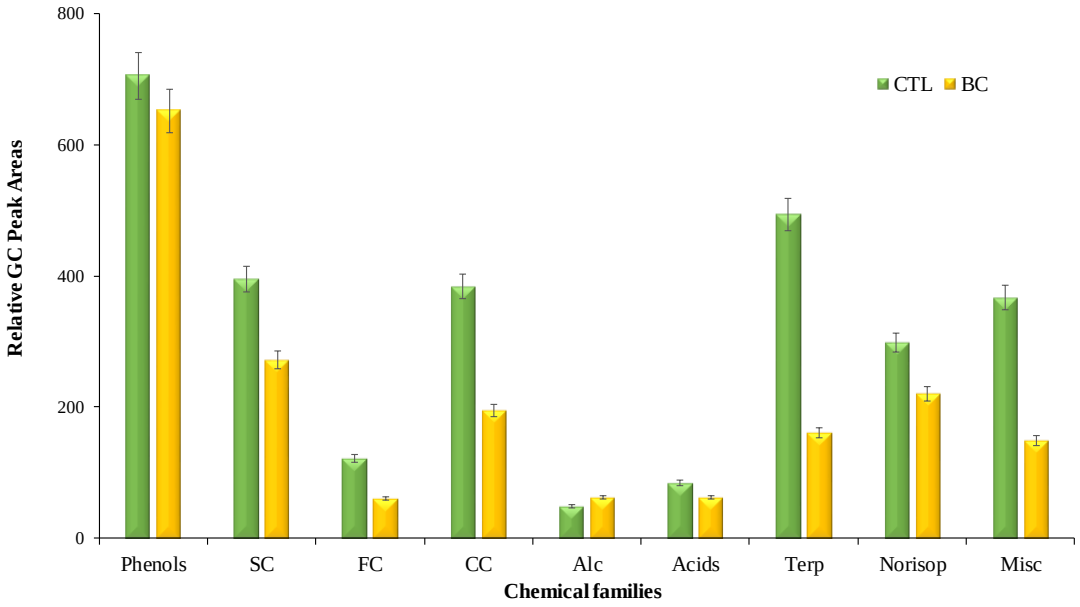

Figure 3. 1. 4 - Major chemical families identified in BC and CTL urine samples. Legend: SC-sulfur compounds; FC-furanic compounds; CC-carbonyl compounds; Alc-alcohols; Terp-terpenoids; Norisop-norisoprenoids; Misc-miscellaneous.

The data obtained was processed using an appropriate software (NIST, 2005; Mass Spectral Search Program V.2.0d) which provides quality matching using advanced spectral matching algorithms background subtraction. Regarding the chemical families, for the CTL group, all chemical families had a higher impact when compared with the BC group, with exception of alcohols. The highest contribution for the volatile profile was obtained for phenols in the CTL group, whereas in the BC group was alcohols. The main VOMs identified in these families were phenol, 4-methyl-phenol and 2-ethyl-1-hexanol, respectively. These metabolites were already reported in literature in biological matrices such as urine [127,128], cancer cell lines [44,135], saliva [51], and exhaled breath [135,136]. Table 3.1.3 presents the identification of VOMs in urine samples, as well as their minimum and maximum of relative GC peak areas, together with their frequency of occurrence for each group.

Table 3. 1. 3 - Identification of metabolites of urine samples from BC patients and healthy volunteers (CTL) through GC-qMS, minimum (Min) and maximum (Max) relative peak areas, variation of relative peak areas regarding to BC group and frequency of occurrence FO (in %) of VOMs from BC and CTL groups.

Peak ID	IEC[a]	RT (min)[b]	VOM[c]	Relative GC Peak Areas				Variation	FO %
				BC		CTL			
				Min	Max	Min	Max		
A1	47	4.72	methanethiol	2.00	78.56	2.20	97.14	↓	100
A2	68	5.03	furan	1.75	108.12	3.48	121.92	↓	100
A3	43,58	5.24	acetone	1.70	77.42	1.01	107.71	↓	100
A4	82	6.09	2-methyl-furan	1.02	69.71	1.44	354.65	↓	96
A5	43	6.40	ethyl acetate	0.23	1.72	1.59	1.71	↑	18
A6	43, 72	6.69	2-butanone	0.30	8.98	0.55	49.61	↓	100
A7	96	8.07	2,5-dimethyl-furan	0.39	60.24	1.15	178.30	↓	77
A8	43, 86	8.79	2-pentanone	3.54	192.37	1.32	99.72	↑	96
A9	95, 110	10.94	1-(2-furanyl)-ethanone	0.07	103.73	0.38	102.50	↑	99
A10	91	11.22	toluene	1.18	65.82	0.32	122.29	↓	100
A11	41, 55	11.37	3,7-dimethyl-1,6-octadiene	0.12	21.71	0.15	19.71	↑	83
A12	43, 57	11.71	3-hexanone	0.21	17.13	0.27	17.39	↓	99
A13	94	12.70	dimethyl disulfide	16.26	693.52	11.84	726.00	↓	100
A14	44, 56	13.13	hexanal	0.42	20.98	0.18	12.96	↑	99
A15	97, 113	13.48	3-methyl-tiophene	0.19	17.45	0.08	6.50	↑	69
A16	43, 71	15.49	4-heptanone	12.03	380.27	17.71	133.36	↑	99
A17	121, 93	17.81	α-terpinene	0.03	18.19	0.07	128.23	↓	93
A18	111	18.12	1,4-cineole	0.03	4.03	0.15	43.21	↓	92
A19	43, 58	18.57	2-heptanone	0.13	8.44	0.56	46.99	↓	100
A20	68, 93	18.73	D-limonene	0.03	2.49	0.05	4.49	↓	86
A21	91	19.36	propyl-benzene	-	-	0.02	2.28	-	37
A22	43, 72	19.63	3-methyl-2-heptanone	0.12	1.80	0.09	14.56	↓	82
A23	43	19.76	eucalyptol	0.13	2.86	0.11	4.62	↓	59
A24	105	20.10	1,2,4-trimethylbenzene	0.02	6.51	0.06	17.76	↓	93
A25	81	20.62	2-pentylfuran	0.08	31.71	0.19	92.69	↓	94
A26	41, 55	20.84	2-hexenal	0.09	1.83	0.05	5.46	↓	11
A27	93	21.20	γ-terpinolene	0.01	23.50	0.09	120.92	↓	79

Peak ID	IEC[a]	RT (min)[b]	VOM[c]	Relative GC Peak Areas				Variation	FO %
				BC		CTL			
				Min	Max	Min	Max		
A28	79	22.19	3,8-p-menthadiene	0.16	6.96	0.16	8.61	↓	58
A29	119	22.69	p-cymene	0.15	350.42	13.54	483.82	↓	100
A30	93, 121	23.16	α-terpinolene	0.11	2.87	0.05	75.80	↓	65
A31	105	23.31	1-ethyl-2-methylbenzene	0.07	20.50	0.05	9.36	↑	76
A32	117, 132	24.11	2-methyl-1-propenylbenzene	0.05	67.38	0.03	45.29	↑	75
A33	119	24.49	1-methyl-2-isopropylbenzene	0.04	69.76	0.11	2.09	↑	46
A34	43,58	24.71	2-octanone	0.08	3.15	0.12	5.22	↓	85
A35	114	25.84	2-methoxythiophene	0.06	19.56	0.30	11.53	↑	100
A36	105	26.86	1,2,3-trimethylbenzene	0.30	104.18	0.80	65.63	↑	96
A37	159	26.96	α-ionene	0.02	12.64	0.12	52.61	↓	86
A38	83, 55	27.73	3-ethylcyclopentanone	0.17	1.26	0.07	7.68	↓	70
A39	128, 113	28.05	2-methyl-5-(methylthio)-furan	0.25	47.44	0.25	27.96	↑	97
A40	45, 55	28.33	2-heptanol	1.95	1.95	0.38	0.40	↑	4
A41	72	28.72	3,4-dimethyl-2-hexanone	0.07	2.43	0.18	3.34	↓	65
A42	99	29.10	allyl Isothiocyanate	0.04	420.94	0.01	27.32	↑	48
A43	159	29.26	1,2,3,4-tetrahydro-1,6,8-trimethylnaphthalene	0.09	12.66	0.03	57.10	↓	56
A44	121	29.47	2,6-dimethyl-2,4,6-octatriene	0.22	0.22	0.03	0.47	↓	7
A45	126	29.83	trimethyl trisulfide	0.56	227.20	6.79	704.86	↓	99
A46	113, 128	30.46	2-methoxy-5-methyl-tiophene	0.11	116.90	0.47	61.27	↑	99
A47	121	30.86	3,4-dimethyl-2,4,6-octatriene	0.07	1.99	0.07	0.97	↓	18
A48	133	31.07	1,3-diethyl-5-methyl-benzene	0.02	7.46	0.05	10.98	↓	25
A49	58, 43	31.21	2-nonanone	0.04	9.61	0.16	4.44	↑	90
A50	82	32.03	2-cyclohexen-1-one	0.52	16.77	0.22	22.70	↓	94
A51	57	32.39	nonanal	0.07	8.52	0.13	5.53	↑	45
A52	159	32.43	1,2,3,4-tetrahydro-1,5,8-trimethylnaphthalene	0.03	24.07	0.34	170.12	↓	94
A53	132	33.19	1-ethenyl-3,5-dimethyl-benzene	0.12	182.80	0.06	150.56	↑	37
A54	105	33.55	(1-methylbutyl)-benzene	0.09	0.77	0.11	1.47	↓	14
A55	132	33.79	1-methyl-4-(1-methylethenyl)-benzene	0.17	62.11	0.48	367.81	↓	90
A56	159	34.04	1,2,3,4-Tetrahydro-1,4,6-trimethyl naphthalene	0.77	100.00	0.38	264.72	↓	92
A57	73	34.84	3,7-dimethyl-3-octanol	0.19	37.82	0.16	23.66	↑	80
A58	117, 132	35.41	1-ethenyl-4-ethyl-benzene	0.05	249.39	0.20	414.60	↓	58
A59	57	35.98	1-octen-3-ol	0.03	2.38	0.09	1.76	↑	56
A60	96	36.77	furfural	1.19	48.59	0.27	148.62	↓	83

Peak ID	IEC[a]	RT (min)[b]	VOM[c]	Relative GC Peak Areas				Variation	FO %
				BC		CTL			
				Min	Max	Min	Max		
A61	119	36.92	prehnitene	1.59	60.02	0.97	399.65	↓	90
A62	59	37.20	2,6-dimethyl-7-octen-2-ol	0.15	331.07	0.57	184.17	↑	93
A63	43,60	37.33	acetic acid	1.90	93.59	3.45	291.59	↓	97
A64	57	38.49	2-ethyl-1-hexanol	2.97	125.41	1.12	58.98	↑	99
A66	192	39.69	vitispirane I	1.53	317.57	0.95	328.65	↓	99
A67	192	39.85	vitispirane II	0.61	131.23	0.41	119.19	↑	99
A68	105	40.02	benzaldehyde	0.28	7.88	0.21	9.71	↓	82
A69	130	40.10	2-(methylthio)-tiophene	0.30	14.62	0.22	13.52	↑	66
A70	56	42.58	1-octanol	0.09	25.62	0.15	4.66	↑	94
A71	110	43.23	5-methyl-2-furancarboxaldehyde	0.62	30.63	0.27	122.84	↓	99
A72	81	43.54	1-terpinenol	0.19	2.53	0.13	46.41	↓	39
A73	131	43.88	2-methyl-benzofuran	0.03	26.22	0.07	5.24	↑	82
A74	58	44.55	2-undecanone	0.02	3.76	0.17	2.54	↑	58
A75	71	44.62	menthol	0.23	183.84	0.44	123.43	↑	73
A76	84	44.90	4-terpineol	0.03	3.48	0.12	15.35	↓	48
A77	111	46.26	isomaltol	0.11	4.53	0.27	36.11	↓	73
A78	71, 81	47.22	menthomenthol	1.08	550.81	0.41	340.85	↑	76
A79	105, 77	47.54	acetophenone	0.14	8.59	0.27	56.57	↓	83
A80	93	48.16	2,6-dimethyl-5,7-octadien-2-ol	0.06	14.21	0.09	40.14	↓	52
A81	59	48.26	4-methyl-1,4-hexadiene	n.d.	n.d.	0.14	3.29	-	3
A82	56	48.55	1-nonanol	0.40	1.43	0.41	0.65	↑	6
A83	134	48.88	1-ethenyl-4-methoxy-benzene	0.09	1.77	0.14	3.08	↓	14
A84	111	49.95	3-thiophenecarboxaldehyde	0.25	15.77	0.60	16.73	↓	94
A85	145	50.31	2-methyl-3-phenyl-2-propenal	0.14	27.65	0.49	123.30	↓	99
A86	59	50.48	α-terpinol	0.70	2.06	0.10	5.54	↓	20
A87	109	51.12	phellandranal	0.03	4.73	0.09	100.55	↓	66
A88	82	52.06	D-carvone	0.55	372.76	0.28	142.95	↑	85
A89	157	52.57	1,2-dihydro-1,1,6-trimethylnaphthalene	2.04	500.14	6.89	629.74	↓	97
A90	120	54.31	methyl salicylate	0.25	25.81	0.55	99.17	↓	82
A91	133	54.63	p-isopropylbenzaldehyde	0.06	8.30	0.27	87.43	↓	89
A92	132	55.30	1-Isopropenyl-4-methylbenzene	0.09	2.48	0.33	356.82	↓	48
A93	121	55.69	4-hydroxy-acetophenone	0.02	3.78	0.13	81.92	↓	68
A94	69	57.00	β-damascenone	0.25	63.23	2.06	51.48	↑	99

Peak ID	IEC[a]	RT (min)[b]	VOM[c]	Relative GC Peak Areas				Variation	FO %
				BC		CTL			
				Min	Max	Min	Max		
A95	135	58.81	p-cymen-8-ol	0.09	1.60	0.30	7.71	↓	21
A96	60, 73	59.09	hexanoic acid	0.18	3.49	0.28	7.29	↓	97
A97	109, 124	59.38	2-methoxy-phenol	0.60	233.16	1.89	216.04	↑	97
A98	123, 138	59.90	4-methoxy-3-methyl-phenol	0.01	6.78	0.19	1.80	↑	18
A99	123, 138	65.73	2-methoxy-5-methylphenol	0.05	22.64	0.14	83.85	↓	68
A100	73, 88	66.07	2-ethyl-hexanoic acid	0.14	0.17	0.44	5.96	↓	10
A101	60, 73	66.27	heptanoic acid	0.11	25.48	0.23	26.03	↓	96
A102	175	68.94	1-ethyl-3,5-diisopropyl-benzene	0.14	299.80	0.24	91.62	↑	99
A103	94	69.82	phenol	16.94	637.24	8.40	427.47	↑	99
A104	60,73	72.54	octanoic acid	0.29	35.90	0.44	95.05	↓	99
A105	119,169	73.01	1-methyl-7-isopropylnaphthalene	2.07	2.07	0.60	13.57	↓	13
A106	107	73.18	4-methyl-phenol	3.32	254.95	9.60	281.82	↓	100
A107	60,73	75.84	nonanoic acid	0.90	43.87	0.42	61.49	↓	96
A108	107	75.99	4-ethyl-phenol	0.14	20.30	0.11	99.24	↓	28
A109	135,150	76.95	thymol	0.10	177.78	0.41	81.58	↑	96
A110	132	77.26	4-(2,3,6-Trimethylphenyl)-2-butanone	0.24	6.95	0.12	9.10	↓	65
A111	60,73	78.28	decanoic acid	0.27	15.20	0.20	16.64	↓	97
A112	135	78.62	p-tert-butyl-phenol	9.49	177.87	6.11	160.77	↑	97
A113	173	78.74	3,3,5,6-Tetramethyl-1-indanone	0.10	121.92	0.69	120.24	↑	86
A114	105	81.90	benzenecarboxylic acid	0.04	168.85	0.34	87.80	↑	92
A115	117	82.31	indole	0.02	24.02	0.06	7.43	↑	85
A116	105	83.26	benzophenone	0.08	4.55	0.15	3.17	↑	76

[a]IEC - ion extraction chromatogram; [b]Retention time (min); [c] VOM- volatile organic metabolite.

It can be observed that most of metabolites were identified in all samples with a FO higher than 90 %, where the maximum relative area was obtained for dimethyl disulfide both in CTL and BC groups, followed by phenol in the CTL group and 1,2-dihydro-1,1,6-trimethylnaphthalene in the BC group. Regarding the relative area values obtained for the BC group, most VOMs were down-regulated with exception of 2-pentanone, 1-(2-furanyl)-ethanone, hexanal, 4-heptanone, 2-methoxythiophene, 1,2,3-trimethylbenzene, 2-methyl-5-(methylthio)-furan, among others, as shown in Table 4.1.3. Some of these have been already reported in literature, namely ethylbenzene and acetic acid were found as discriminant when associated with CTL in a study conducted by Ahmed et *al.* [137] where the authors investigated the possibility of faecal VOMs as potential diagnostic biomarkers for inflammatory bowel disease. Silva et *al.* [127,128] also reported phenols as the major chemical family identified in urine from the oncologic group. Furthermore, Raman et *al.* [138] studied faecal VOMs in obese humans and identified also acetic acid and phenol as major metabolites. On the other hand, Priscilla et *al.* [59] also identified *p*-xylene, *o*-xylene, acetic acid, phenol and *p*-tert-butyl-phenol in urine from cancer patients, some of them having higher values in cancer patients.

Multivariate statistical analysis of urine metabolomic profile

The statistical analysis was performed using the Metaboanalyst 4.0 [134] web server as described in the experimental section. Only the VOMs with FO higher than 90% were considered for the statistical analysis, in a total of 53 VOMs as presented in Table 3.1.3. Initially, data was transformed by log transformation and mean centering approaches, before being subjected to multivariate statistical analysis. Partial least square-discriminant analysis (PLS-DA) was used as a supervised clustering method to verify the existence of an altered metabolite pattern. Additionally, this type of statistical analysis takes into account the variance/covariance between samples of groups where the samples are classified into different groups. Regarding the results obtained, a good discrimination was achieved (54.3%, total variance) between BC and CTL urine samples suggesting the occurrence of metabolic alterations in the groups under study (Figure 3.1.5 A). Then, the top ten metabolites (*p*-cymene, 1-methyl-4-(1-methylethenyl)-benzene, 4-heptanone, trimethyl trisulfide, acetic acid, dimethyl disulfide, 2-pentylfuran, 1,2-dihydro-1,1,6-trimethylnaphthalene, α-terpinene and 2-methyl-3-phenyl-2-propenal) with the highest contribution to group discrimination were selected with variable importance in projection (VIP >1) (Figure 3.1.5 B).

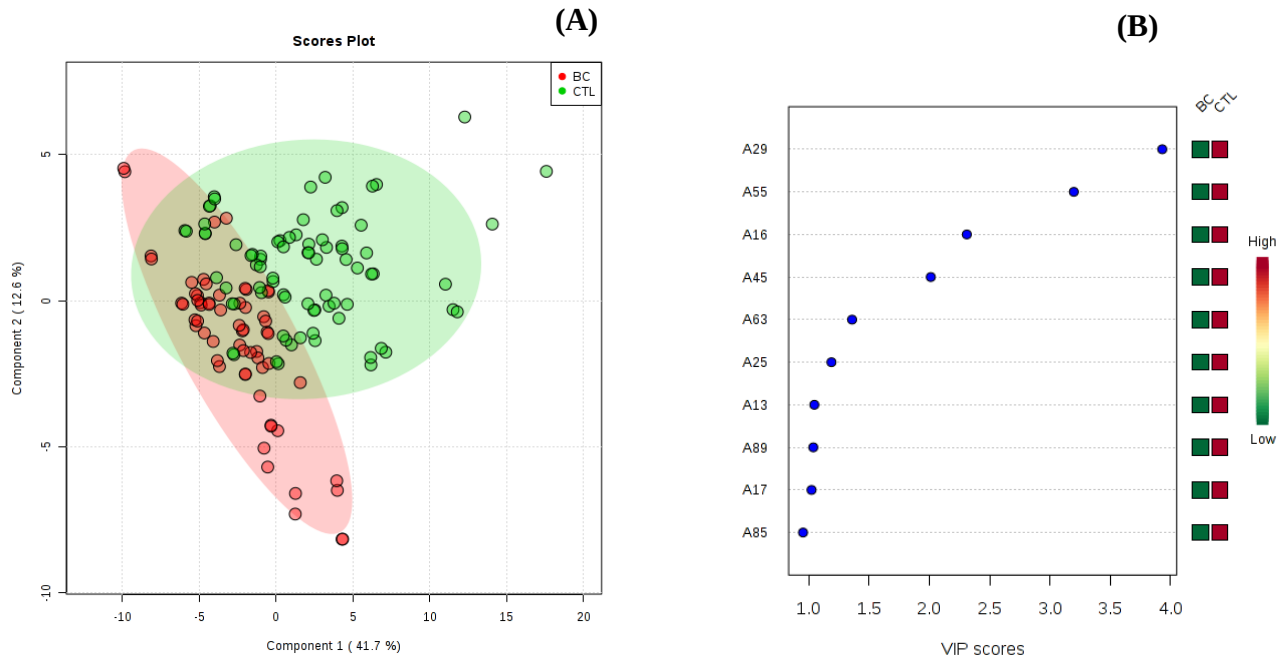

Figure 3. 1. 5 - (A) Score plots of partial least square discriminant analysis (PLS-DA) and (B) VIP scores selected by the PLS-DA analysis. For identification please see Table 3.1.3.

Furthermore, the orthogonal partial least squares-discriminant (OPLS-DA) analysis was applied on urine metabolomic profile dataset in order to maximize the separation of BC and CTL groups. Significant group separation was observed in OPLS-DA score plot between BC and CTL groups indicating intrinsic metabolic alterations in each group (Figure 3.1.6 A). To attest the robustness of the model, a random permutation test with 1000 permutations was performed with OPLS-DA (Figure 3.1.6 B). The permutation test yielded an R^2 (represents goodness of fit) as 0.833 and a Q^2 (represents predictive ability) of 0.742 indicating that the model is not over fitted and have a relative good predictive ability to distinguish between study groups. Additionally, the OPLS-DA uses class information allowing to show which variables are responsible for class discrimination using the predictive information of the first component. The main advantage of OPLS-DA when compared to PLS-DA is that a single component is used as a predictor for the class, while the other components describe the variation orthogonal to the first predictive component [139]. To further evaluate the VOMs predictive value to discriminate between BC patients and CTL, a ROC curve analysis was generated using the top ten metabolites identified by VIP values (Fig. 3.1.6 C and D). This type of analysis is used for the classification of true positives and false positives and the predictive ability is measured using the area under the curve (AUC) [140,141]. Xia *et* al. [108] reported that a value of AUC between 0.9 - 1.0 is excellent, between 0.8 – 0.9 is good, and comparing the results obtained, these were very good (AUC = 0.842). The AUC can be interpreted as the probability that a randomly select diseased subject is ranked as more probable to be diseased than a randomly select healthy

subject [108]. The greater AUC value gives us indication of the effectiveness to separate the control (CTL) group from the BC group.

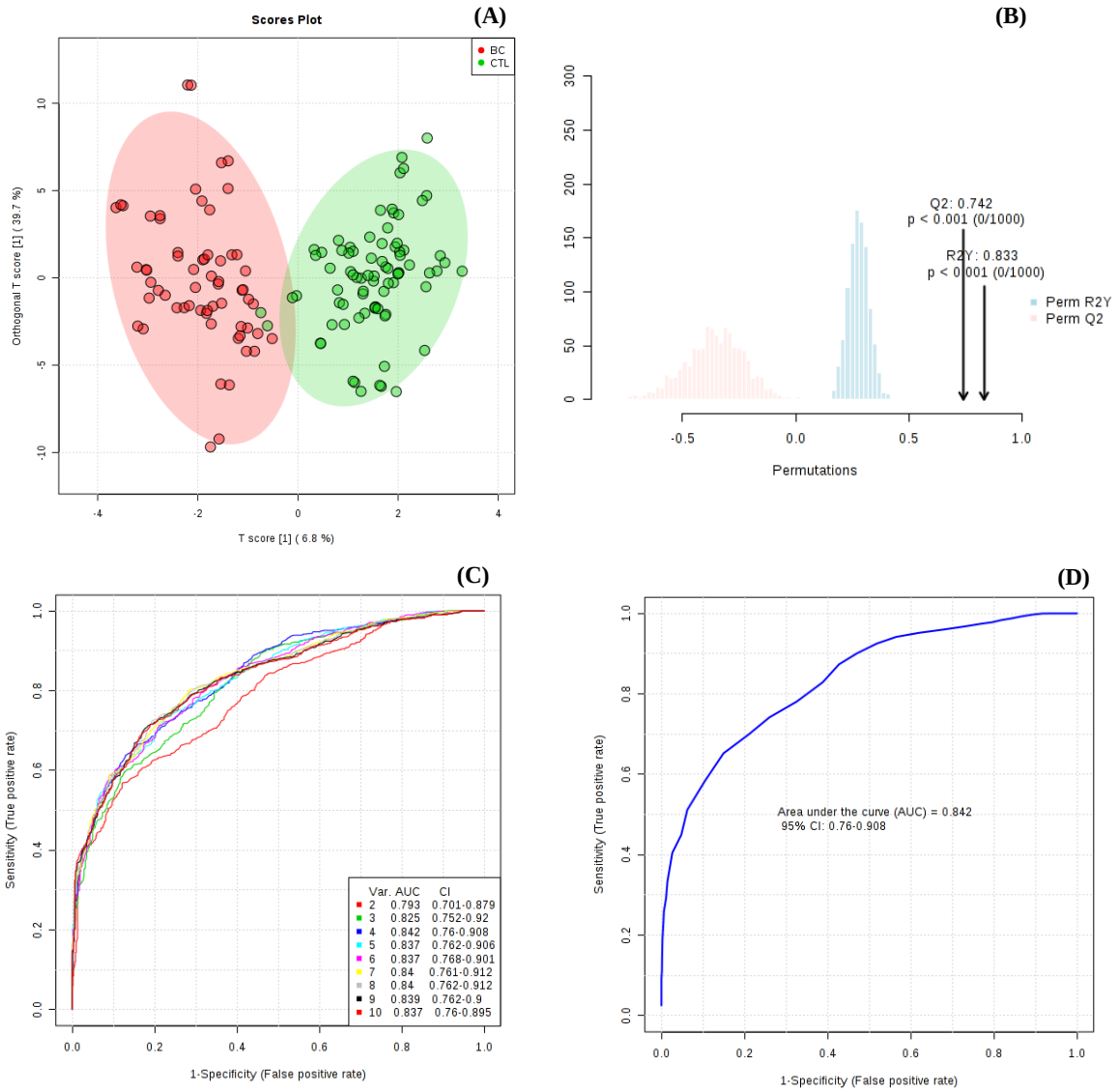

Figure 3. 1. 6 - (A) Loading score plots of orthogonal projection to OPLS-DA analysis, (B) model validation by permutation test based on 1000 permutations of VOMs obtained by GC-qMS of urine samples from the 2 groups under study, (C) ROC curves for the predictive model with a combination of metabolites calculated from the logistic regression analysis using the ten metabolites selected by the VIP (> 1.0) values, and (D) ROC curve for the top 4 metabolites with the highest ability to discriminate BC patients against CTLs.

Furthermore, random forest (RF) was carried out to determine the ability of metabolites to accurately classify the study subjects into their corresponding groups.

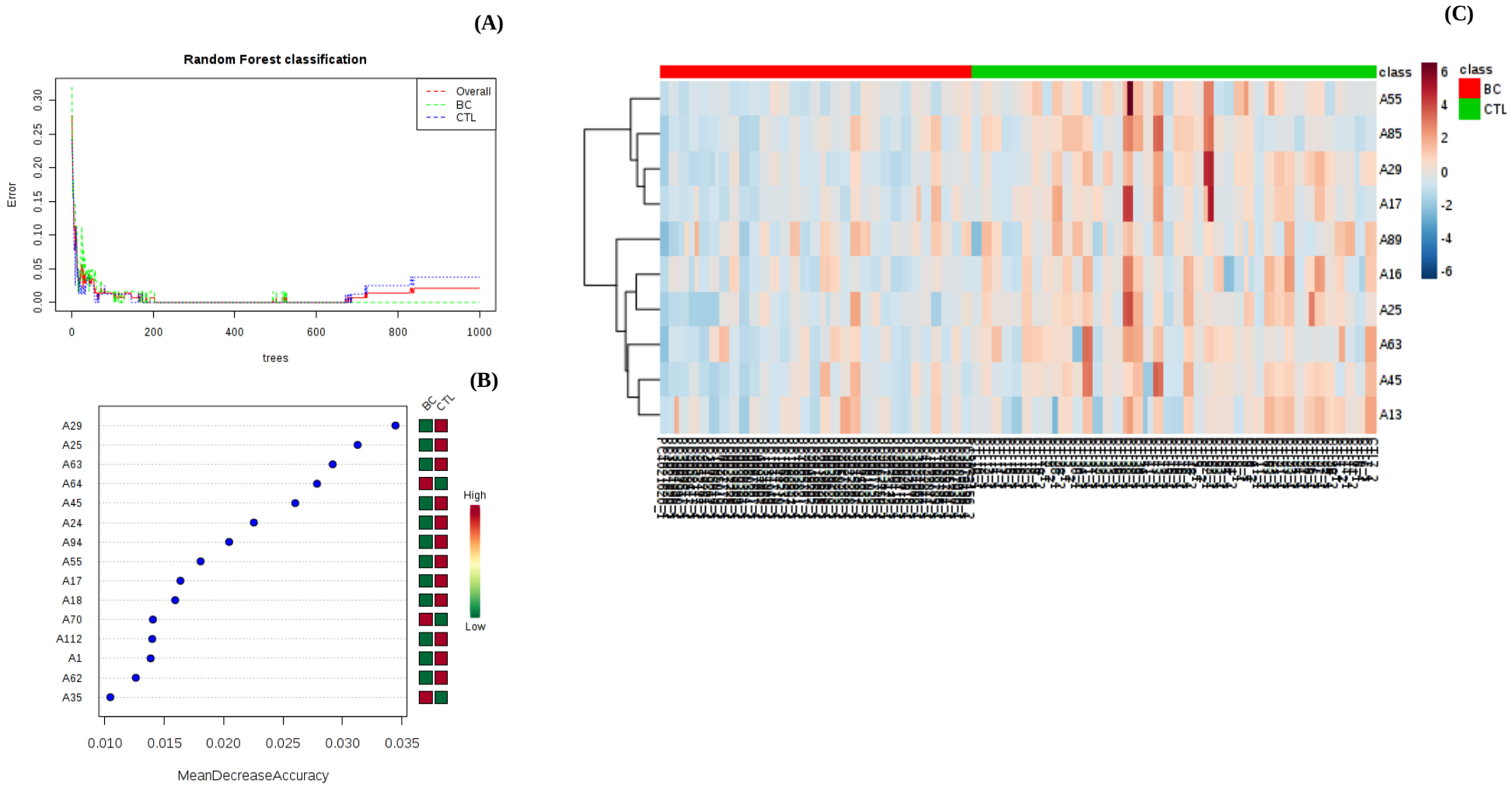

Figure 3. 1. 7 - (A) RF classification of urine metabolites from BC and CTL groups which indicates and overall error of 0.0 for BC and 0.03 for CTL, (B) Features ranked by their contributions to classification accuracy (Mean Decrease Accuracy) and (C) Heat map visualization and hierarchical clustering analysis using the 10 metabolites with significance ($p<0.05$) by Pearson´s distance analysis.

It uses an ensemble of classification trees (in this case 1000), each of which is grown by random feature selection from a bootstrap sample at each branch. RF classified BC and CTL groups with an accuracy of 100 and 90 % respectively, with an out of bag (OOB) classification error rate of 0.0211 (Figure 3.1.7 A-C) where the major contribution of VOMs for classification are shown in Figure 3.1.7 B. The OOB error rate is a measure of the prediction error of machine learning models such as RF. OOB is calculated by bootstrapping data, training data on one part and calculating error on the remaining. Since its calculation is for 1000s of trees, variances of error estimate are reduced. Tests have shown that OOB error is a reliable as a cross validation method [142].

Additionally, the heat map was constructed with the ten selected VOMs by VIP > 1, using Pearson´s correlation, providing intuitive visualization of the data set and was used to identify samples or features that are unusually high or low (Figure 3.1.7 C). Here, it could be observed that, for the majority of VOMs from the CTL group, the levels were higher than in the BC group.

Finally, the metabolic pathway analysis was performed using the metabolites with VIP values higher than 1, in order to explore which pathways were altered in the groups under study. In tis type of analysis, the matched nodes show varied heat map colors and is based on *p* value while the node radius is determined based on the pathway impact values. Then, the pathway name is shown as mouse-over tooltip. If we click the corresponding node it will launch the corresponding pathway view, showing which metabolites allowed the matching. Figure 3.1.8 A and B show the pathway impact for BC and CTL groups, respectively. It can be observed that the pathways with the highest impact were the pyruvate and sulfur metabolism due to acetic acid that was found lower in BC group when compared with CTL group.

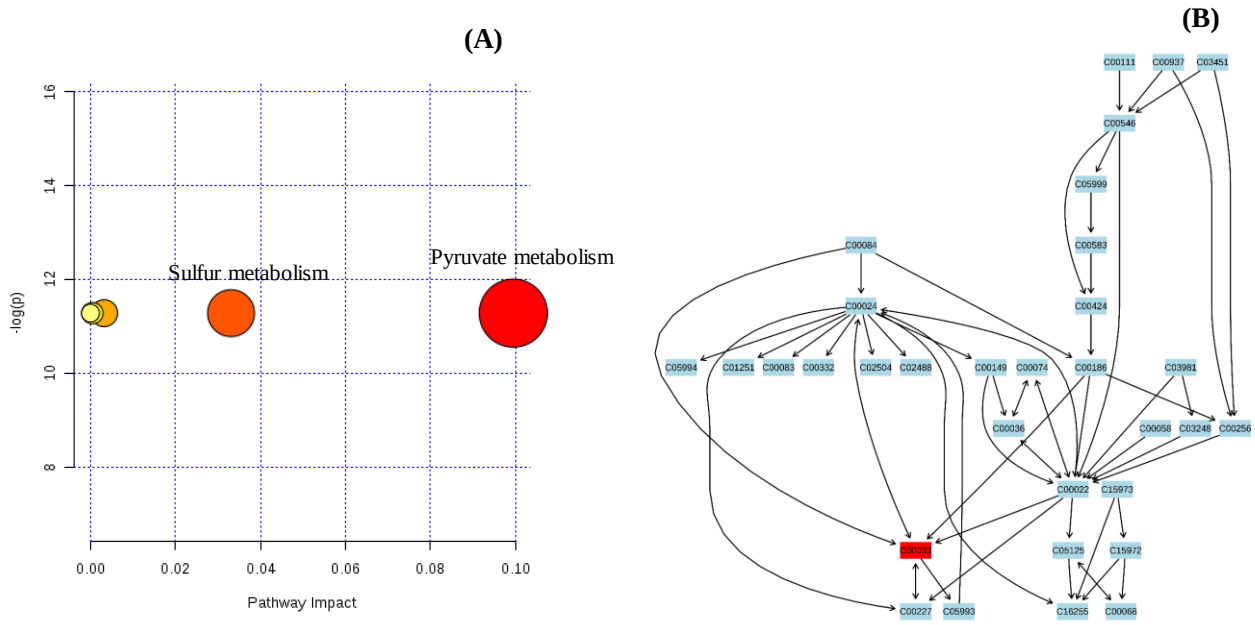

Figure 3. 1. 8 - (A) The metabolome view map of significant altered metabolic pathways observed in urine samples from BC and CTL groups and (B) the pyruvate metabolism. The map was generated using reference map by KEGG. C00033 represents acetic acid.

