DM

MALDI-TOF MS-based urinary proteomic/peptidomic signature of breast cancer as innovative diagnostic approach

MASTER DISSERTATION

Patrícia Maria Gonçalves Sousa
MASTER IN APPLIED BIOCHEMISTRY

UNIVERSIDADE da MADEIRA
A Nossa Universidade
www.uma.pt

September | 2022

MALDI-TOF MS-based urinary proteomic/peptidomic signature of breast cancer as innovative diagnostic approach

MASTER DISSERTATION

Patrícia Maria Gonçalves Sousa
MASTER IN APPLIED BIOCHEMISTRY

ORIENTATION
Rosa Maria de Sá Perestrelo Gouveia

CO-ORIENTATION
José de Sousa Câmara

MALDI-TOF MS-based urinary proteomic/peptidomic signature of breast cancer as innovative diagnostic approach

PATRÍCIA MARIA GONÇALVES SOUSA

Master's in applied Biochemistry

Dissertation presented to the UNIVERSITY OF MADEIRA to meet the requirements necessary to obtain a Master´s degree in Applied Biochemistry, carried out under the scientific guidance of Doctorate Rosa Maria de Sá Perestrelo Gouveia, Researcher of the Center of Chemistry of Madeira, and Professor Doctorate José de Sousa Câmara, Assistant Professor with Habilitation of Center of Chemistry of Madeira.

Adviser: Rosa Maria de Sá Perestrelo Gouveia
Postdoctoral Senior Researcher, Centro de Química da Madeira

Co-adviser: José de Sousa Câmara
Assistant Professor with Habilitation, Faculdade de Ciências Exatas da Universidade da Madeira

Universidade da Madeira

Setembro 2022

Abstract

The development of a rapid and high-sensitive breast cancer (BC) diagnostic method has been increasingly investigated by many researchers, to have a more effective therapy, through the early detection of this disease that affects millions of women worldwide. Therefore, the discovery of specific biomarkers is one of the topics in clinical medicine, however several challenges remain in the development of optimal sample preparation for proteomic/peptidomic analysis of urine, due to its highly variable content, as well as the presence of various proteins in low abundance or modified forms. Thus, this investigation aims to establish potential BC urinary protein biomarkers using one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (1D SDS-PAGE) coupled with matrix-assisted laser desorption/ionization time-of-flight mass spectrometric (MALDI-TOF MS) for BC diagnosis and monitoring the efficiency of therapy. The results of the Lowry´s assay demonstrated that total protein concentration increased after precipitation and that healthy control group (HCs) have higher total protein concentrations than BC patients. Related to MALDI-TOF MS analysis, results revealed that four protein/peptide ion signatures (m/z 1046.5, 1062.4, 1237.7 and 1727.9), with variable importance in projection (VIP) > 1, allowed the discrimination between BC patients and HCs. The discrimination efficiency and accuracy of BC urine proteins/peptides were ascertained by receiver operating characteristic (ROC) curve analysis that allowed the identification of some features with high sensitivity and specificity (99.6%). Therefore, the obtained data revealed MALDI-TOF MS as a powerful tool to explore proteomic/peptidomic biosignatures, due to its speed, sensitivity, and mass accuracy, which allow establishing novel disease biomarkers. Nevertheless, a deep study using a higher number of samples must be carried out to confirm and consolidate the data obtained.

Keywords: Breast cancer, urine, 1D SDS-PAGE, MALDI-TOF MS, biomarkers, proteomic/peptidomic.

Resumo

O desenvolvimento de um método de diagnóstico rápido e de alta sensibilidade e especificidade para o cancro da mama (BC) tem sido um desafio crescente para muitos investigadores. O diagnóstico precoce permitirá uma maior eficiência da terapia, duma patologia que afeta milhões de mulheres em todo o mundo. A descoberta e identificação de biomarcadores específicos é um dos tópicos da medicina clínica, porém ainda subsistem vários desafios no desenvolvimento de técnicas otimizadas de preparação da amostra para a análise proteómica/peptidómica da urina, devido ao seu conteúdo altamente variável, bem como à presença de várias proteínas em baixa abundância ou sob formas modificadas. Assim, com este estudo objetivou-se identificar potenciais biomarcadores do BC com base nos perfis de proteína urinária de pacientes BC (n=56) e grupo de controlo saudável (HCs, n=54), utilizando a eletroforese em gel de poliacrilamida-dodecil sulfato de sódio unidimensional (1D SDS-PAGE) combinada com a espetrometria de massa por ionização e dessorção a laser assistida por matriz (MALDI-TOF MS), como estratégia útil para diagnóstico e evolução da terapia. Os resultados do ensaio de Lowry demonstraram que a concentração total de proteína aumentou após a precipitação e que o HCs tem concentrações de proteína total mais elevadas que os pacientes BC. Relacionados com a análise MALDI-TOF MS, os resultados revelaram que quatro assinaturas de iões de proteína/peptídeo (m/z 1046.5, 1062.4, 1237.7 and 1727.9), com importância variável na projeção (VIP) > 1, permitiram a discriminação entre pacientes BC e HCs. A eficiência e precisão da discriminação do padrão de proteínas/peptídeos na urina de pacientes BC foi verificada pela análise da curva caraterística de operação do recetor (ROC) que permitiu a identificação de algumas caraterísticas com alta sensibilidade e especificidade (99.6%). Assim, os dados obtidos revelaram a MALDI-TOF MS como uma ferramenta poderosa para explorar o padrão urinário proteómico/peptidómico, e desta forma discriminar entre diferentes grupos em estudo possibilitando a deteção de potenciais biomarcadores de uma forma relativamente rápida altamente sensível e com elevada precisão de massa. No entanto, deverá ser realizado um estudo aprofundado utilizando um maior número de amostras a fim de confirmar e consolidar os dados obtidos.

Palavras-chave: cancro da mama, urina, 1D SDS-PAGE, MALDI-TOF MS, biomarcadores, proteómica/peptidómica.

Acknowledgements

First, I would like to thank Doctorate Rosa Perestrelo for accepting to be my supervisor throughout the realization of this thesis, that wouldn't have been possible without this vote of confidence. Thank you for all the teachings, support, and friendship.

To Professor Doctorate José Câmara, my co-supervisor, for the help and incentives during the realization of this thesis.

To Professor Irene Camacho with whom, despite not having worked directly, she always had a friendly word and, mainly, was always available to help me in the analytical part.

Thanks to my lab seniors for the friendship and good moments spent together during work hours. I learned a lot from all of you!

Special thanks to my family, for accompanying me in this process, mainly to my parents and sister who collaborated for my good performance and, although it has been a complex and challenging route, they didn´t let me give up during all the process of realization.

To my close friends, mostly Fátima Oliveira, who was for sure an indispensable help, contributing with support and companionship.

To all those involved in my life during the performance of this work, my sincere thanks! From now on I'll work hard, and I'll always do my best to be the Patrícia Sousa who only gives pride and happiness.

This work was supported by FCT-Fundação para a Ciência e a Tecnologia (Base Fund UIDB/00674/2020), and Programmatic Fund UIDP/00674/2020, Portuguese Government funds), and by ARDITI-Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação, through the project M1420-01-0145-FEDER-000005 – CQM^{+} (Madeira 14-20 Program).

Contents

List of figures

List of tables

Abbreviations

1D SDS-PAGE: One-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis

2DE: Two-dimensional gel electrophoresis

AKI: Acute kidney injury

AUC: Area under the curve

BC: Breast cancer

BSA: Bovine serum albumin

CA15-3: Cancer antigen 15-3

CEA: Carcinoembryonic antigen

CHCA: α-cyano-4-hydroxycinnamic acid

ConA: Concanavalin A

COVID-19: Coronavirus disease 2019

CV: Coefficient of variation

Da: Daltons

DHB: 2,5-dihydroxybenzoic acid

DNA: Deoxyribonucleic acid

ER: Estrogen receptor

FWHM: Full width at half-maximum

GC: Gas chromatography

GLOBOCAN: Global cancer

HCA: Hierarchical cluster analysis

HCs: Healthy control group

HER 2: Human epidermal growth factor receptor 2

HNM: Hospital Dr. Nélio Mendonça

HAS: Human serum albumin

HWS: Half window size

IARC: International agency research on cancer

IEF: Isoelectric focusing

LC-MS/MS: Liquid chromatography tandem mass spectrometry

LOD: Limit of detection

LOQ: Limit of quantification

Lum: Luminal

Lys-C: Endoproteinase LysC

m/z: mass-to-charge ratio

MALDI-TOF MS: Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry

MBC: Male breast cancer

MS/MS: Tandem mass spectrometry

MW: Molecular weight

NMR: Nuclear magnetic resonance

OPLS-DA: Orthogonal partial least squares-discriminant

PCA: Principal component analysis

PIP: Prolactin-induced protein

PLS-DA: Partial least squares discriminate analysis

PR: Progesterone receptor

PRM: Pyrogallol red molybdate

RNA: Ribonucleic acid

ROC: Receiver operating characteristic

SA: 3,5-dimethoxy-4-hydroxycinnamic acid (or sinapic acid)

SELDI-TOF MS: Surface-enhanced laser desorption/ionization time-of-flight mass spectrometry

SNIP: Statistics-sensitive non-linear iterative peak-clipping

SNR: Signal-to-noise ratio

TIC: Total ion current

UV-vis: Ultraviolet-visible

WHO: World health organization

1. INTRODUCTION

Cancer is a leading cause of death and an important barrier to increasing life expectancy in every country [1], which represents one of the major public health concerns worldwide. It is characterized by a large group of heterogeneous diseases that can affect any part of the body, usually sites such as the breast, lung, colon and rectum, prostate, liver, skin and stomach. Cancer comprises a large group of related diseases that can start in almost any organ or tissue of the body, when cells produce new cells, and the old or abnormal ones don't die when they should. As these abnormal cells grow out of control, they can crowd out normal cells and spread into surrounding tissues, to invade adjoining parts of the body and/or extend to other organs [2].

According to World Cancer Research Fund International [3], an estimated 18.1 million cancer cases around the world in 2020. Of these, 9.3 million cases were in men and 8.8 million in women, being the most predominant the lung, prostate and colorectum cancers in men, and lung, breast and colorectum cancers in women.

In this sequence, the global cancer burden is expected to reach around 3 millions of diagnosed cases by 2040, a 47% rise from 2020, with a larger increase in transitioning (64% to 95%) versus transitioned (32% to 56%) countries due to demographic changes, although this may be further exacerbated by increasing risk factors associated with globalization and a growing economy [4].

However, even though cancer incidence and mortality has increased, part is due to the coronavirus disease 2019 pandemic (COVID-19), which has negatively affected cancer diagnosis and treatment. The closure of health care setting and the fear of exposure to the virus has led to reduced access to care and, consequently, a short-term drop in cancer incidence, followed by an increase in the mortality rate [5].

1.1. Breast cancer

According to GLOBOCAN series of the International Agency for Research on Cancer (IARC) [4], the breast cancer (BC) is the most commonly cancer in women, with an estimated 2.3 million new cases (11.7%), followed by lung (11.4%), colorectal (10.0%), prostate (7.3%) and stomach (5.6%) cancers, and the second most common cause of cancer death in women, which comprises 6.9% deaths (Figure 1).

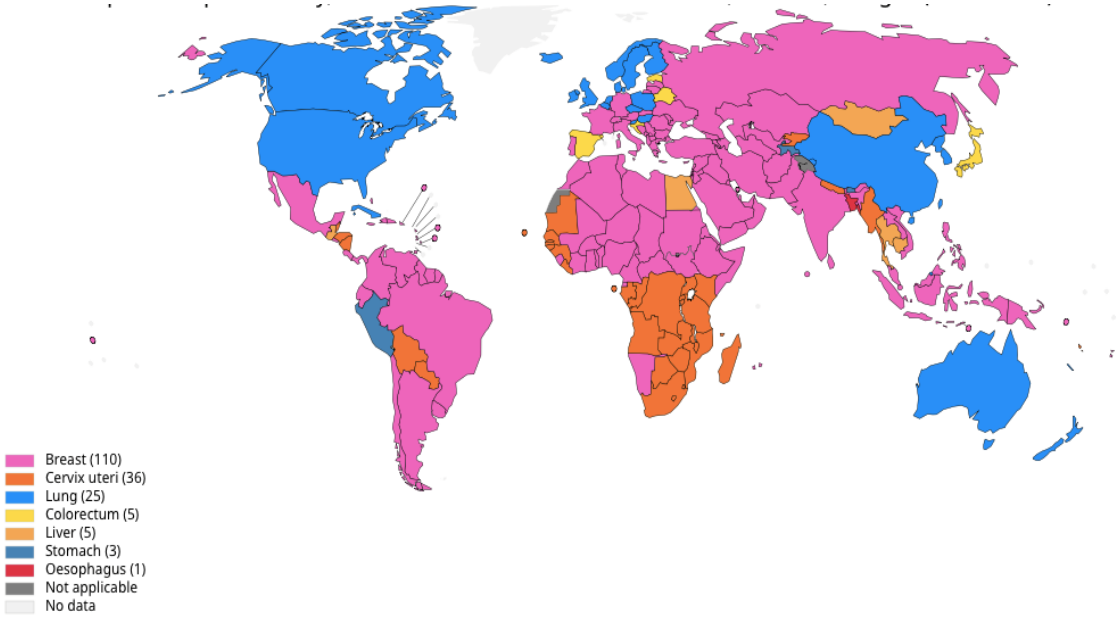


Figure 1 – Leading cause of cancer deaths among women worldwide, 2020. SOURCE: GLOBOCAN 2020.

The breast is composed of two main types of tissues, glandular tissues and stromal (supporting) tissues, as well as a lymphatic tissue-immune system tissue that removes cellular fluids and waste [7]. There are several types of tumors that may develop within different areas of the breast, being a large part of a benign character, for example fibrocystic change, a non-cancerous condition in which women develop cysts (accumulated packets of fluid), fibrosis (formation of scar-like connective tissue), lumpiness, and areas of thickening, tenderness, or breast pain [2]. However, in the case of a malignant tumor, it can lead to cancers, from which it usually begins in the cells that line the ducts or even in the lobules, while a small number start in the other tissues [11].

This type of cancer refers to cancers originating from breast tissue, most commonly from the inner lining of milk ducts or the lobules that supply the ducts with milk [6] and is more often in women than in men, although the rarity in men is due to delays in diagnosis which leads to frequent detection of the disease in more advanced stages [7].

Male BC is a rare medical condition, accounting for only about 1% of all BCs. Statistics from the American Cancer Society suggest that yearly, about 2600 new cases of BC in men are diagnosed causing, approximately 440 deaths (in comparison, almost 40000 women die of breast cancer each year) [8]. Even though breast tissue is very

similar in both sexes, composed essentially of fatty tissue, breast cells and ducts, men´s breasts have fewer ducts and lobules compared to women, since women´s breasts mature during puberty and, consequently, they develop working lobules and milk ducts to produce and carry milk after childbirth.

In other words, even if women have many more breast cells than men, the main reason they develop more BC is because their breast cells are constantly exposed to the growth-promoting effects of the female hormone's estrogen and progesterone.

Even though there are several diagnostic and screening tools available for , namely mammography, ultrasound, or tumor markers, which help the detection and subsequent improvement in survival rates, certain limitations to combating it persist. The main handicap of most of these methods is their low sensitivity and specificity. In addition, over detection may cause harm through unnecessary labelling and treatment of patients as having a cancer that, without screening, might never have been diagnosed [9].

BC screening refers to testing otherwise-healthy women for BC in an attempt to achieve an earlier diagnosis under the assumption that early detection will improve outcomes. Several screening tests have been employed including clinical and self-breast exams, mammography, genetic screening, ultrasound, and magnetic resonance imaging.

While biopsies are the definitive tests for diagnosing cancer, they are invasive, and can be uncomfortable and anxiety-inducing. Thus, instead of immediate biopsy after an abnormal screening, new technologies to diagnose breast cancer should focus on decreasing discomfort as well as increasing test accuracy [10], which they could have a large economic potential and quality of life impact. In addition, the World Health Organization (WHO) recommends organized, population-based mammography screening every 2 years for women at average risk for BC aged 50 to 69 years in well-resourced settings [11].

For this reason, the increasing worldwide incidence of BC disease requires the need to research multidisciplinary approaches in different areas, as well as investigate new clinical diagnostic tests to improve the efficacy of therapies and increase survival rates.

Proteomics emerged as a promising approach in disease profiling for the pursuit of new biomarkers in biological matrices, such as cell extracts, tissues, or biological fluids, which improve our chances for developing diagnostic markers for early detection of cancer and maybe provided the solution that we are looking for.

One of the biggest challenges in conducting proteomic research has been to discover and characterize biomarkers that could be used for better identification of patients who are at risk of the disease becoming more aggressive or for disease progression that may provide targets for clinical intervention or even for the earlier administration of proper treatment [12]. These biomarkers can serve as early warning indicators for disease, help to monitor disease progression, and predict receptivity to treatment.

Thus, a considerable advancement in proteomics is the quantification of biomarkers, with high sensitivity and specificity, provided by new and powerful platforms, from body fluids (e.g., urine, blood, seminal fluid, saliva) and tissues [13], with the objective of favoring not only physicians in clinical decision-making, but also the patient, to promote the diagnosis and follow-up of therapy [14]. The Figure 2 summarizes the evolution of BC research and illustrates how technological progress and scientific focus have worked in parallel to support the discovery of markers, through new technologies, for the detection of BC.

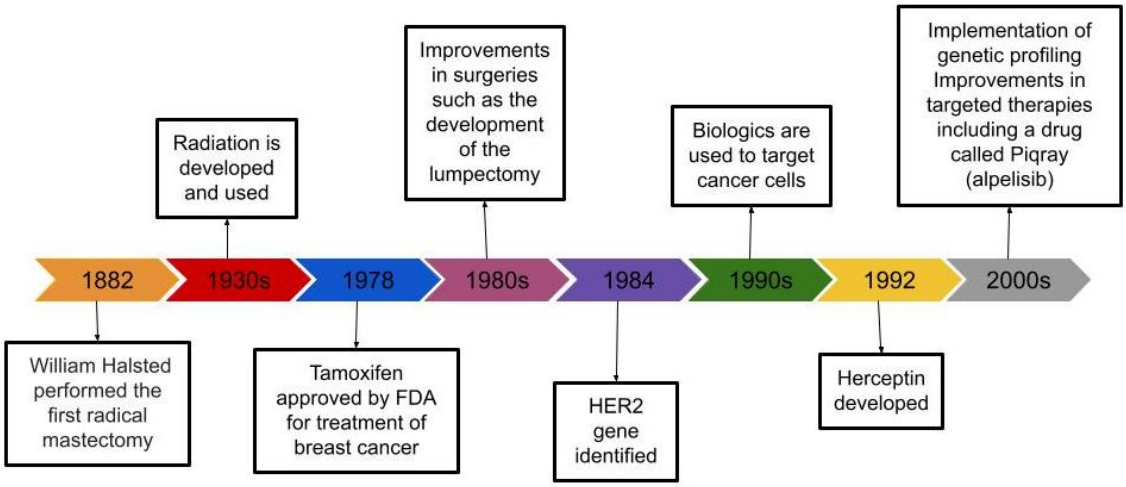


Figure 2 – Timeline of technological advances and scientific progress in BC.

1.2. Search for BC diagnostic biomarkers using high-throughput technologies

In recent years, all “omics” disciplines have demonstrated potential interest as large-scale biomarker discovery platforms. These include genomic, transcriptomic, proteomic, metabolomic, imaging and all variants thereof [15].

Genomics aims to characterize and quantify the complete set of all genes within a cell of an organism, to provide information about the influence of their interactions [16]. Thus, the approach consists of measuring gene expression through methods based

on high-throughput DNA sequencing, bioinformatics, genetic analysis and functional identification to analyze the structure and function of the whole genome, which forms the “omics” framework of systems biology [17].

In the study of biomarkers, genomics possesses an enormous potential to allow the discovery of new biomarkers and diagnostic tests. However, the analysis is limited by its inability to provide complete information on cellular, subcellular and intercellular functions [18]. These limitations have led to the development in the biology of interdisciplinary systems that integrate genomic/epigenomic, transcriptomic, proteomic and metabolomic data.

Since genomic changes may not necessarily lead to biological phenotypic changes, the transcriptome study, which translates into the complete set of RNA transcripts produced by the genome in a specific cell, has played an important role in the characterization of gene expression at the level of mRNA, since the actively transcribed RNA is highly dynamic and reflects the diversity of cell types and their regulatory mechanisms [19]. Thus, transcriptomics, through RNA sequencing, identifies and quantifies differentially expressed genes of cancer cells, whose transcriptional patterns are anomalous and, therefore, could discover biomarkers for the disease [20]. Additionally, this technology emphasizes the concept of time and space, which allows not only to identify the phenotypic attribution of cells, but also to distinguish subtypes of diseases and discriminate the various reactions caused by drugs, as well as to describe the survival rates of patients [21].

Although transcripts are dynamic, they don´t truly reflect the functional phenotype of a cell as they are not their final genetic products [15], that is, they are not able to capture the full spectrum of protein-encoding genes, so proteins and metabolites are closer to the active processes of the disease than DNA and RNA.

Initially, metabolomic studies were limited in number, due to technical difficulties regarding metabolite stability, metabolite identification and the potential introduction of sample variation due to the collection and preparation [22]. However, in recent years, metabolomic approaches have been increasingly used to identify new diagnostic biomarkers for numerous types of cancer, including BC, as the study of bioactive metabolites allows for directly regulate biological processes and phenotypes.

This technology performs qualitative and quantitative analysis of all low molecular weight metabolites (<1000 Da) in a specific physiological period [23] through high-throughput mass spectrometry (MS), cluster index analysis and data

processing, combined with information modelling and systems integration, to screen and identify the differential metabolites correlated to the disease phenotype [24].

The main technology in metabolomics to study bodily fluid samples was gas chromatography-mass spectrometry (GC-MS). Zhou *et al.* [25] used this technology to detect 76 urinary metabolites in 50 patients with bladder cancer and 35 HCs. Through the validation of differential metabolites in a urine cluster of 96 individuals, a panel of combinatorial biomarkers was useful for the early diagnosis of bladder cancer. Park *et al* [26] investigated, via plasma metabolomic profiles, potential biomarkers for the early diagnosis of BC through the MS technique coupled to liquid chromatography (LC). In another study, Xu *et al.* [27] demonstrated that the combination of targeted and non-targeted metabolomic methods can offer advantages in batch metabolite detection and show higher sensitivity and high throughput performance in the known qualitative and quantitative analysis of metabolites. In this way, using an ultra-high-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) system operated in the multiple reaction monitoring (MRM) mode, was proposed for urine metabolomics analysis.

Additionally, metabolic change is an important feature of cancer, i.e., proliferating cancer cells exhibit a different metabolic behavior from normal differentiated cells, thus allowing a better understanding of the metabolic pathways in various types of cancer. This means that, to maintain proliferation and survival after treatment, cancer cells must adjust their metabolism and nutritional needs [28] and, consequently, can further affect the expression of cell surface markers through a variety of functional signalling pathways [29]. For this reason, metabolomics studies are not limited to the discovery of specific biomarkers, but also to the exploration of characteristic pathological pathways and therapeutic targets that may provide a more accurate direction for the development of novel medications [30, 31].

Compared to other types of “omics” data, proteomics data have the advantage of providing a quantitative abundance of individual proteins across the cognate and dynamic variables which evolve over time and among individual subjects [32].

Hence, as proteins are the functional molecules in an organism, responsible for numerous biological processes, and can be most ubiquitously affected in disease, therapy response, and recovery, proteomics holds special promise in detecting pathological conditions, predicting the efficacy of treatment, and tailoring personalized medicine [33].

High-throughput proteomics is still an underdeveloped field compared to genomics and transcriptomics, which represents a major obstacle for protein-based studies to drastically alter the clinical approach to breast cancer, and its contribution to oncology is probably not yet fully understood.

However, in the last two decades, quantitative proteomics has rapidly evolved both technologically and strategically, allowing researchers to explore the complexity of protein interaction networks in a wide variety of situations, but also to formulate new hypotheses to be further functionally tested [20].

Although there are so many approaches, which can be made for the treatment of BC such as surgery, radiation therapy chemotherapy, hormonal therapy and recently nanotechnology and gene therapy, it becomes increasingly necessary to develop even more effective screening and treatment programs.

1.2.1. Proteomic technology

Advances in the field of genomics have directed to the improvement of the proteomic-based strategy in the study of diseases since its constant evolution has allowed a better understanding of their molecular basis.

Macromolecules, in general, and proteins are highly dynamic molecules [34], so while the genome is essentially static, the proteome of an organism is more complex and dynamic [35], which varies according to cell type and functional state. That is, whereas the genetic code is made up of four nucleotides and the sequence of these is identical in every cell of an organism, proteins are built from 20 different amino acids, which undergo post-translational modifications (PTMs) and numerous possibilities of protein-protein interactions, besides that they come in different isoforms [36].

Proteins perform most of the work of living cells, since they are involved in the reading, copying, and organizing of genetic code in the DNA, as well as being key agents in almost all biological processes, such as metabolism, bio signalling, gene regulation, protein synthesis, solute transport cross membranes, immune function, and photosynthesis, which provide the structural scaffolding in cells.

Mechanistically, proteins can be subjected to extensive functional regulation by various processes such as proteolytic degradation, PTMs, involvement in complex structures, and compartmentalization [34]. Thus, understand how cells works requires understanding how proteins function [37]. That being the case, it is necessary to

highlight this large group of biomolecules, which is seen as new targets in understanding cancer processes, because they are effectors that drive cellular behavior and thence potential candidates for biomarkers.

The term “proteomics” was first used by Marc Wilkins, in 1996, to denote the need for new terminology that could reflect the growing interest in studying the entire protein complement of a genome [38], to indicate all time and condition-specific proteins that are simultaneously produced by a cell or a tissue [35].

This term results as a “product” of the genomic era, which can be seen as a reflection of a new way of understanding and approaching all descriptive biological sciences, thanks to the dynamic role of proteins, molecules that are essential to sustain life. Therefore, proteomics emerged as a different approach based on discoveries of genomic technology that promised to provide a more accurate picture of the impact of specific genes and proteins on health and disease [39].

Proteomics is the large-scale study of the proteome, which refers to the complete set of proteins, encoded by the genome, expressed by a cell or tissue, including its location, interactions, PTMs and turnover, at a particular time. Then, the proteome is an ever-changing swarm of modified proteins that differs from cell to cell, which comprises the changes or modifications produced in the native protein when organisms are subjected to a multitude of changes.

In this way, it is possible to reveal the biological status of a given sample, such as cultured cells, a tissue, or an organism, through the identification of its proteomic profiles, as well as its regulatory molecular mechanisms or disease.

Currently, proteomics has focused on an aggregate of independent proteins related to a specific disease, since it is less susceptible to the influence of genetic and environmental “noise” than the level of a single marker protein [40]. At the same time, the search for diagnostic biomarkers requires the measurement of a high number of proteins in several samples, followed by a confirmation of the differential abundance of proteins in clinically useful samples [33].

As protein biomarkers more accurately reflect the pathogenic phenotype, seeing that they are the endpoint of biological processes [40], a proteome profile, including proteins and their abundances, provides a better overview of the physiological and biochemical states of a sample than static DNA information [40], making it possible to develop clinical trials capable of predicting the ideal or targeted therapeutic approach

for some types of cancer. Therefore, they are considered reliable predictors of the state of the disease and the clinical outcome.

Additionally, cancer cells provide the biomarker material that can lead to their own detection, as biomarker proteins, both overly abundant proteins and variant proteins, can be detectable in the circulation as the free, shed proteins or as new autoantibodies to such proteins, so to reveal the presence of tumors or the level of tumor burden [41].