This metabolite according to the human metabolome database, is normally found in most tissues (liver, kidney, among others) and in several biofluids, namely saliva [51] and urine [127]. It results from several reactions namely from pyruvate metabolism, ethanol degradation and aspartate metabolism.

Based on the results obtained, a successful discrimination of tissue samples was achieved, according to the group showing that the volatomic tissue profile can be a useful approach to identify potential BC biomarkers. Also, these results suggested the possibility to identify endogenous metabolites as a platform to discover- potential BC biomarkers and paves a way to investigate the related metabolomic pathways to improve the BC diagnostic tools.

Conclusions

This work described the development and optimization of and HS-SPME with GC-qMS methodology tandem with chemometric tools to analyze the VOMs of urine samples from BC and CTL groups. The CCD model was used in order to optimize the HS-SPME conditions and the results showed that the extraction temperature and extraction time were the factors that affected the analytical response. The best extraction conditions were reached by adding 15% (w/v) of NaCl to urine samples, and then extracting the VOMs using a CAR/PDMS fiber coating for 75 min at 50 °C. The developed HS-SPME method was applied to urine samples to discriminate between BC and CTL groups,

showing promising results that are essential to be explored. From the GC-qMS analysis, 116 VOMs were identified and multivariate statistical analysis revealed some metabolites significantly altered in BC patients. Of the metabolites identified, 3-methyl-thiophene, 4-heptanone, α-terpinene, 2-pentylfuran, *p*-cymene, trimethyl trisulfide, 1-methyl-4-(1-methylethenyl)-benzene, acetic acid, 2-methyl-3-phenyl-2-propenal and 1,2-dihydro-1,1,6-trimethylnaphthalene showed the utmost sensitivity and specificity to discriminate BC patients from healthy controls. The analysis of the plots leads to a metabolomic pattern comprising an array of several biochemical pathways altered in BC patients. The metabolic pathway analysis indicated that the discriminatory metabolites could be originated from several dysregulated pathways in BC such as pyruvate and sulfur metabolism. These results suggested the possibility to identify endogenous metabolites as a platform to discovery potential BC biomarkers and paves a way to investigate the related metabolomic pathways to improve the diagnostic tools of BC.

3.2| ^{1}H NMR metabolomic urinary profile in breast cancer: a feasibility study

(Submitted to Metabolites, 453682)

Abstract

BC remains the second leading cause of death among women worldwide. An emerging approach based on the identification of endogenous metabolites (EMs) and the establishment of the metabolomic fingerprint of biological fluids constitutes a new frontier in medical diagnostics and a promising strategy to differentiate cancer patients from healthy individuals.

In this work we aimed to establish the urinary metabolomic patterns from 40 BC patients and 38 CTLs using ^{1}H-NMR, as a powerful approach to identify a set of BC-specific metabolites which might be employed in the diagnostic of BC. OPLS-DA analysis was applied to ^{1}H-NMR processed data matrix. Metabolomic patterns distinguished BC from CTL urine samples suggesting unique metabolite profiles for each investigated group. A total of 10 metabolites exhibited the highest contribution towards discriminating BC patients from CTLs (VIP >1, $p < 0.05$). The discrimination efficiency and accuracy of the urinary EMs were ascertained by ROC curve analysis that allowed the identification of some metabolites with the highest sensitivities and specificities to discriminate BC patients from healthy controls. The metabolomic pathway analysis indicated several metabolism pathways disruptions including amino acids and carbohydrates metabolism, in BC patients. The obtained results support the high throughput potential of NMR-based urinary metabolomics patterns in discrimination BC patients from CTL. Further investigations could unravel novel mechanistic insights into disease pathophysiology, monitor disease recurrence and predict patient response towards therapy.

Keywords: Breast cancer; ^{1}H NMR; Urine; Metabolomics.

Introduction

The global cancer burden is estimated to have risen to 18.1 million new cases and 9.6 million deaths in 2018 [143], being the second leading cause of death after cardiovascular diseases. This increasing trend is due to several factors, including ageing, population growth as well as the changing prevalence of certain causes of cancer linked to social and economic development [143]. Cancers of the lung and female breast are the leading types worldwide in terms of the number of new cases; for each of these types, approximately 2.1 million diagnoses were diagnosed in 2018, contributing about 11.6 % of the total cancer incidence [143]. BC is also the leading cause of cancer death in women (15.0 %), followed by lung cancer (13.8 %) and colorectal cancer (9.5 %), which are also the second and third most common types of cancer, respectively [143]. Early stage BC patients have higher 5-year survival rates than those diagnosed at later stages. Improved early BC detection methods could reduce patient mortality and improve therapeutic responses and prognosis [123]. Common methods of routine surveillance for BC include periodic mammography, self- or physician performed examination, and blood tests of tumor markers, including cancer antigens (CA 15.3, CA 27.29), carcinoembryonic antigen (CEA), tissue polypeptide specific-antigen, and human epidermal growth factor receptor 2 (the shed form). CA 15.3 and CEA represent the most widely used tumor markers [122]. An additional factor that contributes to the poor prognosis of patients diagnosed with BC is the fact that the diagnosis is often delayed due to limitations in screening tests [144]. A recent approach is metabolomics that studies a subset of small molecules derived from the global or targeted analysis of metabolic profiles from biological samples, such as blood, urine [145–148], cells, or tissue [135,149–152] representing a valuable tool in the detection of several diseases, including cancer. Several reports have already demonstrated the importance of studying the metabolome as a mean to discover a set of metabolites to be used as cancer biomarkers [127,128,153]. The study of specific metabolites to identify cancer fingerprints or signatures can aid in cancer detection and prognosis as well as the assessment of pharmacodynamic effects of therapy [149]. The most common approaches in metabolomics involve GC-MS [146], LC-MS [154] or NMR [67,123,127,128].

MS includes a separation stage using LC or GC and can discriminate between compounds based on mass-to-charge (*m*/*z*) ratio in charged particles. Although compared with NMR, MS exhibits a greater sensitivity, sample preparation is laborious and dependent on metabolite chemical properties [155]. In addition, MS lacks accuracy and precision producing an enhanced resolution profile with numerous peaks. However, only approximately 5% of these peaks are associated with known components [156]. Recently, non-invasive sampling strategies, including exhaled breath, urine and saliva, have emerged as attractive and useful approaches when coupled to high-throughput techniques such as NMR (^{1}H NMR) [150,157,158]. NMR spectroscopy is a particularly appealing platform based

on the energy absorption and re-emission of the nuclei due to the magnetic field, where the selected isotopes having a momentum in an external magnetic field will give a signal. The most targeted isotope in samples with a biological origin is hydrogen (via ^{1}H NMR) given its abundance in biological samples. Furthermore, other isotopes that are less abundant, such as carbon (via ^{13}C NMR) or phosphorous (via ^{31}P NMR) can also be used to provide additional information regarding specific metabolites[27]. The result is an NMR spectrum, and each separate resonance presented in the spectrum corresponds to a part of the spectral pattern of a unique metabolite and provides structural information that can simplify the identification of an unknown metabolite. The spectral peak areas generated by each molecule are used as an indirect measure of the amount of the metabolite present in the sample, not only allowing the quantification of the concentration of a metabolite but also providing information regarding its chemical structure. This technique presents several advantages, such as being a rapid, robust, cost-effective, highly reproducible, non-destructive, and fully quantitative. In addition, it requires no prior compound separation or derivatization generating spectra from a biological sample within few minutes being useful for conducting metabolomic studies on biofluids [27,159,160]. Zhang et al. [120] reviewed the most common methods for NMR spectroscopy based metabolite profiling, data processing and analysis. Bingol et al. [161] recently overview some advances in metabolomics field and contribution to targeted and untargeted approaches. A variety of studies have been conducted by NMR in disease research using urine [123,150], saliva [158,162], serum [163], plasma [164,165], tissue [166–169], and in cell culture [170–172].

The aim of this study was to evaluate the ability of ^{1}H NMR spectroscopy combined with multivariate statistical tools to differentiate and discriminate the urinary metabolomic patterns from BC patients and CTLs as a powerful approach to identify a set of BC-specific metabolites which might be employed in the diagnostic of BC.

Materials and Methods

Reagents

3-(Trimethylsilyl)propionic-2,2,3,3-*d4* acid sodium salt (TSP) and deuterium oxide (D_2O) were supplied by Acros Organics (Geel, Belgium).Potassium dihydrogen phosphate (KH_2PO_4), sodium azide (NaN_3) and potassium deuteroxide solution (KOD) were purchased from Panreac (Barcelona, Spain) and Sigma Aldrich (St. Louis, MO, USA).

Urine collection

To investigate the urinary NMR profile, 40 urine samples from female patients with BC (*n*=40, age range 40-74 years, average 59 ± 10 years) and 38 urine samples from healthy female volunteers without any known pathology (*n*=38, age range 40-72 years, average 53 ± 8 years; control group - CTL) were obtained from the Haemato-Oncology Unit of Hospital Dr. Nélio Mendonça (Funchal, Portugal) and Blood Donors Bank of the same Hospital, respectively (Table 3.2.1), were randomly selected among the volunteers. Patients did not receive any neoadjuvant chemotherapy or radiation therapy prior to sample collection. Healthy controls were age- and gender-matched patients, and had no declared history of cancer or gastrointestinal symptoms. Exclusion criteria included pregnancy, inflammatory conditions, mental disorders, gastrointestinal tract disorders, hypertension, uncontrolled bacterial, viral, or fungal infection and diabetes mellitus.

Table 3. 2. 1 - List of collected urine samples from BC patients and CTLs. SD: standard deviation.

Sample group	**N. subjects**	**Age range/years**	**Mean Age ± SD**
Breast Cancer(BC)	*n* = 40	40-74	59 ± 10
Control (CTL)	*n* = 38	40-72	53 ± 8

The study was approved by the Ethic Committee of Hospital Dr. Nélio Mendonça (Approval no. S.1708625/2017). Written informed consent for the study was obtained from all participants. Each individual (either patient or healthy volunteer) provided a sample of morning urine (after overnight fasting) in a 20 mL sterile container. The samples were aliquoted into 4 mL glass vials and frozen at -80ºC until use in the experiments.

Urine samples were thawed and centrifuged (8000 rpm for 5 min) to remove any suspended cells and other precipitated material. Then, 540 µL of urine was mixed with 60 µL of a buffer solution (KH_2PO_4, 1.5 M in D_2O) containing 0.1% of TSP-*d4* (used as chemical shift reference) and sodium azide (NaN_3, 2 mM). The pH was adjusted to 7.00 ± 0.02 by adding small amounts of KOD.

NMR measurements

NMR spectral acquisition was performed using a Bruker Advance II Plus NMR spectrometer equipped with a 400 MHz magnet UltraShield™ 400 Plus at 300K. All NMR spectra acquisition and pre-processing were performed under the control of a workstation with TopSpin 3.5pl7 (Bruker BioSpin). For each sample, a standard 1D nuclear overhauser enhancement spectroscopy (NOESY) pulse sequence (noesypr1d) was used and solvent signal suppression was achieved by presaturation during relaxation and mixing time (SW 4807.692 Hz, TD 64 K data points, relaxation delay 5 s, 128

scans). The shimming was calibrated automatically. Also, all spectra were processed using a line broadening (1.0 Hz) and baseline automatically corrected. The NMR spectrum of each sample was aligned with reference to the TSP signal at δ 0.00 ppm. Spectral regions within the range of 0.94 to 10 ppm were analyzed after excluding the sub-region δ 4.55-6.05 to remove variability arising from water suppression and possible cross-relaxation effect on the urea signal via solvent exchanging protons. As already known, the TSP signal may be affected by proteins or other macromolecules present in samples [173] and for that reason in the preparation of urine samples before NMR analysis, the step of centrifugation was taken into account and the rotations per minute (rpm) was used in order to remove any proteins present in samples.

The analysis of NMR spectral data was performed using the Chenomx NMR Suite 8.2 (Chenomx Inc., Alberta, Canada) and relative concentrations (in mM) of metabolites were determined using the 400MHz library from Chenomx NMR Suite 8.2, which compares the integral of a known reference signal (TSP) with signals derived from a library of compounds containing chemical shifts and peak multiplicities.

In addition, the identication of selected metabolites was also cross checked from the Human Metabolome Database (HMDB) [174] and literature [175]. Regarding the metabolites that were not available in the library, identification was accomplished by running a standard solution and the relative concentration was calculated manually. This software not only allows the identification of compounds but also access their quantification based on advanced algorithms turning into a very straightforward tool to analyze NMR spectra.

Statistical analysis

Statistical analyses were performed using the web server Metaboanalyst 3.0 [134] where sample specific normalization allowed the manual adjustment of relative concentrations based on biological inputs (i.e., volume, mass), and row-wise normalization allowed the general-purpose adjustment for differences among samples. Regarding data transformation and scaling were accomplished using two different approaches to make features more comparable, raw data were scaled using mean-centring and cubic root transformation, with sample normalization by the sum.

Then, multivariate statistical analyzes namely, PCA, PLS-DA and OPLS-DA were applied to the urinary metabolomic profile dataset to provide insights into the separations between the groups. Furthermore, hierarchical cluster analysis by K-means of the 2 groups in this study was performed and Pearson´s correlation was used to generate the heat map using the 32 metabolites to identify clustering patterns. RF classification was performed to determine the ability of the metabolites to accurately classify the study subjects into their corresponding groups. Moreover, the ROCs were

attained to verify which metabolites had the highest sensitivity/specificity for a BC diagnosis. Finally, the metabolites were used for the metabolic pathway analysis to identify the most relevant metabolic pathways involved in the BC and CTL groups.

Results and Discussion

Urinary metabolomic pattern based on ^{1}H NMR

^{1}H NMR analysis was performed according to the procedure described in the Methods section. A representative first dimension urine ^{1}H NMR spectrum, referenced to TSP (0. 0 ppm), from a BC patient is shown in Figure 3.2.1, and metabolites are indicated based on their chemical shifts. Table 2 represents the identification of metabolites as well as their minimum and maximum relative concentrations (mM) for each group and the respective percentage of occurrence (FO). Each sample analysis was performed in triplicate and the relative standard deviation (RSD) was lower than 2 %.

For most metabolites, the FO was greater than 90 % with the following exceptions: valine, glutamine, carnitine, trigonelline, 4-cresol sulphate and hypoxanthine for the BC group; α-hydroxyisobutyrate, trimethylamine N-oxide, hypoxanthine and glycine for the CTL group. Regarding relative concentrations (mM), the highest level was obtained for creatinine followed by hippurate in the BC group and citrate in the CTL group. In addition, taurine and mannitol presented higher levels in BC group, respectively. It can also be highlighted that the majority of metabolites were down-regulated with regard to the BC group, except α-hydroxybutyrate, cis-aconitate, mannitol, hippurate and 3-methylhistidine, that were up-regulated, also identified by Carrola et al. [150]. As shown in Figure 3.2.1, 33 metabolites were identified and quantified. The main metabolites identified in urine mainly resulted from tricarboxylate (e.g., citrate, *cis*-aconitate), methane (e.g., dimethylamine, trimethylamine *N*-oxide) and amino acid metabolism (e.g., hippurate, glycine). The most intense signals in urine were obtained from creatinine, creatine, hippurate, citrate and trimethylamine *N*-oxide.

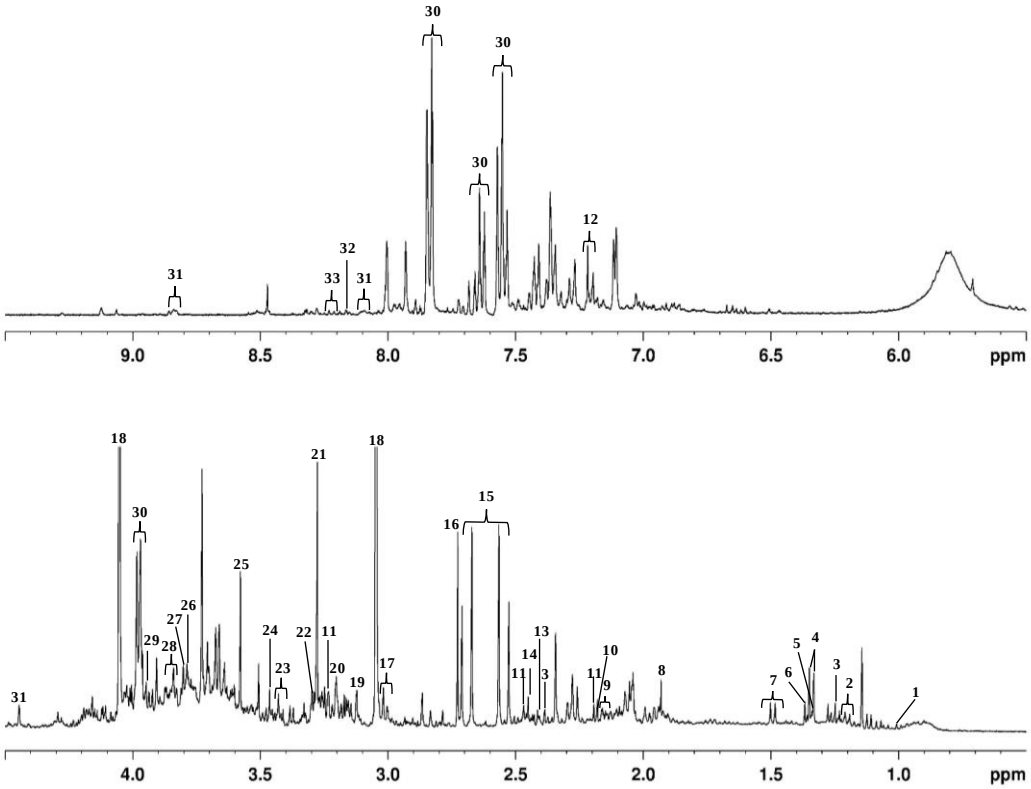

Figure 3. 2. 1 - Typical 400 MHz representative urine ^{1}H NMR spectrum from a BC patient, referenced to TSP (δ 0. 0 ppm). For peak identification see Table 3.2.2.

These metabolites were already identified in several studies that use urine cancer sample [150,169,176]. Trimethylamine *N*-oxide is produced in the liver by intestinal bacteria from dietary quaternary amines, such as choline and carnitine trough trimethylamine (TMA) via flavin-containing monooxygenase (FMO3), and the levels in urine or plasma are used to determine FMO3 deficiency [177–181]. Creatinine is subsequently produced via a biological system involving creatine, phosphocreatine, and adenosine triphosphate (ATP), whereas hippurate and citrate are derived from phenylalanine metabolism and the citrate cycle [182]. The concentration of creatinine is age and sex dependent, decreasing with age and varying throughout the day. Normally, the concentration of creatinine is increased in males when compared with females given the increased body mass index [183]. Creatinine production from the muscles is proportional to the total muscle mass and muscle catabolism. In individuals with a relatively low muscle mass, including children, women, and cancer patients, serum creatinine levels are reduced for a given glomerular filtration rate (GFR), which is the flow rate of filtered fluid through the kidney, thus providing information on kidney function [184].

Table 3. 2. 2 - Metabolites identified in urine samples from BC patients and CTL through ^{1}H NMR spectroscopy, relative concentrations (mM), variation of relative concentration regarding to BC group, K-S values for BC and CTL groups and frequency of occurrence (FO in %) of metabolites for each group.

Peak n.	Metabolite	δ (ppm)	Relative concentrations (mM)						K-S values		*p*-value (< 0.05)	Variation	FO (%)	
			BC			CTL								
		Multiplicity	Min	Max	Average	Min	Max	Average	BC	CTL			BC	CTL
1	valine	1.030 (d)	0.10	1.24	0.67	0.26	1.63	0.95	0.166	0.180	4.127 e-04	↓	68	95
2	α-hydroxybutyrate	1.188 (d)	0.35	4.89	2.62	0.47	2.06	1.26	0.186	0.175	9.201 e-06	↑	95	100
3	β-hydroxyisovalerate	1.256 (s), 2.352 (s)	0.08	1.91	0.99	0.57	2.89	1.73	0.153	0.984	5.048 e-05	↓	100	98
4	lactate	1.319 (d)	0.54	5.39	2.97	1.57	10.31	5.94	0.107	0.069	1.780 e-06	↓	100	98
5	threonine	1.317 (d)	0.61	5.53	3.07	1.60	8.81	5.21	0.098	0.076	1.599 e-05	↓	93	98
6	α-hydroxyisobutyrate	1.346 (s)	0.07	1.77	0.92	1.13	6.64	3.89	0.149	0.092	2.221 e-03	↓	100	78
7	alanine	1.470 (d)	0.24	5.58	2.91	0.62	8.51	4.57	0.098	0.097	3.421 e-04	↓	100	100
8	acetate	1.901 (s)	0.28	4.31	2.3	1.10	6.23	3.66	0.178	0.141	8.737 e-03	↓	96	95
9	glutamine	2.105 (t)	0.34	11.00	5.67	1.47	20.67	11.07	0.193	0.098	5.407 e-03	↓	67	93
10	acetone	2.221 (s)	0.07	4.10	2.08	0.49	4.84	2.66	0.164	0.084	3.367 e-02	↓	92	100
11	carnitine	2.407 (s), 2.447 (s), 3.213 (s)	0.09	1.94	1.01	0.38	6.25	3.32	0.164	0.104	9.002 e-05	↓	80	100
12	4-cresol sulphate	2.350 (s), 7.200 (d)	0.49	7.09	3.79	1.22	11.91	6.57	0.156	0.097	4.315 e-07	↓	84	100
13	pyruvate	2.362 (s)	0.14	4.75	2.44	1.90	7.36	4.63	0.173	0.174	8.880 e-06	↓	100	100
14	succinate	2.390 (s)	0.07	2.58	1.33	0.55	4.85	2.70	0.122	0.188	1.020 e-02	↓	96	100
15	citrate	2.526 (d), 2.688 (d)	3.17	80.34	41.76	9.14	113.56	61.35	0.157	0.146	6.825 e-07	↓	100	100
16	dimethylamine	2.715 (s)	0.19	11.97	6.08	4.49	15.73	10.11	0.199	0.132	2.488 e-04	↓	100	95
17	α-oxoglutarate	2.996 (t)	0.52	12.35	6.43	1.79	15.55	8.67	0.107	0.144	1.172 e-08	↓	92	95
18	creatinine	3.033 (s), 4.047 (s)	3.29	216.36	109.82	104.24	381.60	242.92	0.129	0.175	1.705 e-04	↓	100	100
19	cis-aconitate	3.114 (d)	0.15	22.15	11.15	1.91	16.44	9.18	0.148	0.149	4.671 e-07	↑	100	95
20	choline	3.189 (s)	0.11	2.93	1.52	0.77	3.85	2.31	0.167	0.164	8.094 e-03	↓	96	100
21	betaine	3.252 (s)	0.08	4.13	2.11	0.46	4.61	2.53	0.152	0.154	8.105 e-03	↓	100	93
22	trimethylamine N-oxide	3.254 (s)	0.52	18.04	9.28	1.70	33.81	17.76	0.146	0.082	1.418 e-02	↓	92	85
23	taurine	3.251 (t)	1.05	21.79	11.42	2.59	18.66	10.63	0.200	0.166	n.s.	↑	92	98
24	4-hydroxyphenylacetate	3.438 (s)	0.14	4.63	2.38	0.38	5.95	3.17	0.188	0.087	3.503 e-04	↓	88	90

Peak n.	Metabolite	δ (ppm)	Relative concentrations (mM)						K-S values		*p*-value (< 0.05)	Variation	FO (%)	
			BC			CTL								
		Multiplicity	Min	Max	Average	Min	Max	Average	BC	CTL			BC	CTL
25	glycine	3.553 (s)	1.81	33.87	17.84	1.93	45.26	23.59	0.153	0.168	1.401 e-02	↓	100	88
26	mannitol	3.671 (m)	2.07	78.45	40.26	13.85	63.76	38.81	0.143	0.153	n.s.	↑	92	100
27	guanidoacetate	3.786 (s)	1.71	21.30	11.51	7.10	29.50	18.30	0.122	0.118	4.738 e-03	↓	92	100
28	serine	3.840 (q)	1.71	40.85	21.29	15.00	71.74	43.37	0.104	0.190	9.040 e-05	↓	92	95
29	creatine	3.920 (s)	1.15	28.13	14.64	4.20	46.93	25.56	0.222	0.141	n.s.	↓	92	98
30	hippurate	3.957 (d), 7.538 (t), 7.628 (t), 7.823 (d)	2.10	130.18	66.14	5.56	70.01	37.79	0.229	0.163	3.085 e-04	↑	96	93
31	trigonelline	4.428 (s), 8.073 (m), 8.834 (m)	0.10	6.2	3.15	0.44	17.79	9.12	0.136	0.104	1.296 e-03	↓	88	100
32	3-methylhistidine	8.081 (s)	0.32	16.91	8.62	0.56	12.51	6.64	0.220	0.141	n.s.	↑	100	93
33	hypoxanthine	8.203 (s)	0.23	2.56	1.4	0.28	2.77	1.53	0.228	0.151	n.s.	↓	64	80

Legend: (s), singlet; (d), duplet; (t), triplet; (m), multiplet; (dd), double duplet; Min – minimum concentration; Max: maximum concentration; n.s. none significant between the groups under analysis; K-S- Kolmogorov-Smirnov tests.

Multivariate statistical analysis of urinary metabolomic profile

The first step before performing multivariate statistical analysis was to verify the normal distribution of the urinary metabolomic profile dataset using the Kolmogorov-Smirnov test (Table 3.2.2). All samples under analysis exhibited a normal distribution within each assigned group ($p > 0.05$).

To obtain a reliable dataset to apply multivariate analysis, the dataset was evaluated to exclude the metabolites that had an FO < 90 %. The dataset used to perform the statistical analysis also excluded creatinine as mentioned above as its concentration is dependent on age, gender and disease status, decreasing with age and varying throughout the day. Regarding creatinine relative concentration obtained in this study and the respective differences between groups under study, this metabolite might be considered a potential artefact given that creatinine values may be altered, as the generation of creatinine may not be simply a product of muscle mass but influenced by muscle function, muscle composition, activity, diet and health status [185]. Based on this, the dataset composed of 32 metabolites and 70 samples (32 BC and 38 CTL) that fulfilled this condition was subject to principal component analysis (PCA). PCA is an unsupervised method performed to visualize the similarities/differences between urine samples profiles of groups in this study. In this step, the samples were analyzed individually, e.g., without classification according to the groups. PCA score plot from urine samples are presented in Figure 3.2.2.

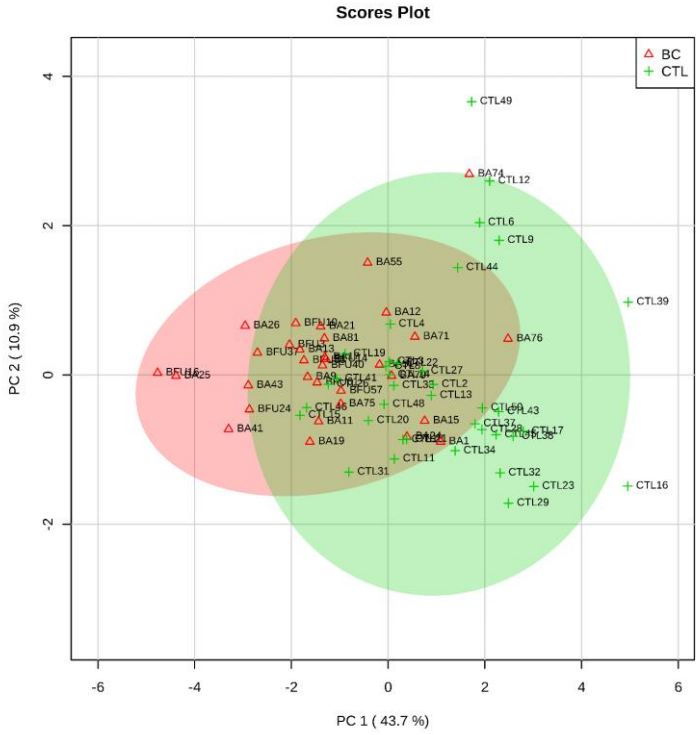

Figure 3. 2. 2 - Scores plot of PCA from urine samples based on 32 metabolites. For numbers identification see Table 3.2.2.

Although the projection of the variance between samples was performed without classification, it is possible to observe that the PCA of urine samples from BC patients and those from CTL presented a tendency for the formation of two clusters across the first principal component (PC1) that explains 54.6 % of the total variance. Most of the metabolites exhibited enormous importance in the variance projection of samples. Then, the partial least square-discriminant analysis (PLS-DA) was used as a supervised clustering method to maximize the separation between the groups and demonstrated that the samples had a tendency to be grouped according with health condition of subject (BC and CTL) through its variance/covariance along the first component. Ten differently expressed metabolites that exhibited a variable importance in projection (VIP) score greater than 1 were identified: creatine, glycine, serine, dimethylamine, trimethylamine N-oxide, α-hydroxyisobutyrate, mannitol, glutamine, cis-aconitate and trigonelline (Figure 3.2.3 A-B).

(A) **(B)**

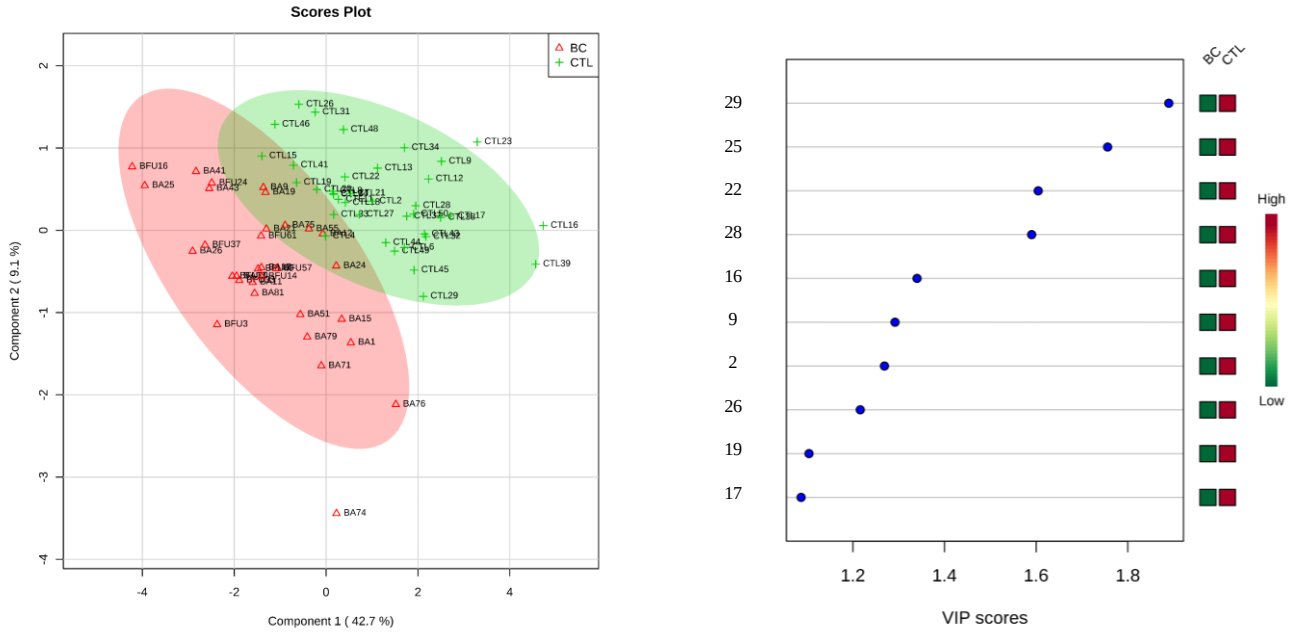

Figure 3. 2. 3 - Score plots of PLS-DA (A) and (B) VIP values of metabolites obtained by ^{1}H NMR analysis of urine samples from the 2 groups in study. For number identification see Table 3.2.2.

Many of these metabolites were already identified in various cancer types, including lung [150,186], breast, ovarian [123,144], bladder [187], and gastric cancers [188], in previous reports. Additionally, Zhou et al. [189] performed a metabonomics study using serum and urine from BC patients based on NMR where citrate, phenylacetylglycine and guanidoacetate exhibited significance in the discrimination of BC patients from CTLs. In addition, Slupsky et al. [123] accomplished a study using urine from breast and ovarian cancer patients to discover metabolites for an early diagnosis.