Despite considerable efforts directed at early detection, no single marker or combination of markers provides the sensitivity and specificity necessary for the early detection of BC. Therefore, MS-based proteomic methods should be improved to explore proteins with lower abundance in biological samples and, consequently, obtain biomarker candidates.

In fact, the biggest hurdle in biomarker research is to translate the laboratory findings into real clinical applications through rigorously, well-designed clinical trials [42].

1.2.1.1. Biomarkers discovered through proteomics

Recently, proteomics has been investigating a wide range of techniques and strategies to discover the change in the internal environment in the human body, with the aim to provide results with valuable information that can serve the clinician for high-throughput diagnostic and prognostic applications. For this reason, it is considered one of the dynamic and innovative tools that can confirm, complement, or even provide more elaborate information [34], aiming to understand the relationship between regulatory mechanisms of proteins at the molecular level and phenotypic behavior, such as development and progression of the disease. In addition, an essential goal for applying proteomics to study cancers is to adapt its high-throughput tools for regular use in clinical laboratories for the purpose of diagnostic and prognostic categorization of cancers, as well as in assessing various cancer therapeutic regimens (Figure 3) [34].

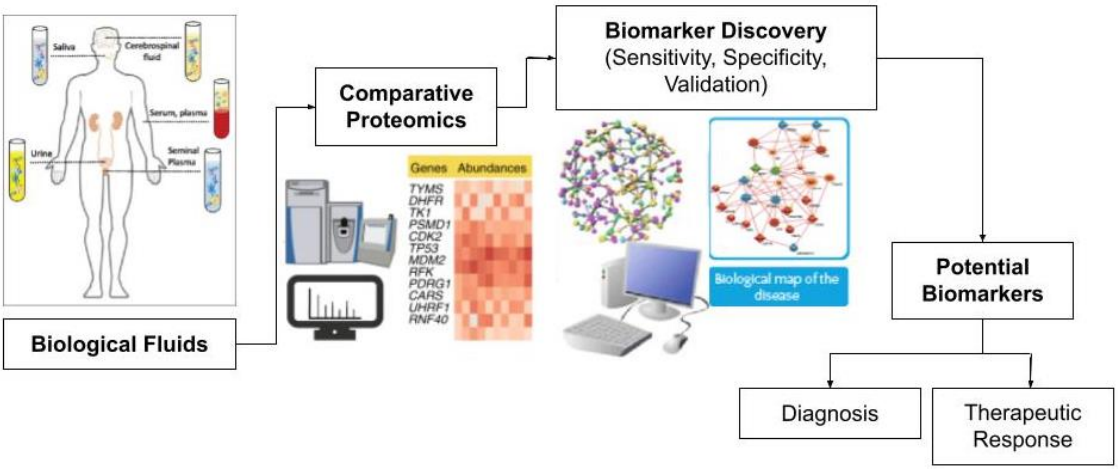


Figure 3 – Diagram of a typical procedure for proteomics biomarker discovery and their applications in disease management.

Therefore, several studies have focused on the research and identification of critical diagnostic and prognostic biomarkers for cancer. The diagnosis and characterization of BC involve the evaluation of marker proteins found inside or on the surface of tumor cells [43].

There are already, at present, a myriad of protein biomarkers in use in clinical diagnosis for the detection and prediction of diseases, but there are no validated plasma/serum biomarkers for BC. Only a few histological markers for diagnosis and prognosis, which include the estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) [44].

The levels of these biomarkers can influence how a person with BC is treated in the clinic, that is, it not only helps in the prognosis of the tumor but also helps in deciding its treatment. Consequently, it is essential to determine the status of these hormone receptors, which act on the cell surface and help the cells to grow and divide, to provide right treatment to the right patient. Depending on their expression profile, BC can be classified into four main subtypes: luminal A (lumA), luminal B (lumB), triple-negative (TNBC) and HER2-related BCS [45, 46].

Depending on the hormone status, breast carcinomas can be divided into several different categories, through existing scoring systems that allow to recognize and classify particular image features, without the need for tissue stains.

Though they are widely used and well-validated, they can involve costly preparation and variable interpretation. Additionally, discordances between histology

and expected biomarker findings can promptly repeated testing to address biological, interpretative, or technical reasons for unexpected results.

Clinical biomarkers are extremely important in the diagnostic investigation and treatment of BC, to early detect its pre-invasive state before metastasis. Therefore, it is increasingly necessary to develop new strategies for the ultrasensitive detection of potential markers, to use them as a promising tool to detect BC at the early stage.

Surface antigens, often adopted for diagnostic and classification purposes in oncology, are usually optimal biomarker candidates because they are accessible for both antibody-based diagnostic and pharmacological interventions [47]. For this reason, previous studies have researched the development of new BC biomarkers, in different sample types, selecting appropriate and well-defined patient cohorts that address a specific clinical issue.

Li *et al.* [48] performed a meta-analysis of 13 case-control studies in a malignant and benign tumor, to corroborate the potential of the association between serum or salivary CA15-3 and CEA for BC monitoring. The results showed that the high levels of these two tumor markers were associated with the tumor stage, usually detected in patients with breast malignancies, since the combination of both increased sensitivities. Gangadharan *et al.* [49] characterized the expression pattern of prolactin-induced protein (PIP), an aspartyl proteinase, from malignant and non-malignant breast tissues. The results demonstrated that PIP transcription was negatively regulated in early-stage invasive BC samples. Park *et al.* [26] used metabolic profiles to analyze cancerous and non-cancerous plasma samples by means of LC-MS followed by receiver operating characteristic (ROC) curve analysis. The abundance level of four specific metabolites (L-octanoylcarnitine, 5-oxoproline, hypoxanthine and docosahexaenoic acid) was associated with clinical-pathological features of BC. Recent studies have shown that microRNAs are potential cost-effective and non-invasive biomarkers, in addition to being fast and reproducible, for the early diagnosis of BC. Thus, Cano *et al.* [50] analyzed the levels of miR-99a-5p expression in primary tumors and plasma from BC patients using the real-time quantitative PCR and the ROC analysis, to reveal its diagnostic potential. Artigues *et al.* [51] also used real-time quantitative PCR and ROC curve analysis to evaluate the expression of miR-30b-5p in plasma and breast tissue from cohorts of patients and healthy volunteers.

Furthermore, in a relatively recent study, Palma *et al.* [52] revealed a clinical interest in a new class of newly identified long non-coding RNAs, the circular RNAs

(circRNAs), making them ideal candidates for use as highly stable and non-invasive diagnostic, prognostic, and predictive biomarkers of BC. These molecules have been described as key regulators of carcinogenesis and breast progression, moreover they play a role in therapeutic agent resistance.

However, due to their complexity of large cohorts, that is, large-scale datasets, more sophisticated approaches are required for the development of biomarker panels. Therefore, over the last few years, the constant commitment and interest facing the inherent challenges of proteomics, such as material limitations and variability, sample sensitivity and degradation, low to high abundance, a plethora of post-translational modifications and unlimited tissue, developmental and temporal, diseased state, and drug responses variations [36], has been crucial for the field of proteomics to express its full potential.

1.2.2. Urinary proteomic

Body fluids, especially blood and urine, have been regarded as significant source of biomarkers in the proteomic analysis, which could be used for the early diagnosis and state forecast of clinical diseases. This is an emerging and promising field of research made possible through recent advances in high-throughput methods, that have the ability for deciphering the complex molecular events involved in cancers.

One of the biggest challenges to analyze human body fluids is the high degree of variability, both intra- as well as inter-individual. Different approaches for this problem have risen to date and hence there are several sample preparation methods for urinary proteomics. The effectiveness of urine proteomics has been recognized year by year because urine reflects various biological phenomena in our body [53] and, for that reason, is an attractive biospecimen for biomarker discovery.

Urine is a desirable material for the diagnosis and classification of diseases because of the convenience of its collection in large amounts in a non-invasive way, in addition to its lower complexity than blood-derived biospecimens, comprised primarily of shed cells, debris, and secreted components from the urinary tract, as well as blood components that have passed through glomerular filtration and renal tubule reabsorption. Therefore, the urine proteome is expected to contain a variety of potential biomarkers related to specific renal or urogenital organ pathophysiology as well as more systemic alterations in distant organs, that is, to overall health [54]. However, urine

proteomic analysis has lagged that of serum/plasma, due to limitations in their depth and confidence of identification, since only 4430 urinary proteins had been identified through recent improvements in pre-analytical concentration techniques [55, 56]. Moreover, the composition and the concentration of urine components are influenced by several factors, such as hydration and elimination frequency, which are not easily controllable within the clinical setting [57], and it is extremely crucial to normalize the samples to make meaningful comparisons among subjects [58].

Therefore, with the advent of modern and, highly sensitive MS, many researchers have focused on the proteomics of urine, because it reflects various biological phenomena in our body [55], in addition to the fact that, compared to other biofluids, it presents an easy accessibility, in large quantities, without the use of invasive procedures, as well as it does not undergo significant proteolytic degradation when compared to them (which explain the fact that it can be stored for several years at –20ºC without suffering significant changes) [59, 60], and, also, the proteins and peptides excreted in urine are generally stable and less complex in comparison to plasma/serum [16]. For this reason, it thus becomes one of the most attractive biofluids in clinical proteomics, creating a possibility to control the level of biomarkers in a chosen time course.

Currently, the identification of proteins with altered expression in cancer is possible because of advances in the procedures and proteomic strategies available in sample preparation, protein separation and MS-based analysis [16, 61]. Thus, with the appropriate tools it is possible to perform the identification and quantification of many proteins or peptides of interest for the discovery of new candidates for protein biomarkers. Two-dimensional gel electrophoresis (2-DE), protein arrays and MS, in conjunction with advanced bioinformatics, stand out.

This approach has been successfully used by Adachi *et al.* [62], in 2006, who reported the first urinary proteome result from high-resolution MS, identifying 1543 proteins. Three years later, in 2009, Kentsis *et al.* [63], employed the same methodology, identify 2362 proteins. In 2011, 1823 urinary proteins were identified using SDS-PAGE and lectin enrichment followed by liquid chromatography tandem mass spectrometry (LC-MS/MS) [64]. Gel-free methods have also been used for urine proteome profiling. For instance, Li *et al.* [65], applied a multidimensional LC-MS/MS method and identified 1310 urine proteins. There have been many efforts to identify more urinary proteins in recent years using different separation approaches coupled with

MS. To date, the human urine PeptideAtlas database contains a total of 23739 peptide corresponding to 2487 proteins [66]. However, most of these global urine proteome profile studies were mainly focused on HCs and taking this into account, it would be valuable to have a comprehensive urine proteome database derived from HCs and disease conditions as a reference resource for guiding urine protein biomarker discovery [67].

In the biomarker discovery process, since urine plays a key role in clinical proteomics, it is crucial to comprehensively profile the normal urinary proteome as a baseline reference. In a urine proteomic study, a total of 6085 proteins were identified as a proposed reference proteome, indicating that implementation of high proteome depth within a short turnaround analytical time could accelerate discoveries of new biomarkers [32]. Besides urine being a true non-invasive sampling source with great potential for biomarker discovery, it is necessary to determine the range of the variations in a large number of samples, in order to confirm whether proteins differentially expressed between control and disease groups represent actual differences, or are just physiological variations among the individuals, which in turn hinders the validation phase in the discovery of potential biomarkers to screen for diseases such as cancer. Leng *et al.* [68] provided a highly efficient workflow by measuring variation among 497 urines collected from 167 HCs, establishing reference intervals for 2000 proteins.

1.2.3. Proteomic analysis

Through the rapid experimental advances in proteomics, which have boosted new methods of bioinformatics analysis, the respective technology has been widely used to study various areas of research, with an emphasis on microbiology, cell and molecular biology, plant sciences, marine sciences, food sciences, cancer, and immunology [69], to understand the complexity of the host-pathogen interaction process.

Several techniques have been applied to biological samples (e.g., urine, exhaled breath, tissue) to explore the possible mechanisms underlying cancer, including gas chromatography (GC), LC, tandem mass spectrometry (MS/MS), nuclear magnetic resonance (NMR) spectroscopy and MALDI-TOF MS [59]. In this regard, MALDI has gained popularity focusing on the global composition of proteins in biological systems

becoming useful to search for potential biomarkers for the detection and prognosis of diseases, since the proteins have a major stake in the initiation, progression, sustenance, and completion of cellular processes, and have also demonstrated their vital roles in cancer processes.

A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic aberrations, and response to an exposure or intervention, including therapeutic interventions [70].

Thereby, the discovery of better biomarkers is critical for accurate clinical assessments, encompassing diagnosis, prognosis, guide molecularly targeted therapy and monitoring activity and therapeutic response across a wide spectrum of diseases. Many current biomarker studies are focused on cancers, the second leading cause of noncommunicable disease mortality worldwide [71]. However, although proteomics methods based on MS hold special promise for the discovery of novel biomarkers that might form the foundation for new clinical blood tests, the lack of well-established methods for validation represents certain adversity, as well as their inadequate specificity and sensitivity.

Additionally, a recent study showed that the single use of existing urinary biomarkers is not accurate enough to predict cancer [72, 73]. To overcome this issue, many complementary research and technologies have been developed and optimized, to the identification of predictive biomarkers of drug resistance, candidate biomarkers for diagnosis and prognostic biomarkers [74], through the construction of a comprehensive biomarker pipeline from six essential process components: candidate discovery, qualification, verification, research assay optimization, biomarker validation, and commercialization [75]. Thus, it is necessary for researchers to shift their focus from a single biomarker to a panel of biomarkers that measure levels or changes in myriad molecular species which may even integrate existing clinical indicators [32].

MALDI-TOF instruments have been reported sensitive and robust for clinical trials [76]. In MALDI, each protein tends to pick up a single proton, which means that the mass-per-charge (m/z) ratio of the protein is its mass. The detector counts the number of ions that the mass analyzer – in this case, time of flight (TOF) – resolves [77].

However, in the studies based on MS analyses of biological samples like urine, the application of enrichment strategies seems to be necessary for generating good quality mass spectra [78]. Highly abundant proteins, as well as the presence of lipids

and salts, mask other low abundant compounds, including cancer-related biomarkers [79]. Therefore, many different strategies have been proposed for urines pretreatment.