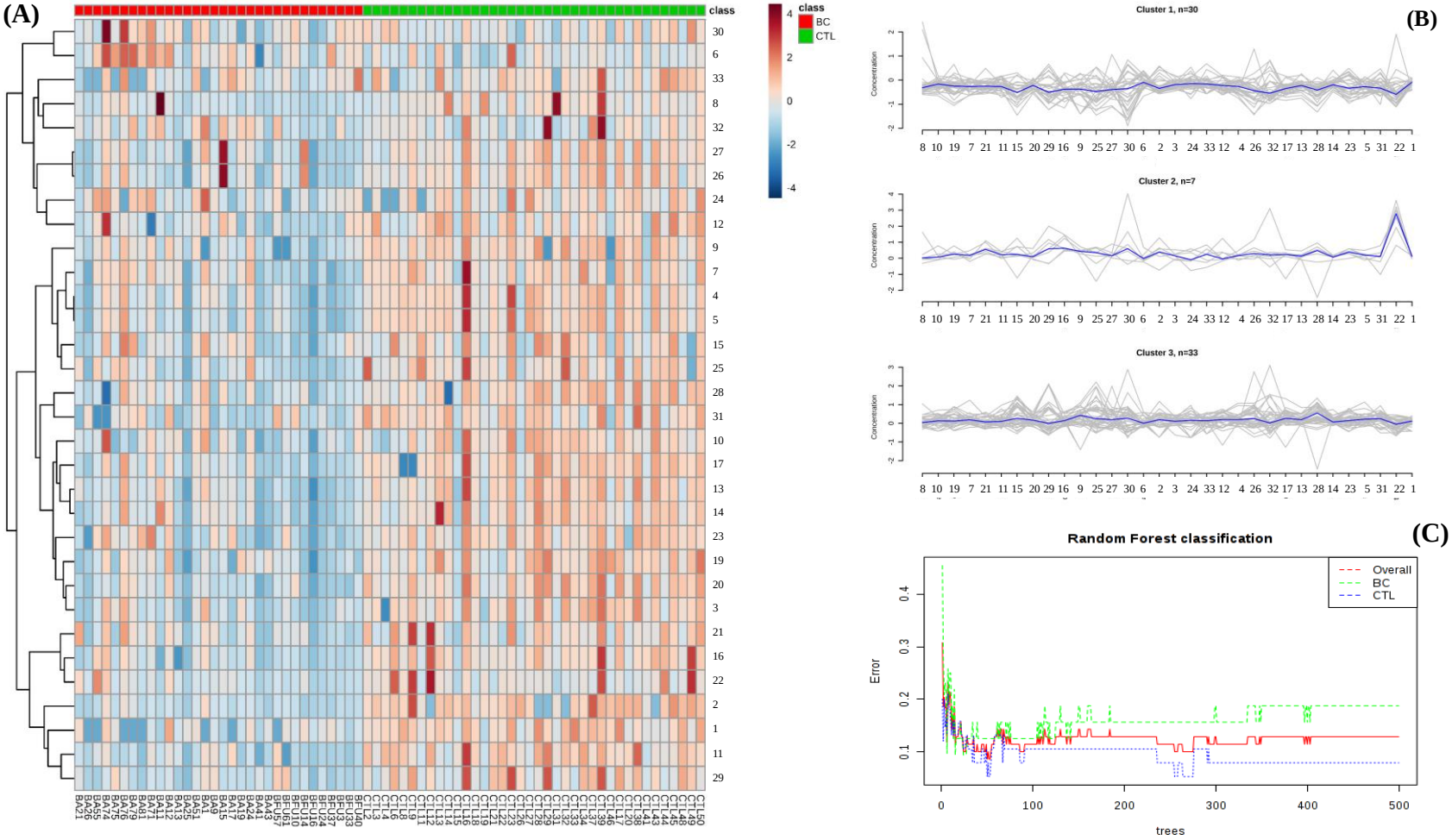

Figure 3. 2. 4 - (A) Heat map visualization and hierarchical clustering analysis and was constructed using the 32 metabolites by Pearson distance analysis; (B) clustering analysis by K-means of the 2 groups; (C) RF classification of urine metabolites from BC and CTL groups which indicated and overall error of 0.19 for BC and 0.08 for CTL.

The authors found that certain intermediates of the tricarboxylic acid cycle and metabolites relating to energy metabolism, amino acids, and gut microbial metabolism were perturbed. With regards to aminoacids as raw materials of protein synthesis and catabolism products in vivo their changes whether in composition and concentration can reflect the metabolic status of patients [189]. Moreover, the heat map was constructed using Pearson´s correlation, providing intuitive visualization of the data set. The heat map contains the 32 metabolites and was used to identify samples or features that are unusually high or low (Figure 3.2.4 A). As noted in Figure 3.2.4, the higher relative concentrations for the majority of metabolites were found in the CTL group whereas the lowest relative concentrations were noted in the BC group. K-means clustering analysis was also performed as a non-hierarchical clustering technique. The method initially creates k random clusters (k is the number of groups in the study). The programme then calculates the mean of each cluster (indicated by the shaded line). If an observation is closer to the centroid of another cluster, then the observation is made by a member of that cluster. This process is repeated until none of the observations are reassigned to a different cluster (Figure 3.2.4 B). For this analysis, 8 BC and 25 CTL samples were included in cluster 1; 1 BC and 6 CTL samples were included in cluster 2; 23 BC and 7 CTL samples were included in cluster 3.

RF analysis was performed to determine the ability of metabolites to accurately classify the study subjects into their corresponding group. The method uses an ensemble of classification trees (in this case 1000), and each of tree is grown by random feature selection from a bootstrap sample at each branch. Class prediction is based on the majority vote of the ensemble. These out-of-bag (OOB) data are then subsequently used as a test sample to obtain an unbiased estimate of the classification error (OOB error). The average value was 0.13, indicating that the accuracy of classification was lower (approximately 18 %) for the BC than for the CTL group at 8 % (Figure 3.2.4 C).

Additionally, orthogonal partial least squares-discriminant (OPLS-DA) analysis was applied to the urinary metabolomic profile dataset to maximize the separation between the CTL and BC groups as presented in Figure 3.2.5 (A and B). OPLS-DA uses class information to demonstrate variables responsible for class discrimination using the predictive information of the first component. The main advantage of OPLS-DA when compared with PLS-DA is that a single component is used as a predictor for the class, whereas the other components describe the variation orthogonal to the first predictive component [139].

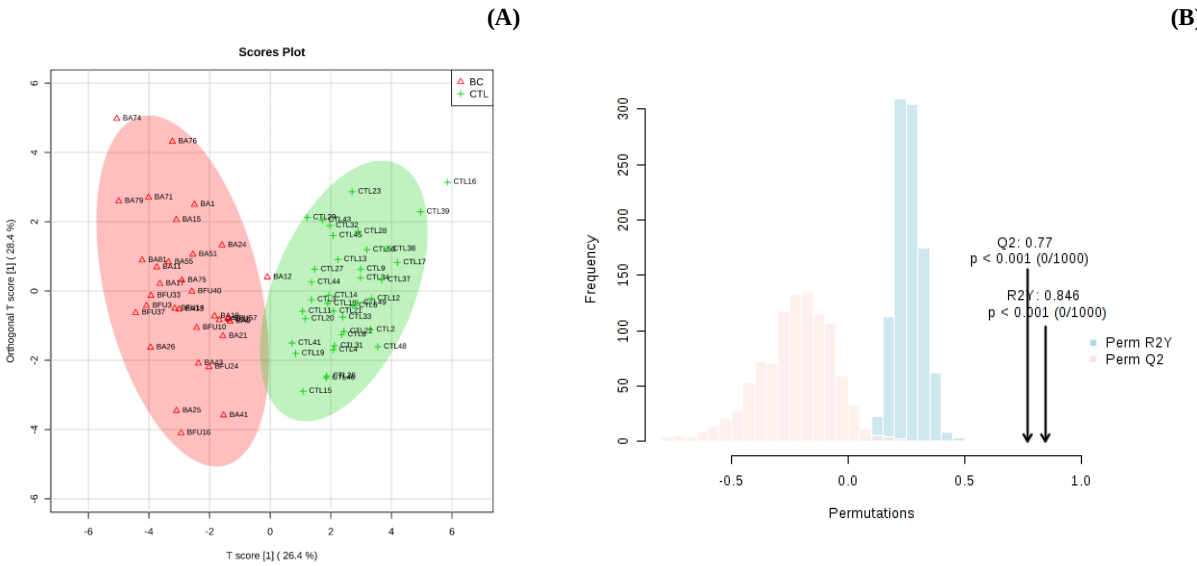

Figure 3. 2. 5 – (A) Score plots of OPLS-DA analysis and (B) model validation by permutation tests based on prediction accuracy of metabolites obtained by 1H NMR analysis of urine samples from the 2 groups in study. For number identification see Table 3.2.2.

To attest the robustness of the model, a random permutation test with 1000 permutations was performed with OPLS-DA (Figure 3.2.5 C). The permutation test yielded R^2 (represents goodness of fit) as 0.846 and Q^2 (represents predictive ability) as 0.770 indicating that the model is not over fitted and have a relative good predictive ability to distinguish between study groups. Moreover, receiver operating characteristic curves (ROCs) was generated for the two groups (CTL-BC) using the 32 identified metabolites and is presented in Figure 3.2.6.

As noted in the Figure, as the number of metabolites increases the area under the curve (AUC) also increases. Thus, using only 2 metabolites, the AUC value obtained was 0.92 for the CTL-BC demonstrating the higher sensitivity/specificity to distinguish the groups. The metabolites with significance were creatine, glycine, serine, α-hydroxyisobutyrate and trimethylamine N-oxide. In addition, using all 32 metabolites it can be achieved the highest area under the curve. These results are in accordance with the literature, where Xia et al. [108], an AUC value between 0.9 and 1.0 is excellent, and a value between 0.8 and 0.9 is good, comparing the results, the values obtained were very good. A greater AUC value indicates a greater ability to distinguish the CTL from the BC group.

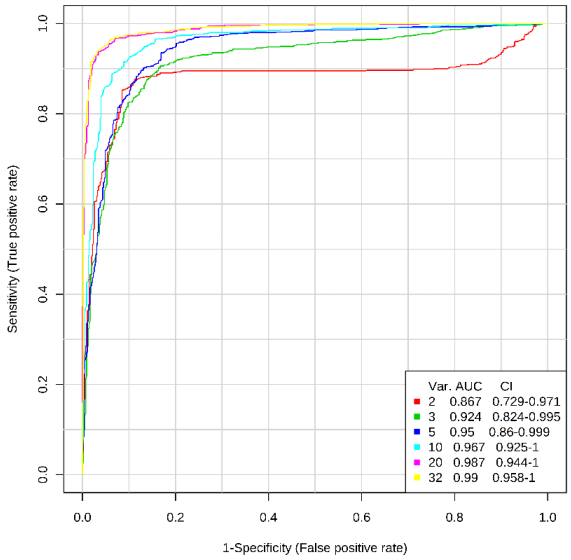

Figure 3. 2. 6 - ROC curves for the predictive model. A combination metabolites model calculated from the logistic regression analysis using the 32 identified metabolites for distinguishing BC patients from CTL.

The AUC can be interpreted as the probability that a randomly selected diseased subject is classified as diseased than a casually selected healthy subject [108].

Finally, the metabolic pathway analysis was performed to determine which pathways were altered in the groups under study.

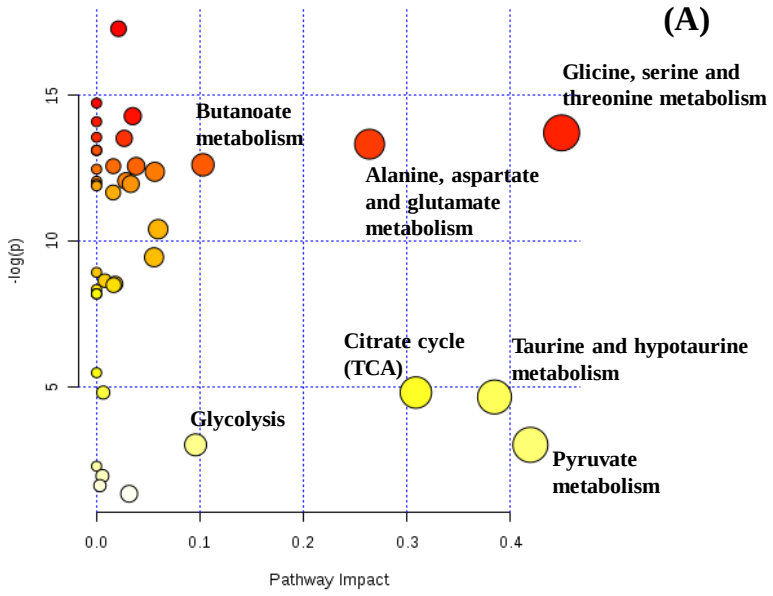

Figure 3. 2. 7 - The metabolome view map (A) of significant altered metabolic pathways observed in urine from BC and CTL groups and (B) metabolic pathways with highest impact that include the most promising potential BC biomarkers identified in this study.

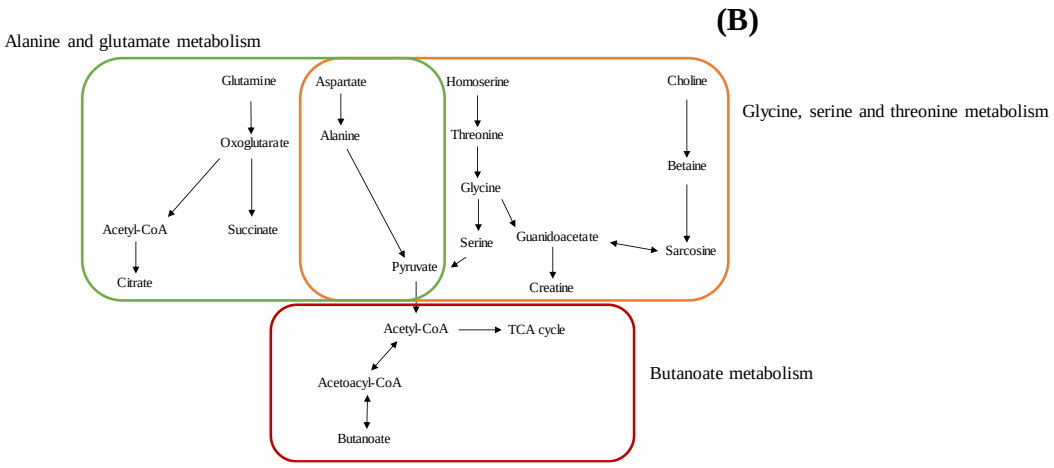

Figure 3. 2. 7 - The metabolome view map (A) of significant altered metabolic pathways observed in urine from BC and CTL groups and (B) metabolic pathways with highest impact that include the most promising potential BC biomarkers identified in this study.

Figure 3.2.7 (A and B) presents the impacted pathways in the CTL-BC groups, respectively. The most impacted metabolic pathways were glycine metabolism, glutamate metabolism; butanoate metabolism; glycolysis, citrate cycle (TCA cycle), taurine metabolism and pyruvate metabolisms, indicated by the red and yellow colors. It can be highlighted that the pathway with highest impact was the glycine metabolism (Figure 3.2.7 B). Based on these results, a successful differentiation and discrimination of samples was achieved between CTL and BC group. The results indicate that the ^{1}H NMR urinary profile represents a useful approach to identify potential BC biomarkers.

Conclusions

This study assessed the metabolomic urinary profile in BC patients in active and follow-up stages compared with that in healthy volunteers using ^{1}H NMR combined with multivariate statistical tools (PCA, PLS-DA and OPLS-DA) were applied to two groups (BC and CTL). Thirty-three metabolites were identified and quantified using Chenomx software. Multivariate statistical analysis revealed some metabolites were significantly altered in BC patients. Of the metabolites detected, creatine, glycine, serine, dimethylamine, trimethylamine N-oxide, α-hydroxyisobutyrate, mannitol, glutamine, cis-aconitate and trigonelline exhibited the highest sensitivities and specificities to discriminate BC patients from healthy controls. Plot analysis revealed a metabolomic biosignature comprising an array of several biochemical pathways altered in BC patients. Metabolic pathway analysis indicated that the discriminatory metabolites potentially originated from several dysregulated pathways in BC: glycine, metabolism, glutamate metabolism; butanoate metabolism; glycolysis, citrate cycle (TCA

cycle), taurine metabolism and pyruvate metabolism. These results suggested the possibility of identifying endogenous metabolites as a platform to discover potential BC biomarkers and paves a way to investigate related metabolomic pathways to improve the diagnostic tools of BC.

3.3| Volatile metabolomic signature of human breast cancer cell lines

(Silva et al.; Scientific Reports **2017***; 7:43969)*

Abstract

BC remains the most prevalent oncologic pathology in women, causing huge psychological, economic and social impacts on our society. Currently, the available diagnostic tools have limited sensitivity and specificity. Metabolome analysis has emerged as a powerful tool for obtaining information about the biological processes that occur in organisms and is a useful platform for discovering new biomarkers or make disease diagnosis using different biological samples. VOMs from the headspace of cultured BC cells and normal human mammary epithelial cells, were collected by HS-SPME and analyzed by GC–MS, thus defining a volatile metabolomic signature. Sixty VOMs were identified in all cell lines, belonging to different chemical families. From these, 2-pentanone, 2-heptanone, 3-methyl-3-buten-1-ol, ethyl acetate, ethyl propanoate and 2-methyl butanoate were detected only in the headspace of cultured cancer cell lines. Multivariate statistical methods were used to verify the volatile metabolomic differences between BC cell lines and normal cells to find related volatile metabolites that could be associated with BC, providing comprehensive insight into VOMs as potential cancer biomarkers. The establishment of the volatile fingerprint of BC cell lines presents an opportunity to identify endogenous VOMs to discover biomarkers of BC and the related metabolomic pathways, thereby improving the available BC diagnostic tools.

Keywords: BC cell lines; HMEC cells; HS-SPME; GC-MS; VOMs; biomarkers.

Introduction

Although there has been a sustained decline in mortality rates over recent decades, BC continues to be the most prevalent malignancy among women worldwide and is a major cause of female deaths. A number of associated factors, including age, gender, ethnicity, lifestyle (tobacco, alcohol, diet, and lack of exercise) and genetics, such as mutations in the tumor suppressor genes BRCA1 (mutation on chromosome 17) and BRCA2 (mutation on chromosome 13), have been identified as the most common inherited causes of susceptibility to BC. The BRCA1 and BRCA2 genes encode very large proteins that are expressed in a wide variety of different tissues and are implicated in processes such as DNA repair and recombination, checkpoint control of the cell cycle, and transcription. Families with a high incidence of BC may carry mutations in one or both of these genes, or alternatively, members of these families may have similar lifestyle habits and may have been affected by similar environmental factors [190,191]. Furthermore, other important gene mutations related to BC development include mutations in ATM (ataxia-telangiectasia mutated), TP53 (tumor protein p53), and PTEN (phosphatase and tensin homolog deleted on chromosome ten) [192–194]. Most of these are rare and often do not increase the risk of BC as much as do mutations in the BRCA genes. Detection of BC at an early stage is extremely important to reduce the burden of disease because earlier detection leads to better patient outcomes, as metastatic states are avoided [107,195]. Currently, there is no single screening test that is totally reliable, and a number of tests can be combined to help detect early stage BC. Current screening techniques include self-examination for lumps or nodes, mammography, ultrasound, magnetic resonance imaging and biopsy using a fine needle or similar instrument to aspirate or otherwise remove a sample of fluid or cells from any suspicious lump or node for microscopic examination [113]. These methods are generally invasive, time-consuming and require special medical skills. Therefore, there is a need for non-invasive, accurate and rapid screening tests for early detection. In this context, it is useful to identify VOMs that may prevent or predict the occurrence of metastasis before it manifests in the patient.

Metabolomics has emerged as a powerful tool for understanding biological processes that occur in humans, and it has mostly been based on the analysis of biofluids, such as blood, saliva, or urine, to discover new cancer biomarkers or to diagnose a disease [196]. The use of cell culture metabolomics enables both the discovery of novel biomarkers of pathological conditions and investigation of the related metabolomic pathways (Figure 3.3.1). Many of the metabolic processes in the body, such as lipid peroxidation, energy metabolism through glycolysis and amino acid catabolism are common to all living cells [196]. It is believed that some metabolic pathways might be up- or down-regulated in cancer cells and, therefore, metabolome analysis may reveal differences between biological samples based on metabolic profiles or fingerprints. Indeed, cancer cells have an

altered metabolism compared with normal cells that may lead to the production of specific compounds [197]. In recent years, several studies have reported the analysis of cancerous cell lines to find potential cancer biomarkers [198–201]. The most recent techniques include the use of nanomaterial-based sensors [198], electrochemical sensors [202], or thermal desorption coupled with gas chromatography mass spectrometry [152]. However, most of these techniques are expensive and time-consuming. In this work, HS-SPME which was developed by the Pawliszyn group [203] in the early 90´s and consists of a fiber coated with different polymers extracting a wide range of chemical compounds, was selected as an extraction technique. This technique is superior to other extraction techniques, in that it is rapid, easy to use, sensitive and does not require a concentration step before analysis.

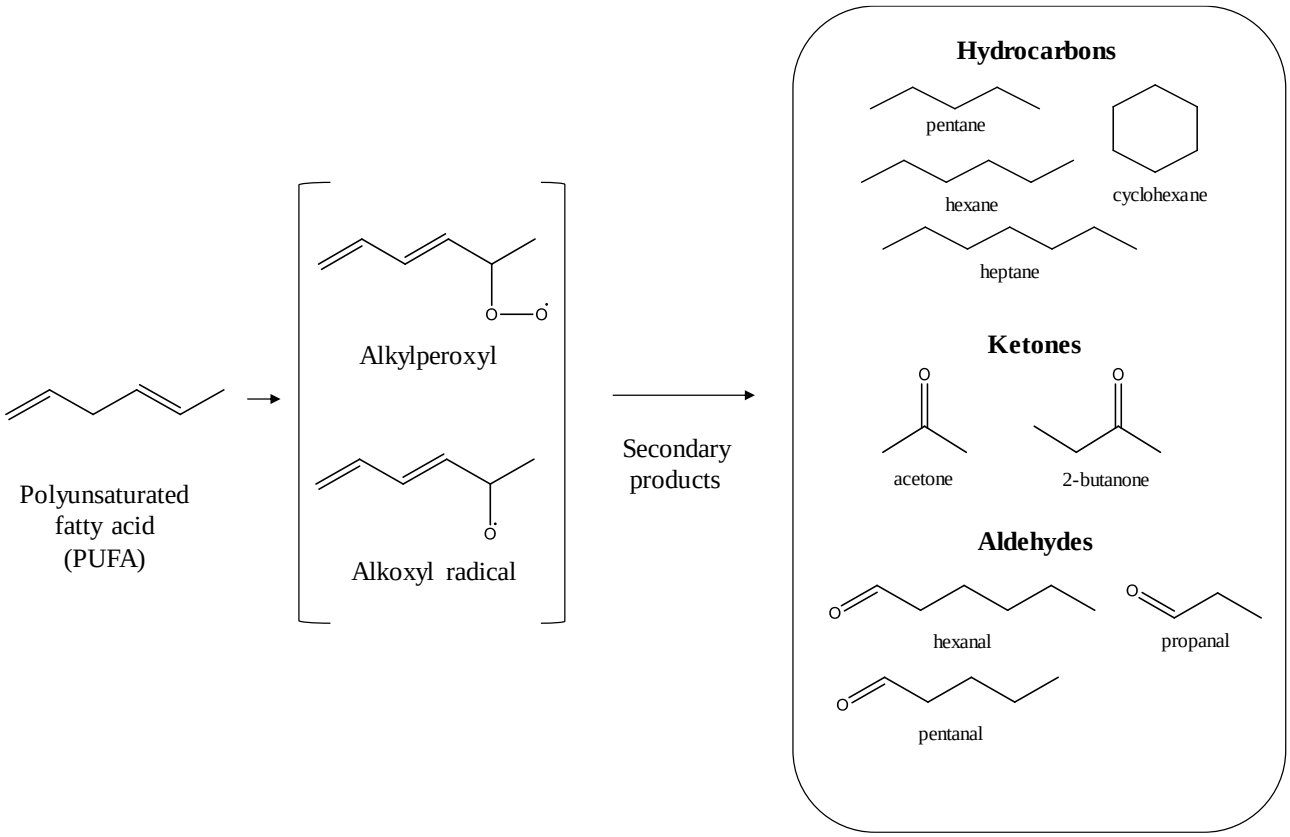

Figure 3. 3. 1 - Formation of some intermediate products of lipid peroxidation. Adapted from Li et al. [204].

In this study, a comparative analysis of the volatile metabolomic signature of BC cell lines (T-47D, MDA-MB-231, MCF-7) and normal human mammary epithelial cells (HMEC) was carried out, in order to identify BC-specific VOMs and to identify a set of biomarkers that could hopefully be correlated with VOMs released *in vivo* by BC cells. This finding will improve the knowledge about the origin of VOMs and providing comprehensive information as potential BC biomarkers. This strategy can help reveal novel BC biomarkers that might expand the current understanding of this

multi-factorial disease. The GC–qMS analyzes allow specific identification of VOMs, while multivariate statistical analysis is able to differentiate and discriminate oncologic from normal cells providing proof-of-principle for the detection of different volatile metabolomic patterns in target cells.

Materials and Methods

Reagents and materials

PBS was purchased from Sigma-Aldrich (St. Louis, MO, USA), sodium chloride was obtained from Panreac (Barcelona, Spain), the SPME holder for manual sampling of SPME fiber [50/30 µm divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS)] and the glass vials were purchased from Supelco (Bellefonte, PA, USA). The SPME fiber was conditioned according to manufacturer's instructions. Before each daily analysis, the fiber was conditioned for 10 min in the injector port to prevent carryover. T75 glass flasks were purchased from Ningbo (Ja-Hely Technology, China).

Cell lines and cultivation conditions

The human breast adenocarcinoma cell line MCF-7and human breast carcinoma cell lines T-47D and MDA-MB-231 were purchased from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany). MCF-7 was grown in 90% RPMI 1640 (Life technologies, Gibco®) supplemented with 15% fetal bovine serum (FBS, Life technologies, Gibco®), 1% Antibiotic-Antimycotic solution (AA, Life technologies, Gibco®), 1% MEM Non-Essential Amino Acids solution (Life technologies, Gibco®), 1 mM sodium pyruvate (Sigma-Aldrich, St. Louis, MO, USA) and 10 µg/mL human insulin (Sigma-Aldrich, St. Louis, MO, USA); T-47D cell line was grown in 85% RPMI 1640 supplemented with 15% fetal bovine serum (FBS), 1% Antibiotic-Antimycotic solution and 10 µg/mL human insulin, while the MDA-MB-231 cell line was grown in 85% RPMI 1640 supplemented with 15% fetal bovine serum (FBS) and 1% Antibiotic-Antimycotic solution. Human mammary epithelial cells (HMEC) were purchased from Life technologies (Gibco®) and grown in HUMEC serum-free medium supplemented with 20 µg/mL of Antibiotic-Antimycotic solution (Life technologies, Gibco®). All cells were incubated in a humidified atmosphere containing 5% CO_2 and 95% air at 37ºC. Culture media was changed every 2 days and the cultures were passaged with 0.25% trypsin-EDTA (Life technologies, Gibco®) when 80% of confluence was achieved.

VOMs extraction from cell cultures

To extract VOMs from cell cultures, glass flasks were treated with collagen to promote cell adherence. Briefly, the glass flasks were covered with a collagen solution (0.2 mg/mL) for 30 min and then washed with PBS (3 times). The cells were then cultured in the T75 flasks for 48 h. After this period, volatile metabolites were extracted using a DVB/CAR/PDMS SPME fiber exposed in the headspace of the flasks for 45 min at 37 ºC, followed by injection into the GC injection port for 10 min to allow the desorption of VOMs from the fiber. After these extractions, cell-free aliquots were collected from the flasks holding 10 mL of the culture medium with growing cells. They were centrifuged to remove any suspended cells, and then 1 mL aliquots were adjusted to pH 2, 7 or10 with 1M HCl or 1M NaOH [205]. After the addition of 200 mg NaCl and subsequent stirring (0.5 mm × 0.1 mm bar) at 800 rpm, the vials were capped with PTFE septa through which the SPME fiber was inserted in the headspace of the vial and placed in a thermostatic bath at 37 ºC for 45 min. After this, the fiber was withdrawn into the needle and injected in the GC port (250 ºC) over 10 min, when the analytes were thermally desorbed and transferred to the analytical column. Control headspace samples were also collected from flasks containing only empty media treated with the same incubation conditions to determine the contribution to the background. The analyzes were performed in triplicate.

GC-MS analysis

VOMs in the headspace were analyzed using an Agilent Technologies 6890N Network gas chromatograph system (Palo Alto, CA, USA) equipped with a BP-20 fused silica column (60 m × 0.25 mm I.D. × 0.25 µm film thickness, SGE, Dortmund, Germany) interfaced with an Agilent 5975 quadrupole inert mass selective detector. The following oven temperature profile was set: (a) 5 min at 45 ºC; (b) increase temperature until 150 ºC, at a rate of 2 ºC min^{-1} (hold for 10 min); (c) 150 ºC for 10 min; (d) increase temperature until 220 ºC, at a rate of 7 ºC min^{-1}; and (e) 220 ºC for 10 min. Column flow was constant at 1.0 mL/min using helium (He, N60, Air Liquide, Portugal) as the carrier gas. The injection port was maintained at 250ºC and operated in the splitless mode. Regarding MS analyzes, the operating temperatures of the transfer line, quadrupole and ionization source were 270, 150 and 230ºC, respectively. The electron impact mass spectra were recorded at 70 eV and the ionization current was 10 µA, and data acquisition was performed in scan mode (30-200 *m/z*). The identification of metabolites was performed comparing mass spectra with the Agilent MS ChemStation Software (Palo Alto, CA, USA) equipped with the NIST05 mass spectral library with a similarity threshold higher than 80 %, or with commercially standards when available. All

experiments were performed in triplicate and the results were expressed as the mean ± standard deviation.

Statistical analysis

Statistical tests were performed using the StatSoft STATISTICA 12.0 (2013) software (Tulsa, OK, USA). Differences in VOMs between groups were tested with *one-way* ANOVA, and $p < 0.05$ was considered as statistically significant. PCA, PLS and LDA were carried out on VOMs selected by *ANOVA* to evaluate differences among the studied groups. PCA was performed in order to obtain differentiation between samples under study without classification and PLS was used as a supervised linear pattern recognition algorithm for data classification of samples. PCA and PLS were performed through variables values scale by unit standard deviation with convergence criterion (0.0001) and leave-one-out cross validation for accuracy confirmation. For LDA analysis a backward selection method was used with a $p < 0.05$ through Wilks test. For cross validation a leave-one-out strategy was used.

Results and discussion

VOMs associated with normal breast cells (HMEC) and BC cell lines (T-47D, MDA-MB-231 and MCF-7 cells) were investigated. The cell lines for the present study were chosen based on their different molecular characteristics (Table 3.3.1): namely, the expression of the estrogen receptor (ER), the progesterone receptor (PR), and the human epidermal growth factor receptor 2 (HER2). It is well known that BC is heterogeneous and that its prognosis and treatment depends on the molecular subtype of the cancer cells. The VOMs arising from the cellular metabolism were studied using HS-SPME/GC-MS: a) by direct analysis of the headspace of the culture flasks after cell growth (these results are hereinafter designated as "Cells"); and b) by analysis of the volatile metabolites from the culture media at different pH values. From the analysis of chromatograms, it was possible to identify 60 VOMs belonging to distinct chemical groups, namely, alkanes, aldehydes, ketones, acids, alcohols and benzene derivatives.

Table 3. 3. 1 - Characteristics of investigated breast cells.

Cell line	Molecular subtype	Classification		
		ER	PgR	HER2
HMEC	Epithelial	Low	Low	Low
MCF-7	Luminal A	+	+	-
T-47D	Luminal A	+	+	-
MDA-MB-231	Triple negative	-	-	-

ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; PgR, progesterone receptor.

VOMs signature of BC cell lines and breast normal cells

We identified twenty-six VOMs belonging to several chemical groups (Figure 3.3.2): namely, alkanes, aldehydes, ketones, acids, alcohols and benzene derivatives (Table 3.3.2).

Table 3. 3. 2 - Identification of VOMs from investigated human mammary epithelial cells and human BC cell lines by HS-SPME/GC-MS.

Peak n.	RT (min)	Abbreviation	Ion	Compound	HMEC ($\times 10^5$)	T-47D ($\times 10^5$)	MDA-MB-231 ($\times 10^5$)	MCF-7 ($\times 10^5$)
2	4.356	HEXA	57	hexane	0.44	-	-	-
4	5.127	M4HEPT	43, 70	4-methyl-heptane	4.83	-	-	-
5	5.739	ACTONE	43, 58	acetone	9.24	501.87	10.82	22.00
6	6.843	ETHATE	43, 61	ethyl acetate	-	8.56	3.16	25.55
8	8.716	ETHPPATE	57, 44	ethyl propanoate	-	18.11	2.17	19.66
9	9.337	PENTONE2	43, 86	2-pentanone	-	-	-	22.74
10	9.994	DECA	57, 44	decane	3.36	354.35	4.28	12.80
12	12.341	M2BTATE	57	2-methyl butanoate	-	12.83	-	13.88
17	15.865	ETHBNZ	91, 106	ethylbenzene	12.1	4.02	2.38	10.23
18	16.646	DMBNZ13	91, 106	1,3-dimethylbenzene	6.32	3.81	2.30	5.23
19	17.395	BUTOL1	56, 41	1-butanol	1.21	2.84	2.88	5.94
20	19.379	HPTONE2	43, 58	2-heptanone	-	4.96	16.46	40.00
22	20.157	DODEC	57, 43,	dodecane	14.19	21.29	19.93	54.42
25	23.765	M33BUTOL1	41, 56	3-methyl-3-buten-1-ol	-	-	-	3.93
26	24.155	STYENE	104, 78	styrene	64.50	139.81	39.93	125.93
27	25.511	TMBNZ124	105	1,2,4-trimethylbenzene	3.88	-	-	3.92
29	26.207	CHEXONE	55, 42,	cyclohexanone	2.11	20.23	12.83	8.99
32	33.450	TTDECANE	57	tetradecane	21.92	-	-	-
33	33.598	CHEXOL	57, 82	cyclohexanol	-	99.42	74.31	210.4
34	35.176	B13DMEBNZ	175, 190	1,3-bis(1,1-	4.56	9.42	7.22	31.67
36	39.305	E2HEXOL1	57, 43	2-ethyl-1-hexanol	28.32	950.78	184.81	1803.82
38	41.341	BNZAL	106, 77	benzaldehyde	2.81	-	-	-
45	48.761	ACTPONE	105, 77	acetophenone	4.44	-	-	-
47	53.752	NPTENE	128	naphthalene	3.60	3.80	3.70	5.61
54	70.775	PHOL	94	phenol	3.93	3.35	3.96	5.48
60	79.610	DT24BPHOL	191	2,4-di-*tert*-butylphenol	51.57	73.53	71.04	160.00

From these VOMs, 13 were found to be common in all studied breast cells (both normal and cancerous), 5 were present only in normal breast cells (HMEC), and 2 compounds were identified only in the MCF-7 breast cell line. As can be observed in Figure 3.3.2, the MCF-7 cell line demonstrated the most complex volatile metabolomic signature in terms of number of the identified VOMs and total GC peak areas compared with the other cell lines. Moreover, for all BC cell lines, the major chemical group identified was the higher alcohols, represented mainly by 2-ethyl-1-hexanol and cyclohexanol. These VOMs have already been reported in previous studies using BC cell lines [124,206] and in urine [127,128] from cancer patients. It is believed that their endogenous origin is as hydrocarbon metabolism byproducts [204,207]. The obtained data indicated that the levels of both

VOMs (2-ethyl–1-hexanol and cyclohexanol) were higher in all investigated BC cells than HMEC. This might be due to the production of lipid peroxidation biomarkers with hydroxylase that are mediated by cytochrome P450 [199,206].