One promising approach in the identification of biomarkers is the use of proteomic technology, which allows us to study proteins in a high-throughput fashion and, consequently, greatly increases the chances of identifying single or even combinations of protein biomarkers.

The MS has become the technique of choice in proteomics research for identifying and quantifying proteins in biological samples, which consists of an ion source, a mass analyzer, and a detector. The main principle behind MS-based analysis is the ionization of chemical compounds into charged molecules or molecule fragments and the subsequent measurement of their m/z ratios with confirmation to identify by fragmentation-based sequencing [80]. The two main MS-based proteomic approaches are top-down and bottom-up approaches.

In a protein profile of a sample, obtained by MS, it is possible to identify the protein by the m/z of the peak and determine its amount from the intensity value of the peak. That being the case, in top-down proteomics, which concerns analysis of intact proteins, derives protein sequence information directly from the analysis of intact proteins without the need for enzymatic digestion [80, 81, 82]. From this reason, this mode could be treated as a fast method for molecular mass measurements and identification of proteoforms with PTMs [83]. However, the measurement accuracy in the MS decreases as the mass of the protein increases, also, the PTMs further complicate the identification of proteins, because the sequence of amino acids remains unchanged, but the mass is changed [77]. Therefore, this approach includes difficulties in protein separation, solubility, MS interpretation and analysis, and quantification, which requires the development of new algorithms for signal deconvolution [80].

In contrast, the bottom-up strategy involves initial proteolytic digestion of the entire sample, which breaks proteins into smaller units, called peptides, subsequent separation of the protein-peptide mixture, using gel electrophoresis or chromatography, and final identification by MS/MS. The proteins are digested into thousands of peptides by adding proteases such as trypsin and Endoproteinase LysC (Lys-C), which cleaves the proteins at predictable amino acid locations, resulting in better accuracy. This mode, mostly known as “shotgun” technology that allows the detection of a large number of proteins (over 10000 proteins [84]) in a single run, since the resulting peptides are

smaller and easier to analyze, it is one of the most popular techniques applied on a wide scale in quantitative proteomics.

To identify the proteins, bioinformatic tools are used to compare the detected masses from the MS with the theoretical masses of proteins from the genome of the organism [77].

1.2.3.1. Pre-analytical factors and sample quality

Gel electrophoresis is widely used in combination with MALDI-TOF MS to analyze the proteome of biological samples to discover proteins that are differentially expressed between two groups, thus obtaining a set of potential biomarkers. Nevertheless, the dearth of information about the urine proteome has been due to historically labor-intensive methods, such as gel electrophoresis, which is a highly time-consuming and repetitive process and it is prone to error, thus threatening reliability and reproducibility, as well as the presence of variables.

Pre-analytical variables include biological (patients-related) and technical factors (related to sample selection, handling, collection, and storage) [85]. These analytical factors include sample processing procedures (sample preparation, tryptic digestion, use of internal standard) and analytical features related to the technological platform.

Several studies have widely demonstrated how variables, such as temperature, freeze-thaw cycles, tubes, as well as the use of anticoagulants and protease inhibitors, can have dramatic implications for clinical proteomics analysis [86-89], being the most frequent source of errors leading to false positive and/or negative [90].

In this context, owing to its increased sensitivity and ease of use, MALDI-TOF MS has also become a promising method to check sample quality before biomarker studies [85].

1.2.3.2. Urine sample preparation for proteomic analysis

The condition of the urine is essential for the proteomic determination by MS and software used in the analysis, since human urine is a very complex matrix comprising 95% water and a mixture of water-soluble components such as urea, sodium chloride (NaCl) and potassium chloride (KCl), various amounts of organic acids such as oxalic and citric acid, and other constituents that include phosphates, uric acid, and

creatinine, which may influence the research of potential biomarkers. In addition, certain colour components, such as urochrome, the hemoglobin degradation product responsible for the yellowish colour, urobilin, an orange-brown pigment, and uroerythrin, which has a pink colour [91], may also interfere with the respective analysis, as well as in the protein concentration measurement by the Lowry method.

Based on these difficulties, past studies concluded that urine sample preparation plays an important role in urinary proteomic research. Therefore, before MS analysis, to avoid signal suppression and protein loss during their identification, it is advisable to use methods for protein extraction, isolation, and cleaning of the urine samples with the aim to decrease complexity, reduce the high salt content and remove small molecules, which renders it more amenable to extensive analysis.

In most cases, the regular urinary sample preparation process includes protein extracted by organic solvent precipitation, protein re-dissolving and overnight digestion [92].

Therefore, the standard procedure for sample desalting is protein precipitation with a suitable organic solvent (e.g., methanol, chloroform) to reach representative samples with a certain proportion and purity of the extracted proteins. The deficiency of the precipitation procedure is that a small portion of the yellow components and salts remains in the pellet. Washing the protein pellet with water or a mixture of organic solvent with water can partially improve the technique, but on the other hand, can lower the amount of protein yield [93]. Other equally used extraction methods are ultracentrifugation, dialysis with subsequent lyophilization and ultrafiltration. However, as technology has advanced and in accordance with the properties of proteins, there is an increasing tendency to separate them using a combination of multiple techniques, which can maintain the activity and structure of proteins better and achieve higher resolution.

After the isolation of proteins from urines, the protein purification can be performed in different ways, including, chromatography, electrophoresis, membrane separation, precipitation, magnetic separation, molecular imprinting, microfluidic chip, and crystallization, among others. Then, the ultimate step of in-depth analysis of the urinary proteome depends on the efficacy of protein digestion, which directly correlates with the ability to dissociate protein complexes, denature tertiary structures, and solubilize hydrophobic membrane-bound proteins [94].

Furthermore, to increase the detection depth, protein or peptide fractionation prior to MS-analysis typically include approaches that combine one- or two-dimensional (1D or 2D, respectively) gel electrophoresis with highly accurate MS and/or LC-based.

Thus, to discover potential biomarkers of urinary proteins, pre-analytical processing and fractionation methodologies are emphasized to concentrate urinary proteins and, in turn, improve the detection of low-abundance proteins.

1.3. One- and two-dimensional electrophoresis

The combination of proteomic techniques with computational analyses enabled a breakthrough in protein sciences for biomarker discovery based on the simultaneous analysis of thousands of proteins. So, the identification of proteins with altered expression in cancer is possible because of advances in sample preparation, protein separation and MS-based analysis [95]. Following the separation of a mixture of proteins, the resulting fractions consist of fewer proteins or peptides and, therefore, the preliminary steps in the preparation of the sample are crucial for further identification of potential biomarkers for early cancer detection [96].

The invention of 1D SDS-PAGE, for separating proteins according to their molecular weight offered for the first time a display method allowing many proteins to be analyzed by a single analytical event within complex biological samples [39]. Thenceforth the SDS-PAGE followed by protein staining has become the core analytical separation technique in proteomic studies, which is used as a fundamental laboratory procedure to determine the composition and characterization of protein-containing mixtures at all stages of a purification scheme.

However, it was the introduction of 2D-PAGE that truly raised awareness that for understanding biological mechanisms, proteomes rather than proteins should be investigated. The 2-DE technology allows a global profile of a sample proteome by the simultaneous resolution of hundreds to thousands of proteins in a single gel, and allows the identification of proteoforms, which constitute excellent candidates for biomarkers.

Widely used 2D-PAGE separates proteins based on their charge (the first dimension) and size (the second dimension), where proteins are presented as stains in a biaxial plane. Although the 2-DE technology can resolve a large quantity of proteins, it has its limitation, when the sample is relatively complex like urine and serum, as well as

its inability to resolve multiple classes of proteins (e.g., transmembrane proteins) and its low yield that does not allow application in clinical settings [97].

Thereby, 1D-PAGE is still used to separate complex macromolecule samples, because it is a simple and reliable quantification method, whereby proteins are resolved and migrated according to their molecular weights. Besides that, the facilities necessary for separation in 1D gel electrophoresis are not sophisticated and expensive in comparison to 2D separating technique facilities. Nevertheless, 1D gel electrophoresis is a tool that cannot be comprehensively used in proteomic studies, since it is time-consuming, labor-intensive, and semi-quantitative.

The complementarity of the analytical platforms in evidence provides a greater understanding of proteomic data, improving the biological understanding of proteins, their interactions with the environment and levels of expression in cells, especially in disease states.

By employing 1D SDS-PAGE to separate the urinary proteins, followed by MS for protein identification, it is possible with these techniques, to simultaneously analyze up to 10000 protein spots, thus avoiding the troublesome use of isoelectric focusing (IEF) in 2-DE.

1.4. Mass spectrometry-based proteomics

Surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) and MALDI-TOF MS, firstly described in 1985 by Karas *et al.* [98], are two well recognized soft ionization techniques for obtaining protein profiles from biological samples [99, 100].

In the last 20 years, MALDI-TOF MS has become a popular and versatile method for biochemical and clinical investigations, providing a robust, fast, and affordable identification of proteins from 1- or 2-DE gels and, consequently, high-throughput support to clinics.

It is a versatile analytical technique to detect and characterize mixtures of organic molecules, which consists of a soft ionization process by means of a laser that reaches the analyte mixed with a solution of a matrix [e.g., α-cyano-4-hydroxycinnamic acid (CHCA), 2,5-dihydroxybenzoic acid (DHB) and 3,5- dimethoxy-4-hydroxycinnamic acid (SA)] in an organic solvent, able to absorb energy in the form of UV light. The mixture is first deposited onto a sample plate known as a target, in which

the solvent, under vacuum, evaporates and the sample is cocrystallized. After a bombardment by a pulsed laser beam (UV or infrared radiation), matrix molecules that are energetically ablated from the surface of the sample, transfer protons to the analyte, resulting in the formation of intact gas-phase molecular ions (which usually carry a single positive charge) (Figure 4).

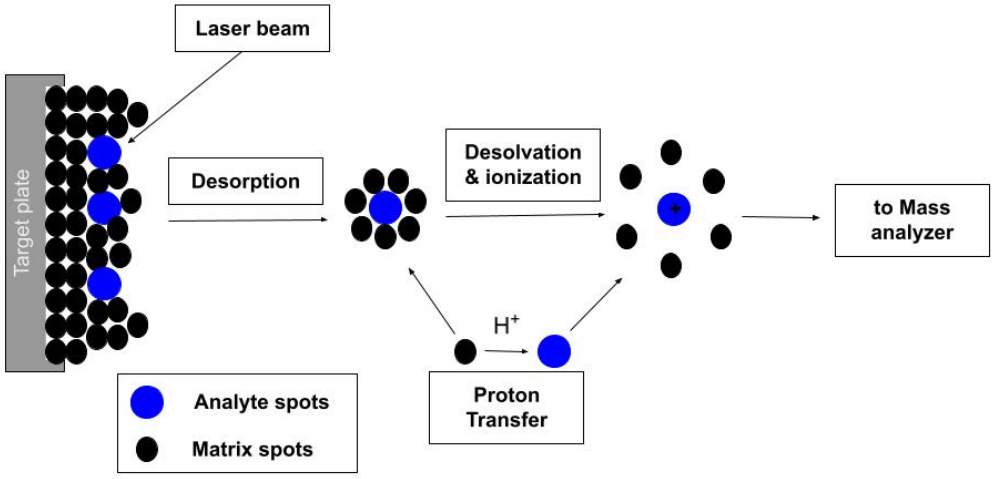


Figure 4 – Schematic diagram of MALDI principle.

After the ionization process, the masses can be analyzed by the TOF MS in which it accelerates the gas-phase ions in a high-voltage electric field, which transmits a constant amount of kinetic energy that will cause the smallest of the ions to travel the fastest and, consequently, are separated by mass (the heavier the ions the longer the time of detection). The time between ionization and detection, and the known constant kinetic energy of the electric field, allow the spectrometer to be able to calculate and record the m/z ratio of each ion it detects. The TOF reflector mode analyzer reflection mode provides greater resolution because it is equipped with a longer flight path, ion mirrors and electric fields that refocus ions by their masses [101].

These proteomic technologies are relatively recent and have been increasingly extensively used in deciphering protein differential expression in human cancers [102].

Several papers have reported that peptide profiling is of great importance in the diagnosis of different kinds of cancer, such as lung, gastric, colorectal, ovarian, and leukemia [103]. In that case, several studies have been developed using MALDI in diverse diseases, as well as in BC.

Comparing with normal urine proteome analysis from other published studies, Kiprijanovska *et al.* [104] identified 11 distinct proteins, through 2D PAGE/MALDI-

TOF MS profiling, present in urines from 8 patients with prostate cancer, among them three are involved in processes of angiogenesis, tumor growth or metastasis. Wang *et al.* [105] identified 76 different urinary polypeptides, using MALDI-TOF MS to analyze peptide expression profiles, that were applied in urines from 34 patients with lung adenocarcinoma and 36 HCs, to serve as potential biomarkers to noninvasively detect lung adenocarcinoma. Additionally, Tantipaiboonwong *et al.* [106], to identify preliminary urinary protein markers related to cancer, used 2-DE method for a good separation (and provide the highest number of protein spots) and identified 9 potential biomarkers by MALDI Q-TOF MS analysis, that were expressed in urines obtained from HCs and lung cancer patients.

Also, in the study of discover promising biomarkers for acute kidney injury (AKI), Jung *et al.* [107] employed proteomic techniques in 10 urines following 3 phases, which revealed 3 protein biomarkers for early, accurate prediction of AKI in preterm infants.

Compared to previous studies, Theodorescu *et al.* [108], used different methods, precisely, capillary electrophoresis combined with a MALDI-TOF MS, to identify 22 potential biomarkers of urothelial carcinoma present in urines of 46 patients with this genitourinary disease and 33 HCs.

Kreunin *et al.* [109] also demonstrated that the application of different methods from those mentioned above, such as the combination of concanavalin A (Con A) affinity chromatography and nano-LC-MS/MS, allowed the identification of potential biomarkers of bladder cancer from human urine obtained noninvasively.

Another studies, related to the search for molecular markers for their potential clinical use, which aid in the early diagnosis and treatment of various diseases, in particular cancer, as well as in the improvement of the management and advancement with oncological medical care, were reported by several researchers, using serum samples for this purpose. Among them, Zografos *et al.* [110] identified 42 proteomic MBC biomarkers, through 2-DE separation and MALDI-TOF MS methods, between 8 HCs and 11 patients with MBC. It was also reported, by Ding *et al.* [111], five-peptide with high sensitivity and specificity for the diagnosis of bladder cancer, through MALDI-TOF MS technique, that were applied in serum peptide profiles of 67 hematuria patients and 64 HCs.

Yang *et al.* [112], using the same methods, identified 9 differentially expressed proteins in serum between 50 patients with urosepsis, which were divided into two

groups: uroseptic patients (non-shock groups) and severe uroseptic patients (shock groups), that revealed to be molecular markers of early diagnosis and prognosis of infection.

To conclude, several researchers have as main objective the study of proteomic/peptidomic profiles of various biospecimens obtained from cancer patients in order to identified biomarkers for the purpose of diagnostic and prognostic categorization of cancers, as well as in assessing various cancer therapeutic regimens.

1.5 Objectives of the thesis

The aim of this thesis was to perform the screening of protein/peptide content in urine from BC patients and HCs, through an optimized workflow to promote urine protein quality, using 1D SDS-PAGE combined with MALDI-TOF MS approach (Figure 5). This procedure allows to establish the urinary proteomic/peptidomic signature of BC and to unravel the main differences by comparing with HCs pattern. This will allow establishing specific proteomic patterns for each studied group which will be useful to discriminate between them and identify pattern biomarkers for BC diagnosis. It is a non-invasive method that can be potentially applied in a clinical environment as a diagnostic approach or as a complementary way to current diagnostic methods. Moreover, this high-throughput strategy provides several advantages, including the capacity to increase the detection sensitivity of low abundance proteins and thus enlarge the screening range for protein identification, previously separated by 1D SDS-PAGE, through peptide mass mapping. Specific objectives are:

- Optimization of collection of protein/peptide spots with different solvents (methanol, acetonitrile, acetone, trichloroacetic acid) along with digestion methods to process the analyses by MALDI-TOF MS.
- Optimization of 1D SDS-PAGE parameters, through gel modification to obtain the most appropriate concentration of polyacrylamide, where techniques for protein concentration and removal of salts and other contaminants from urine were evaluated.
- Establishment of urine signature of BC using MALDI-TOF MS through the proteomic/peptidomic analysis with the aim of eventually discovering potential biomarkers for screening, pre-diagnosis and, consequently, to provide a more efficient treatment for BC, improving the quality of life of patients and decreasing the mortality rate.
- Statistical analysis (one-way ANOVA, principal component analysis) to analyze the generated matched peak list sets for BC biomarkers research.

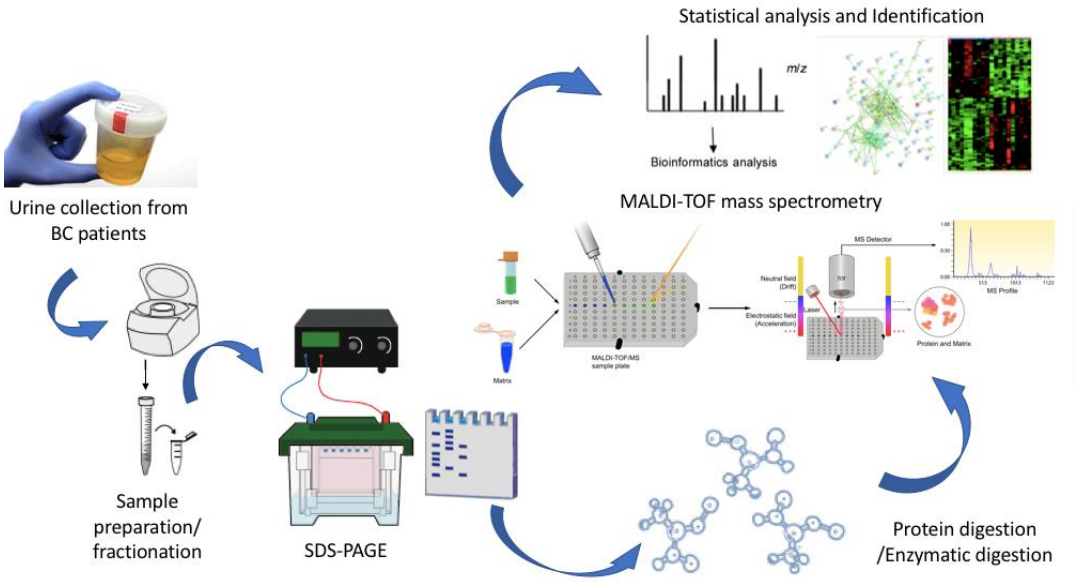


Figure 5 – Workflow of the urinary proteomic/peptidomic signature.

2. MATERIALS AND METHODS

2.1. Reagents

All chemicals used in this trial were of analytical grade. Bovine serum albumin (BSA, ≥98%) was supplied from Acros Organics (Geel, Belgium). Thiourea (CH_4N_2S, ≥99%), DL-1,4-dithiothreitol (DTT, 99%), trichloroacetic acid (TCA, 99%), trifluoroacetic acid (TFA, 99.5%) and acrylamide (C_3H_5NO, ≥ 99%) were obtained from Acros Organics (New Jersey, USA). Urea (CH_4N_2O, > 99%) was supplied from MERCK Schuchardt (Hohenbrunn, Germany). Trisodium citrate dihydrate ($C_6H_5Na_3O_7.2H_2O$, 99%), copper(II) sulfate ($CuSO_4$, 99%), bromophenol blue (pH 3.0-4.6), formic acid (CH_2O_2, 98-100%), N,N,N´,N´- tetramethylethylenediamine (TEMED, $C_6H_{16}N_2$, 99.5%) and ammonium persulfate (APS, $(NH_4)_2S_2O_8$), 98%) were from Merck (Darmstadt, Germany). Sodium carbonate (Na_2CO_3, ≥99.5%), N,N´-methylenebis(acrylamide), sodium dodecyl sulfate ($NaC_{12}H_{25}SO_4$, 99%) and acetic acid ($C_2H_4O_2$, ≥99.8%) were obtained from Sigma-Aldrich (St. Louis, USA). Sodium hydroxide (NaOH, 98.9 %) was supplied from Fisher Scientific (Loughborough, UK). Glycerol ($C_3H_8O_3$, >99.5%) and ammonium hydrogen carbonate (NH_4HCO_3, >99%) were obtained from Fluka (Buchs, Switzerland). Methanol (CH_3OH, ≥99.8%) was obtained from Chem-Lab NV (Zedelgem, Belgium). Acetonitrile (CH_3CN, ACN, 99.5%) was from Riedel-deHaën (Seelze, Germany). Trypsin (≤100%) for protein digestion was provided from Sigma-Aldrich (St. Louis, MO, USA). Sodium citrate ($(C_3H_4(OH)(COONa)_3$, 99%) was from May & Baker LTD (Dagenham, England). Tris Base ($(CH_2OH)_3CNH_2$, 99%) was supplied from Fisher Scientific (Geel, Belgium). Tris hydrochloride for molecular biology ($C_4H_{11}NO_3.HCl$, ≥99%) was supplied from PanReac AppliChem (Darmstadt, Germany). Glycine ($C_2H_5NO_2$, ≥98.5%) was provided from Fisher BioReagents (Geel, Belgium). Ethylenediaminetetra-acetic acid (EDTA) ($C_{10}H_{16}N_2O_8$, 99.3%) was from VWR Chemicals BDH (Belgium). Chloroform ($CHCl_3$, 99.2%) was from VWR International (Fontenay-sous-Bois, France). The water used was obtained from a Milli-Q water purification system (Millipore, Bedford, MA, USA).

2.2. Subjects selection and urine collection

To investigate the BC proteomic/peptidomic profile, 56 urines from patients diagnosed with BC (age range 40-78 years, average 59 ± 1) were compared with a group of 54 HCs (age range 44–68 years, average 56 ± 1). The visual characteristics of urines were examined based on color and turbidity, which appeared to range from very dilute to very concentrated. To prevent technical and analytical variation caused by handling, all urines were collected, processed, and stored following the same procedural conditions, and by the same laboratory person until a final protocol was established, in a total set of 110 samples. Thus, these samples were frozen at −80°C, thawed for 3 h at room temperature (25 ± 1 ºC), and then supernatants were divided into 2 mL aliquots and lyophilized. To achieve a representative urinary proteome that portrays the group pattern sample variability, the frozen aliquots were completely thawed, mixed well, and then equal volumes (2 mL) of each donor supernatant were pooled into the BC or HCs. In this study, no protease inhibitors were utilized, since some studies have shown that they diminish protein identification and may interfere with the subsequent digestion procedure in untargeted urine proteomics [113].

These samples were obtained at Unit of Hematology-Oncology of Hospital Dr. Nélio Mendonça (HNM) at Funchal, Portugal, according to Table 1. All subjects of this study were informed about the investigation, in which they signed an informed consent to participate in the study (Appendix, page 117), prior to sample collection, and the research was already approved by the Ethics Commission of HNM. Epigenetic factors were also considered, as well as relevant information about the clinical characteristics of the study population, which were provided by the doctors of both institutions that collaborated in the study.

In addition, all the women enrolled in the study were non-smokers. Using the TNM (tumor, node, and metastasis) staging approach, the examined BC cases included fifteen of stage IA, nine of stage IIA, seven of stage IIIA, ten of stage IIB, nine of stage IIIB and six of stage IIIC.

Table 1 – Characteristics of subjects which participated in the study.

Samples	**BC**	**HCs**
Number	56	54
Age (range, average)	(40-78, 59)	(44-68, 56)
Disease stage (number of samples)	IA (15) IIA (9) IIIA (7) IIB (10) IIIB (9) IIIC (6)	Not applicable

2.3. Urine sample precipitation

In this study, four different extraction methods were compared on urine samples from BC patients and HCs. The chloroform/methanol-based method (method A), ACN/TFA-based method (method B), TCA/acetone-based method (method C) and TCA-based method (method D) were tested to ascertain the most fitting extraction method for the 1D SDS-PAGE analysis of the urines. All extraction methods were performed in triplicate.

Method A – For the chloroform/methanol precipitation, which was adapted from Saito *et al.* [114], lyophilized urines were resuspended in 500 µL of methanol and 125 µL of chloroform. The mixture was vortexed and centrifuged for 15 min at 13400 rpm at room temperature (25 ± 1 ºC). Then, the supernatant was discarded, and 500 µL of methanol was added to the Eppendorf's. Subsequently, it was vortexed vigorously and centrifuged, in the same conditions as the previous ones. The upper phase was carefully removed and discarded, followed by the precipitate air-dried. Finally, the precipitate was resuspended in 100 µL of 8M urea/50 mM Tris-HCl buffer, pH 8.0, with subsequent stirring.

Method B – The ACN/TFA precipitation method was adapted from Chertov *et al.* [115] with some modifications. Briefly, urines were resuspended in 10 mM phosphate buffer (pH 7.5) and added with two volumes of cold ACN/0.1% TFA. The mixture was then vortexed and centrifuged at 13400 rpm for 15 min. The supernatant was removed, and the pellet was vacuum dried, before being resuspended in 100 µL of 8M urea/50 mM Tris-HCl buffer, pH 8.0, with subsequent stirring.

Method C – The TCA/acetone-based precipitation method was adapted and modified from Tantipaiboonwong *et al.* [107]. Succinctly, two volumes of cold 20% TCA in acetone (–20ºC) were added after lyophilized urines were resuspended in 100 mM phosphate buffer (pH 7.5). After being kept at –20ºC for the night, the mixture was centrifuged for 15 min at 13400 rpm to obtain a pellet. The pellet was then centrifuged, as previously mentioned, after being washed twice with cold acetone. After removing the supernatant and vacuum-drying the pellet, it was then resuspended in 100 µL of rehydration buffer solution (7M urea, 2M thiourea, 0.1% 120 mM DTT).

Method D – The TCA-based extraction method was adapted and modified from Jonscher *et al.* [116] and Arunambiga *et al.* [117]. To summarize, 2 mL TCA was added to lyophilized urines and then incubated overnight at 4ºC. Following centrifugation at 13400 rpm for 30 min, the supernatant was removed, and the pellets washed with 90% ice-cold acetone for subsequent incubation on ice for 15 min. Samples were then centrifuged at 13400 rpm for 15 min and, after removing the supernatants, the pellets were allowed to air dry before resuspension in 100 µL of rehydration buffer solution.

2.4. Total protein concentration in urine (Lowry´s test)

For the determination of total protein concentration in urine, using the Lowry method, 2.5 mL of Lowry´s reagent solution was added into Eppendorf's containing 30 µL of urine and 470 µL of H_2O. The Lowry's reagent solution was prepared in a 1:50 ratio, 2 mL of 0.5% (w/v) $CuSO_4$ and 1% (w/v) $KNaC_4H_4O_6.4H_2O$ solution into 50 mL of 5% (w/v) Na_2CO_3 and 0.1N NaOH. Then, this mixed solution was rest for 10 min. After that time, 250 µL of Folin´s reagent (1:1 v/v Folin-Ciocalteu: H_2O) was added to the homogenate, mixed, and rested for 30 min in the absence of light. Later, the sample´s absorbance was measured at 750 nm, with the aid of quartz cuvettes (Hellma®, Germany) in a ultraviolet-visible (UV-Vis) spectrophotometer (Lambda 25 da PerkinElmer, United Kingdom), using H_2O as blank. This procedure was performed for urines obtained from HCs and BC patients before and after precipitation. To construct the calibration curve, the same procedure was applied to several BSA concentrations in urine, ranging from 9 to 277 µg/mL.