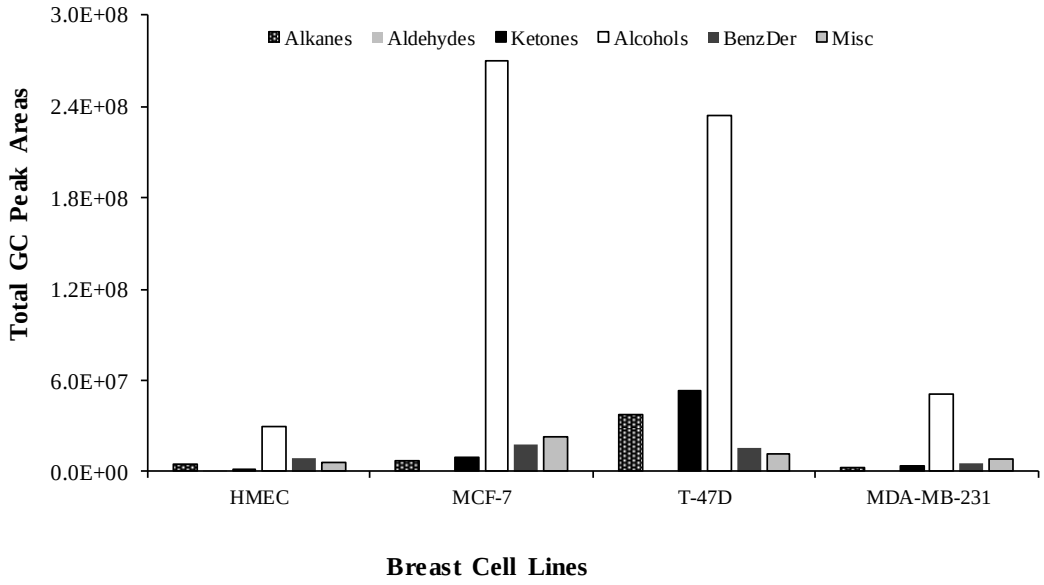

Figure 3. 3. 2 - Distribution of VOMs identified in cultured breast cell lines.

Similar results were reported by Peled and collaborators [208] when studying genetic mutations in lung cancer cells, and by Davies and collaborators [209], who compared the volatile profile from the headspace of lung cancer cells with genetic mutations in TP53 and KRAS. Most of the identified VOCs were common to all BC cell lines and normal human mammary epithelial cells, but six of them, 2-pentanone, 2-heptanone, 3-methyl-3-buten-1-ol, ethyl acetate, ethyl propanoate, and 2-methyl butanoate, were detected only in BC cell lines. This finding justifies a more detailed investigation to evaluate of these six VOMs as BC biomarkers.

The influence of pH on the VOMs identified from culture media

The pH is one of the parameters that influences the extraction efficiency of VOMs and therefore it is required an optimization step. This was accomplished by the assessment of volatiles from culture media at different pH. We evaluated the effect of pH on the volatile signature obtained from culture media. At pH 2, the MCF-7 cells had the highest total GC peak area with acids (hexanoic acid, octanoic acid and 2-ethyl-hexanoic acid) being the most dominant chemical group. Aldehydes (benzaldehyde and 3,4-dimethyl-benzaldehyde) were the most predominant chemical group in the T-47D and MDA-MB-231 cells (Figure 3.3.2; Table 3.3.2). At pH 7, for MCF-7 cells, alkanes, ketones and alcohols were found the dominant chemical groups, which were represented by dodecane, 2-heptanone, and 2-ethyl-1-hexanol. For the other breast target cells alcohols (cyclohexanol) were the

most representative chemical group. Finally, at pH 10 the main chemical group identified for MCF-7 cells were alkanes, ketones and alcohols (2-ethoxy-2-methyl-propane, acetone and 2-ethyl-1-hexanol). For T-47D and MDA-MB-231 cells, alcohols represented by cyclohexanol, presented the major contribution. As previously mentioned, it is believed that cancer cells have altered metabolisms leading to different volatile metabolomic patterns. This was observed in our study, where we identified some differences between BC cell lines and normal cells (Table 3.3.2).

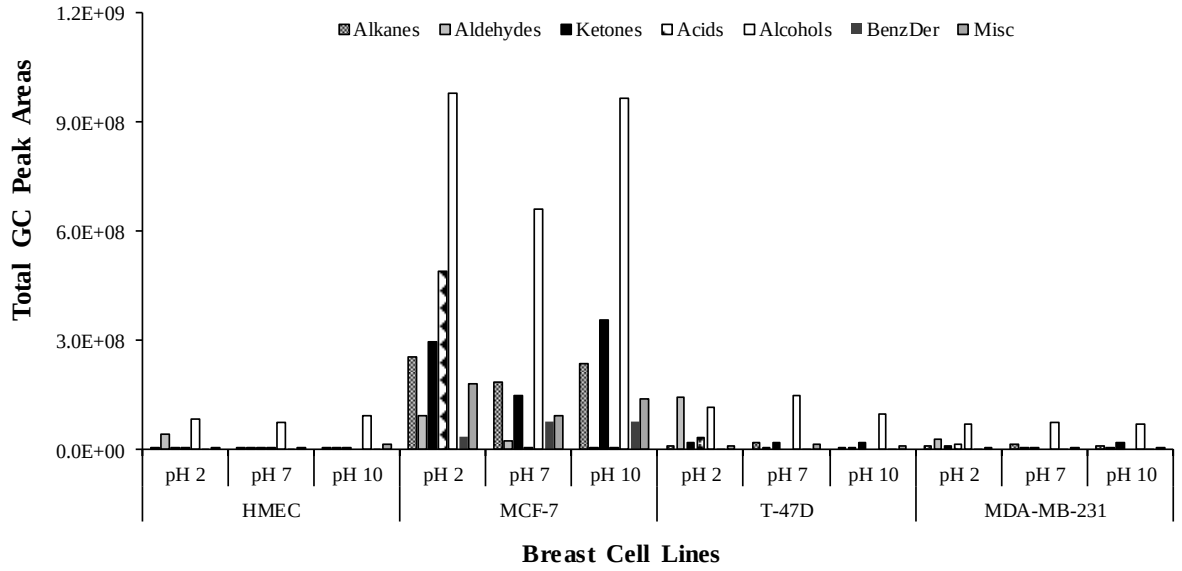

Figure 3. 3. 3 - Distribution by chemical families of VOMs identified in culture media at different pH conditions of breast cell lines.

Several VOMs were found to be common in all breast cell lines for all conditions, including, 2-ethoxy-2-methyl-propane, acetone, 2-methyl-2-propanol, cyclohexanol, 1,3-bis(1,1-dimethylethyl)-benzene and 2-ethyl-1-hexanol which had higher levels in BC cells. Ethyl acetate was only present in the T-47D cell line (Table 3.3.2). Kwak and collaborators [205] described a similar study using melanoma cells and identified higher concentrations of acetone in cancer cells. The metabolomic origin of most VOMs is still unknown, as they rely on a variety of endogenous pathways and exogenous sources. Huang and collaborators [206] reported that cyclohexanol and 2-ethyl-1-hexanol had lower concentrations in BC cells and suggested that they were generated by endogenous hydrocarbon metabolism. Hydrocarbons can be metabolized to aldehydes or ketones in the body via alcohol dehydrogenase (ADH) and cytochrome P450 activities [199]. The higher activity of cytochrome P450 may explain why BC cell lines have less cyclohexanol than normal breast cells [206]. It can also induce a variety of biological responses, including the biotransformation of alkanes, alkenes and aromatic compounds [210]. Furthermore, Philips et al. suggested that breast diseases are associated with increases in oxidative stress and a higher activity of P450 [211]. Nevertheless, 2-

ethyl-1-hexanol was found at higher levels in BC cells than in to normal cells. According to the human metabolome database, 2-ethyl-1-hexanol is involved in cell signaling, membrane integrity/stability and energy storage and it was also detected in lung cancer cell lines [152] at increased levels when compared with the medium. At pH 10, the levels for most of the VOMs were higher in BC cells than in normal breast cells, including those of acetone, 2-pentanone, cyclohexanol, 2-ethyl-1-hexanol and acetophenone.

PCA and PLS-DA analyzes of VOMs

To verify the significance of the identified VOMs from the headspaces of the culture flasks and the cell culture media at different pH conditions, a *one-way* ANOVA test was applied to analyze the data matrix. From the identified VOMs, a total of 23 (from cultured flask headspace), 52 (from culture media at pH 2), 34 (from the culture media at pH 7) and 43 (from the culture media at pH 10) showed significant differences ($p < 0.05$) (Figure 3.3.4).

Principal component analysis (PCA) was performed for each condition to identify variables to differentiate the VOMs pattern of the HMEC cells from those of the BC cell lines (MCF-7, T-47D and MDA-MB-231 cells), and from the VOMs patterns obtained from the culture media at different pH values (Figure 3.3.5). The differentiation between the above conditions was shown as the loading scores plot of the two principal components of the PCA. The PCA analysis is an unsupervised projection method used to visualize the dataset that displays the similarities and differences between groups and, in this case, demonstrated that the variables (scaled by standard deviation) used were sufficient to describe subsets with similar characteristics. These results demonstrated that the scores from the cancer cell lines and those from the normal breast cells exhibited separate trends in the plots. Figure 3.3.5 A shows the loading scatterplots of the PCA obtained from the analysis of the VOMs in the headspace of cultured flasks. It can be observed that 3 groups were formed, where HMEC cells was clearly differentiated from BC cell lines, which showed greater differentiation from MCF-7 cell lines across the PC1 and from T-47D and MDA-MB-231 across the PC2. Interestingly, BC cell lines formed two separated groups according molecular subtype (luminal A versus triple negative). However, no differentiation was achieved between T-47D and MDA-MB-231 cell lines, which formed a single group, perhaps this grouping of the two cell lines might be due to the fact that they have similar molecular characteristics. The variables that explain the differentiation between cell lines are represented in Figure 3.3.5 B.

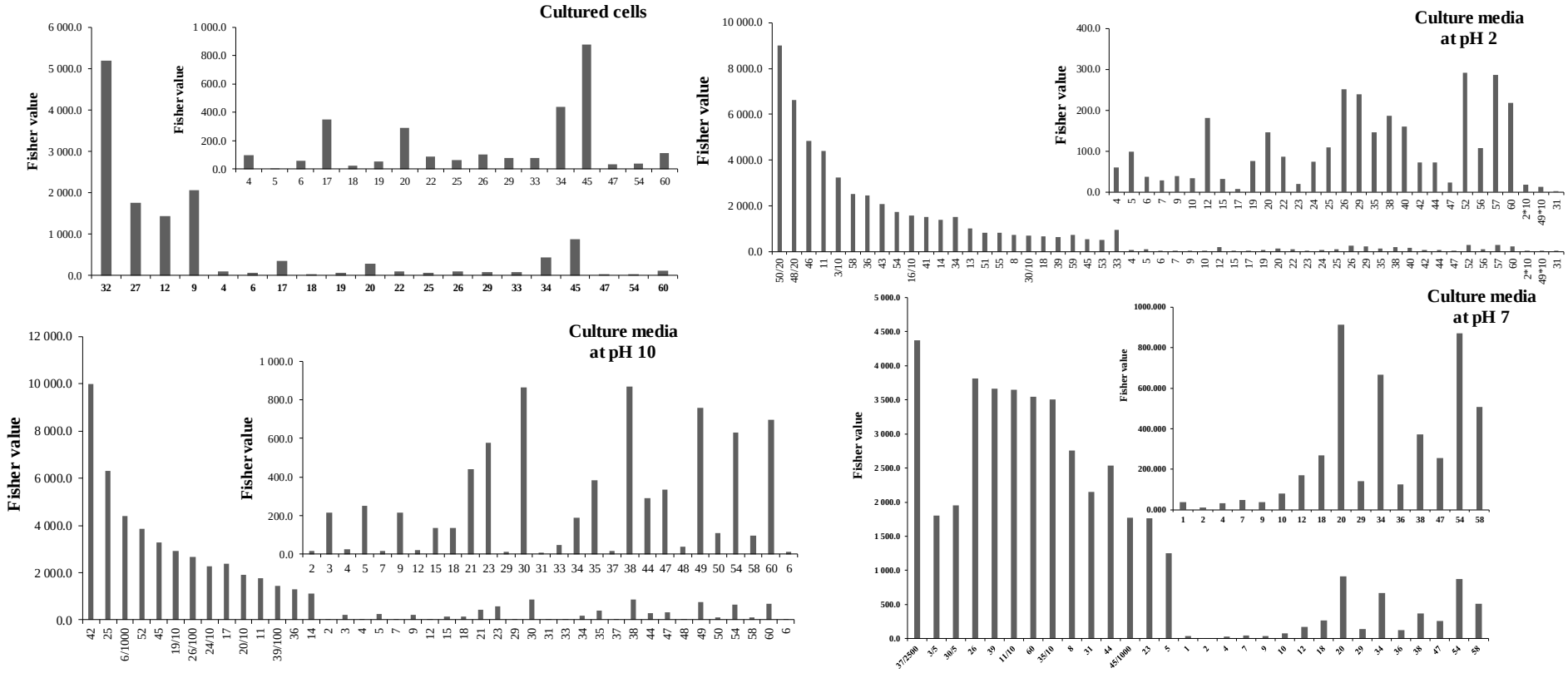

Figure 3. 3. 4 - One-way analysis of variance (ANOVA) test in breast cell lines in studied conditions (for numbers correspondence please see Table 3.3.2).

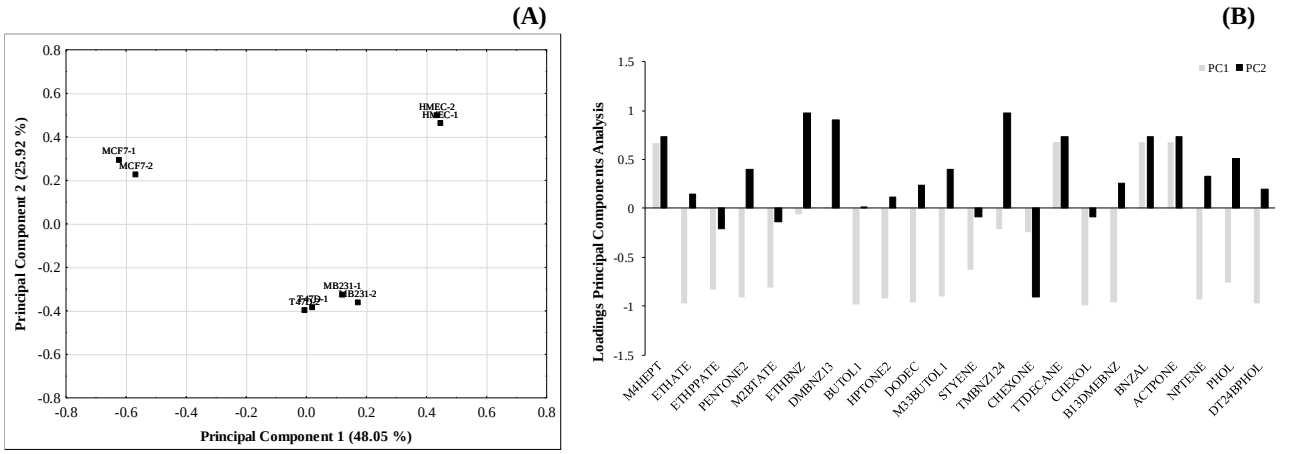

Figure 3. 3. 5 - (A) Separation of breast cancer cell lines and normal cells based on PCA scores scatter plot and (B) Lineplot of principal component values obtained using selected compounds by significance of one-way ANOVA ($p < 0.05$) obtained from the analysis of 4 types.

The PCs values of MCF-7 cell lines were influenced by most of variables used in this test. On the other hand, PCs values of T-47D and MDA-MB-231 cell lines were highly influenced by cyclohexanone, 1,2,4-trimethylbenzene, ethylbenzene and 1,3-dimethylbenzene for PC1, and by cyclohexanone for PC2. The HMEC cells were influenced by 4-methyl-heptane, tetradecane, benzaldehyde and acetophenone for PC1 values, and by 1,2,4-trimethylbenzene, ethylbenzene, 1,3-dimethylbenzene and phenol for PC2 values.

Concerning the other tested conditions, the loading scatterplots of the PCA obtained from the analysis of VOMs from cell culture media at pH 2, pH 7 and pH 10, and respective influence of variables, are showed in Figure 4.3.6. Surprisingly, four groups were formed encompassing all breast cell lines in study under pH 2 and pH 7, where MCF-7 was differentiated from other cell lines mainly across PC1, while HMEC, T-47D and MDA-MB-231 were differentiated from each other through PC2. Under pH 10, the pattern of differentiation between cell lines is similar to obtained for headspace, in which 3 groups (HMEC, T-47D/MDA-MB-231 and MCF-7) were formed. The differentiation between cell lines obtained under pH 2 and pH 7 may be due to the alterations of molecular components released under more acidic conditions than those normally present in the culture medium (pH 7.3).

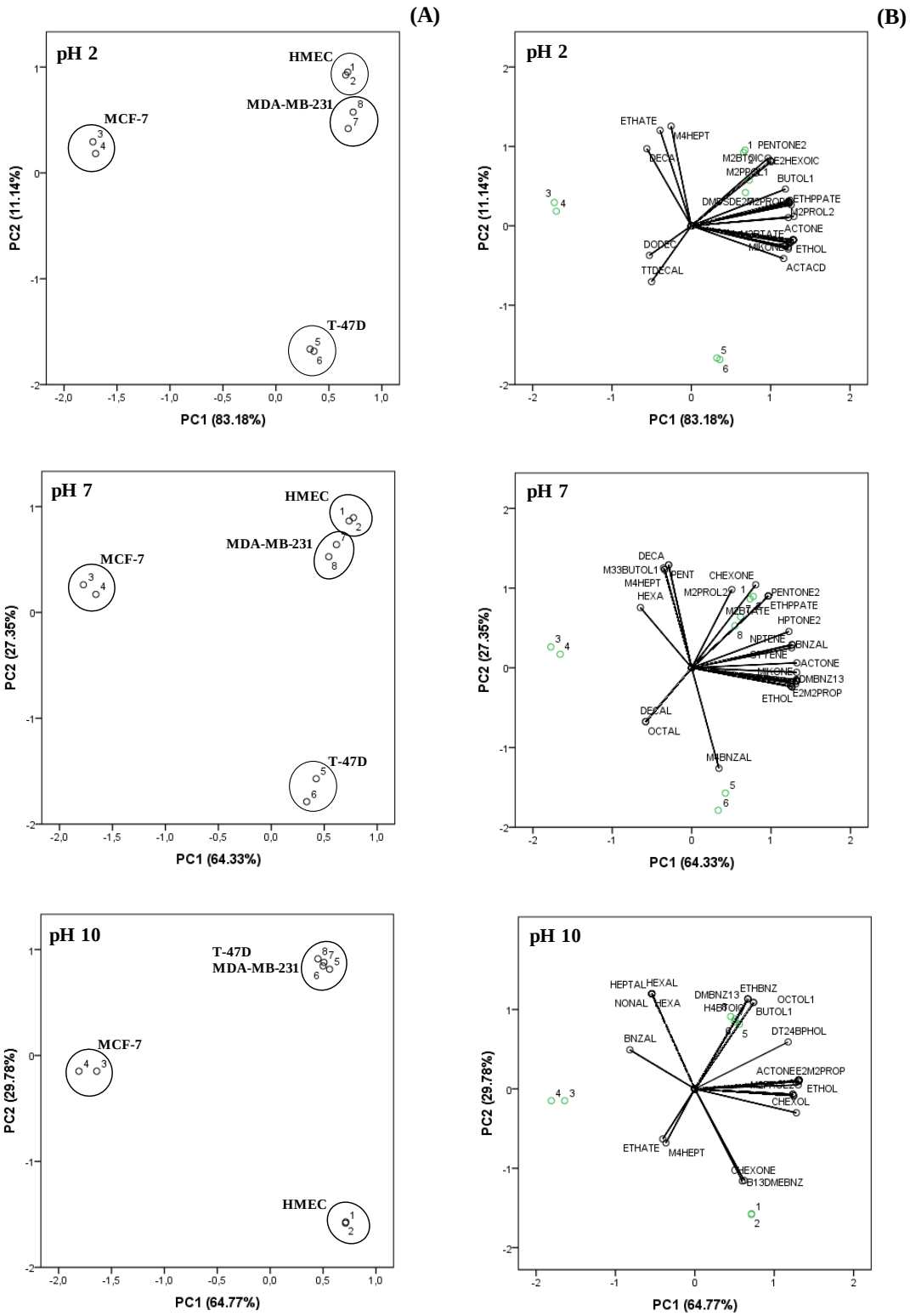

Figure 3. 3. 6 - (A) Loadings of variables on the PC1 × PC2 plane of culture media at different pH conditions and (B) Biplots of data obtained from different conditions using selected compounds by significance of one-way ANOVA ($p < 0.05$) obtained from the analysis of 4 types of breast cell lines.

However, for differentiation and discrimination between normal breast cell lines and oncological breast cell lines based on the VOMs emitted as close to reality as possible in human cell tissues, partial least squares analysis (PLS) and linear discriminant analysis (LDA) were performed only with data from headspace of cell cultures. The statistical data summary of PLS and LDA are described in Tables 3.3.3 and 3.3.4, respectively.

Table 3. 3. 3 - Statistical data summary of Partial Least Square Analysis.

PC	R^2X	R^2X(Cumul.)	Eigenvalues	R^2Y	R^2Y(Cumul.)	Q^2	Q^2(Cumul.)	Significance	Interations
1	0.561557	0.561557	14.60037	0.14286	0.14286	-0.03171	-0.03171	S	9
2	0.306333	0.867891	7.96463	0.14286	0.28571	-0.05208	-0.08543	NS	7

Table 3. 3. 4 - Statistical data summary of Linear Discriminant Analysis.

Function	Eigenvalue	Canonical R	Wilk's Lambda	Chi-Sqr.	df	*p*-value
1	111309.5	0.999996	1.34E-08	81.56097	4.000000	1.11E-16
2	667.2	0.999251	1.50E-03	29.27062	1.000000	6.29E-08

Sample classification by PLS showed that the differentiation between cell lines was explained through one single component. PLS loading lineplot are presented in Figure 3.3.7, which can be observed four centroids corresponding to each cell lines.

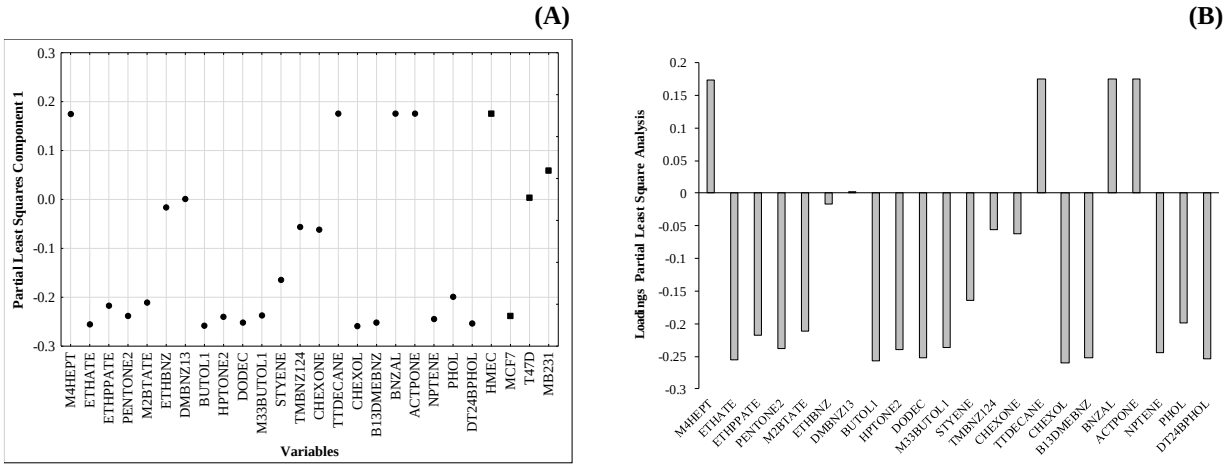

Figure 3. 3. 7 - (A) Partial Least Square Analysis (PLS) scatter plot and (B) Lineplot of selected compounds by significance of one-way ANOVA ($p < 0.05$) obtained from the analysis of 4 types of breast cell lines using cultured headspace analysis.

Similar to obtained in the PCA, HMEC centroid was clearly differentiated from oncologic breast cell lines, and MCF-7 (triple negative type) was distinguished from two luminal A type cell lines. On the other hand, T-47D and MDA-MB-231 remain very close to each other, which PLS values was

0.0588 and 0.0037, respectively. Regarding the influence of variables on PLS values of cell lines, HMEC was highly influenced by 4-methyl-heptane, tetradecane, benzaldehyde and acetophenone, T-47D and MDA-MB-231 were influenced by cyclohexanone, 1,2,4-trimethylbenzene, ethylbenzene and 1,3-dimethylbenzene, and MCF-7 was influenced by the remaining VOMs.

The linear discriminant analysis (LDA) was applied as a supervised pattern recognition method in order to discriminate statistically the cell lines under study, where samples were grouped according to molecular type as follows: N (HMEC), BL (T-47D and MDA-MB-231) and BTN (MCF-7). The LDA scatterplot of cell lines classification according to canonical functions were showed in Figure 3.3.8.

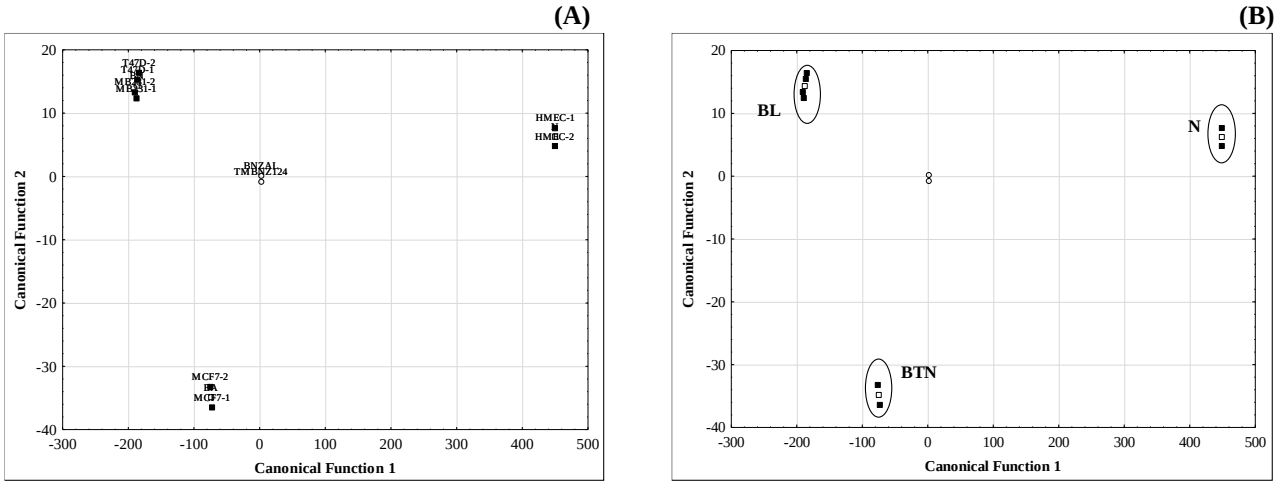

Figure 3. 3. 8 - (A) Linear discriminant analysis (LDA) scatter plot of cultured headspace from breast cell samples (B) Classification of breast cells according to the canonical discriminant functions. Legend: BL-breast luminal; N- normal; BTN- breast triple negative.

The cell lines samples formed three clearly defined groups with a classification rate of 100%. Recognition ability, calculated as the percentage of members of the data set that were correctly classified, and prediction ability, calculated as the percentage of members that were correctly classified, were 100% in all cases. After applying LDA with backward removal ($p < 0.05$) of variables, only two VOMs proved to be significant for discrimination between three defined previously, namely 1,2,4-trimethylbenzene and benzaldehyde. These compounds have been already identified in cancer cell studies by Brunner et al. [212] using PTR-MS, by Filipiak et al. [152] in lung cancer cells and Mochalski et al. [213] with human hepatocellular carcinoma cells where it was observed an increase in the release of this compound. Moreover, these two VOMs appear to be promise biomarkers due to fact that achieve a successful discriminant classification of samples according to molecular type of breast cell lines, demonstrating that volatile metabolomic signature of breast cells can be a useful approach to identify potential BC biomarkers for early diagnosis of BC.

Conclusions

This study demonstrated that HS-SPME/GC-MS is a simple, rapid, sensitive and solvent-free method that can be used to establish the volatile metabolomic patterns of normal and cancer breast cells. In addition, this study showed the potential of screening the *in vitro* VOMs associated with BC to identify potential volatile biomarkers to be used in early diagnosis. The headspace of culture media of normal and cancer cell lines was analyzed at different pH conditions. Sixty VOMs were identified as belonging to several chemical groups: namely, alkanes, aldehydes, ketones, acids, alcohols and benzene derivatives. Most of the identified VOMs are common to all BC cell lines and normal human mammary epithelial cells, but six of them, 2-pentanone, 2-heptanone, 3-methyl-3-buten-1-ol, ethyl acetate, ethyl propanoate, and 2-methyl butanoate, were detected only in the headspace of cancer cell lines. Multivariate statistical data obtained in this study revealed that combining *in vitro* assays with HS-SPME/GC-MS is a useful strategy to differentiate and discriminate the volatile metabolomic signature of normal cells and BC cell lines according to molecular type, thus contributing to the discovery of novel biomarkers of BC and investigations of the related metabolomic pathways, and thereby improving the diagnostic tools for BC.

3.4| Volatomic pattern of breast cancer and cancer-free tissues as a powerful strategy to identify potential biomarkers

(Silva et al.; Analyst **2019***; 144:4153)*

Abstract

BC is ranked as the fifth amongst all cancers, remains at the top of women's cancers worldwide followed by colorectal, lung, cervix, and stomach cancers. The main handicap of most of the screening/diagnostic methods is based on their low sensitivity and specificity and the invasive behavior of most sampling procedures. The aim of this study was to establish the volatomic pattern of BC and cancer-free (CF) tissues (n = 30) from the same patients, as a powerful tool to identify a set of volatile organic metabolite (VOM) potential BC biomarkers which might be used together or complement with the traditional BC diagnostics strategies, through the integration of chromatographic data, obtained by solid-phase microextraction followed by gas chromatography-mass spectrometry (SPME/GC-qMS), with chemometric tools. A total of four metabolites: limonene, decanoic acid, acetic acid and furfural presented the highest contribution towards discrimination of BC and CF tissues (VIP > 1, $p < 0.05$). The discrimination efficiency and accuracy of BC tissue metabolites was ascertained by ROC curve analysis that allowed the identification of some metabolites with high sensitivity and specificity. The results obtained with this approach suggest the possibility of identifying endogenous metabolites as a platform to find potential BC biomarkers and pave the way to investigate the related metabolomic pathways in order to improve BC diagnostic tools. Moreover, deeper investigations could unravel novel mechanistic insights into the disease pathophysiology.

Keywords: Volatile pattern; Breast cancer; Tissue; Metabolomics; Chemometric tools

Introduction

BC is ranked as the fifth amongst all cancers remaining at the top of women´s cancers worldwide followed by colorectal, lung, cervix, and stomach cancers according to IARC, contributing with more than 11.6 % of all cancers [3]. Although BC is a multifactorial disease, with highly variable clinical behavior and response to therapy, it can be curable in early stages. Furthermore, there is still the need for the development of new methodologies to aid or monitor the disease together with the current diagnostic tools, namely mammography, ultrasound or tumor markers. Moreover, before BC treatment, a complex and time-consuming analysis is required that uses many different assays, such as the determination of histological type and grading, evaluation of estrogen receptor (ER), progesterone receptor (PgR) and human epidermal growth factor receptor 2 (HER-2), among others [122]. The main handicap of most of these screening/diagnostic methods is their low sensitivity and specificity and the invasive procedure required to obtain the samples [123]. Taking into account these aspects, research is being directed towards the use of new tools that can support the clinicians in BC treatment and follow-up [124]. In this sense, in recent years, metabolomic studies have emerged as a powerful tool to investigate the changes and/or metabolic responses of living systems to stimuli or genetic modifications [39]. The metabolome profile represents the unbiased quantitative and qualitative analysis of the complete set of metabolites present in cells, body fluids and/or tissues [125]. To date, beyond the most used biological specimens (e.g., urine, saliva, blood), BC tissues have been used in metabolomics with the aim of discriminating cancer from normal tissues suggesting that metabolomic profiles differ within molecular subtypes of BC [66,214]. The metabolome coverage in BC tissues can be maximized by combining different technologies for metabolic profiling, namely gas chromatography mass-spectrometry (GC-qMS), and the results can be used to classify BC helping to identify new prognostic and predictive markers and to discover new targets for future therapeutic interventions [126]. Among them, the study of volatile organic metabolites (VOMs) present in biological samples, namely in saliva, urine, breath and tissues can be useful for cancer diagnosis, in particular for BC. The most common procedure used in extraction of volatile compounds in biological samples is solid-phase microextraction (SPME), normally in headspace mode (HS-SPME) being used in several biologic matrices [44,59,127,128]. GC-qMS was used to screen salivary volatiles for putative BC as an exploratory study involving geographically distant populations [51], also to establish the metabolomic signature of human BC cell lines [44] and to discriminate different types of cancer based on urinary volatomic biosignatures [59]. In the first study, up to 120 VOMs from distinct chemical families, with significant variations among the groups, were identified [51], whereas Silva et al. [44] and Porto-Figueira et al. [59] identified 60 and 130 VOMs in BC cell lines and urine, respectively. On the other hand, Budczies et al. [77,82] used gas

chromatography time of flight mass spectrometry (GC-TOFMS) framework to evaluate the glutamate enrichment as a new diagnostic opportunity in BC and to accomplish comparative metabolomics of estrogen receptor positive (ER^+) and estrogen receptor negative (ER^-) samples. Budczies et al. [82] identified 19 VOMs and the GC–TOFMS based analysis of metabolites present in BC tissues revealed significantly differences in central metabolism in the more aggressive ER^- compared to the ER^+ type. The detected changes included the metabolism of glutamine with a decrease in concentration of glutamine and an increase in concentration of glutamate and 2-hydroxyglutaric acid [82]. In turn, Dougan et al. [66] used GC-MS to evaluate the detectability, reliability, and distribution of metabolites measured in pre-diagnostic plasma samples in a pilot study of women listed in the Northern California site of the BC Family Registry. In this study, 661 VOMs were detected, 338 (51 %) of them were found in all samples, and 490 (74 %) in more than 80 % of samples. The aim of this study was to establish the volatomic pattern of BC tissue and CF tissue samples collected after surgery, from the same patient (to minimize the interference of epigenetic and external factors) in order to find a set of volatile metabolites to be used as potential BC biomarkers, using HS-SPME/GC-qMS combined with multivariate statistical tools. This high-throughput strategy might have the potential to be applied in a clinical environment as a diagnostic approach or as a complementary way with the current diagnostic methods to improve the diagnostic decision.