2.5. One-dimensional gel electrophoresis

The protein content of the precipitated sample was analyzed by 1D SDS-PAGE. Ten µL urines from HCs and BC patients were separated under denaturing conditions using a 1 mm thickness resolving gel containing 12% (w/v) acrylamide/bis-acrylamide and a staking gel containing 4% acrylamide/bis-acrylamide (Sigma-Aldrich, St. Louis, MO, USA), as well as 7 µL of protein ladder (Bio-Rad), containing fragments of known size, 10-250 kDa range, which served as a standard for estimating the size of the fragments of the respective samples. After pouring the 12% resolving gel solution into the separating gel and overlaid with H_2O was to prevent contact with air (oxygen), it was allowed to polymerize for 10-15 min to form a gel. Later, the H_2O was removed, and the 4% stacking gel solution was added, where a comb was inserted at the end and left to polymerize. For the preparation of urines, these were diluted 1:1 (v/v) with SDS sample buffer (8% w/v SDS, 1M Tris-HCl (pH 6.8), 40% v/v glycerol, 14.7 M β-mercaptoethanol, 0.5 EDTA, 0.01% bromophenol blue). That is, 10 µL of each sample with 10 µL of loading buffer were incubated on heating block set to 95°C for 8 min before loading into hand cast gels (Table 2). The gel was mounted in the electrophoresis apparatus using binder clips and then, pouring running buffer into the upper and lower chambers of the electrophoresis apparatus, to proceed with the loaded samples and molecular weight markers in wells, separately. The 1D SDS-PAGE gel was run in a Bio-Rad´s PowerPac Basic power supply (BIO-RAD, USA) at 20 mA per gel. After completion of electrophoresis, the protein bands in the gel were stained with Coomassie brilliant blue R-250 solution (0.5% Coomassie Brilliant Blue G-250 in 30% methanol, 10% acetic acid and 59.5% H_2O). Then, it was covered with aluminum foil, to avoid contamination and to prevent the evaporation of the solution and allowed to incubate overnight at room temperature. After staining by Coomassie Brilliant Blue, the stained solution was carefully removed and the gels were washed at least two – three times with a destained solution (30% v/v methanol, 10% v/v acetic acid, 60 % H_2O), until reduce the background staining. Lastly, the gels were stored in the destaining solution until the spots of interests were manually picked. Additionally, the gels were analyzed with the software Gel-Analyzer 19.1 [118].

Table 2 – Gel solutions composition for 1D SDS-PAGE.

	12% Resolving Gel (mL)	4% Stacking Gel (mL)
Water (H_2O)	1	2
Acrylamide (30%)	3.75	0.9
Resolving buffer[1]	4.5	-
Stacking buffer[2]	-	2.5
10% APS	0.1	0.05
TEMED	0.01	0.005

[1]4.54 g Trizma-base + 0.1 g SDS;
[2]1.51 g Trizma-base + 0.1 g SDS.

2.6. Protein digestion

Three different digestion approaches (methods I, II and III) have been compared to determine the optimal procedure, combined with the methods described above, for the subsequent analysis MALDI-TOF MS. In this way, protein spots were manually excised from the Coomassie-stained gels and transferred into 1.5 mL eppendorfs.

Method I – The gel fractions underwent two 15 min washes in 200 µL of a mixture of 50% ACN/25 mM NH_4HCO_3 (pH 8.0) and then once more in 200 µL of 100% ACN. The pieces were then allowed to dry under vacuum before being washed again with 10 µL of 25 mM NH_4HCO_3 containing 0.1 mg of trypsin and left to incubate overnight at 37 ºC. The next day, the gel fragments were rinsed twice with 50 µL of a 50% ACN/5% TFA solution, dried and then resuspended with 25 µL of TFA.

Method II – For the second digestion method, the gel fractions were washed once for 30 min in 200 mM NH_4HCO_3 (pH 7.8) at 37°C, and then afterwards these fractions were dehydrated by incubation in 200 mM NH_4HCO_3, pH 7.8/ ACN (4:6 v/v) at 37 °C for 30 min. Following incubation in ACN for 2 min, the gel pieces were then swollen in 0.1 mg of trypsin (from Porcine Pancreas, Sigma-Aldrich, St. Louis, USA) in 50 mM NH_4HCO_3 (pH 7.8) and incubated overnight at 37 °C. The tryptic peptide mixture was subsequently recovered into fresh 1.5 mL eppendorfs and a second peptide extraction from these gel pieces was carried out with 0.1% TFA in 60% ACN for 5 min. The peptide-containing liquid fractions were pooled, dried under vacuum, and resuspended in 25 µL 0.5% TFA/5% ACN after the supernatant was collected.

Method III – Briefly, 250 µL of each solution (50% ACN in H_2O and 50% ACN/50 mM NH_4HCO_3) were added to the gel particles and stirred for 5 and 30 min, respectively. After discard the supernatant, 250 µL of 50% ACN/10 mM NH_4HCO_3

were added and vortexed. The gel pieces were hydrated with minimum volume required of trypsin solution for 1h at 25 ± 1 ºC, topped up with 10 mM NH_4HCO_3 solution until the gel pieces were covered and incubated overnight at 37 °C. Peptides were extracted from gel fragments with consecutive shaking once with 20% TFA/60% ACN and twice with 0.1% TFA. Supernatants were combined in a fresh eppendorf and resuspended in 50 µL of H_2O.

2.7. MALDI-TOF MS analysis

For MALDI-TOF MS analysis, the digested samples were premixed with MALDI matrix solution (10 mg/mL SA in methanol) at the ratio of 1:1. Then, 1 µL of each mixture was applied in triplicate on to a MTP 384 ground steel target plate (Bruker Daltonics, Billerica, MA) and then, were allowed to dry at room temperature (25 ± 1ºC). Mass spectra were obtained on the Autoflex maX MALDI-TOF/TOF mass spectrometer (Bruker Daltonics) in the reflector and linear positive-ion modes across the m/z ranges of 600 – 3500 Da and 5 – 20 kDa, respectively. Each mass spectrum was acquired by 500 laser shots, with intensity of 75%, collected across the whole sample. The following ion source parameters were used: ion source 1, 19.22 kV; ion source 2, 17.00 kV. Other settings for MALDI-TOF MS analysis were as follows: pulsed ion extraction, 170 ns and lens, 8.53 kV. FlexControl 3.4 software (Bruker Daltonics, Bremen, Germany) was acquired for the acquisition and processing of the spectra. A standard BSA calibration mixture in the ranges of 600 to 3500 Da and 5 to 20 kDa was used for mass calibration. The calibration curve was constructed by the same procedure applied to several BSA concentrations in urine (5.00, 10.00, 15.00, 20.00, 25.00 and 50.00 mg/mL).

2.8. Statistical analysis

The background signals were subtracted and the mass spectra components (m/z list, peak area, SNR (signal-to-noise ratio) and peak relative intensity) were analyzed by Mass-Up software, which is available from the project homepage on http://sing.ei.uvigo.es/mass-up. The software provides an analytical framework for the design of models for automatic classification based on variations in MALDI-TOF MS spectra from BC patient and HCs, as well as for the identification of potential biomarkers that are indicative of a particular disease state. Each sample was spotted on

the MALDI plate three times, resulting in a total of 330 MALDI mass spectra; so, to let the software group all analyzed sample replicates into one biological replicate, spectra grouping function was applied for each preparation, that were subsequently used for data analysis. The peak list calculation was performed within the range from 600 to 3500 m/z. According to standard procedures in the field [119], preprocessing of the acquired raw spectra of mzXML input files as a critical stage was performed to define the optimal model allowing the discrimination of the urines, wherein the principal workflow consists of transforming, smoothing, baseline correction, intensity calibration, spectra alignment, peak detection and ultimately converted to a data set containing aligned m/z and intensity before further analysis. Therefore, the corresponding raw spectra of each sample acquired from the MALDI experiments were pretreated/preprocessed by Mass-Up v1.0.14 open-source software [120], using the following parameters: (I) Intensity transformation: square root; (II) Smoothing: Savitzky Golay; (III) Baseline correction: Snip; (IV) Standardization: Total Ion Current (TIC); (V) Peak Detection: MALDIquant with a signal-to-noise ratio (SNR) of 3 and a Half Window Size of 50. Then, the peak matching was performed with the following configuration: (I) Intra-sample matching using MALDIquant with a tolerance of 0.05 without consensus spectrum generation; (II) Inter-sample matching using MALDIquant with a tolerance of 0.05. Finally, peak lists containing m/z and peak areas from Mass-Up Software were exported to Microsoft Excel (data format CSV). For each BC patient, the average m/z and area were calculated for peaks with S/N≥10 in 10 replicate mass spectra. Data statistical analysis was also performed using the MetaboAnalyst 5.0 (University of McGill, Montreal, Canada). Significant differences among the BC patients and HCs were assessed with *t-test* (for normal variable distribution). P-value <0.05 was considered statistically significant. The principal component analysis (PCA) test was used to obtain a representation from the data projection. The orthogonal projection to latent structures discriminant analysis (OPLS-DA) was applied on the urine proteomic/peptidomic profile dataset to verify the distribution of the variables for the groups under study and to provide insights for biomarkers research. ROC curve was generated to assess the diagnostic value of the potential BC biomarkers.

3. RESULTS AND DISCUSSION

Apparently, proteome changes play a role in BC carcinogenesis, as recently shown by Ferrari *et al.* [121], emphasizing the importance of identifying potential protein/peptide biomarkers in urine to mitigate metastasis. As a result, to discover potential biomarkers for the early diagnosis of BC and, consequently, to reduce the morbidity and mortality caused by this condition, the current study involves protein extraction, protein separation by 1D SDS-PAGE, followed by protease digestion for subsequent MS analysis. The different experimental parameters in different used methods were optimized previously.

3.1. Optimization step

In a preliminary investigation, the different protein extraction and digestion methods were tested on 2 mL of urine from an HCs and a BC patient, since, as to be expected, the amount of urine collected for analysis plays a significant effect on the number of isolated proteins.

Firstly, the two lyophilized urines from each subject were used to compare each precipitation method, and the final experiment was conducted in triplicate to evaluate reproducibility.

The only difference between the methods is the kind of organic solvent that is added, which causes the proteins in the patient sample to precipitate out of solution, and after centrifugation, the proteins form a pellet at the bottom of the tube. Thus, before precipitation, the concentration of urinary protein is normally low (100-200 mg/L) [122], but after precipitation, the concentration of urines, including all the proteins as well other compounds, such as urea and cell debris, increases. On the other hand, the concentration of non-protein interfering substances is high relative to the protein concentration and very variable, not to mention the inorganic ion content is also high. For this reason, although the precipitation is a crucial and quick step in the preparation of samples in the analytical phases, all these factors affect the precision and accuracy, where the sample remains somewhat unclean and that can cause issues with matrix effects (ion suppression) in MS methods potentially leading to loss of both sensitivity and reproducibility [123].

Having said that, to select the best precipitation method among the four, the spectra of the two urines were studied using the MALDI-TOF MS technique. Based on the results shown in Figures 6 and 7, it was determined that the chloroform/methanol precipitation method was the most effective, based solely on the fact that it provided a high mass resolution good spectrum in both reflector and linear positive-ion mode over the mass range of 700-2000 Da and 5-20 kDa, respectively, in comparison to the other methods (Figures A1-A7 – Appendix, pages 96-99). Moreover, the protein band patterns and intensities were different among the four precipitations, since the urine proteins extracted by chloroform/methanol precipitation showed greater amount of protein in the gel than the other precipitations.

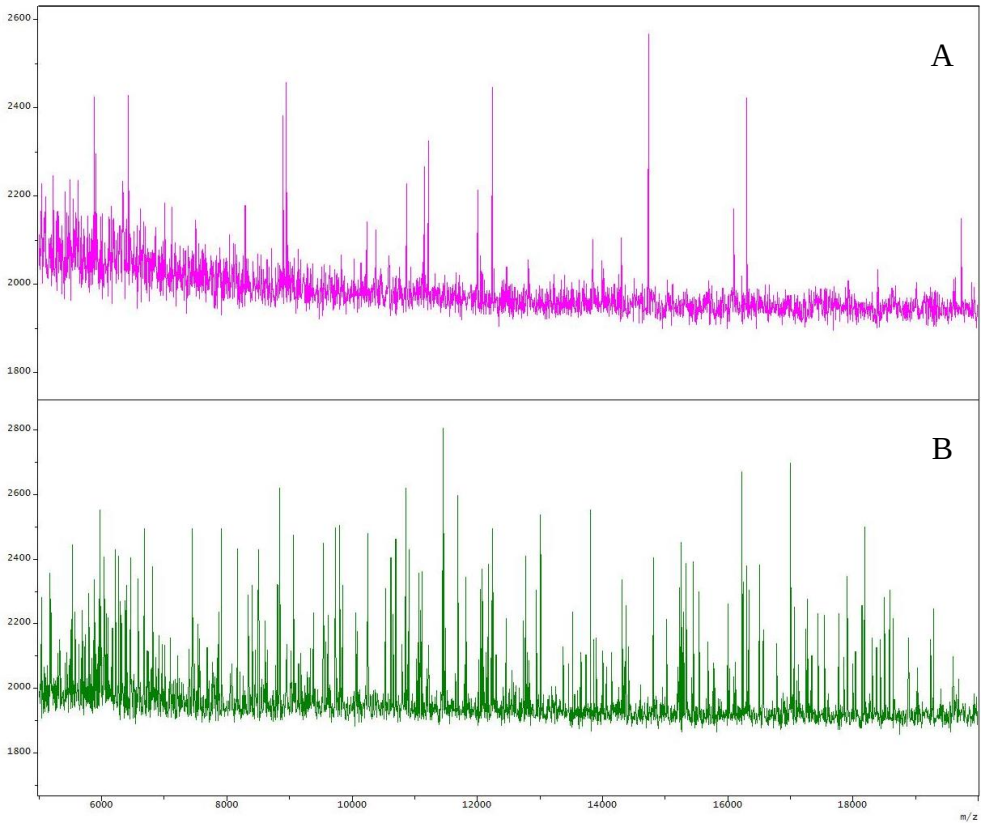


Figure 6 – MALDI-TOF MS spectra in linear positive-ion mode across the m/z range of 5 to 20 kDa of samples from (A) HC and (B) BC patient using chloroform/methanol precipitation.