Materials and Reagents

Reagents and materials

Sodium chloride (NaCl), hydrochloric acid (HCl) and 4-methyl-2-pentanol were supplied by Panreac (Barcelona, Spain) and Sigma Aldrich (St. Louis, MO, USA), respectively. Phosphate buffer saline (PBS) was also purchased from Sigma-Aldrich (St. Louis, MO, USA). The C_8-C_{20} alkanes solution (concentration of 40 mg/L in *n*-hexane) was purchased from Fluka (Buchs, Switzerland). The digital stirring plate (Cimarec™) was supplied by Thermo Scientific (Waltham, MA, USA) while SPME holder for manual sampling and 75 µm carboxen/polydimethylsiloxane fiber (CAR/PDMS) were purchased from Supelco (Bellefonte, PA, USA).

Subjects and tissue collection

To investigate the BC tissue metabolomic profile, 30 samples from patients with breast cancer (BC, *n*=30, age range 44-85, average 67), and 30 samples from cancer-free tissue (CF, *n*=30, age range 44-85, average 67) without malignant infiltration were resected from each patient. The resected samples were divided into the active carcinoma and cancer-free tissue outside the tumor margin and were immediately frozen in liquid nitrogen, in a total set of 60 samples. The tissues were stored at -

80°C until extraction. These samples were obtained at the Pathologic Anatomy Unit of Hospital Dr. Nélio Mendonça (Funchal, Portugal) according to Table 3.4.1.

Table 3. 4. 1 - List of collected tissue samples from breast cancer patients and CF individuals.

Samples	**BC tissue**	**Cancer-free**
Number	30	30
Age (range, median)	(44-85, 65)	(44-85, 65)
Histological grade (number of samples)	IA (5) IIA (10) IIIA (1) IIB (7) IIIB (5) IIIC (2)	Not applicable

All experiments were performed in accordance with the standard Guidelines from Declaration of Helsinki and approved by the institutional ethics committee of Hospital Dr Nélio Mendonça and University of Madeira. All the participants of this study were informed about the investigation and a signed informed consent was obtained from all human participants prior to sample collection. Using the TNM (tumor, node, and metastasis) staging approach, the examined BC cases included five of stage IA, ten of stage IIA, one of stage IIIA B, seven of stage IIB, five of stage IIIB and two of stage IIIC.

Extraction of metabolites from breast tissues

The HS-SPME extraction conditions were based on a developed method previously established in our laboratory [127,128]. Briefly, tissue samples were thawed and then portions of 100 mg were weighted into 20 mL vials together with 17 % NaCl (w/v), 1000 µL of ultrapure water and 100 µL of the internal standard (IS, 4-methyl-2-pentanol, 1.6 mg/L). The pH was adjusted to 2 with small amounts of HCl 5M. Then, the vial was capped with a Teflon (PTFE) septum using a screw cap and the SPME fiber was introduced and exposed into the headspace during 75 min at 50 °C at 800 rpm (0.5 mm × 0.1 mm bar). After this period, the fiber was removed from the vial and inserted into the GC injection port and VOMs extracted were desorbed for 10 min at 250 °C. Each sample was analyzed in duplicate and blanks were performed before each analysis.

Gas chromatography quadrupole mass-spectrometry (GC-qMS) conditions

After the extraction procedure, the SPME fiber with the analytes was inserted into the injection port of an Agilent Technologies 6890N Network gas chromatograph system (Palo Alto, CA, USA) where the VOMs were desorbed at 250 °C for 10 min. The gas chromatograph was equipped with a 60 m × 0.25 mm I.D.× 0.25 µm film thickness, BP-20 (SGE, Dortmund, Germany) fused silica capillary column and interfaced with an Agilent 5975 quadrupole inert mass selective detector. The following oven temperature profile was set: (a) 5 min at 45 °C; (b) increase temperature until 150 °C, at a rate of 2 °C min^{-1} (hold for 10 min); (c) 150 °C for 10 min; (d) increase temperature until 220 °C, at a rate of 7 °C min^{-1}; and (e) 220 °C for 10 min for a total GC run time of 87.5 min. The column flow was constant at 1.3 mL min^{-1} using Helium (He, N60, Air Liquide, Portugal) as carrier gas. The injection port was operated in the splitless mode and held at 250 °C. For the 5975 MS system, the operating temperatures of the transfer line, quadrupole and ionization source were 270, 150 and 230 °C, respectively, while electron impact mass spectra were recorded at 70 eV ionization voltage and the ionization current was 10 µA. Data acquisition was performed in the scan mode (30–200 m/z). The electron multiplier was set to the auto tune procedure. Metabolites identification was accomplished by manual interpretation through single ion monitorization (SIM) of spectra and matching against the Agilent MS ChemStation Software, equipped with a NIST05 mass spectral library with a similarity threshold higher than 80% and comparison with commercially available standard samples when available. A series of C_8–C_{20} n-alkanes were analyzed using the same extraction procedure to establish the kovat indices (KI), and to confirm the identity of the VOMs by comparison with the literature. The analyzes were performed in triplicate and the results expressed by mean ± standard deviation.

Statistical Analysis

Statistical analysis was performed using the web server Metaboanalyst 4.0 [134]. The multivariate statistical analysis, namely the principal component analysis (PCA), the partial least squares-discriminant analysis (PLS-DA) and the orthogonal projections to latent structures discriminant analysis (OPLS-DA) were applied on tissue metabolomic profile dataset to provide insights into the groups under study. The metabolites with VIP scores higher than 1.0 were selected by the PLS-DA analysis as well as for the pathway analysis. Furthermore, hierarchical cluster analysis by K-means of the two groups in study was carried out and Pearson´s correlation was used to build the heat map with the aim of identifying BC clustering patterns. Moreover, Random forest (RF) classification was carried out to determine the ability of VOMs to accurately classify the study subjects into their corresponding groups. The receiver operating characteristic curves (ROC) were attained to verify

which metabolites had the highest sensitivity/specificity for a potential BC diagnosis. Finally, the selected metabolites were used for the metabolic pathway analysis to identify the most relevant metabolic pathways involved in BC.

Results and Discussion

Tissue metabolomic pattern based on GC-qMS

A total of twenty-nine VOMs were identified in BC tissue and CF tissue samples which were classified in several chemical families, namely phenols, benzene derivates, carbonyl compounds, acids, alcohols and furanic compounds as presented in Figure 3.4.1.

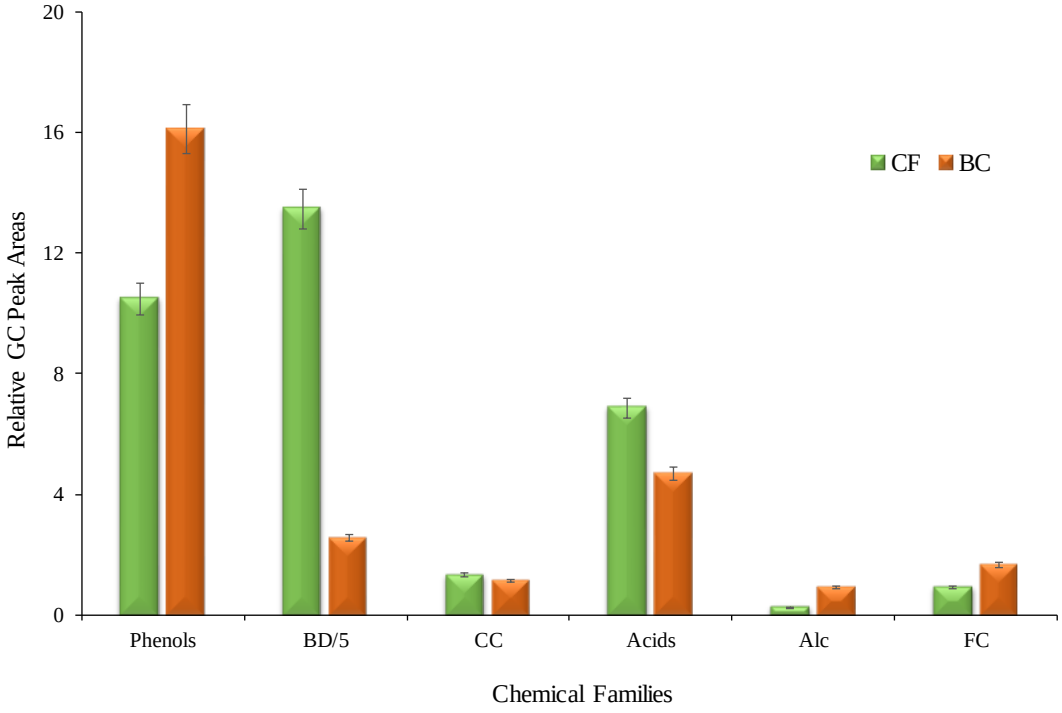

Figure 3. 4. 1 - Major chemical families identified in BC and CF tissue samples. Legend: BD-benzene derivates; CC-carbonyl compounds; Alc-alcohols; FC-furanic compounds.

Data was processed using software (NIST, 2005; Mass Spectral Search Program V.2.0d) which provides quality matching using advanced spectral matching algorithms background subtraction and KI comparison. The KI were determined through the injection of a C_8-C_{20} alkane solution in order to confirm the identity of the VOMs by comparison with the literature (e.g., pherobase). It can be observed in Table 3.4.2 that the KI of the identified VOMs, when compared to the ones found in literature for a similar column (BP-20), are closer to one another, ensuring that a good identification was achieved. The highest contribution for the volatile profile was from phenols in the BC group, whereas for the CF group was from benzene derivates and acids. The main VOMs identified for these

families were phenol, toluene and acetic acid, respectively. These metabolites were already identified in several reports in literature in biological matrices such as urine [127,128], cancer cell lines [44,135], saliva [51], and exhaled breath [135,136]. In Table 3.4.2 is represented the identification of VOMs, as well as the minimum and maximum relative GC peak areas for each metabolite and group.

It can be observed that most metabolites were identified in all samples with a frequency of occurrence (FO) higher than 90 %, where the maximum relative area was obtained for *p*-tert-butyl-phenol in the CF group and phenol for the BC group. For the CF group, the major VOMs identified include acetic acid, phenol and *p*-tert-butyl-phenol.

Some of these have been already reported in literature, namely acetic acid was found as discriminant when associated with CTL in a study conducted by Ahmed et al. [137] that investigated the possibility of faecal VOMs as potential diagnostic biomarkers for inflammatory bowel disease. Mochalski et al. [206] investigated the emission of volatile compounds from gastric cancer tissues and non-cancerous tissues with the aim of identifying characteristic chemical patterns associated with gastric cancer.

Furthermore, Silva et al. [127,128] also reported phenols as the major chemical family identified in urine from the oncologic group. Furthermore, Raman et al. [138] studied faecal VOMs in obese humans and identified also acetic acid and phenol as major metabolites. On the other hand, Priscilla et al. [59] also identified *p*-xylene, *o*-xylene, acetic acid, phenol and *p*-tert-butyl-phenol in urine from cancer patients, with some of them presenting higher values in cancer patients. Regarding the relative area values obtained for the BC group, most VOMs were down-regulated with exception of acetic acid, toluene, octanal and nonanal.

Table 3. 4. 2 - Identified metabolites in tissue samples from BC patients and cancer-free (CF) through GC-qMS, minimum (Min) and maximum (Max), variation of relative peak areas (relative to internal standard, RSD<10%)regarding to BC group and their frequency of occurrence FO (in %). Identification mode and the Kóvats index for each identified VOM and the literature values for a similar GC column.

Peak n.	IEC[a]	RT (min)[b]	ID[c]	KI_{cal}[d]	KI_{lit}[e]	VOM[f]	Relative GC Peak Areas				Variation	FO (%)
							CF		BC			
							Min	Max	Min	Max		
A1	41, 69	9.77	MS	990	983	methyl isobutyrate	0.03	1.81	0.02	1.66	↓	71
A2	91	11.24	Std, MS	1029	1042	toluene	0.03	7.62	0.01	17.17	↑	100
A3	44, 56	13.12	Std, MS	1073	1075	hexanal	0.03	6.82	0.04	2.09	↓	99
A7	41, 70	18.48	Std, MS	1172	1168	heptanal	0.04	0.17	0.05	0.19	↑	13
A8	68, 93	18.73	Std, MS	1176	1198	limonene	0.01	4.46	0.01	1.03	↓	92
A10	43, 33	19.84	Std, MS	1187	-	2-methyl-1-propanol	0.01	1.44	0.01	0.51	↓	79
A11	81	20.69	MS	1197	1229	2-pentyl-furan	0.01	0.07	0.01	0.13	↑	77
A12	105, 120	21.39	MS	1209	-	trimethylbenzene	0.01	5.41	0.01	1.70	↓	64
A13	119, 134	23.18	Std, MS	1237	1234	*o*-cymene	0.01	0.84	0.01	0.33	↓	24
A14	42, 55	24.36	MS	1254	1255	1-pentanol	0.01	0.03	0.01	0.23	↑	50
A15	43, 56	24.90	Std, MS	1262	1280	octanal	0.01	0.24	0.01	9.19	↑	34
A16	56	30.20	Std, MS	1338	1360	1-hexanol	0.01	0.05	0.01	1.17	↑	27
A17	41, 57	31.89	Std, MS	1362	1385	nonanal	0.02	0.48	0.17	2.36	↑	52
A18	117, 132	34.45	MS	1396	-	*p*-cymenene	0.01	0.76	0.01	0.15	↓	100
A19	43, 60	36.42	Std, MS	1428	1450	acetic acid	0.62	42.75	0.10	12.28	↓	100
A20	96	36.61	Std, MS	1431	1455	furfural	0.01	5.87	0.04	18.03	↑	100
A21	57	38.53	Std, MS	1461	1487	2-ethyl-1-hexanol	0.01	0.50	0.01	3.20	↑	2
A22	43, 57	38.77	Std, MS	1465	1484	decanal	n.d.[g]	n.d.	0.14	0.15	-	72
A23	77, 106	39.81	Std, MS	1480	1495	benzaldehyde	0.01	0.36	0.01	0.33	↓	54
A24	41, 56	42.63	MS	1525	1553	1-octanol	0.01	0.17	0.01	1.57	↑	74
A25	77, 105	47.36	MS	1602	1607	acetophenone	0.01	0.19	0.01	0.42	↑	79
A26	60, 73	58.70	Std, MS	1802	1829	hexanoic acid	0.01	1.25	0.02	5.77	↑	100
A27	94	69.54	Std, MS	1968	1965	phenol	0.51	23.32	0.45	44.08	↑	39
A28	60, 73	72.22	Std, MS	2004	2013	octanoic acid	0.01	0.25	0.01	6.51	↑	76

Peak n.	IEC[a]	RT (min)[b]	ID[c]	KI_{cal}[d]	KI_{lit}[e]	VOM[f]	Relative GC Peak Areas				Variation	FO (%)
							CF		BC			
							Min	Max	Min	Max		
A29	107	73.00	MS	2009	2017	4-methyl-phenol	0.01	1.08	0.01	3.72	↑	45
A30	60, 73	75.65	Std, MS	2027	-	nonanoic acid	0.01	0.18	0.01	0.64	↑	29
A31	155, 170	75.71	MS	2027	-	1,6,7-trimethyl-naphthalene	0.02	0.41	0.01	0.37	↓	39
A32	60, 73	78.15	Std, MS	2043	-	decanoic acid	0.01	0.11	0.02	0.52	↑	100
A33	135, 150	78.49	Std, MS	2045	-	*p-tert*-butyl-phenol	0.02	53.76	0.03	20.71	↓	100

[a] IEC - ion extraction chromatogram; [b] Retention time (min); [c] metabolite identification using a standard compound (St) or mass spectra of the NIST library search (MS); [d] Kovat index relative n-alkanes (C_8–C_{20}) on a BP-20 capillary column; [e] Kovat index relative reported in literature for equivalent capillary column; [g] n.d.- not detected

Multivariate statistical analysis of tissue metabolomic profile

The statistical analysis was performed using the Metaboanalyst 4.0 [134] web server as described in the experimental section. Only the VOMs with FO higher than 90 % were considered for the statistical analysis, in a total of 8 VOMs (toluene, hexanal, limonene, *p*-cymenene, acetic acid, furfural, hexanoic acid and decanoic acid). Initially, data were transformed by log transformation and mean centering approaches, before being subjected to multivariate statistical analysis. Partial least square-discriminant analysis (PLS-DA) was used as a supervised clustering method to verify the existence of an altered metabolite pattern. Additionally, this type of statistical analysis takes into account the variance/covariance between samples of groups where the samples are classified into different groups. Regarding the results obtained, a good discrimination was achieved (80.5 %, total variance) between BC and CF tissue samples suggesting the occurrence of characteristic metabolic alterations in the groups under study (Figure 3.4.2 A). Then, the top four metabolites (limonene, decanoic acid, acetic acid and furfural) with the highest contribution for group discrimination were selected with variable importance in projection (VIP > 1) (Figure 3.4.2 B).

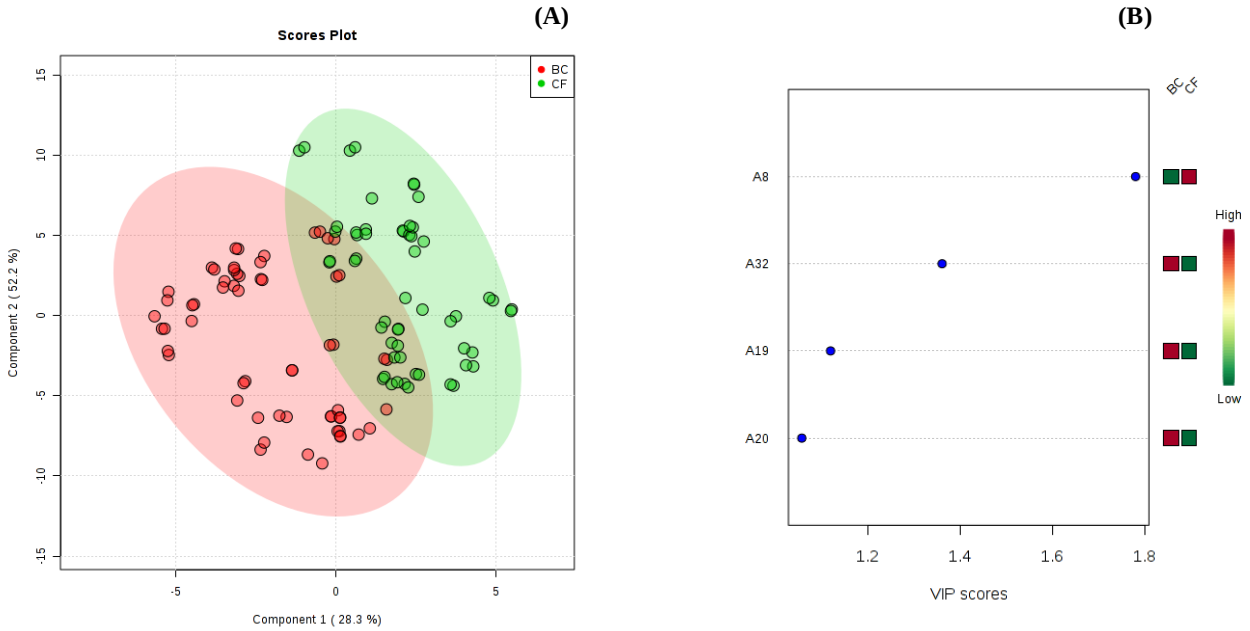

Figure 3. 4. 2 - (A) Loading score plots of PLS-DA and (B) VIP scores of tissue samples from BC and CF subjects. For identification please see Table 3.4.2.

Moreover, orthogonal partial least squares-discriminant (OPLS-DA) analysis was applied on tissue metabolomic profile dataset to maximize the separation of BC and CF groups. Significant group separation was observed in OPLS-DA score plot between BC patients and CF group indicating intrinsic metabolic alterations in each group (Figure 3.4.3 A). To attest the robustness of the model, a random permutation test with 1000 permutations was performed with OPLS-DA (Figure 3.4.3 B). The permutation test yielded R^2 (represents goodness of fit) as 0.717 and Q^2 (represents predictive

ability) as 0.691 indicating that the model is not over fitted and have a relative good predictive ability to distinguish between study groups.

Although in literature there are no reports regarding the volatile profile of human tissues, many of these metabolites were already identified in previous reports, namely in lung [150,186], breast, ovarian [123,144], bladder [187], and gastric cancers [188]. The OPLS-DA uses class information allowing to show which variables are responsible for class discrimination using the predictive information of the first component. The main advantage of OPLS-DA when compared to PLS-DA is that a single component is used as a predictor for the class, while the other components describe the variation orthogonal to the first predictive component [139].

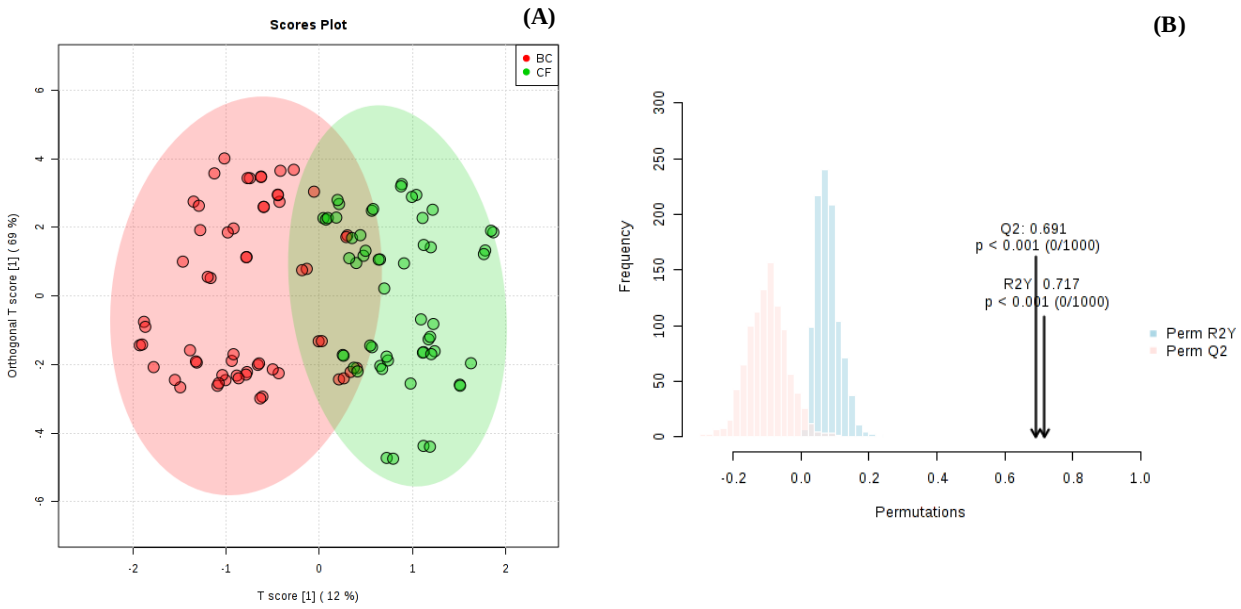

Figure 3. 4. 3 - (A) Loading score plots of OPLS-DA and (B) model validation by permutation test based on 1000 permutations of VOMs obtained by GC-qMS of tissue samples from the 2 groups under study.

To further evaluate the predictive value of the metabolites to discriminate between BC patients and CF, a receiver operating characteristic curve (ROC) analysis was generated using the top four metabolites identified by VIP values (Figure 3.4.4 A and B). This type of analysis is used for the classification of true positives and false positives and the predictive ability is measured using the area under the curve (AUC) [140,141]. According to Xia et al. [108], an AUC between 0.9 - 1.0 is excellent and between 0.8 – 0.9 is good. Based on this classification, the results obtained were, thus, very good (AUC = 0.966).

ROC curves were generated by Monte-Carlo cross validation (MCCV) using balanced sub-sampling. In each MCCV, twothirds (2/3) of the samples were used to evaluate the feature

importance. The top 3 important features were then used to build classification models which were then validated on 1/3 of the samples that were left out. This procedure was repeated multiple times to calculate the performance and confidence interval of each model. The AUC can be interpreted as the probability that a randomly selected diseased subject is ranked as more probable to be diseased than a randomly selected healthy subject [108]. A greater AUC value indicates the effectiveness to separate the CF group from the cancer group (BC).

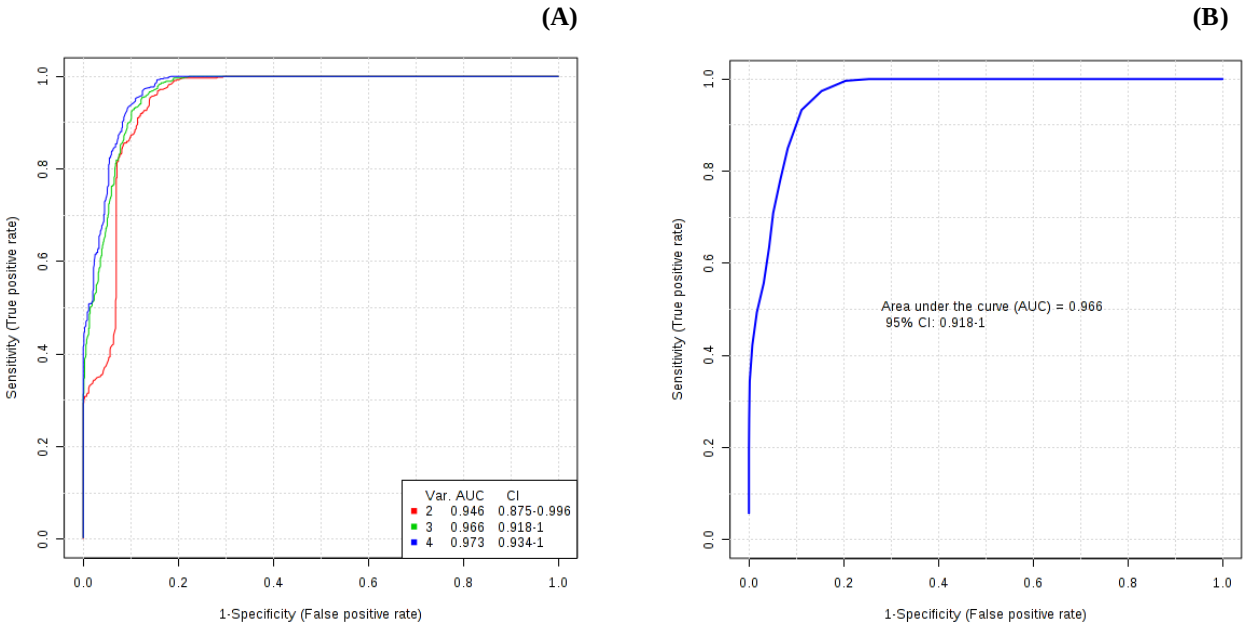

Figure 3. 4. 4 - ROC curves for the predictive model. (A) A combination metabolites model calculated from the logistic regression analysis using the 4 metabolites selected by the VIP (> 1.0) values, (B) ROC curve for the top 4 metabolites (limonene, decanoic acid, acetic acid and furfural) with highest ability to discriminate BC patients against the CF.

Additionally, a 10-fold cross validation was used to generate a logistic regression model and the performance was calculated according to the equation:

log it (P) = log $(P/(1 - P))$ = – 0.077 + 0.966 limonene – 0.752 acetic acid – 0.076 furfural – decanoic acid

where P *is* Pr(y = 1|x). The threshold (or cut-off) for the predicted P was 0.61. Figure 3.4.5 A and B show the results obtained for the predicted probabilities using the OPLS-DA model and the average of the predictive accuracy for the same model, where it can be observed that the model allowed a good classification of samples (>90%).

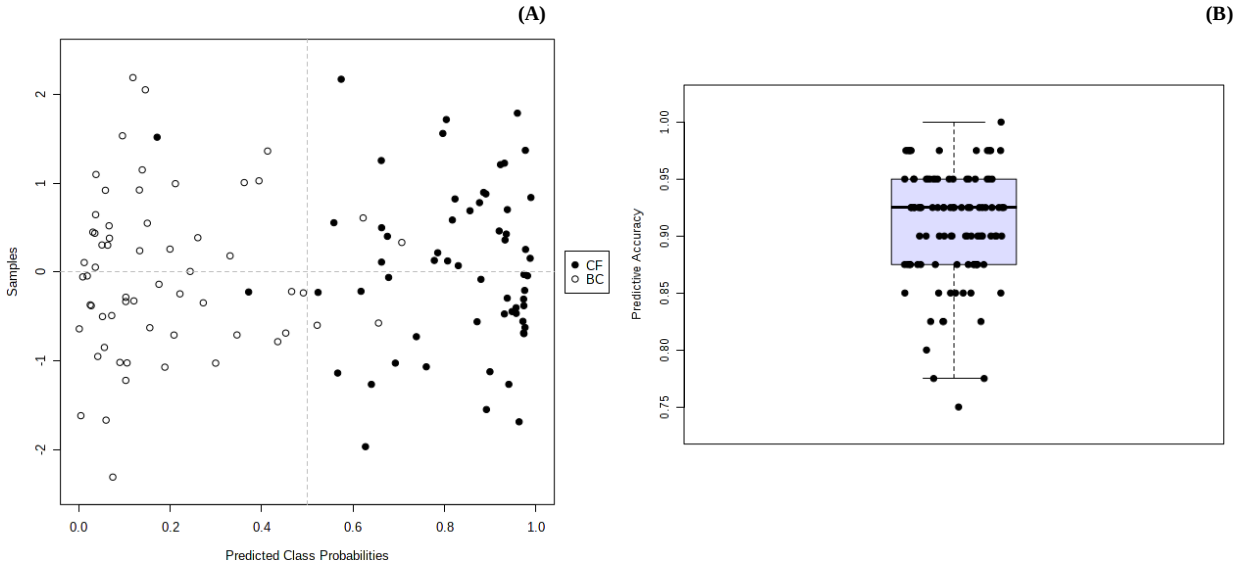

Figure 3. 4. 5 - (A) Plot of the predicted class probabilities for all samples using the OPLS-DA biomarker model based on AUC and (B) box plot of the predictive accuracy (with an average of 0.908) of the biomarker model based on 100 cross validations.

Moreover, 20 samples without known labels were processed together with the ones with known labels in order to obtain the probability of class labels. Most of the cases were classified in their respective groups with the exception of O10, O5 and O18 with a probability score ranging from 0.982 to 0.595.

Also, the heat map was constructed with selected VOMs by VIP > 1, using Pearson's correlation, providing intuitive visualization of the data set and the correlations between samples and VOMs (Figure 3.4.6).

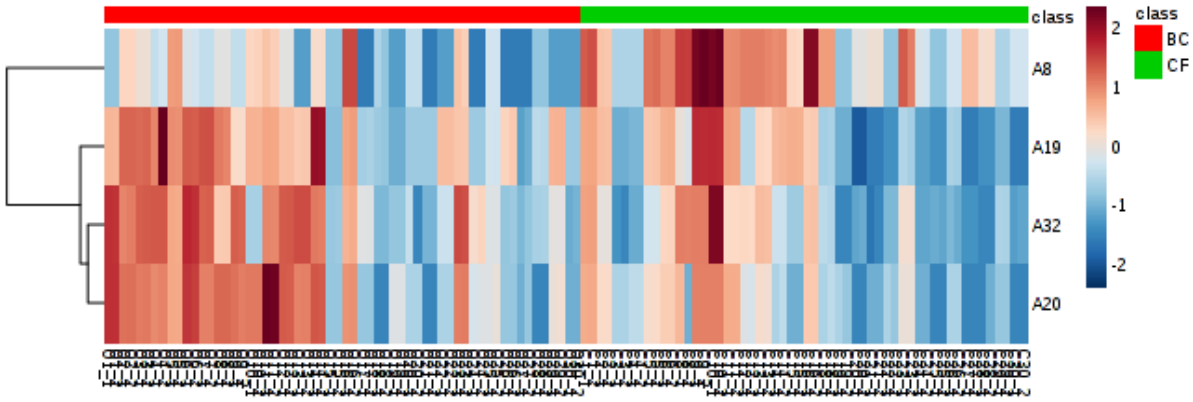

Figure 3. 4. 6 - Heat map visualization and hierarchical clustering analysis using the four metabolites with significance ($p < 0.05$) by Pearson's distance analysis (A8 – limonene; A19 – acetic acid; A32 – decanoic acid: A20 – furfural).

Finally, the metabolic pathway analysis (Figure 3.4.7 A) was performed in order to explore which pathways changed in BC. In Figure 3.4.7 B, we can observe the utmost important metabolomics routes producing the most significant volatile metabolites in BC (indicated by KEGG accession number). It can be observed that the pathway with highest impact was the pyruvate and sulfur metabolism due to acetic acid. This metabolite according to human metabolome database is normally found in most tissues (liver, kidney, among others) and in several biofluids, namely saliva [51] and urine [127]. Moreover, limonene was also included as a significant metabolite belonging to the class of monoterpenoids involved in monooxygenase activity through cytochrome P450 and the mevalonate pathway [215,216].

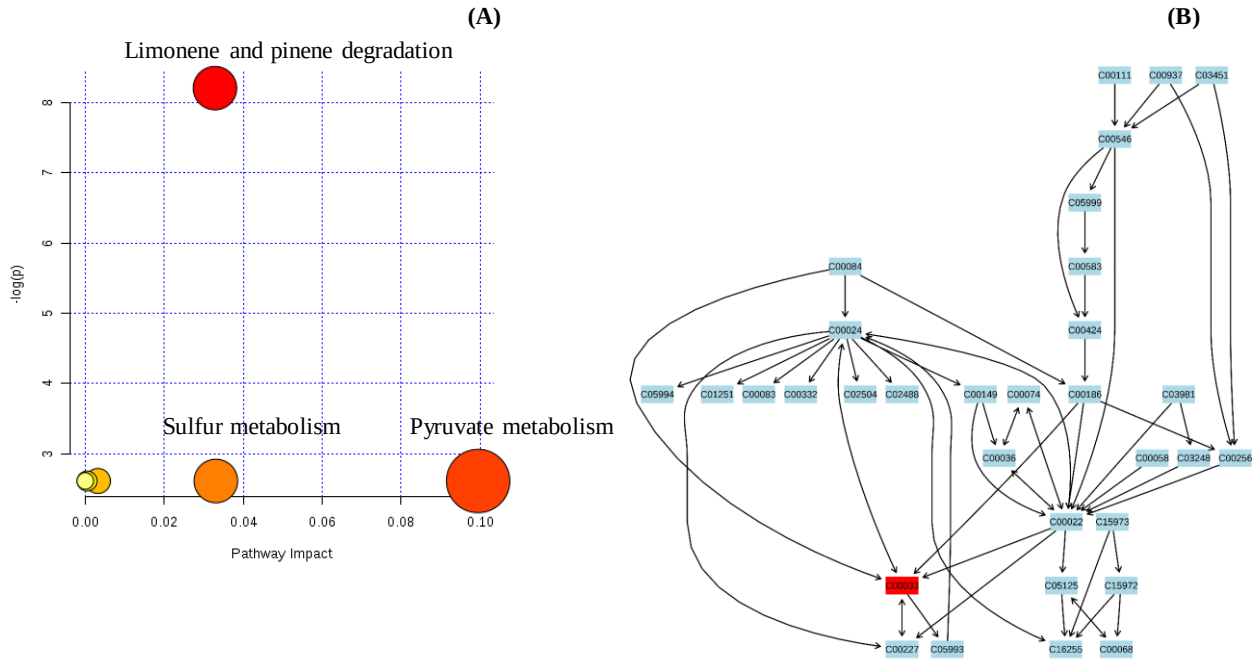

Figure 3. 4. 7 - (A) The metabolome view map of significant altered metabolic pathways observed in tissue samples from BC and CF groups and (B) the pyruvate metabolism. The map was generated using reference map by KEGG; C00033 represent acetic acid.

Based on the results obtained, a successful discrimination of tissue samples was achieved, according to the group showing that the volatomic tissue profile can be a useful approach to identify potential BC biomarkers.

Conclusions

This study has enabled the untargeted assessment of the metabolomic tissue profile from BC patients when compared with CF tissues using GC-qMS combined with multivariate statistical tools (PLS-DA and OPLS-DA). Twenty-nine metabolites were identified and multivariate statistical

analysis revealed some metabolites significantly altered in BC patients. Of the metabolites identified, limonene, decanoic acid, acetic acid and furfural showed the highest sensitivity and specificity to discriminate BC patients. Also, the analysis of the plots leads to a metabolomic pattern comprising an array of some biochemical pathways altered in BC patients. The metabolic pathway analysis indicated that the discriminatory metabolites could originate from several dysregulated pathways in BC such as those involved in pyruvate and sulphur metabolism, and limonene degradation.

In addition, the analysis of cancer and non-cancer tissue from the same subject can also aid to balance the effect of external interferents, such as diet or environmental exposure and help to identify potential biomarkers. These results suggest a possibility of identifying endogenous metabolites as a platform to discover potential BC biomarkers and paves a way to investigate the related metabolomic pathways to improve the diagnostic tools of BC.

3.5| Integrated metabolomics based on GC-MS and NMR data as powerful strategy to search potential breast cancer biomarkers

(Submitted to European Journal of Cancer, EJC-D-19-00135)

Abstract

BC is leading at the top of women´s diseases, and as a multifactorial disease, there is the need for the development of new approaches to aid clinicians on monitoring BC treatments. In this sense, metabolomic studies have become an essential tool allowing the establishment of interdependency among metabolites in biological samples. The goal is to capture the main changes and the overall physiological status in biochemical pathways to enlighten sites of perturbations in diseases, such as cancer. In this study, we used the combination of NMR and GC-qMS based metabolomic analyzes of urine and breast tissue samples from BC patients and cancer-free individuals. The data was processed using multivariate statistical tools to obtain a panel of metabolites (and their metabolic profiles) that can discriminate malignant from healthy status assisting in the diagnostic field.

Tissue and urine samples were collected from BC patients (urine: *n*=31; tissue *n*=30) and cancer-free (CF) subjects (urine *n*=40 tissue *n*=30) and analyzed by NMR and GC-qMS methodologies. The OPLS-DA model showed a clear separation between BC patients and controls for both class of samples. Namely, for urine samples, the goodness of fit was $R^2Y = 0.946$ and predictive ability was $Q^2 = 0.910$ while for tissue the $R^2Y = 0.888$; $Q^2 = 0.813$ had a good predictable accuracy. The discrimination efficiency and accuracy of BC tissue and urine metabolites was ascertained by ROC curve analysis that allowed the identification of some metabolites with high sensitivity and specificity. The metabolomic pathway analysis identified several dysregulated pathways in BC patients, including those related with pyruvate and glutamine metabolism. Moreover, possible correlations between urine and tissue metabolites were investigated and five metabolites were found significant using a dual platform approach. Overall, this study suggests that an improved metabolic profile combining NMR and GC-qMS may be useful to achieve more insights regarding the mechanisms underlying cancer.

Keywords: Breast cancer; Tissue; Urine; NMR; GC-qMS; Metabolomics; Chemometric tools.

Introduction

Up to date, BC is leading at the top of women´s diseases, accounting with around 2.1 million cases diagnosed in 2018 and expected to increase 1.1 million by 2040, according to IARC [3]. Being a multifactorial disease, with highly variable clinical behavior and response to therapy, it can be curable when detected in early stages. Although the extensive investigation on new therapy targets and diagnosis, there is the need for the development of new approaches to aid clinicians to monitor BC treatments and follow-up together with the current diagnostic tools, namely mammography, ultrasound or tumor markers [124]. Nowadays, the available diagnostic tools have supported in BC early detection leading to the improved of survival rates. In this regard, metabolomics (metabolomic profiling) studies along with the related area of metabonomics emerged as powerful approaches to study the metabolic changes in several diseases including cancer [67,217] involving a comprehensive analysis of all metabolites present in biological systems (metabolome) [107,218,219]. Indeed, metabonomics refers to the "quantitative measurement of the multiparametric metabolic responses of living systems to patho-physiological stimuli or genetic modifications" [220] and represents a subset of metabolomics. So, whereas metabolomics is the analysis of the biochemical profile of an organism under normal conditions, metabonomics is focused on the analysis of the changes in that biochemical profile under diseased conditions or genetic modifications. Moreover, metabolomics plays an important role in disease profiling being a promising approach for the pursuit of new biomarkers in biological matrices, such as cell extracts, tissues or biological fluids. As a fast growing field that focus on the investigation of metabolites present in biological systems, it reflects the altered metabolism and the physiological status [125]. Normally these studies are combined with analytical techniques being the most popular NMR spectroscopy and MS that have gained attention in this field as strong tools for the identification of potential biomarkers in a variety of clinical fields [161,217]. Usually, MS includes a separation stage LC or GC and can discriminate between compounds based on mass-to-charge (m/z) ratio in charged particles. Regarding NMR, it is an appealing technique that allows the investigation of metabolism due to its advantages as non-destructive, non-invasive, highly reproducible, giving information about biological samples environment offering both qualitative and quantitative measure. When compared with NMR, MS exhibits a greater sensitivity, although sample preparation is laborious and dependent on metabolite chemical properties [155]. On the other hand, MS lacks accuracy and precision producing an enhanced resolution profile with several peaks. In addition, these platforms can be used together, often applied to complex samples and coupled with advanced chemometric tools integrating the datasets obtained by the analytical techniques. If used in combination, these techniques enable the identification of a more comprehensive panel of metabolites involved in metabolic alterations and help unraveling the possible correlations and the underlying

mechanisms induced by a disease [217,221,222]. A variety of studies have been conducted by NMR and DART-MS using this approach to find serum biomarkers for BC [67]. Another study was developed by Chen et al. [217] that used the dual platform of NMR and MS methods to establish the urinary metabolomic profile of bipolar disorder (BD) subjects with a diagnosis purpose. Marshall et al. [223] combined direct infusion electrospray ionization mass spectrometry (DI-ESI-MS) and NMR to analyze the impact of neurotoxins involvement in dopaminergic cell death which is relevant to Parkinson´s disease. Wei et al. [71] used NMR, LC–MS and statistical analysis to predict the response to chemotherapy in the neoadjuvant setting using serum samples from 28 patients with breast cancer. In turn, Falegan et al. [224] used urine and serum samples from renal cell carcinoma (RCC) patients with the aim of distinguishing between stages of the disease and also to make a distinction between benign renal tumors and RCC. In summary, multiplatform approaches are useful tools to achieve a comprehensive analysis of the broad variety of metabolic alterations in cancer progression and development.

The current work describes the combination of NMR and GC-qMS based metabolomic analyzes of urine and breast tissue samples from BC patients and cancer-free individuals combined with multivariate statistical tools in order to obtain a panel of metabolites (and their metabolic profiles) that can discriminate malignant from healthy status, thus assisting in the diagnostic field.

Materials and Methods

Reagents and materials

All solvents and chemicals used in the experimental work were analytical grade.

For the NMR analysis, 3-(trimethylsilyl)propionic-2,2,3,3-*d4* acid sodium salt (TSP) and deuterium oxide (D_2O) were supplied by Acros Organics (Geel, Belgium) while potassium dihydrogen phosphate (KH_2PO_4), sodium azide (NaN_3), phosphate buffer solution (PBS), potassium deuteroxide solution (KOD) were purchased from Panreac (Barcelona, Spain) and Sigma Aldrich (St. Louis, MO, USA), respectively. Methanol (MeOH) and chloroform ($CHCl_3$) were purchased from Thermo Fisher Scientific (Waltham, Massachusetts, USA).

For the GC-qMS analysis, sodium chloride (NaCl), hydrochloric acid (HCl) and 4-methyl-2-pentanol (internal standard, IS) were supplied by Panreac (Barcelona, Spain) and Sigma Aldrich (St. Louis, MO, USA), respectively. The SPME holder for manual sampling together with 75 µm carboxen/polydimethylsiloxane (CAR/PDMS) fiber was purchased from Supelco (Bellefonte, PA, USA).

Subjects and sample collection

Urine samples

Urine samples (first urine morning) from BC patients (*n*=30) were collected at the Haemato-Oncology Unit from Dr. Nélio Mendonça Hospital, while the urine collection from CTLs (*n*=40) (Table 3.5.1) was carried out in Blood Transfusion Medicine Service in the same Hospital.

Table 3. 5. 1 - List of collected urine samples from breast cancer BC patients and CTLs.

Sample group	N. subjects	Age range/years	Mean Age
Breast Cancer (BC)	*n* = 30	44-85	67
Control (CTL)	*n* = 40	43-80	64

Participants were instructed to collect the first urine morning (after the rejection of the first urine stream) into a sterile cup. The collected urine from either patients or healthy volunteer were aliquoted into 4 mL glass vials and stored at -80 °C for further analysis by GC-MS and NMR. Prior to analysis, all urine samples were centrifuged at 4000 rpm for 20 min at 4 °C, and the supernatant used for the analysis. All the analyzes were performed in triplicate. The research was approved by the Ethics Committee of Funchal Central Hospital Dr. Nélio Mendonça (Approval no. S.1708625/2017), and have been performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. All the participants were fully informed of the objectives of the study and signed the informed consent.

Tissue samples

Regarding the tissue samples, 30 samples from patients with breast cancer (BC, *n*=30, age range 44-85, average 67), and 30 samples from cancer-free tissue (CF, *n*=30, age range 44-85, average 67) without malignant infiltration were resected from each patient. The resected samples were divided into the active carcinoma and cancer-free tissue outside the tumor margin and were immediately frozen in liquid nitrogen, in a total set of 60 samples. The tissues were stored at -80 °C until extraction. With regard to tissue, the analysis of cancer and non-cancer tissue from the same subject can also aid to balance the effect of external interferents, such as diet or environmental exposure. These samples were obtained at the Pathologic Anatomy Unit of Hospital Dr. Nélio Mendonça (Funchal, Portugal) according to Table 3.5.2.

Table 3. 5. 2 - List of collected tissue samples from BC patients and CF individuals.

Samples	**BC tissue**	**Cancer-free**
Number	30	30
Age (range, median)	(44-85, 65)	(44-85, 65)
Histological grade (number of samples)	IA (5) IIA (10) IIIA (1) IIB (7) IIIB (5) IIIC (2)	Not applicable

The resected BC tissues were classified using the TNM (tumor, node, and metastasis) staging approach which included five cases of stage IA, ten cases of stage IIA, one case of stage IIIA B, seven cases of stage IIB, five cases of stage IIIB and two of stage IIIC. The research was approved by the Ethics Committee of Funchal Central Hospital Dr. Nélio Mendonça (Approval no. S.1708625/2017), and have been performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. All the participants were fully informed of the objectives of the study and signed the informed consent.

Sample preparation

Urine samples by HS-SPME/GC-qMS

Prior to HS-SPME, urine samples were thawed and then 4 mL of urine was placed into 8 mL vials together with 17 % NaCl (w/v) and 100 µL of the, 4-methyl-2-pentanol (IS = 1.6 mg/L). The pH was adjusted to 2 with small amounts of HCl 5M. Then, the vial was capped with a Teflon (PTFE) septum using a screw cap and the CAR/PDMS fiber was introduced and exposed into the headspace during 75 min at 50 °C at 800 rpm (0.5 mm × 0.1 mm bar). After this period, the fiber was removed from the vial and inserted into the GC injection port and the extracted VOMs were desorbed for 10 min at 250 °C. Each sample was analyzed in triplicate and blanks were performed before each analysis.

Urine samples by NMR

Before NMR analysis, urine samples were thawed and centrifuged (8000 rpm for 5 min) to remove any suspended cells and other precipitated material. Then, 540 µL of urine was mixed with 60 µL of

a buffer solution (KH_2PO_4, 1.5 M in D_2O) containing 0.1 % of TSP-d4 (used as chemical shift reference) and 2 mM NaN_3. The pH was adjusted to 7.00 ± 0.02 by adding small amounts of KOD.

Tissue samples by HS-SPME/GC-qMS

Tissue samples were thawed and then portions of 100 mg were weighted into 20 mL vials together with 17 % NaCl (w/v), 1000 µL of ultrapure water and 100 µL of the 4-methyl-2-pentanol (IS = 1.6 mg/L). The pH was adjusted to 2 with small amounts of HCl 5M. Then, the vial was capped with a Teflon (PTFE) septum using a screw cap and the SPME fiber was introduced and exposed into the headspace during 75 min at 50 °C at 800 rpm (0.5 mm × 0.1 mm bar). After this period, the fiber was removed from the vial and inserted into the GC injection port and VOMs extracted were desorbed for 10 min at 250 °C. Each sample was analyzed in duplicate and blanks were performed before each analysis.

Tissue samples by NMR

The intact frozen tissues were weighed and portions of 100 mg were transferred into a glass vial. Then, 5 mL of a PBS was added to remove any blood residues from the samples. After this, 5 mL of cooled MeOH and $CHCl_3$ were added prior to homogenization and vortex. The vials were placed at -20 °C and the vials were vortexed three times every 10 min. After this procedure, the vials were centrifuged at 4000*g* for 15 min at 4 °C. The upper phase (methanol) containing the polar metabolites was placed into another vial and the MeOH was removed under a nitrogen stream before lyophilisation. At the time of the NMR analysis, 540 µL of D_2O was added to the obtained extract and mixed with 60 µL of a buffer solution (KH_2PO_4, 1.5 M in D_2O) containing 0.1 % of TSP-d4 (used as chemical shift reference) and 2mM NaN_3. The pH was adjusted to 7.00 ± 0.02 by adding small amounts of KOD.

GC-qMS conditions

After the extraction procedure, the SPME fiber with the analytes was inserted into the injection port of an Agilent Technologies 6890N Network gas chromatograph system (Palo Alto, CA, USA) where the VOMs were desorbed at 250 °C for 10 min. The gas chromatograph was equipped with a 60 m × 0.25 mm I.D.× 0.25 µm film thickness, BP-20 (SGE, Dortmund, Germany) fused silica capillary column and interfaced with an Agilent 5975 quadrupole inert mass selective detector. The following oven temperature profile was set: (a) 5 min at 45 °C; (b) increase temperature until 150 °C, at a rate of 2 °C min^{-1} (hold for 10 min); (c) 150 °C for 10 min; (d) increase temperature until 220 °C, at a rate of 7 °C min^{-1}; and (e) 220 °C for 10 min for a total GC run time of 87.5 min. The column

flow was constant at 1.3 mL min^{-1} using Helium (He, N60, Air Liquide, Portugal) as carrier gas. The injection port was operated in the splitless mode and held at 250 °C. For the 5975 MS system, the operating temperatures of the transfer line, quadrupole and ionization source were 270, 150 and 230 °C, respectively, while electron impact mass spectra were recorded at 70 eV ionization voltage and the ionization current was 10 µA. Data acquisition was performed in the scan mode (30–200 m/z). The electron multiplier was set to the auto tune procedure. Metabolites identification was accomplished by manual interpretation through single ion monitorization (SIM) of spectra and matching against the Agilent MS ChemStation Software, equipped with a NIST05 mass spectral library with a similarity threshold higher than 80% and comparison with commercially available standard samples when available. A series of C_8–C_{20} n-alkanes were analyzed using the same extraction procedure to establish the kovat indices (KI), and to confirm the identity of the VOMs by comparison with the literature. The analyzes were performed in triplicate and the results expressed by mean ± standard deviation.

NMR measurements

NMR spectral acquisition was performed using a Bruker Advance II Plus NMR spectrometer equipped with a 400 MHz magnet UltraShield™ 400 Plus at 300K. All NMR spectra acquisition and pre-processing were performed under the control of a workstation with TopSpin 3.5pl7 (Bruker BioSpin). For each sample, a standard 1D ^{1}H NMR spectrum was acquired using a "noesypr1d" (Bruker library) water suppression pulse sequence with water irradiation during relaxation delay and mixing time (SW 4807.692 Hz, TD 64 K data points, relaxation delay 5 s, 128 scans). The shimming was calibrated automatically. Also, all spectra were processed using a line broadening (1.0 Hz) and baseline automatically corrected. The NMR spectrum of each sample was aligned with reference to the TSP signal at δ 0.00 ppm. Spectral regions within the range of 0.94 to 10 ppm were analyzed after excluding the sub-region δ 4.55-6.05 to remove variability arising from water suppression and possible cross-relaxation effect on the urea signal via solvent exchanging protons. Each sample analysis was performed in triplicate and the relative standard deviation (RSD) was lower than 2%.
The analysis of NMR spectral data was performed using the Chenomx NMR Suite 8.2 (Chenomx Inc., Alberta, Canada) and relative concentrations were determined using the 400MHz library from Chenomx NMR Suite 8.2, which compares the integral of a known reference signal (TSP) with signals derived from a library of compounds containing chemical shifts and peak multiplicities. Regarding the metabolites that were not available in the library, identification was accomplished by running a standard solution and the relative concentration was calculated manually. This software not only

allows the identification of compounds but also access their quantification based on advanced algorithms turning into a very straightforward tool to analyze NMR spectra.

Statistical Analysis

Statistical analysis was performed using the web server Metaboanalyst 4.0 [134]. The multivariate statistical analysis, namely the orthogonal projections to latent structures discriminant analysis (OPLS-DA) were applied on tissue and urine metabolomic dataset to provide insights into the groups under study for each analytical platform. The metabolites with VIP scores higher than 1.0 were considered significant and used for further analysis. The receiver operating characteristic curves (ROC) were also attained to verify which metabolites had the highest sensitivity/specificity for a potential BC diagnosis. The selected metabolites were used for the metabolite set enrichment analysis (MSEA) to identify significant patterns of metabolite concentration changes. MSEA uses a collection of predefined metabolite pathways and disease states obtained from the HMDB.

Furthermore, to inspect the correlations between urine and tissue metabolites, samples from the same individuals were matched and the correlation matrices obtained for the results for each analytical platform (GC-qMS and NMR) separately. The matrices were generated by calculating the Pearson´s correlation coefficient between each pair of variables from either NMR or GC-MS. The results were generated and plotted using MATLAB R2018b Academic version (MathWorks, Natick, MA, US).

Results and Discussion

Tissue and urine metabolomic pattern based on GC-qMS and ^{1}H NMR spectroscopy

The combined datasets of tissue and urine samples composed of 32 metabolites (GC-qMS) towards 24 metabolites (NMR) in tissue samples and 52 metabolites (GC-qMS) towards 33 metabolites (NMR) in urine samples, were subjected to statistical analysis using the Metaboanalyst 4.0 [134] server in order to obtain a preliminary information about data projection. Prior to multivariate analysis, all datasets were scaled to unit variance to select which metabolites were used for further analysis (*t*-test, $p < 0.05$) and then the selected metabolites were autoscaled so that each variable had the same weight. OPLS-DA statistical analysis was generated to compare the metabolic profiles between BC patients and controls, maximizing the class discrimination. The quality of the model was evaluated by R^2Y and Q^2 values, which gave the variance explained and predicted for the model, respectively, thus confirming that the model was effective with a good predictable accuracy.

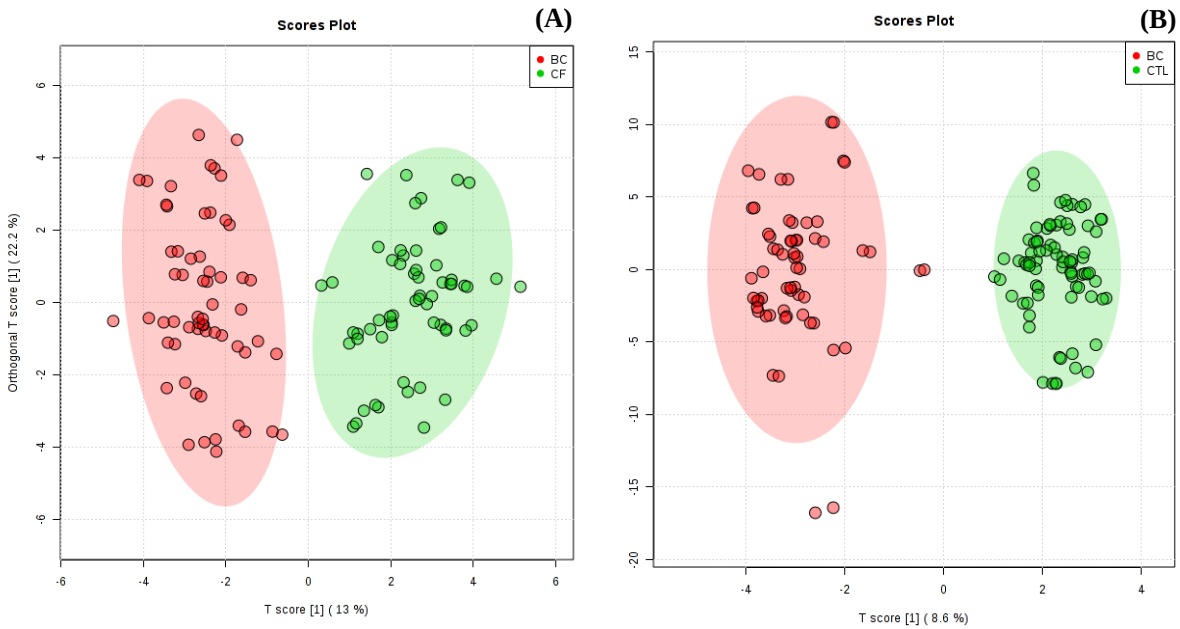

Figure 3. 5. 1 - Loading score plots of orthogonal projection to OPLS-DA analysis of tissue (A) and urine samples (B) model validation by permutation test based on 1000 permutations from the 2 groups under study.

Figure 3.5.1 (A) and (B) show the results obtained for the multivariate analysis. (OPLS-DA) for tissue (R^2Y = 0.888; Q^2 = 0.813) and urine (R^2Y = 0.946 and Q^2 = 0.910) samples, respectively. As observed, the obtained OPLS-DA score plot showed a clear separation between BC patients and controls for both sort of samples. Additionally, the metabolites from each biological matrix with VIP values higher than 1.0 were used for the pathway analysis to verify which pathways were the relevant involved in BC.

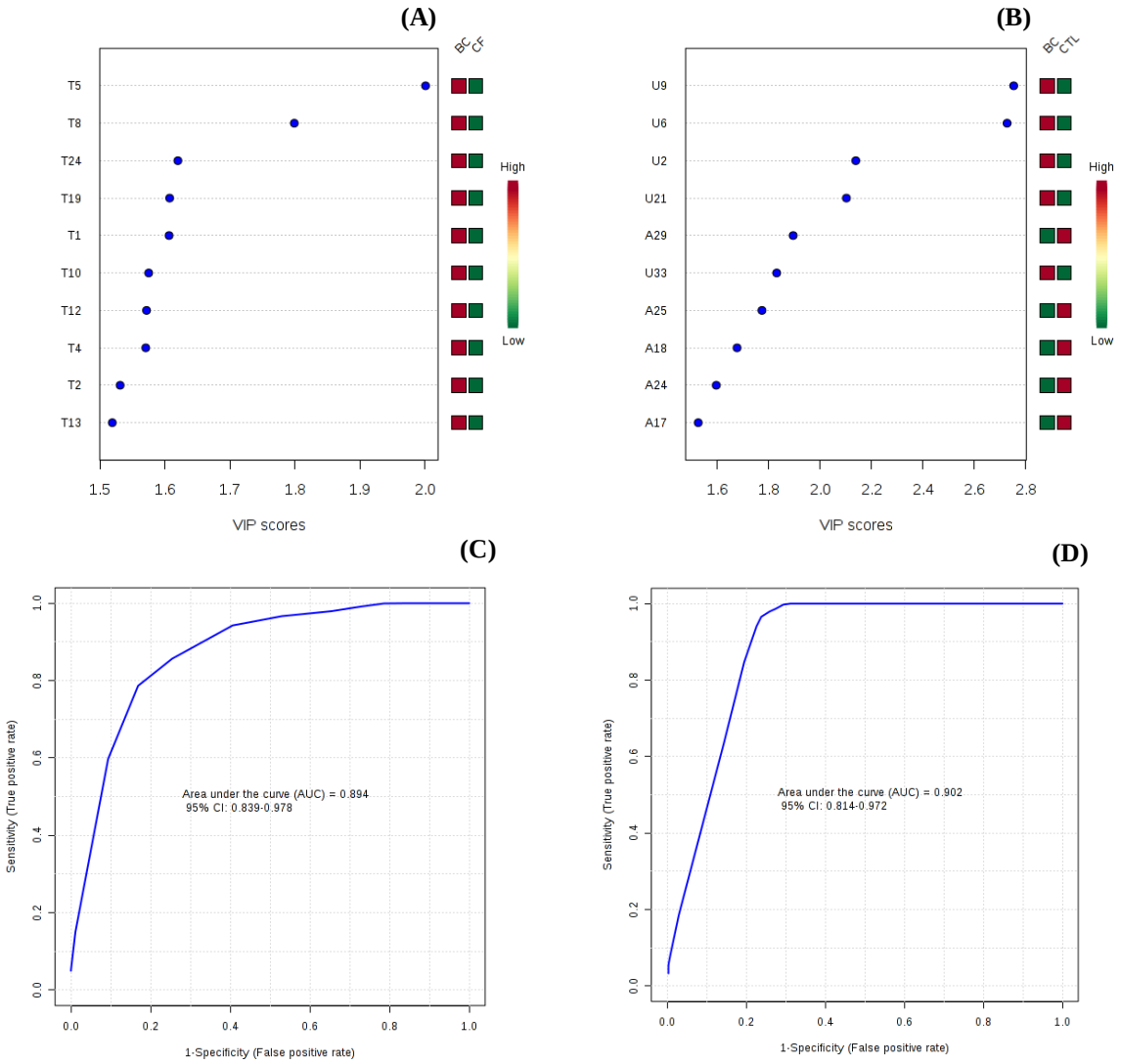

Figure 3. 5. 2 - Top ten significant features from tissue (A) and urine (B) samples based on VIP value and ROC curves for tissue (C) and urine (D) samples using the selected metabolites by VIP values.

Figure 3.5.2 A-B includes the top ten metabolites with the highest importance in the projection which comprised the majority of metabolites arising from the tissue analysis by NMR when compared with the VIP values from urine analysis that derived from the GC-qMS technique. Moreover, the ROC (Figure 3.5.2 C-D) curves for each type of biological sample (urine and tissue) were constructed using the metabolites with higher VIP values as described previously. Regarding the results obtained, we can observe that in the case of tissue with five metabolites, namely lactate, glutamate, taurine, *o*-phosphocholine, valine.

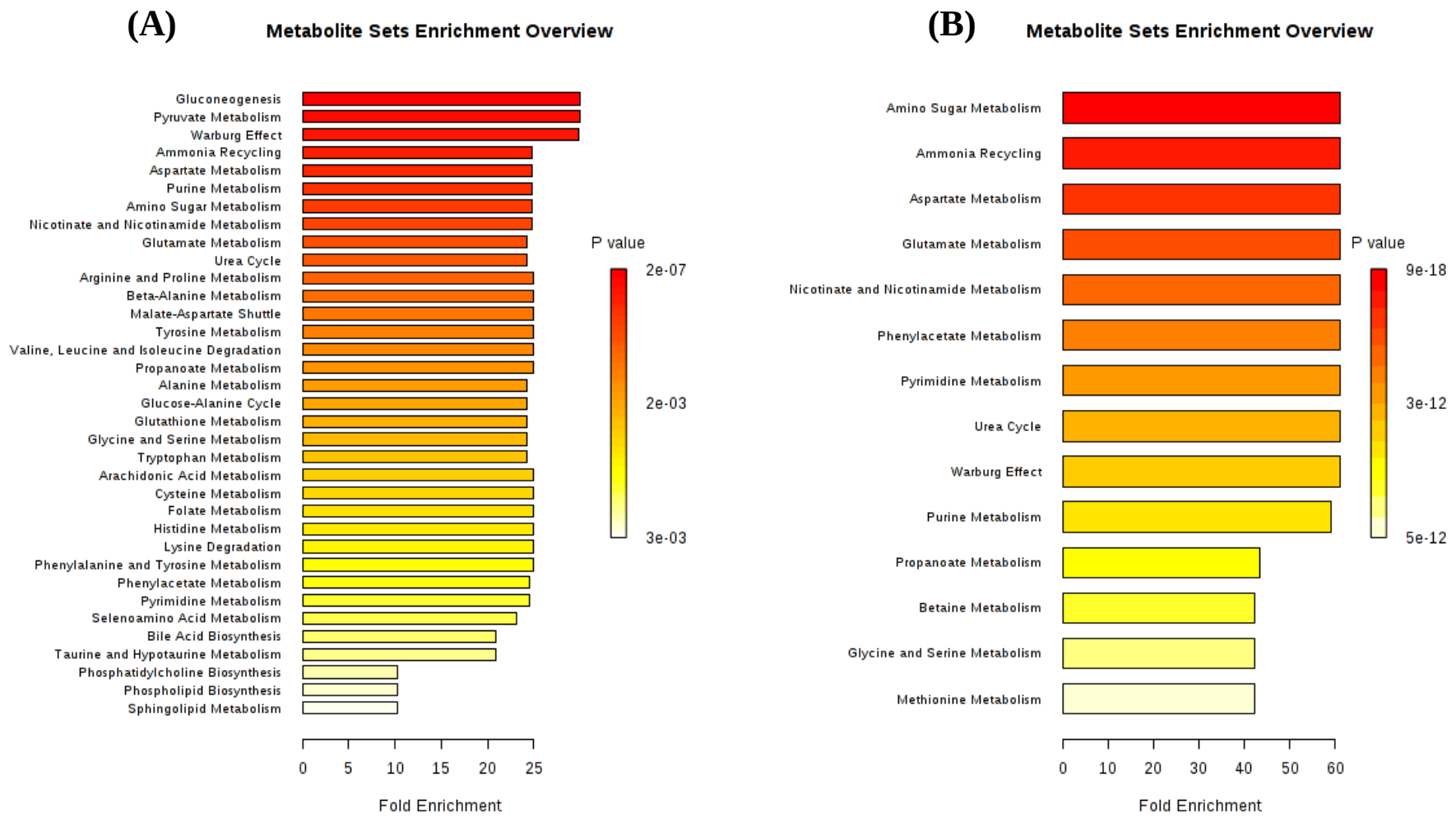

Figure 3. 5. 3 - Summary plots for MESA analysis overview of tissue (A) and urine (B) samples and respective pathways associated.

The values obtained for the AUC, namely 0.894 for tissue versus 0.902 for urine, were enough to provide a good sensitivity and specificity whereas for urine only three metabolites, namely glutamine, α-hydroxyisobutyrate and α-hydoxybutyrate were necessary (Figure 3.5.2 C-D).

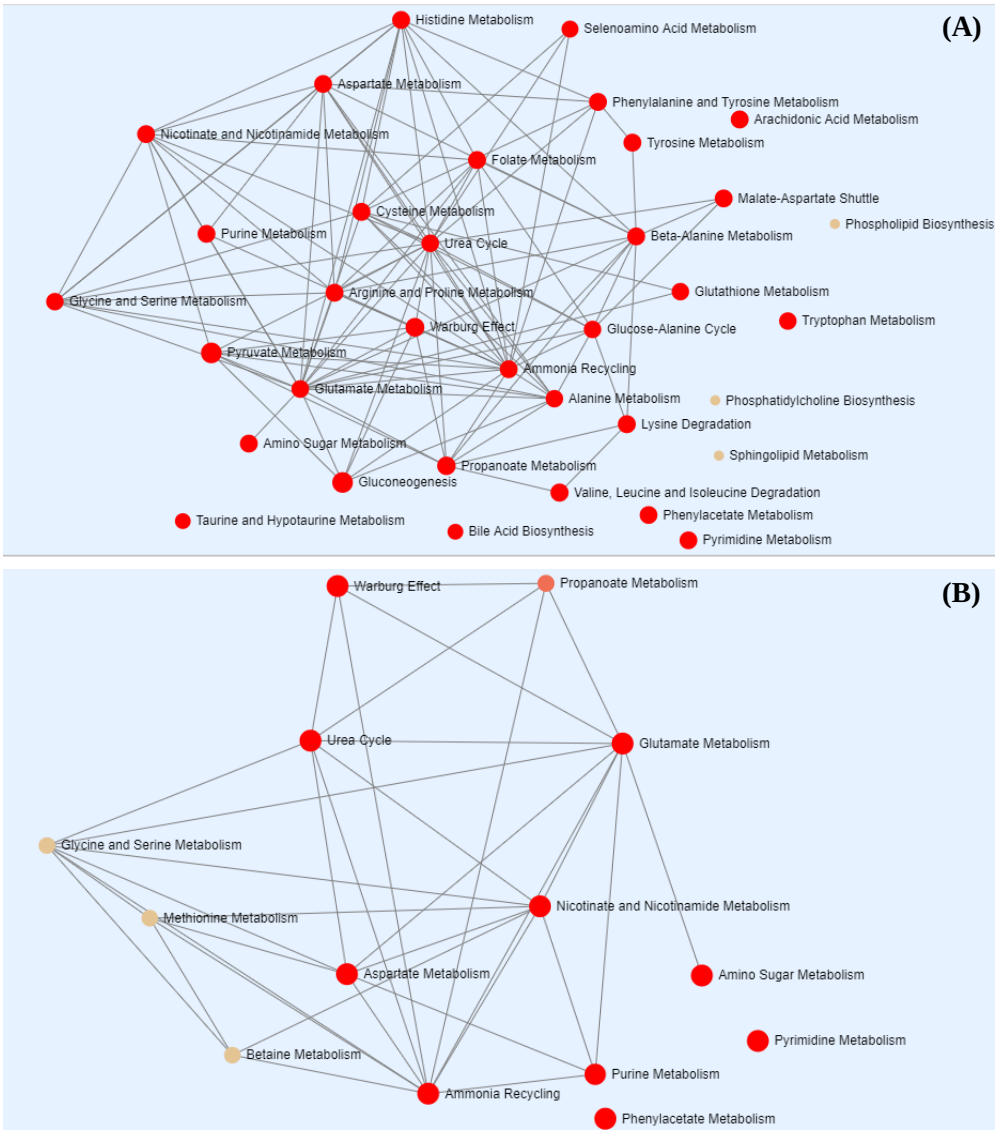

Figure 3. 5. 4 - The metabolome view map of significant altered metabolic pathways observed in tissue (A) and urine (B) samples from BC and CTL groups resulted from MESA.