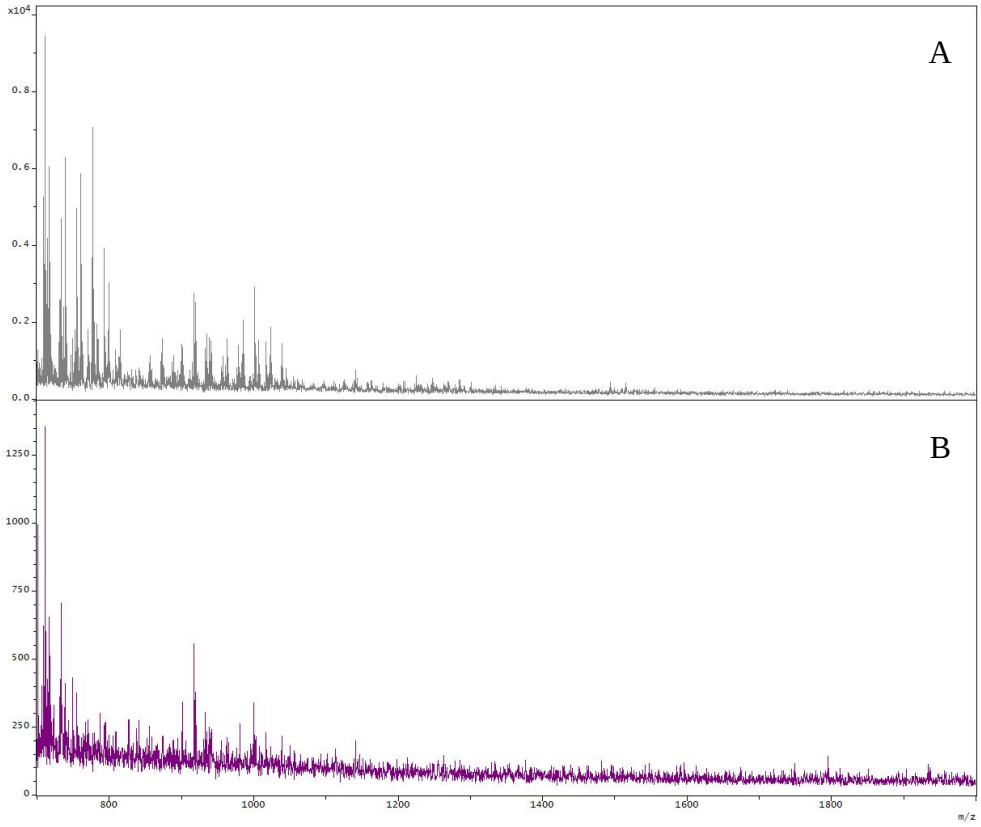


Figure 7 – MALDI-TOF MS spectra in reflector positive-ion mode across the m/z range of 700 to 2000 Da of samples from (A) HC and (B) BC patient using chloroform/methanol precipitation.

Then, three different protein digestion methods were evaluated, before and after precipitation, to also analyze the role of precipitation in the preparation of urine samples. As shown in Figures 8 and 9, the analysis of urines revealed differences in the protein/peptide composition and the intensity when comparing the different methods of digestion. It was concluded that the method II proposed had the best MS fragmentation and sensitivity, which improved the interpretation of MS data. Hence, among the tested protein extraction methods, chloroform/methanol precipitation in combination with digestion method II consistently led to the highest yield in terms of total extracted protein as measured by the Lowry´s assay and largest number of proteins/peptides identified with high confidence, as well as being the most reproducible of all methods evaluated, as shown in Figure 10.

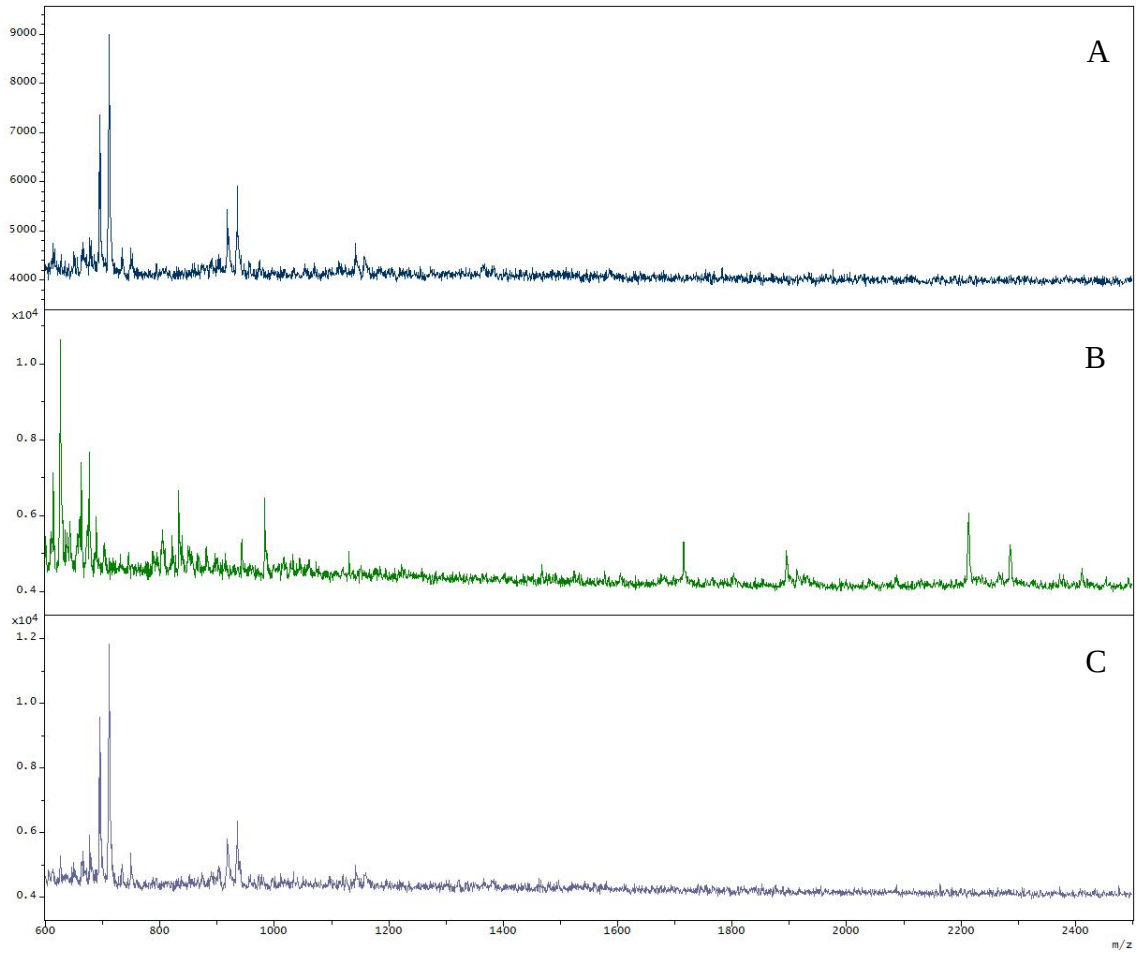


Figure 8 – MALDI-TOF MS spectrum of the BC patient´s non-precipitated sample prepared only by three different digestion methods designated (A) I, (B) II, and (C) III.

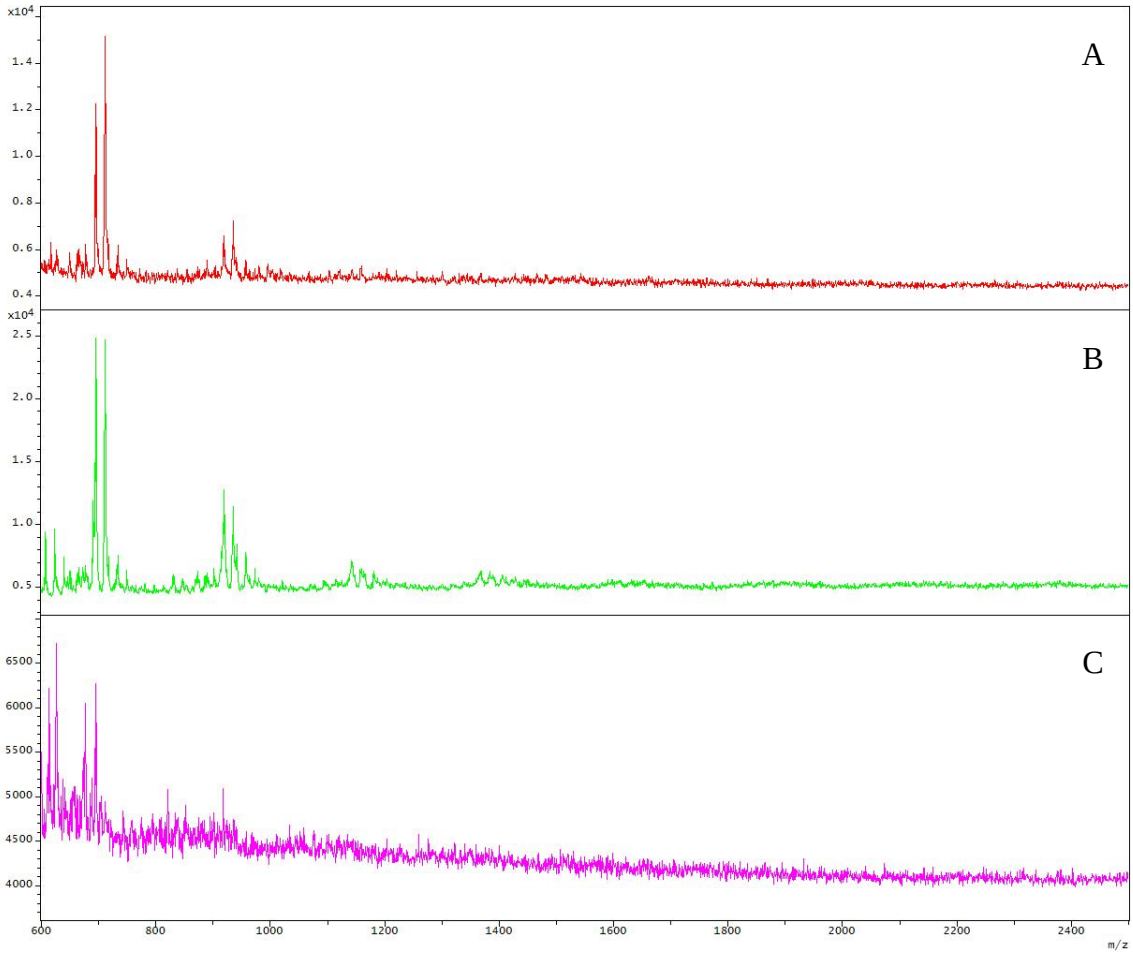


Figure 9 – MALDI-TOF MS spectrum of the BC patient´s sample prepared by chloroform/methanol precipitation, followed by three different digestion methods designated (A) I, (B) II, and (C) III.

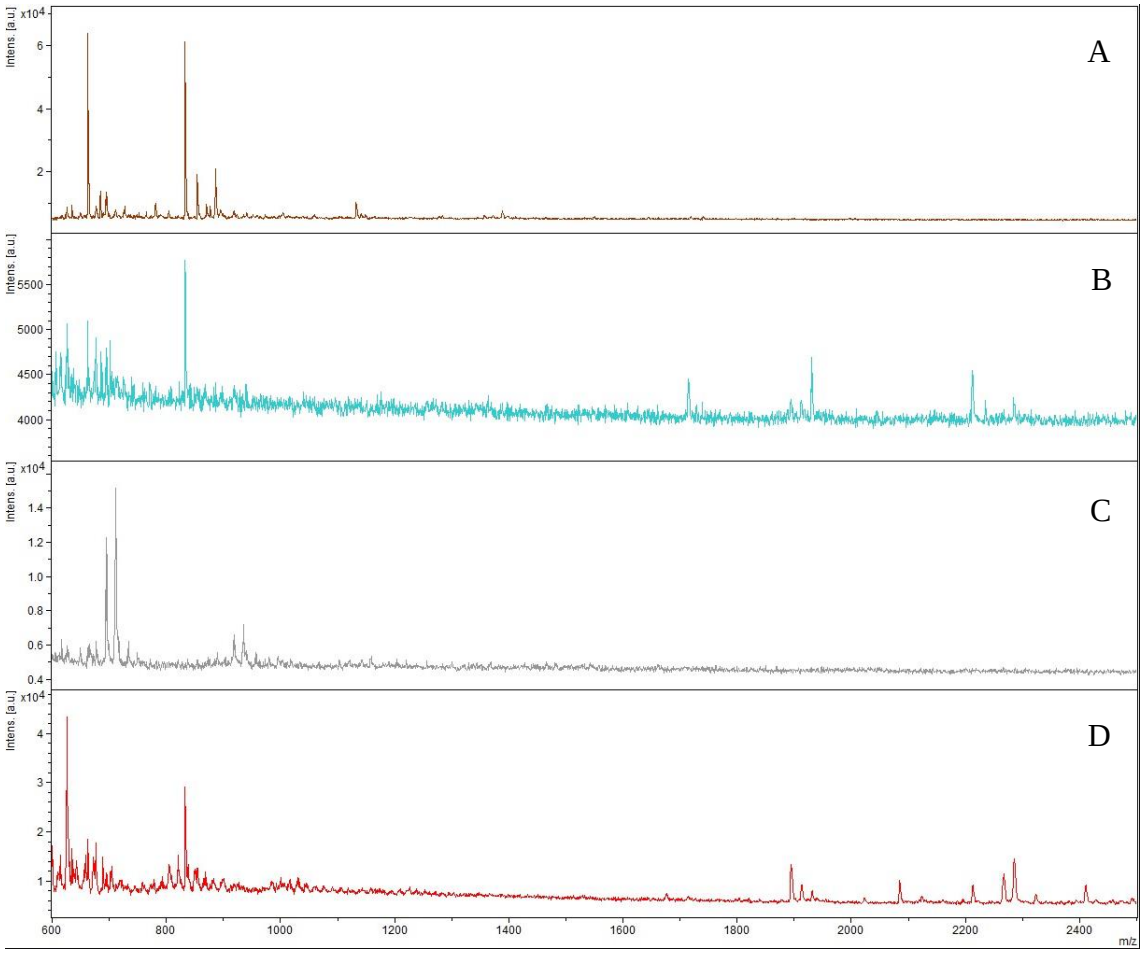


Figure 10 – MALDI-TOF MS spectra of the BC patient's sample prepared by different precipitations. (A) chloroform/methanol precipitation; (B) ACN/TFA precipitation; (C) TCA/acetone; (D) TCA precipitation both followed by the second digestion method (designated method II).