To go further, the metabolic pathway analysis using Metabolite Set Enrichment Analysis (MESA) was performed in order to identify which pathways were affected in both cases. This type of analysis is used as a mean to recognize biologically meaningful patterns that are enriched in metabolomic data. For the tissue analysis, the metabolic pathways mainly affected included those of pyruvate, alanine and glutamine whereas, for urine, only glutamine and alanine metabolisms had changes (Figure 3.5.3 A-B). The metabolites responsible for each pathway were lactate in the case of tissue samples and glutamine and glutamate for urine samples (Figure 3.5.4 A-B). In order to check the possible correlations between urine and tissue metabolites, samples from the same individuals were

matched and the correlation matrices obtained for the results for each analytical platform (GC-qMS and NMR) are presented in Figure 3.5.5.

For this test, only the metabolites with correlation coefficient higher than 0.6 and $p < 0.05$ were used. Using this criterion, for example, there are no correlations for tissue *vs* urine in NMR or tissue NMR *vs* urine GC-qMS. By the Figure, we can observe that five metabolites were found significant as indicated in Table 3.5.3, being that the correlation level increases with the intensity of colour.

Table 3. 5. 3 - Correlations obtained from tissue and urine samples by dual analytical platforms.

Metabolite	**Sample**	**Platform**	**Metabolite**	**Sample**	**Platform**	**Correlation Coefficient**	***p*-value**
acetone	urine	GC-MS	1-pentanol	tissue	GC-MS	0.858	3.37E-27
3-hexanone	urine	GC-MS	1-pentanol	tissue	GC-MS	0.621	6.09E-11
4-heptanone	urine	GC-MS	1-pentanol	tissue	GC-MS	0.645	6.85E-12
2-methyl-5-(methylthio)-furan	urine	GC-MS	decanal	tissue	GC-MS	0.906	1.32E-34
acetate	urine	NMR	1,6,7-trimethyl-naphthalene	tissue	GC-MS	0.656	2.19E-12

Regarding the possible origin of these metabolites, 4-heptanone was already identified in urine samples from cancer patients and healthy controls being hypothesized that it arises from in vivo β-oxidation of 2-ethylhexanoic acid (EHA) from plasticizers, similar to formation of 3-heptanone from valproic acid [127,225]. Acetate is a common metabolite found in most tissues and was also identified in biological specimens such as urine, saliva or faeces. The main pathways where acetate is involved includes pyruvate, aminoacid and aspartate metabolisms [226]. Also, 2-methyl-5-(methylthio)-furan was identified in urine samples from cancer patients [127]. However, to the best of our knowledge the possible origin of this metabolite is not yet reported in literature. Regarding acetone, it was also identified in urine and tissue samples and the main pathways associated with it are the synthesis and degradation of ketone bodies and propanoate metabolism through acetyl-CoA in the liver [227]. The results obtained with this preliminary research suggest the possibility to identify endogenous metabolites using a dual platform to discover potential BC biomarkers and paves a way to investigate the related metabolic pathways to improve the diagnostic tools in BC.

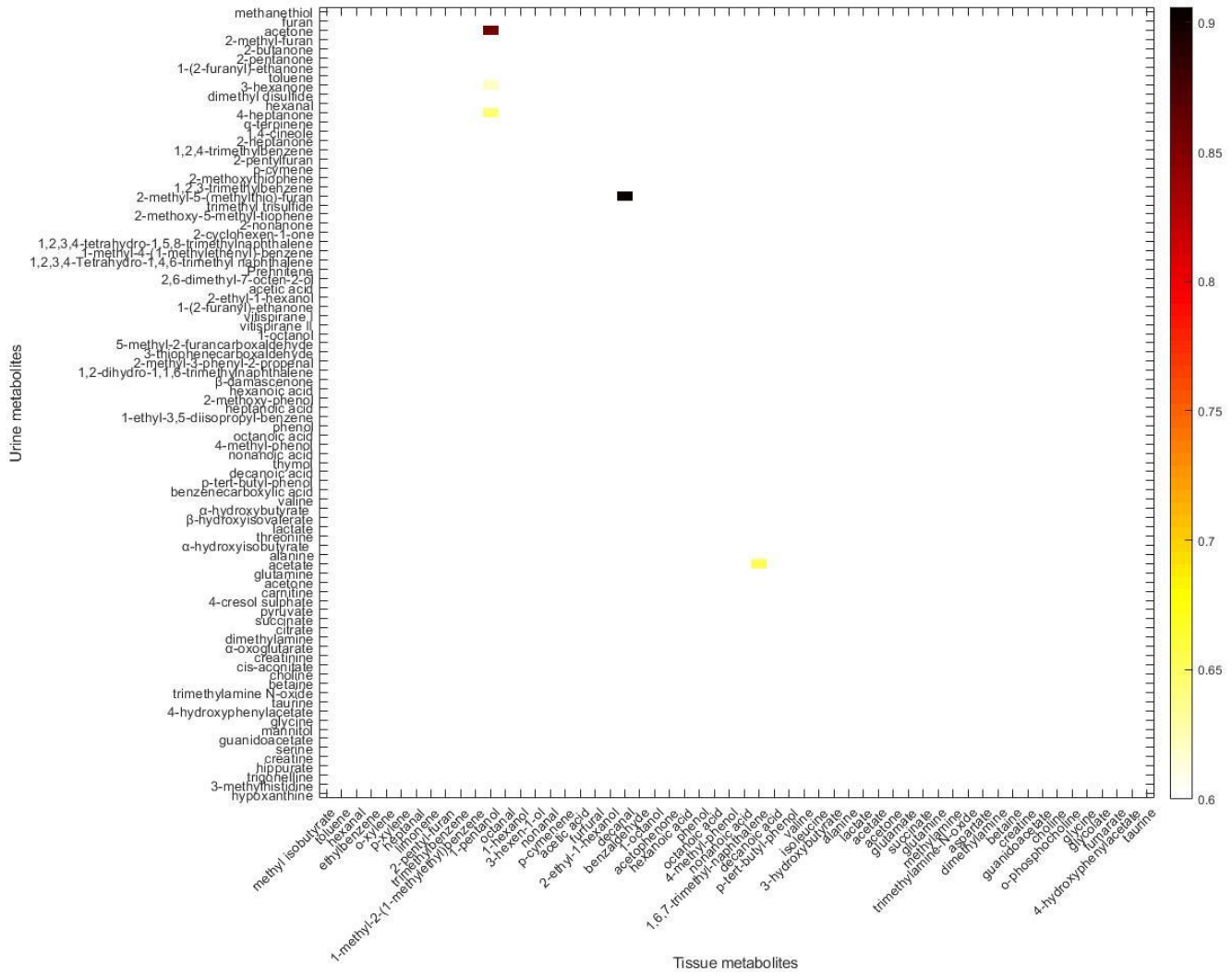

Figure 3. 5. 5 - Correlation analysis of metabolites between tissue and urine samples. The matrices were generated by calculating the Pearson correlation coefficient between each pair of variables from either NMR or GC-qMS.

Conclusions

This research allowed the combination of NMR and GC-MS based metabolomic analyzes of urine and breast tissue samples from BC patients and cancer-free individuals tandem with multivariate statistical tools in order to obtain a panel of metabolites (and their metabolic profiles) that can discriminate malignant from healthy status thus assisting in the diagnostic field. Additionally, using a dual platform approach (NMR and GC-qMS), we could enlarge the panel of identified metabolites showing a promising diagnostic tool for a BC diagnosis. Overall, this study suggests that an improved metabolic profile combining NMR and GC-qMS may be useful to achieve more insights regarding cancer mechanisms. Nevertheless, due to the challenge of identifying metabolites linked to BC, further studies are needed for an ampler understanding of the mechanisms underlying BC.

SECTION 4| Integrated Discussion

The improvement of diagnostic/screening methods are important achievements in cancer research besides the intensive investigation undertaken on the discovery of new biomarkers. With the emergence of metabolomic studies in the last decade that successfully overcomed the knowlegde about pathological processes such as cancer, there is the hypothesis that metabolic alterations result in qualitative and quantitative alterations on metabolites.

Although the sample size in a exploratory research is usually small, it is essential to analyze a larger number of sampels in order to verify and validate the results and to increase the statistical ability of the results. Moreover, the alterations on metabolites levels will be reflected in both genetic and environmental factors, leading to the increase of the susceptibility to cancer [217]. In the case of BC, numerous metabolomic studies have be conducted using biological samples, such as urine [63,189], serum [71,189] or tissues [228] by many analytical approaches, being the most commom the NMR and MS in single or in dual application [144].

Thus, the work described through this thesis represents a contribution on the study of BC metabolome, by analyzing urine, cell lines and breast tissue in order to gain a deeper understanding about the metabolic alterations occuring in BC and eventually find potential metabolites for BC diagnosis. The studies in this thesis were carried out using GC-qMS and NMR spectroscopy and their results are presented in Chapters 3.1 to 3.5. In general, the achieved results confirm that urine might be a source of valuable information for the search of BC biomarkers.

In the first study, described in Sub-section 3.1, the main goal was the development of a statistical design (CCD) and the optimization of significant variables of SPME procedure for the isolation of VOMs from urine of BC patients (n=30) and CTL (n=40). This type of optimization takes into account the possible interations between the variables. The establishment of the urinary volatomic composition, through GC-qMS analysis, improved the identification of VOMs potential BC biomarkers useful to be used together or to complement the current BC diagnostics tools. For that purpose, several extraction influencing parameters (e.g., SPME coating, NaCl amount, extraction time and temperature) were tested. The developed HS-SPME method was applied to urine samples to discriminate between BC and CTL groups, showing promising results that are essential to be explored. Additionaly, from the GC-qMS analysis, 116 VOMs were identified and multivariate statistical analysis revealed some metabolites significantly altered in BC patients. Of the metabolites identified, 3-methyl-thiophene, 4-heptanone, α-terpinene, 2-pentylfuran, *p*-cymene, trimethyl trisulfide, 1-methyl-4-(1-methylethenyl)-benzene, acetic acid, 2-methyl-3-phenyl-2-propenal and 1,2-dihydro-1,1,6-trimethylnaphthalene showed the utmost sensitivity and specificity to discriminate BC patients from CTL. Some of these metabolites have been already identified in urine from renal cell carcinoma (RCC) patients as being increased in RCC patients [229]. The metabolic pathway

analysis indicated that the discriminatory metabolites could be originated from several dysregulated pathways in BC such as pyruvate and sulfur metabolisms. These results suggested the possibility to identify endogenous metabolites as a platform to discovery potential BC biomarkers and paves a way to investigate the related metabolomic pathways to improve the diagnostic tools of BC.

The second study (Sub-section 3.2) consisted of an extension of the abilities of urine on the field of BC biomarker discovery. For that purpose, the urinary metabolomic pattern was obtained from BC patients and CTL using ^{1}H NMR spectroscopy, as a powerful approach to identify a set of BC-specific metabolites which might be employed in the BC diagnostic. OPLS-DA was applied to ^{1}H-NMR processed data matrix. Metabolomic patterns distinguished BC from CTL urine samples suggesting unique metabolite profiles for each investigated group. The discrimination efficiency and accuracy of the urinary metabolites were ascertained by ROC curve analysis that allowed the identification of some metabolites with the highest sensitivity and specificity to discriminate BC patients from CTL. The metabolomic pathway analysis indicated several metabolism pathways disruptions including aminoacids and carbohydrates metabolism, in BC patients. The obtained results support the high throughput potential of NMR-based urinary metabolomics patterns in the discrimination BC patients from CTL.

Moreover, the third study described in Sub-section 3.3 demonstrated that HS-SPME/GC-MS can be used to establish the *in vitro* volatile metabolomic patterns of normal and cancer breast cells. In addition, this study showed the potential of screening the in vitro VOMs associated with BC to identify potential volatile metabolites to be used in early diagnosis. VOMs from the headspace of cultured BC cells and normal human mammary epithelial cells, were collected by HS-SPME and analyzed by GC–qMS, thus defining the volatile metabolomic signature. Also, the headspace of cultured media of normal and cancer cell lines, containing the exometabolome with extracelular metabolites, was analyzed at different pH conditions. Most of the identified VOMs were common to all BC cell lines and normal human mammary epithelial cells, but 2-pentanone, 2-heptanone, 3-methyl-3-buten-1-ol, ethyl acetate, ethyl propanoate, and 2-methyl butanoate, were detected only in the headspace of cancer cell lines. Multivariate statistical data obtained in this study revealed that combining *in vitro* assays with HS-SPME/GC-qMS is a promising strategy to discriminate the volatile metabolomic signature of normal cells and BC cell lines according to molecular type.

The fourth study (Sub-section 3.4) consisted of the untargeted assessment of metabolomic tissue profile from BC patients when compared with CF tissues using GC-qMS tandem with multivariate statistical tools (PLS-DA and OPLS-DA). Twenty-nine metabolites were identified, and multivariate statistical analysis revealed some metabolites significantly altered in BC patients. Of the metabolites identified, limonene, decanoic acid, acetic acid and furfural showed the highest sensitivity and

specificity to discriminate BC patients. Also, the analysis of the plots leads to a metabolomic pattern comprising an array of some biochemical pathways altered in BC patients. The metabolic pathway analysis indicated that the discriminatory metabolites could be originated from several dysregulated pathways in BC such as those involved in pyruvate and sulphur metabolisms.

In the last study (Sub-section 3.5) the combination of NMR and GC-qMS based metabolomic approach was applied using urine and breast tissue samples from BC patients and cancer-free individuals. The data obtained was processed using multivariate statistical tools in order to obtain a panel of metabolites (and their metabolic profiles) that could discriminate malignant from healthy status. Tissue and urine samples were collected from BC patients (urine: *n*=30; tissue *n*=30) and CF subjects (urine *n*=40; tissue *n*=30) and analyzed. The OPLS-DA model showed a clear separation between BC patients and controls for both classes of samples. The discrimination efficiency and accuracy of BC tissue and urine metabolites was ascertained by ROC curve. Regarding the metabolomic pathway analysis several dysregulated pathways including pyruvate and glutamine metabolisms, in BC patients were identified. Moreover, the correlations between urine and tissue metabolites were investigated being that five metabolites were found significant using a dual platform approach. Overall, this study suggests that an improved metabolic profile combining NMR and GC-qMS may be useful to achieve more insights regarding cancer mechanisms.

In summary, cancer is a complex disease where many metabolic pathways are dysregulated regardless on cancer type. This might be the major handicap of metabolomics studies when searching for specific biomarkers, since the results obtained can be the reflection of the commom carcinogenic processes and not from the type of cancer of interest. One of the important steps is the validation of the potential biomarkers, the requirement of a larger cohort and also the undertake of an external validation [77,230].

SECTION 5| Conclusions and Future Perspectives

The work described in this thesis emphasizes the potential of urine for the pursue of BC biomarkers for a non-invasive diagnosis, since it is the final step of metabolic pathways. Among the biofluids used in metabolomic studies, urine has been shown to contain a wealth of metabolic information that is believed to be changed in disease conditions, including cancer. Moreover, it is abundant, easily collected, stored and given its complexity, is particularly rich in potential disease biomarkers. This makes it an ideal biofluid for the detection or monitoring of disease evolution. However, to complement the data obtained from urine and to enhance a comphreensive analysis of BC metabolome, it is essential to use other biological samples, such as as tissue or cell lines once they can provide additional information regarding urine.

The main conclusions of this thesis:

- ✓ A CCD model was optimized and applied to urine samples from BC and CTL groups, using HS-SPME/GC-qMS in order to obtain the optimal extraction conditions for the VOMs extraction from the target samples. In this context, a CAR/PDMS fiber coating at 50 °C during 75 min with the addition of 15 % NaCl were selected. All the assays to establish the volatomic composition of urine and BC tissue were carried out using the optimized conditions obtained by the CCD model;

- ✓ The establisment of the metabolomic profile from urine samples based on volatile (GC-qMS) and non-volatile (^{1}H NMR) fractions was performed and unveiled the ability to discriminate BC from CTL groups. The metabolites creatine, glycine, serine, dimethylamine, trimethylamine-N-oxide, α-hydroxyisobutyrate, manitol, glutamine, cis-aconitate and trigonelline exhibited the highest sensivitity in the discrimination of both groups by NMR studies, whereas 3-methyl-thiophene, 4-heptanone, α-terpinene, 2-pentylfuran, *p*-cymene, trimethyl trisulfide, 1-methyl-4-(1-methylethenyl)-benzene, acetic acid, 2-methyl-3-phenyl-2-propenal and 1,2-dihydro-1,1,6-trimethylnaphthalene showed the utmost sensitivity and specificity to discriminate BC from CTLs using GC-qMS technique;

- ✓ The establishment of the volatile metabolomic pattern by HS-SPME/GC-qMS of normal and cancer breast cells was obtained. The headspace of culture media of normal and cancer cell lines was analyzed at different pH conditions. Most of the identified VOMs were common to all BC cell lines and HMEC cells, with exception of 2-pentanone, 2-heptanone, 3-methyl-3-

buten-1-ol, ethyl acetate, ethyl propanoate, and 2-methyl butanoate that were detected only in the headspace of cancer cell lines. Multivariate statistical analysis revealed that it was possible to differentiate the volatile metabolomic signature of normal cells and BC cell lines according to molecular type, thus contributing to the discovery of novel biomarkers and the related metabolomic pathways thus improving the diagnostic tools for BC;

- ✓ The untargeted assessment of the metabolomic tissue profile from BC patients and compared with cancer-free (CF) tissues using GC-qMS tandem with multivariate statistical tools (PLS-DA and OPLS-DA). From the identified metabolites, limonene, decanoic acid, acetic acid and furfural showed the highest sensitivity and specificity to discriminate BC patients from CTLs. The discriminatory metabolites could be originated from several dysregulated pathways in BC such as, those involved in pyruvate and sulphur metabolisms. Furthermore, the analysis of cancer and non-cancer tissue from the same subject can also aid to balance the effect of external interferents, such as diet or environmental exposure and help to identify potential biomarkers;

- ✓ Urine and tissue samples from BC and CF subjects were analyzed by different analytical platforms (GC-qMS and NMR) in order to evaluate the correlations between urinary and tissue metabolites, where samples from the same individuals were matched through the integration of data, being that five metabolites (e.g., acetone, 3-hexanone, 4-heptanone, 2-methyl-5-(methylthio)-furan and acetate) were found significant using a dual approach. Moreover, the integration of high-resolution analytical frameworks, namely MS and NMR, appeared as an outcome on metabolomics studies, providing a sensitive, reliable detection and the quantification of metabolites in biological samples and their related metabolic pathways;

- ✓ Furthermore, the metabolic networks obtained by the correlation analysis will be also an attempt to analyze the complex interactions between metabolites as a complementary tool to statistical univariate and multivariate data analysis methods in order to reveal conections between pathways, identifying metabolite differences in physiological state.

The future work challanges encompass:

- ✓ To investigate the metabolic profile of cultured cells from breast tissue and urine samples from the same donor, in order to have a deeper insight into metabolism imbalance;

- ✓ To gain a better understanding of the causes of the dysregulation that occur in BC, by taking into account an external validation with larger cohorts of samples and using statistical tools allowing the selection of significant variables reducing the impact of the less meaningful ones.

References

References

1. Donepudi MS, Kondapalli K, Amos SJ, Venkanteshan P. Breast cancer statistics and markers. J Cancer Res Ther. 2014;10(3):506–11.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7–30.
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
4. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359–86.
5. Allison KH. Molecular Pathology of Breast Cancer. Am J Clin Pathol. 2012;138(6):770–80.
6. DeSantis CE, Bray F, Ferlay J, Lortet-Tieulent J, Anderson BO, Jemal A. International Variation in Female Breast Cancer Incidence and Mortality Rates. Cancer Epidemiol Biomarkers Prev. 2015;24(10):1495–506.
7. Ghoncheh M, Pournamdar Z, Salehiniya H. Incidence and Mortality and Epidemiology of Breast Cancer in the World. Asian Pacific J Cancer Prev Cancer Control West Asia Spec Issue J Cancer Prev Cancer Control West Asia Spec Issue. 2016;17(17):43–6.
8. Verma R, Bowen RL, Slater SE, Mihaimeed F, Jones JL. Pathological and epidemiological factors associated with advanced stage at diagnosis of breast cancer. Br Med Bull. 2012;103(1):129–45.
9. Shah R, Rosso K, Nathanson SD. Pathogenesis, prevention, diagnosis and treatment of breast cancer. World J Clin Oncol. 2014;5(3):283–98.
10. Libson S, Lippman M. A review of clinical aspects of breast cancer. Int Rev Psychiatry. 2014 Feb;26(1):4–15.
11. Maruti SS, Willett WC, Feskanich D, Rosner B, Colditz GA. A prospective study of age-specific physical activity and premenopausal breast cancer. J Natl Cancer Inst. 2008;100(10):728–37.
12. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, et al. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. BMJ. 2016;354:i3857.
13. McDonald JA, Goyal A, Terry MB. Alcohol Intake and Breast Cancer Risk: Weighing the Overall Evidence. Curr Breast Cancer Rep. 2013;5(3).
14. Maskarinec G, Jacobs S, Park S-Y, Haiman CA, Setiawan VW, Wilkens LR, et al. Type II

Diabetes, Obesity, and Breast Cancer Risk: The Multiethnic Cohort. Cancer Epidemiol Biomarkers Prev. 2017;26(6):854–61.

15. Heidegger I, Ofer P, Doppler W, Rotter V, Klocker H, Massoner P. Diverse Functions of IGF/Insulin Signaling in Malignant and Noncancerous Prostate Cells: Proliferation in Cancer Cells and Differentiation in Noncancerous Cells. Endocrinology. 2012;153(10):4633–43.
16. Djiogue S, Nwabo Kamdje AH, Vecchio L, Kipanyula MJ, Farahna M, Aldebasi Y, et al. Insulin resistance and cancer: the role of insulin and IGFs. Endocr Relat Cancer. 2013;20(1):R1–17.
17. Neuhouser ML, Aragaki AK, Prentice RL, Manson JE, Chlebowski R, Carty CL, et al. Overweight, Obesity, and Postmenopausal Invasive Breast Cancer Risk. JAMA Oncol. 2015;1(5):611.
18. Picon-Ruiz M, Morata-Tarifa C, Valle-Goffin JJ, Friedman ER, Slingerland JM. Obesity and adverse breast cancer risk and outcome: Mechanistic insights and strategies for intervention. CA Cancer J Clin. 2017;67(5):378–97.
19. Moore SC, Playdon MC, Sampson JN, Hoover RN, Trabert B, Matthews CE, et al. A Metabolomics Analysis of Body Mass Index and Postmenopausal Breast Cancer Risk. JNCI J Natl Cancer Inst [Internet]. 2018;6(110):588–97.
20. Kos Z, Dabbs DJ. Biomarker assessment and molecular testing for prognostication in breast cancer. Histopathology. 2016;68(1):70–85.
21. Duffy MJ, Harbeck N, Nap M, Molina R, Nicolini A, Senkus E, et al. Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). Eur J Cancer. 2017;75:284–98.
22. Harris LN, Ismaila N, McShane LM, Andre F, Collyar DE, Gonzalez-Angulo AM, et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol. 2016;34(10):1134–50.
23. Jasbi P, Wang D, Cheng SL, Fei Q, Cui JY, Liu L, et al. Breast cancer detection using targeted plasma metabolomics. J Chromatogr B [Internet]. 2019 Jan 15 [cited 2019 Feb 18];1105:26–37.
24. Clish CB. Metabolomics: an emerging but powerful tool for precision medicine. Cold Spring Harb Mol case Stud. 2015;1(1):a000588.
25. Roberts LD, Souza AL, Gerszten RE, Clish CB. Targeted metabolomics. Curr Protoc Mol Biol. 2012;Chapter 30:Unit 30.2.1-24.
26. Klupczyńska A, Dereziński P, Kokot ZJ. Metabolomics in medical sciences--trends,

challenges and perspectives. Acta Pol Pharm. 2015;72(4):629–41.

27. Alonso A, Marsal S, Julià A. Analytical methods in untargeted metabolomics: state of the art in 2015. Front Bioeng Biotechnol. 2015;3:23.
28. Cho K, Mahieu NG, Johnson SL, Patti GJ. After the feature presentation: technologies bridging untargeted metabolomics and biology. Curr Opin Biotechnol. 2014;28:143–8.
29. Claudino WM, Goncalves PH, di Leo A, Philip PA, Sarkar FH. Metabolomics in cancer: A bench-to-bedside intersection. Crit Rev Oncol Hematol [Internet]. 2012;84(1):1–7.
30. Beger RD. A review of applications of metabolomics in cancer. Metabolites. 2013;3(3):552–74.
31. Lécuyer L, Victor Bala A, Deschasaux M, Bouchemal N, Nawfal Triba M, Vasson M-P, et al. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer. Int J Epidemiol. 2018;47(2):484–94.
32. Gu H, Gowda GAN, Raftery D. Metabolic profiling: are we en route to better diagnostic tests for cancer? Future Oncol. 2012;8(10):1207–10.
33. Jové M, Collado R, Quiles JL, Ramírez-Tortosa M-C, Sol J, Ruiz-Sanjuan M, et al. A plasma metabolomic signature discloses human breast cancer. Oncotarget. 2017;8(12):19522–33.
34. Fan Y, Zhou X, Xia T-S, Chen Z, Li J, Liu Q, et al. Human plasma metabolomics for identifying differential metabolites and predicting molecular subtypes of breast cancer. Oncotarget. 2016;7(9):9925–38.
35. Bain JR, Stevens RD, Wenner BR, Ilkayeva O, Muoio DM, Newgard CB. Metabolomics Applied to Diabetes Research: Moving From Information to Knowledge. Diabetes. 2009;58(11):2429–43.
36. Günther UL. Metabolomics Biomarkers for Breast Cancer. Pathobiology. 2015;82(3–4):153–65.
37. Hadi NI, Jamal Q. "OMIC" tumor markers for breast cancer: A review. Pakistan J Med Sci. 2015 Sep;31(5):1256–62.
38. Cappelletti V, Iorio E, Miodini P, Silvestri M, Dugo M, Daidone MG. Metabolic Footprints and Molecular Subtypes in Breast Cancer. Dis Markers. 2017;2017:1–19.
39. McCartney A, Vignoli A, Biganzoli L, Love R, Tenori L, Luchinat C, et al. Metabolomics in breast cancer: A decade in review. Cancer Treat Rev. 2018;67:88–96.
40. Anderson NM, Mucka P, Kern JG, Feng H. The emerging role and targetability of the TCA cycle in cancer metabolism. Protein Cell. 2018;9(2):216–37.
41. Ryan DG, Murphy MP, Frezza C, Prag HA, Chouchani ET, O'Neill LA, et al. Coupling Krebs cycle metabolites to signalling in immunity and cancer. Nat Metab. 2019;1(1):16–33.

42. Ciccarone F, Vegliante R, Di Leo L, Ciriolo MR. The TCA cycle as a bridge between oncometabolism and DNA transactions in cancer. Semin Cancer Biol. 2017;47:50–6.
43. Cífková E, Lísa M, Hrstka R, Vrána D, Gatěk J, Melichar B, et al. Correlation of lipidomic composition of cell lines and tissues of breast cancer patients using hydrophilic interaction liquid chromatography/electrospray ionization mass spectrometry and multivariate data analysis. Rapid Commun Mass Spectrom. 2017;31(3):253–63.
44. Silva CL, Perestrelo R, Silva P, Tomás H, Câmara JS. Volatile metabolomic signature of human breast cancer cell lines. Sci Rep. 2017;7:43969.
45. Le Guennec A, Tea I, Antheaume I, Martineau E, Charrier B, Pathan M, et al. Fast Determination of Absolute Metabolite Concentrations by Spatially Encoded 2D NMR: Application to Breast Cancer Cell Extracts. Anal Chem [Internet]. 2012;84(24):10831–7.
46. Willmann L, Schlimpert M, Hirschfeld M, Erbes T, Neubauer H, Stickeler E, et al. Alterations of the exo- and endometabolite profiles in breast cancer cell lines: A mass spectrometry-based metabolomics approach. Anal Chim Acta. 2016;925:34–42.
47. Cao Y, Wang Q, Gao P, Dong J, Zhu Z, Fang Y, et al. A dried blood spot mass spectrometry metabolomic approach for rapid breast cancer detection. Onco Targets Ther. 2016;9:1389.
48. Martinez-Lozano Sinues P, Landoni E, Miceli R, Dibari VF, Dugo M, Agresti R, et al. Secondary electrospray ionization-mass spectrometry and a novel statistical bioinformatic approach identifies a cancer-related profile in exhaled breath of breast cancer patients: a pilot study. J Breath Res. 2015;9(3):31001.
49. Cala MP, Aldana J, Medina J, Sánchez J, Guio J, Wist J, et al. Multiplatform plasma metabolic and lipid fingerprinting of breast cancer: A pilot control-case study in Colombian Hispanic women. Bathen TF, editor. PLoS One. 2018;13(2):e0190958.
50. Roig B, Rodríguez-Balada M, Samino S, Lam EW-F, Guaita-Esteruelas S, Gomes AR, et al. Metabolomics reveals novel blood plasma biomarkers associated to the BRCA1-mutated phenotype of human breast cancer. Sci Rep. 2017;7(1):17831.
51. Cavaco C, Pereira JAM, Taunk K, Taware R, Rapole S, Nagarajaram H, et al. Screening of salivary volatiles for putative breast cancer discrimination: an exploratory study involving geographically distant populations. Anal Bioanal Chem. 2018;410(18):1–10.
52. Tsutsui H, Mochizuki T, Inoue K, Toyama T, Yoshimoto N, Endo Y, et al. High-Throughput LC–MS/MS Based Simultaneous Determination of Polyamines Including N-Acetylated Forms in Human Saliva and the Diagnostic Approach to Breast Cancer Patients. Anal Chem. 2013;85(24):11835–42.
53. Zhong L, Cheng F, Lu X, Duan Y, Wang X. Untargeted saliva metabonomics study of breast

cancer based on ultra performance liquid chromatography coupled to mass spectrometry with HILIC and RPLC separations. Talanta. 2016;158:351–60.

54. Takayama T, Tsutsui H, Shimizu I, Toyama T, Yoshimoto N, Endo Y, et al. Diagnostic approach to breast cancer patients based on target metabolomics in saliva by liquid chromatography with tandem mass spectrometry. Clin Chim Acta. 2016;452(15):18–26.
55. Cífková E, Holčapek M, Lísa M, Vrána D, Gatěk J, Melichar B. Determination of lipidomic differences between human breast cancer and surrounding normal tissues using HILIC-HPLC/ESI-MS and multivariate data analysis. Anal Bioanal Chem. 2015;407(3):991–1002.
56. Budhu A, Terunuma A, Zhang G, Hussain SP, Ambs S, Wang XW. Metabolic profiles are principally different between cancers of the liver, pancreas and breast. Int J Biol Sci. 2014;10(9):966–72.
57. Kanaan YM, Sampey BP, Beyene D, Esnakula AK, Naab TJ, Ricks-Santi LJ, et al. Metabolic profile of triple-negative breast cancer in African-American women reveals potential biomarkers of aggressive disease. Cancer Genomics Proteomics. 2014;11(6):279–94.
58. Tenori L, Oakman C, Morris PG, Gralka E, Turner N, Cappadona S, et al. Serum metabolomic profiles evaluated after surgery may identify patients with oestrogen receptor negative early breast cancer at increased risk of disease recurrence. Results from a retrospective study. Mol Oncol. 2015;9(1):128–39.
59. Porto-Figueira P, Pereira JAM, Câmara JS. Exploring the potential of needle trap microextraction combined with chromatographic and statistical data to discriminate different types of cancer based on urinary volatomic biosignature. Anal Chim Acta. 2018;1023:53–63.
60. Thomson CA, Thompson PA. Dietary patterns, risk and prognosis of breast cancer. Futur Oncol. 2009;5(8):1257–69.
61. Martineau E, Tea I, Akoka S, Giraudeau P. Absolute quantification of metabolites in breast cancer cell extracts by quantitative 2D 1H INADEQUATE NMR. NMR Biomed. 2012;25(8):985–92.
62. Kim K-J, Kim H-J, Park H-G, Hwang C-H, Sung C, Jang K-S, et al. A MALDI-MS-based quantitative analytical method for endogenous estrone in human breast cancer cells. Sci Rep. 2016;6(1):24489.
63. Singh A, Sharma RK, Chagtoo M, Agarwal G, George N, Sinha N, et al. 1H NMR Metabolomics Reveals Association of High Expression of Inositol 1, 4, 5 Trisphosphate Receptor and Metabolites in Breast Cancer Patients. PLoS One. 2017;12(1):e0169330.
64. Zhao C, Xie P, Wang H, Cai Z. Liquid chromatography-mass spectrometry-based metabolomics and lipidomics reveal toxicological mechanisms of bisphenol F in breast cancer

xenografts. J Hazard Mater. 2018;

65. Huang S, Chong N, Lewis NE, Jia W, Xie G, Garmire LX. Novel personalized pathway-based metabolomics models reveal key metabolic pathways for breast cancer diagnosis. Genome Med. 2016;8(1):34.
66. Dougan MM, Li Y, Chu LW, Haile RW, Whittemore AS, Han SS, et al. Metabolomic profiles in breast cancer:a pilot case-control study in the breast cancer family registry. BMC Cancer. 2018;18(1):532.
67. Gu H, Pan Z, Xi B, Asiago V, Musselman B, Raftery D. Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: Application to the detection of breast cancer. Anal Chim Acta. 2011;686(1):57–63.
68. Jobard E, Pontoizeau C, Blaise BJ, Bachelot T, Elena-Herrmann B, Trédan O. A serum nuclear magnetic resonance-based metabolomic signature of advanced metastatic human breast cancer. Cancer Lett. 2014;343(1):33–41.
69. Lv W, Yang T. Identification of possible biomarkers for breast cancer from free fatty acid profiles determined by GC–MS and multivariate statistical analysis. Clin Biochem. 2012;45(1–2):127–33.
70. Lyon DE, Starkweather A, Yao Y, Garrett T, Kelly DL, Menzies V, et al. Pilot Study of Metabolomics and Psychoneurological Symptoms in Women With Early Stage Breast Cancer. Biol Res Nurs. 2018;20(2):227–36.
71. Wei S, Liu L, Zhang J, Bowers J, Gowda GAN, Seeger H, et al. Metabolomics approach for predicting response to neoadjuvant chemotherapy for breast cancer. Mol Oncol. 2013;7(3):297–307.
72. Playdon MC, Ziegler RG, Sampson JN, Stolzenberg-Solomon R, Thompson HJ, Irwin ML, et al. Nutritional metabolomics and breast cancer risk in a prospective study. Am J Clin Nutr. 2017;106(2):637–49.
73. Cala M, Aldana J, Sánchez J, Guio J, Meesters RJW. Urinary metabolite and lipid alterations in Colombian Hispanic women with breast cancer: A pilot study. J Pharm Biomed Anal. 2018;152:234–41.
74. Tayyari F, Gowda GAN, Olopade OF, Berg R, Yang HH, Lee MP, et al. Metabolic profiles of triple-negative and luminal A breast cancer subtypes in African-American identify key metabolic differences. Oncotarget. 2018;9(14):11677–90.
75. Bathen TF, Geurts B, Sitter B, Fjøsne HE, Lundgren S, Buydens LM, et al. Feasibility of MR Metabolomics for Immediate Analysis of Resection Margins during Breast Cancer Surgery.