3.2. Total protein content by Lowry´s assay

Bearing in mind that the preparation of protein samples is extremely important for the analysis and preliminary screening of biomolecules, the total protein concentration was determined, before and after precipitation, by a chemical method, Folin-Lowry reactions, to determine the higher resolution and facilitate further analysis. According to Waterborg [124], the Lowry´s method is based on both the Biuret and the Folin-Ciocalteau reactions. In the first reaction, the proteins cause the reduction of cooper, under alkaline conditions, which in turn reacts with the Folin reagent (phosphomolybdate-phosphotungstate) into a strong blue pigment (which depends

partly on the tyrosine and tryptophan content) that is detected by UV-Vis spectrophotometry.

Therefore, after the optimization step, for this study, urines from 56 BC patients and 54 HCs were considered. One calibration curve was constructed by using concentrations between 9 to 277 µg/mL with a reproducibility expressed as relative standard deviation (RSD) less than 10%. The total protein calibration curve obtained by Lowry´s assay is illustrated in Figure 11, with BSA as standard.

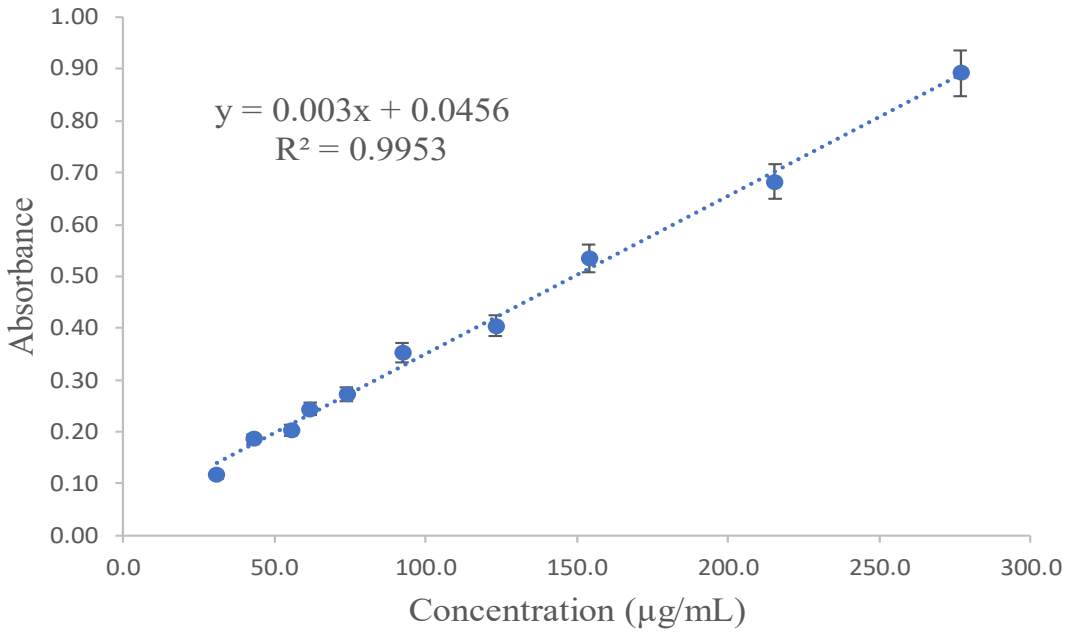


Figure 11 – BSA calibration curve for the Lowry´s assay.

Since earlier studies have demonstrated inter-individual variability in urinary proteins of 20–60% in HCs [125], it is crucial to emphasize that the composition and concentration of proteins in urines are related to intra- and inter-individual variations. However, through the results, the urines showed minimal inter-individual variation between the BC patients and HCs in the total protein concentration. Therefore, most urinary proteins are highly stable between individuals and, hence, serve as a source for the discovery of potential biomarkers.

In addition, there is evidence that BC patients have a less total protein concentration when compared to HCs, since the highest value originated from the control 39 (HC 39, 318±0.001 µg/mL) and the lowest from the patient 36 (BC36, 140±0.001 µg/mL). In the HCs, the total proteins range from 160 to 318 µg/mL, on

average 142 µg/mL, while in BC patients it varies from 140 to 311 µg/mL, in average 115 µg/mL, verifying that the highest concentration value obtained in urine from HCs is slightly higher than any protein concentration compared to the BC patients.

Additionally, the method has high sensibility until 10 µg of protein/mL, and it is preferably used in low protein concentration ranges (10 to 1000 µg/mL of protein). Under those conditions, the obtained values in Table 3 for total protein concentration of non-precipitated urine were within Lowry´s optimal concentration range. Thus, this method is suitable for total protein dosage in urine.

After precipitation, both HCs and BC patients' urines had a total protein concentration above the limit of quantification (LOQ), exactly to what would be expected considering that precipitation should increase the protein concentration by eliminating impurities from the biological sample (Table 3). The lower limit of quantitation (LLOQ) is the lowest amount of an analyte in a sample that can be quantitatively determined with suitable precision and accuracy [126]. However, it is necessary to take into account that UV methods are not accurate and, therefore, the higher protein content, after precipitation, may be related to the quantification not only of proteins but also of other compounds (urea, cell debris and other substances). For this reason, it would be necessary to perform several other protein dosing assays with a considerably larger number of controls as well as BC patients or, simply, use methods to remove the interfering substances to obtain more robust results.

Table 3 – Average protein concentration (µg/mL)± standard deviation of urines from HCs and BC patients.

Samples	Before precipitation	After precipitation
BC patients	115±1.60E-03	319±2.40E-03
HCs	142±1.60E-03	417±3.40E-03

To sum up, after precipitation, on average 31900 µg of protein was obtained from 100 mL and about 638 µg from 2 mL for BC patients, whereas HCs obtained an average of 41700 µg of protein from 100 mL and approximately 833 µg from 2 mL.

3.3. One-dimensional gel electrophoresis

In proteomics research, electrophoretic methods combined with the subsequent identification of proteins by MS are quite beneficial. Therefore, protein stains must be MS-compatible for proteomics studies, which limits the choice of silver stains. Furthermore, the most intriguing proteins in proteomic research are frequently low-abundance proteins with concentrations close to or below the 1D-PAGE detection limit. As a result, the Coomassie Brilliant Blue dye is the best staining approach for this investigation, as it enables the identification of proteins in the μg range. Thus, the most used method to stain the bands on a gel, which is immersing the gel in a solution of Coomassie Brilliant Blue, was used in the current investigation to visualize the entire gel lane.

1D SDS-PAGE is a high throughput technique widely used for determining the MW of proteins in denaturing conditions [127], in which proteins are separated based on their ability to move within an electrical current. The success of the electrophoretic separation depends heavily on selecting the proper polyacrylamide concentration because it can influence the pore size of the gel.

Therefore, 7.5, 10 and 12% polyacrylamide gels were used to observe which one allowed a more effective separation of the proteins present in the urines. As previously mentioned, the 7.5% gel which has a lower concentration of polyacrylamide, allows effective separation of proteins of high molecular weight (MW) (75 – 250 kDa), while on the 10 and 12% gels (containing a higher concentration of acrylamide, which entails a smaller pore size in the gel) the proteins of low MW (10 – 75 kDa) were accurately separated [96]. Taking these results into account, the best insight into the protein composition of the samples was obtained with 12% polyacrylamide gels, which offered the most selective separation of proteins within a specific MW distribution. Thus, the concentration of 12% was selected as optimal for further 1D SDS-PAGE analysis, since the present study was designed to separate as many proteins as possible in the complex biological samples.

Following the 1D SDS-PAGE analysis, the SDS-PAGE patterns of urinary proteins from BC patients and HCs samples are shown in Figure 12 and suggest potential differences in each proteome. The major differences between BC and HCs urine were observed in the range 50 – 75 kDa, where clear bands of high-concentration proteins with higher intensity in BC patients than in HC urines were found. In addition,

control had less or no protein between 60 – 75 kDa compared to BC urines. The software Gel-Analyzer 19.1 [118] was used to identify 536 proteins and evaluate their molecular weights, relative migration distance (Rf) and band intensities (Table A1 – Appendix, pages 103-116). Although SDS-PAGE could not show the major differences between sample preparations because of limited separation in one dimension, it allowed to separate spots that represent the differentially expressed protein, which were excised from stained gels. Thus, for protein profile analysis, we excised around 220 gel spots from selected proteins of interest and were in-gel digested them with trypsin and the resulting peptide mixture analyzed by MALDI-TOF MS, which led to the identification of differentially significant proteins in each group.

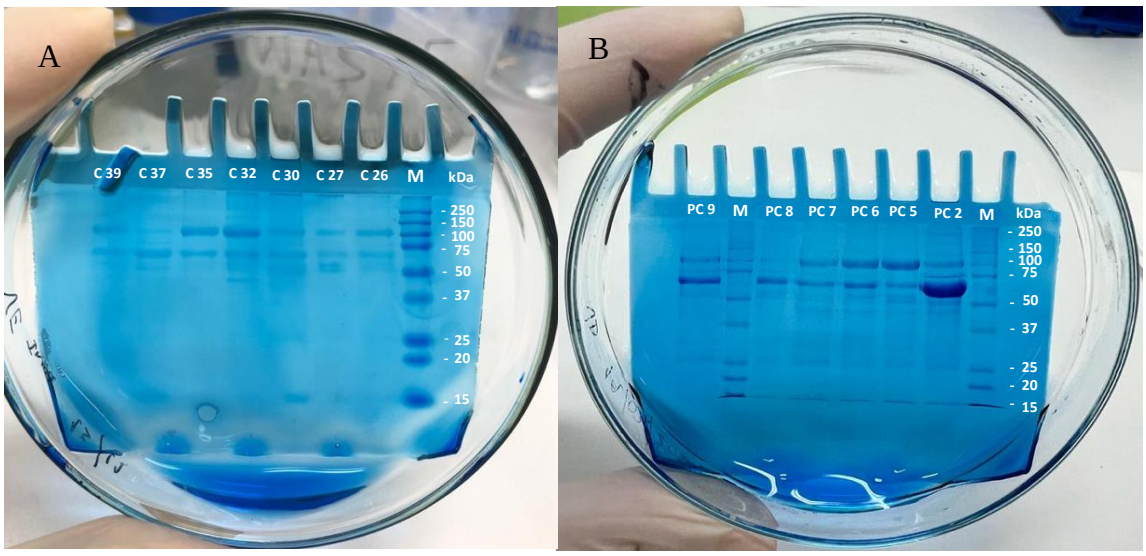


Figure 12 – Representative results of 1D SDS-PAGE analysis from (A) HCs and (B) BC patients.

However, it was challenging to produce a 1D gel with high-quality resolution and reproducibility due to the complexity of human urine. Because of this, in proteomic research, where it is occasionally necessary for the solution of thousands of proteins on a single gel, 2D PAGE provides the highest resolution for protein analysis.

As observed by Yang *et al.* [112], the 2-DE-based proteomics approach was used to compare changed proteins in the shock group with the non-shock group to identify proteins that were differentially expressed by MALDI-TOF MS. Additionally, research published by Robati *et al.* [128], examined the variations in protein expression between cancerous and normal esophageal tissues and found 4 differently expressed proteins using 2D-PAGE and MALDI-TOF MS. Fayazfar *et al.* [129], verified the separation and identification of plasma proteins using 2-DE followed by MS/MS, which

demonstrated to be possible biomarkers for early diagnosis of patients with RCC. Later, Noreen *et al.* [130] stated that spotted more differentially abundant spots on 2D-GE gel, which exhibited eleven proteins, for the high-throughput proteomic analysis. Undoubtedly, the additional study must be conducted using a 2D-PAGE separation of proteins to verify the results.

3.4. Digestion step

Another essential method in proteomics is proteolysis, which involves characterizing the digested peptides for MS identification, confirmation, or deduction. Additionally, the analytical separation by one-dimensional electrophoresis showed that the extracts mainly contained large proteins/peptides (>30 kDa) [131, 132]. For this reason, spots representing the differentially expressed proteins were excised, with the help of a small spatula, from stained gels and in-gel digested by trypsin. This was another proteomic approach, which breaks proteins into smaller units, called peptides. The protein is digested by the addition of protease (trypsin), which cleaves proteins at predictable amino acid locations and improves measurement accuracy with lower mass peptides.

Trypsin was selected for the current investigation because it is the most often used serine protease for protein/peptide identification and cleaves proteins into peptides with an ideal size range for MS between 700 and 1500 Da [130].

3.5. Proteomic/peptidomic profile

To determine the MALDI-TOF MS limit of detection (LOD), several calibration curves were obtained for peak area in function of BSA concentrations (5.00 – 50.00 mg/mL) using SA matrix both in linear and reflector positive-ion modes over a mass range of 30-210 kDa and 600-2000 Da, respectively.

However, it was not possible to detect all the BSA signals (BSA single charged m/z = 66400, BSA double charged m/z = 33200, and BSA triple charged m/z = 22100) in any of the spectra of the different BSA concentrations. For that reason, the m/z values with the largest area around the 66400 and 33200 peaks were considered, since BSA signals are rather broad and constituted by multiple overlapped peaks.

Therefore, firstly, SA matrix was used in a linear mode in a high mass range (30-210 kDa) to correlate the 66400 and 33200 m/z peak areas with BSA concentration. By comparing Figures 13A and 13B, it is evident that there is a good logarithmic correlation, within the analyzed BSA concentration range, particularly for the main BSA peak (66400 m/z). Also, it is verified in all BSA concentrations