Han A, editor. PLoS One. 2013;8(4):e61578.

76. Tang X, Lin C-C, Spasojevic I, Iversen ES, Chi J-T, Marks JR. A joint analysis of metabolomics and genetics of breast cancer. Breast Cancer Res. 2014;16(4):415.
77. Budczies J, Pfitzner BM, Györffy B, Winzer K-J, Radke C, Dietel M, et al. Glutamate enrichment as new diagnostic opportunity in breast cancer. Int J Cancer. 2015;136(7):1619–28.
78. Vettukattil R, Hetland TE, Flørenes VA, Kærn J, Davidson B, Bathen TF. Proton magnetic resonance metabolomic characterization of ovarian serous carcinoma effusions: chemotherapy-related effects and comparison with malignant mesothelioma and breast carcinoma. Hum Pathol. 2013;44(9):1859–66.
79. Euceda LR, Haukaas TH, Giskeødegård GF, Vettukattil R, Engel J, Silwal-Pandit L, et al. Evaluation of metabolomic changes during neoadjuvant chemotherapy combined with bevacizumab in breast cancer using MR spectroscopy. Metabolomics. 2017;13(4):37.
80. Gogiashvili M, Horsch S, Marchan R, Gianmoena K, Cadenas C, Tanner B, et al. Impact of intratumoral heterogeneity of breast cancer tissue on quantitative metabolomics using high-resolution magic angle spinning 1H NMR spectroscopy. NMR Biomed. 2018;31(2):e3862.
81. Choi JS, Baek H-M, Kim S, Kim MJ, Youk JH, Moon HJ, et al. Magnetic resonance metabolic profiling of breast cancer tissue obtained with core needle biopsy for predicting pathologic response to neoadjuvant chemotherapy. PLoS One. 2013;8(12):e83866.
82. Budczies J, Brockmöller SF, Müller BM, Barupal DK, Richter-Ehrenstein C, Kleine-Tebbe A, et al. Comparative metabolomics of estrogen receptor positive and estrogen receptor negative breast cancer: alterations in glutamine and beta-alanine metabolism. J Proteomics. 2013;94:279–88.
83. Dai C, Arceo J, Arnold J, Sreekumar A, Dovichi NJ, Li J, et al. Metabolomics of oncogene-specific metabolic reprogramming during breast cancer. Cancer Metab. 2018;6(1):5.
84. Yu L, Jiang C, Huang S, Gong X, Wang S, Shen P. Analysis of urinary metabolites for breast cancer patients receiving chemotherapy by CE-MS coupled with on-line concentration. Clin Biochem. 2013;46(12):1065–73.
85. Krilaviciute A, Heiss JA, Leja M, Kupcinskas J, Haick H, Brenner H. Detection of cancer through exhaled breath: a systematic review. Oncotarget. 2015;6(36):38643–57.
86. Zhang A, Sun H, Wang P, Han Y, Wang X. Modern analytical techniques in metabolomics analysis. Analyst. 2012;137(2):293–300.
87. Issaq HJ, Van QN, Waybright TJ, Muschik GM, Veenstra TD. Analytical and statistical approaches to metabolomics research. J Sep Sci. 2009 Jun;32(13):2183–99.

88. Dunn WB, Bailey NJC, Johnson HE. Measuring the metabolome: current analytical technologies. Analyst. 2005;130(5):606.
89. Wang JH, Byun J, Pennathur S. Analytical approaches to metabolomics and applications to systems biology. Semin Nephrol. 2010;30(5):500–11.
90. Sas KM, Karnovsky A, Michailidis G, Pennathur S. Metabolomics and diabetes: analytical and computational approaches. Diabetes. 2015;64(3):718–32.
91. Ahad T, Jasia Nissar I, Tehmeena Ahad C, Nissar J. Division of food science and technology, Skuast-k Fingerprinting in determining the adultration of food. J Pharmacogn Phytochem JPP. 2017;6(66):1543–53.
92. M. Narwate JG. Ultra performance liquid chromatography: a new revolution in liquid chromatography. Int J Pharm Drug Anal. 2014;2(1).
93. Yandamuri N, Srinivas Nagabattula KR, Swamy Kurra S, Batthula S, S Nainesha Allada LP, Bandam P. Comparative Study of New Trends in HPLC: A Review. Int J Pharm Sci Rev Res Int J Pharm Sci Rev Res Accept. 2013;23(10):52–7.
94. De Vos J, Broeckhoven K, Eeltink S. Advances in Ultrahigh-Pressure Liquid Chromatography Technology and System Design. Anal Chem. 2016;88(1):262–78.
95. Cacciola F, Farnetti S, Dugo P, Marriott PJ, Mondello L. Comprehensive two-dimensional liquid chromatography for polyphenol analysis in foodstuffs. J Sep Sci. 2017;40(1):7–24.
96. Reichenbach SE, Tian X, Tao Q, Ledford EB, Wu Z, Fiehn O. Informatics for cross-sample analysis with comprehensive two-dimensional gas chromatography and high-resolution mass spectrometry (GCxGC–HRMS). Talanta. 2011;83(4):1279–88.
97. Yu Z, Huang H, Reim A, Charles PD, Northage A, Jackson D, et al. Optimizing 2D gas chromatography mass spectrometry for robust tissue, serum and urine metabolite profiling. Talanta. 2017;165:685–91.
98. Umar A, Luider TM, Foekens JA, Paša-Tolić L. NanoLC-FT-ICR MS improves proteome coverage attainable for ~3000 laser-microdissected breast carcinoma cells. Proteomics. 2007;7(2):323–9.
99. Hendriks MMWB, Eeuwijk FA va., Jellema RH, Westerhuis JA, Reijmers TH, Hoefsloot HCJ, et al. Data-processing strategies for metabolomics studies. TrAC Trends Anal Chem. 2011;30(10):1685–98.
100. Yi L, Dong N, Yun Y, Deng B, Ren D, Liu S, et al. Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. Anal Chim Acta. 2016;914:17–34.
101. Gromski PS, Muhamadali H, Ellis DI, Xu Y, Correa E, Turner ML, et al. A tutorial review: Metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a

shotgun wedding. Anal Chim Acta. 2015;879:10–23.

102. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics. 2006;7(1):142.
103. Sysi-Aho M, Katajamaa M, Yetukuri L, Orešič M. Normalization method for metabolomics data using optimal selection of multiple internal standards. BMC Bioinformatics. 2007;8(1):93.
104. Kohl SM, Klein MS, Hochrein J, Oefner PJ, Spang R, Gronwald W. State-of-the art data normalization methods improve NMR-based metabolomic analysis. Metabolomics. 2012;8(S1):146–60.
105. Sugimoto M, Wong DT, Hirayama A, Soga T, Tomita M. Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. Metabolomics. 2010;6(1):78–95.
106. Xi B, Gu H, Baniasadi H, Raftery D. Statistical analysis and modeling of mass spectrometry-based metabolomics data. Methods Mol Biol. 2014;1198:333–53.
107. Zhang A, Sun H, Qiu S, Wang X. Metabolomics in noninvasive breast cancer. Clin Chim Acta. 2013;424:3–7.
108. Xia J, Broadhurst DI, Wilson M, Wishart DS. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. Metabolomics. 2013;9(2):280–99.
109. Artusi R, Verderio P, Marubini E. Bravais-Pearson and Spearman Correlation Coefficients: Meaning, Test of Hypothesis and Confidence Interval. Int J Biol Markers. 2002;17(2):148–51.
110. Liland KH. Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. TrAC Trends Anal Chem. 2011;30(6):827–41.
111. Köhn H-F, Hubert LJ. Hierarchical Cluster Analysis. In: Wiley StatsRef: Statistics Reference Online. Chichester, UK: John Wiley & Sons, Ltd; 2015. p. 1–13.
112. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognit Lett. 2010;31(8):651–66.
113. Yin S-Y, Wei W-C, Jian F-Y, Yang N-S. Therapeutic Applications of Herbal Medicines for Cancer Patients. Evidence-Based Complement Altern Med. 2013;2013:1–15.
114. Sperandei S. Understanding logistic regression analysis. Biochem Medica. 2014;24(1):12–8.
115. Ivanescu AE, Li P, George B, Brown AW, Keith SW, Raju D, et al. The importance of prediction model validation and assessment in obesity and nutrition research. Int J Obes. 2016;40(6):887–94.
116. Xu Q-S, Liang Y-Z, Du Y-P. Monte Carlo cross-validation for selecting a model and

estimating the prediction error in multivariate calibration. J Chemom. 2004;18(2):112–20.

117. Haddad K, Rahman A, A Zaman M, Shrestha S. Applicability of Monte Carlo cross validation technique for model development and validation using generalised least squares regression. J Hydrol. 2013;482:119–28.
118. Jaki T, Su T-L, Kim M, Van Horn ML. An evaluation of the bootstrap for model validation in mixture models. Commun Stat - Simul Comput. 2018;47(4):1028–38.
119. Lindgren F, Hansen B, Karcher W, Sjöström M, Eriksson L. Model Validation by Permutation tests: Applications to VariableSselection. J Chemom. 1996;10(5–6):521–32.
120. Zhang A, Sun H, Wang P, Han Y, Wang X. Recent and potential developments of biofluid analyses in metabolomics. J Proteomics. 2012;75(4):1079–88.
121. Milosevic M, Jankovic D, Milenkovic A, Stojanov D. Early diagnosis and detection of breast cancer. Technol Heal Care. 2018;26(4):729–59.
122. Duffy MJ. Serum Tumor Markers in Breast Cancer: Are They of Clinical Value? Clin Chem. 2006;52(3).
123. Slupsky CM, Steed H, Wells TH, Dabbs K, Schepansky A, Capstick V, et al. Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers. Clin Cancer Res. 2010;16(23):5835–41.
124. Lavra L, Catini A, Ulivieri A, Capuano R, Baghernajad Salehi L, Sciacchitano S, et al. Investigation of VOCs associated with different characteristics of breast cancer cells. Sci Rep. 2015;5:13246.
125. Koek MM, Jellema RH, van der Greef J, Tas AC, Hankemeier T. Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. Metabolomics. 2011;7(3):307–28.
126. Denkert C, Bucher E, Hilvo M, Salek R, Orešič M, Griffin J, et al. Metabolomics of human breast cancer: new approaches for tumor typing and biomarker discovery. Genome Med. 2012;4(4):37.
127. Silva CL, Passos M, Câmara JS. Solid phase microextraction, mass spectrometry and metabolomic approaches for detection of potential urinary cancer biomarkers--a powerful strategy for breast cancer diagnosis. Talanta. 2012;89:360–8.
128. Silva CL, Passos M, Câmara JS. Investigation of urinary volatile organic metabolites as potential cancer biomarkers by solid-phase microextraction in combination with gas chromatography-mass spectrometry. Br J Cancer. 2011;105(12):1894–904.
129. Musteata FM, Pawliszyn J. Bioanalytical applications of solid-phase microextraction. TrAC Trends Anal Chem. 2007;26(1):36–45.

130. Sun Y, Liu J, Kennedy JF. Application of response surface methodology for optimization of polysaccharides production parameters from the roots of Codonopsis pilosula by a central composite design. Carbohydr Polym. 2010;80(3):949–53.
131. Monteiro M, Carvalho M, Henrique R, Jerónimo C, Moreira N, de Lourdes Bastos M, et al. 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. Eur J Cancer. 2014;50(11):1993–2002.
132. Calejo I, Moreira N, Araújo AM, Carvalho M, Bastos M de L, de Pinho PG. Optimisation and validation of a HS-SPME–GC–IT/MS method for analysis of carbonyl volatile compounds as biomarkers in human urine: Application in a pilot study to discriminate individuals with smoking habits. Talanta. 2016;148:486–93.
133. Monteiro MS, Carvalho M, de Lourdes Bastos M, de Pinho PG. Biomarkers in renal cell carcinoma: a metabolomics approach. Metabolomics. 2014;10(6).
134. Xia J, Wishart DS, Xia J, Wishart DS. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. In: Current Protocols in Bioinformatics. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2016. p. 14.10.1-14.10.91.
135. Filipiak W, Filipiak A, Sponring A, Schmid T, Zelger B, Ager C, et al. Comparative analyses of volatile organic compounds (VOCs) from patients, tumors and transformed cell lines for the validation of lung cancer-derived breath markers. J Breath Res. 2014;8(2):027111.
136. Poli D, Carbognani P, Corradi M, Goldoni M, Acampa O, Balbi B, et al. Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study. Respir Res. 2005;6:71.
137. Ahmed I, Greenwood R, Costello B, Ratcliffe N, Probert CS. Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease. Aliment Pharmacol Ther. 2016;43(5):596–611.
138. Raman M, Ahmed I, Gillevet PM, Probert CS, Ratcliffe NM, Smith S, et al. Fecal Microbiome and Volatile Organic Compound Metabolome in Obese Humans With Nonalcoholic Fatty Liver Disease. Clin Gastroenterol Hepatol. 2013;11(7):868-875.e3.
139. Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. J Chemom. 2006;20(8–10):341–51.
140. Baker SG. The Central Role of Receiver Operating Characteristic (ROC) Curves in Evaluating Tests for the Early Detection of Cancer. JNCI J Natl Cancer Inst. 2003;95(7):511–5.

141. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Casp J Intern Med. 2013;4(2):627–35.
142. Janitza S, Hornung R. On the overestimation of random forest's out-of-bag error. Taguchi Y, editor. PLoS One. 2018;13(8):e0201904.
143. World Health Organization. 2018.
144. Asiago VM, Alvarado LZ, Shanaiah N, Gowda GAN, Owusu-Sarfo K, Ballas RA, et al. Early Detection of Recurrent Breast Cancer Using Metabolite Profiling. Cancer Res. 2010;70(21):8309–18.
145. Kim Y, Koo I, Jung BH, Chung BC, Lee D. Multivariate classification of urine metabolome profiles for breast cancer diagnosis. BMC Bioinformatics. 2010;11 Suppl 2(Suppl 2):S4.
146. Woo HM, Kim KM, Choi MH, Jung BH, Lee J, Kong G, et al. Mass spectrometry based metabolomic approaches in urinary biomarker study of women's cancers. Clin Chim Acta. 2009;400(1–2):63–9.
147. Nam H, Chung BC, Kim Y, Lee K, Lee D. Combining tissue transcriptomics and urine metabolomics for breast cancer biomarker identification. Bioinformatics. 2009;25(23):3151–7.
148. Henneges C, Bullinger D, Fux R, Friese N, Seeger H, Neubauer H, et al. Prediction of breast cancer by profiling of urinary RNA metabolites using Support Vector Machine-based feature selection. BMC Cancer. 2009;9(1):104.
149. Abaffy T, Möller M, Riemer DD, Milikowski C. A Case Report - Volatile Metabolomic Signature of Malignant Melanoma using Matching Skin as a Control. J Cancer Sci Ther. 2011;03(06):140–4.
150. Carrola J, Rocha CM, Barros AS, Gil AM, Goodfellow BJ, Carreira IM, et al. Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of urine. J Proteome Res. 2011;10(1):221–30.
151. Abaffy T, Möller MG, Riemer DD, Milikowski C, DeFazio RA. Comparative analysis of volatile metabolomics signals from melanoma and benign skin: a pilot study. Metabolomics. 2013;9(5):998–1008.
152. Filipiak W, Sponring A, Filipiak A, Ager C, Schubert J, Miekisch W, et al. TD-GC-MS analysis of volatile metabolites of human lung cancer and normal cells in vitro. Cancer Epidemiol Biomarkers Prev. 2010;19(1):182–95.
153. Matsumura K, Opiekun M, Oka H, Vachani A, Albelda SM, Yamazaki K, et al. Urinary Volatile Compounds as Biomarkers for Lung Cancer: A Proof of Principle Study Using Odor Signatures in Mouse Models of Lung Cancer. Pan X, editor. PLoS One. 2010;5(1):e8819.

154. Chen Y, Zhang R, Song Y, He J, Sun J, Bai J, et al. RRLC-MS/MS-based metabonomics combined with in-depth analysis of metabolic correlation network: finding potential biomarkers for breast cancer. Analyst. 2009;134(10):2003.
155. Oakman C, Tenori L, Biganzoli L, Santarpia L, Cappadona S, Luchinat C, et al. Uncovering the metabolomic fingerprint of breast cancer. Int J Biochem Cell Biol. 2011;43(7):1010–20.
156. Bu Q, Huang Y, Yan G, Cen X, Zhao Y-L. Metabolomics: a revolution for novel cancer marker identification. Comb Chem High Throughput Screen. 2012;15(3):266–75.
157. Santini G, Mores N, Penas A, Capuano R, Mondino C, Trové A, et al. Electronic Nose and Exhaled Breath NMR-based Metabolomics Applications in Airways Disease. Curr Top Med Chem. 2016;16(14):1610–30.
158. Figueira J, Jonsson P, Nordin Adolfsson A, Adolfsson R, Nyberg L, Öhman A. NMR analysis of the human saliva metabolome distinguishes dementia patients from matched controls. Mol Biosyst. 2016;12(8):2562–71.
159. Frank A, Möller HM, Exner TE. Toward the Quantum Chemical Calculation of NMR Chemical Shifts of Proteins. 2. Level of Theory, Basis Set, and Solvents Model Dependence. J Chem Theory Comput. 2012;8(4):1480–92.
160. Zhang A, Sun H, Qiu S, Wang X. NMR-based metabolomics coupled with pattern recognition methods in biomarker discovery and disease diagnosis. Magn Reson Chem. 2013;51(9):549–56.
161. Bingol K. Recent Advances in Targeted and Untargeted Metabolomics by NMR and MS/NMR Methods. High-throughput. 2018;7(2).
162. Bertram HC, Eggers N, Eller N. Potential of Human Saliva for Nuclear Magnetic Resonance-Based Metabolomics and for Health-Related Biomarker Identification. Anal Chem. 2009;81(21):9188–93.
163. Tiziani S, Lopes V, Günther UL. Early stage diagnosis of oral cancer using 1H NMR-based metabolomics. Neoplasia. 2009;11(3):269–76.
164. Louis E, Bervoets L, Reekmans G, De Jonge E, Mesotten L, Thomeer M, et al. Phenotyping human blood plasma by 1H-NMR: a robust protocol based on metabolite spiking and its evaluation in breast cancer. Metabolomics. 2015;11(1):225–36.
165. Louis E, Adriaensens P, Guedens W, Bigirumurame T, Baeten K, Vanhove K, et al. Detection of Lung Cancer through Metabolic Changes Measured in Blood Plasma. J Thorac Oncol. 2016;11(4):516–23.
166. Wang H, Wang L, Zhang H, Deng P, Chen J, Zhou B, et al. ^{1}H NMR-based metabolic profiling of human rectal cancer tissue. Mol Cancer. 2013;12(1):121.

167. Lindon JC, Holmes E, Lindon JC, Holmes E. NMR Spectroscopy of Biofluids, Tissues, and Tissue Extracts. In: eMagRes. Chichester, UK: John Wiley & Sons, Ltd; 2008.
168. Beckonert O, Monnerjahn J, Bonk U, Leibfritz D. Visualizing metabolic changes in breast-cancer tissue using1H-NMR spectroscopy and self-organizing maps. NMR Biomed. 2003;16(1):1–11.
169. Sitter B, Bathen TF, Singstad TE, Fjøsne HE, Lundgren S, Halgunset J, et al. Quantification of metabolites in breast cancer patients with different clinical prognosis using HR MAS MR spectroscopy. NMR Biomed. 2010;23(4):424–31.
170. Madhu B, Dadulescu M, Griffiths J. Artefacts in 1H NMR-based metabolomic studies on cell cultures. Magn Reson Mater Physics, Biol Med. 2015;28(2):161–71.
171. Aranibar N, Borys M, Mackin NA, Ly V, Abu-Absi N, Abu-Absi S, et al. NMR-based metabolomics of mammalian cell and tissue cultures. J Biomol NMR. 2011;49(3–4):195–206.
172. Bradley SA, Ouyang A, Purdie J, Smitka TA, Wang T, Kaerner A. Fermentanomics: Monitoring Mammalian Cell Cultures with NMR Spectroscopy. J Am Chem Soc. 2010;132(28):9531–3.
173. Kostidis S, Addie RD, Morreau H, Mayboroda OA, Giera M. Quantitative NMR analysis of intra- and extracellular metabolism of mammalian cells: A tutorial. Anal Chim Acta. 2017;980:1–24.
174. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0—The Human Metabolome Database in 2013. Nucleic Acids Res. 2012;41(D1):D801–7.
175. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, et al. The Human Urine Metabolome. Dzeja P, editor. PLoS One. 2013;8(9):e73076.
176. Monteiro MS, Barros AS, Pinto J, Carvalho M, Pires-Luís AS, Henrique R, et al. Nuclear Magnetic Resonance metabolomics reveals an excretory metabolic signature of renal cell carcinoma. Sci Rep. 2016;6:37275.
177. Cross AJ, Pollock JRA, Bingham SA. Haem, not protein or inorganic iron, is responsible for endogenous intestinal N-nitrosation arising from red meat. Cancer Res. 2003;63(10):2358–60.
178. Ohshima H, Kawabata T. Mechanism of N-nitrosodimethylamine formation from trimethylamine and trimethylaminoxide. IARC Sci Publ. 1978;(19):143–53.
179. Awwad HM, Geisel J, Obeid R. Determination of trimethylamine, trimethylamine N-oxide, and taurine in human plasma and urine by UHPLC–MS/MS technique. J Chromatogr B. 2016;1038:12–8.
180. Chalmers RA, Bain MD, Michelakakis H, Zschocke J, Iles RA. Diagnosis and management of trimethylaminuria (FMO3 deficiency) in children. J Inherit Metab Dis. 2006;29(1):162–72.

181. Mackay RJ, McEntyre CJ, Henderson C, Lever M, George PM. Trimethylaminuria: causes and diagnosis of a socially distressing condition. Clin Biochem Rev. 2011;32(1):33–43.
182. Zhao L, Dong M, Liao S, Du Y, Zhou Q, Zheng H, et al. Identification of key metabolic changes in renal interstitial fibrosis rats using metabonomics and pharmacology. Sci Rep. 2016;6(1):27194.
183. Barr DB, Wilder LC, Caudill SP, Gonzalez AJ, Needham LL, Pirkle JL. Urinary Creatinine Concentrations in the U.S. Population: Implications for Urinary Biologic Monitoring Measurements. Environ Health Perspect. 2004;113(2):192–200.
184. Nankivell BJ. Abnormal Laboratory Results: Creatinine clearance and the assessment of renal function. Aust Prescr. 2001;24(1):15–7.
185. Gowda S, Desai PB, Kulkarni SS, Hull V V, Math AAK, Vernekar SN. Markers of renal function tests. N Am J Med Sci. 2010;2(4):170–3.
186. Wang L, Tang Y, Liu S, Mao S, Ling Y, Liu D, et al. Metabonomic profiling of serum and urine by (1)H NMR-based spectroscopy discriminates patients with chronic obstructive pulmonary disease and healthy individuals. PLoS One. 2013;8(6):e65675.
187. Srivastava S, Roy R, Singh S, Kumar P, Dalela D, Sankhwar SN, et al. Taurine – a possible fingerprint biomarker in non-muscle invasive bladder cancer: A pilot study by 1H NMR spectroscopy. Cancer Biomarkers. 2010 Feb;6(1):11–20.
188. Jung J, Jung Y, Bang EJ, Cho S, Jang Y-J, Kwak J-M, et al. Noninvasive Diagnosis and Evaluation of Curative Surgery for Gastric Cancer by Using NMR-based Metabolomic Profiling. Ann Surg Oncol. 2014;21(S4):736–42.
189. Zhou J, Wang Y, Zhang X, Zhou J, Wang Y, Zhang X, et al. Metabonomics studies on serum and urine of patients with breast cancer using 1H-NMR spectroscopy. Oncotarget. 2017;5(0).
190. Singletary SE. Rating the Risk Factors for Breast Cancer. Ann Surg. 2003;237(4):474–82.
191. Venkitaraman AR. Cancer Susceptibility and the Functions of BRCA1 and BRCA2. Cell. 2002;108(2):171–82.
192. Yoshida K, Miki Y. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. Cancer Sci. 2004;95(11):866–71.
193. Mangone FR, Miracca EC, Feilotter HE, Mulligan LM, Nagai MA. ATM gene mutations in sporadic breast cancer patients from Brazil. Springerplus. 2015;4(1):23.
194. Narod SA. Early-onset breast cancer: what do we know about the risk factors?: A Countercurrents Series. Curr Oncol. 2011;18(5):204–5.
195. Ali I, A. Al-Othman Z, Saleem K, Hussain A, Hussain I. Role of Chromatography for Monitoring of Breast Cancer Biomarkers. Recent Patents Biomarkerse [Internet]. 2011 [cited

2016 Feb 23];1(1):89–97.

196. Suzuki M, Nishiumi S, Matsubara A, Azuma T, Yoshida M. Metabolome analysis for discovering biomarkers of gastroenterological cancer. J Chromatogr B. 2014;966:59–69.
197. Wu W, Zhao S. Metabolic changes in cancer: beyond the Warburg effect. Acta Biochim Biophys Sin (Shanghai). 2013;45(1):18–26.
198. Haick H, Amal, Ding, Liu, Tisch, Xu, et al. The scent fingerprint of hepatocarcinoma: in-vitro metastasis prediction with volatile organic compounds (VOCs). Int J Nanomedicine. 2012;7:4135–46.
199. Hakim M, Broza YY, Barash O, Peled N, Phillips M, Amann A, et al. Volatile organic compounds of lung cancer and possible biochemical pathways. Chem Rev. 2012;112(11):5949–66.
200. Filipiak W, Sponring A, Mikoviny T, Ager C, Schubert J, Miekisch W, et al. Release of volatile organic compounds (VOCs) from the lung cancer cell line CALU-1 in vitro. Cancer Cell Int. 2008;8(1):17.
201. Hanai Y, Shimono K, Oka H, Baba Y, Yamazaki K, Beauchamp GK. Analysis of volatile organic compounds released from human lung cancer cells and from the urine of tumor-bearing mice. Cancer Cell Int. 2012;12(1):7.
202. Zhang Y, Gao G, Liu H, Fu H, Fan J, Wang K, et al. Identification of Volatile Biomarkers of Gastric Cancer Cells and Ultrasensitive Electrochemical Detection based on Sensing Interface of Au-Ag Alloy coated MWCNTs. Theranostics. 2014;4(2):154–62.
203. Pawliszyn J. Theory of Solid-Phase Microextraction. In: Handbook of Solid Phase Microextraction. Elsevier; 2012. p. 13–59.
204. Li J, Peng Y, Duan Y. Diagnosis of breast cancer based on breath analysis: an emerging method. Crit Rev Oncol / Hematol [Internet]. 2013 Jul [cited 2016 May 17];87(1):28–40.
205. Kwak J, Gallagher M, Ozdener MH, Wysocki CJ, Goldsmith BR, Isamah A, et al. Volatile biomarkers from human melanoma cells. J Chromatogr B. 2013 Jul 15;931:90–6.
206. Huang Y, Li Y, Luo Z, Duan Y. Investigation of biomarkers for discriminating breast cancer cell lines from normal mammary cell lines based on VOCs analysis and metabolomics. RSC Adv. 2016;6(48):41816–24.
207. Cederbaum AI. Alcohol metabolism. Clin Liver Dis. 2012 Nov;16(4):667–85.
208. Peled N, Barash O, Tisch U, Ionescu R, Broza YY, Ilouze M, et al. Volatile fingerprints of cancer specific genetic mutations. Nanomedicine Nanotechnology, Biol Med. 2013;9(6):758–66.
209. Davies MPA, Barash O, Jeries R, Peled N, Ilouze M, Hyde R, et al. Unique volatolomic

signatures of TP53 and KRAS in lung cells. Br J Cancer. 2014;111(6):1213–21.

210. Wang C, Sun B, Guo L, Wang X, Ke C, Liu S, et al. Volatile Organic Metabolites Identify Patients with Breast Cancer, Cyclomastopathy, and Mammary Gland Fibroma. Sci Rep. 2014;4:697–702.
211. Phillips M, Cataneo RN, Saunders C, Hope P, Schmitt P, Wai J. Volatile biomarkers in the breath of women with breast cancer. J Breath Res. 2010;4(2):026003.
212. Brunner C, Szymczak W, Höllriegl V, Mörtl S, Oelmez H, Bergner A, et al. Discrimination of cancerous and non-cancerous cell lines by headspace-analysis with PTR-MS. Anal Bioanal Chem. 2010;397(6):2315–24.
213. Mochalski P, Sponring A, King J, Unterkofler K, Troppmair J, Amann A. Release and uptake of volatile organic compounds by human hepatocellular carcinoma cells (HepG2) in vitro. Cancer Cell Int. 2013 Jan;13(1):72.
214. Mochalski P, Leja M, Gasenko E, Skapars R, Santare D, Sivins A, et al. Ex vivo emission of volatile organic compounds from gastric cancer and non-cancerous tissue. J Breath Res. 2018;12(4):046005.
215. Janocha S, Schmitz D, Bernhardt R. Terpene Hydroxylation with Microbial Cytochrome P450 Monooxygenases. In 2015. p. 215–50.
216. Costello Z, Martin HG. A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. npj Syst Biol Appl. 2018;4(1):19.
217. Chen J, Liu Z, Fan S, Yang D, Zheng P, Shao W, et al. Combined Application of NMR- and GC-MS-Based Metabonomics Yields a Superior Urinary Biomarker Panel for Bipolar Disorder. Sci Rep. 2015;4(1):5855.
218. Dona AC, Kyriakides M, Scott F, Shephard EA, Varshavi D, Veselkov K, et al. A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. Comput Struct Biotechnol J. 2016;14:135–53.
219. Trifonova O, Lokhov P, Archakov A. Postgenomics diagnostics: metabolomics approaches to human blood profiling. OMICS. 2013;17(11):550–9.
220. Nicholson JK, Lindon JC, Holmes E. "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;29(11):1181–9.
221. Marshall DD, Powers R. Beyond the paradigm: Combining mass spectrometry and nuclear magnetic resonance for metabolomics. Prog Nucl Magn Reson Spectrosc. 2017;100:1–16.
222. Robertson DG, Watkins PB, Reily MD. Metabolomics in Toxicology: Preclinical and Clinical Applications. Toxicol Sci. 2011;120(Supplement 1):S146–70.

223. Marshall DD, Lei S, Worley B, Huang Y, Garcia-Garcia A, Franco R, et al. Combining DI-ESI–MS and NMR datasets for metabolic profiling. Metabolomics. 2015;11(2):391–402.
224. Falegan O, Ball M, Shaykhutdinov R, Pieroraio P, Farshidfar F, Vogel H, et al. Urine and Serum Metabolomics Analyses May Distinguish between Stages of Renal Cell Carcinoma. Metabolites. 2017;7(1):6.
225. Statheropoulos M, Sianos E, Agapiou A, Georgiadou A, Pappa A, Tzamtzis N, et al. Preliminary investigation of using volatile organic compounds from human expired air, blood and urine for locating entrapped people in earthquakes. J Chromatogr B. 2005;822(1–2):112–7.
226. Gray LR, Tompkins SC, Taylor EB. Regulation of pyruvate metabolism and human disease. Cell Mol Life Sci. 2014;71(14):2577–604.
227. Mediani A, Abas F, Maulidiani M, Abu Bakar Sajak A, Khatib A, Tan CP, et al. Metabolomic analysis and biochemical changes in the urine and serum of streptozotocin-induced normal- and obese-diabetic rats. J Physiol Biochem. 2018;74(3):403–16.
228. Jagannathan NR, Sharma U. Breast Tissue Metabolism by Magnetic Resonance Spectroscopy. Metabolites. 2017;7(2).
229. WANG D, WANG C, PI X, GUO L, WANG Y, LI M, et al. Urinary volatile organic compounds as potential biomarkers for renal cell carcinoma. Biomed Reports. 2016;5(1):68–72.
230. Budczies J, Denkert C. Tissue-Based Metabolomics to Analyze the Breast Cancer Metabolome. In: Recent results in cancer research Fortschritte der Krebsforschung Progres dans les recherches sur le cancer. 2016. p. 157–75.


UNIVERSIDADE da MADEIRA
CQM
CENTRO DE QUÍMICA DA MADEIRA
SESARAM E.P.E.
Serviço de Saúde da RAM E.P.E.
OPERAÇÃO CQM+
OPERATION CQM+
INNO INDIGO
FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR
arditi
agência regional para o desenvolvimento da investigação tecnologia e inovação
MADEIRA 14-20
Programa Operacional da Região Autónoma da Madeira
PORTUGAL 2020
COMPETE 2020
UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional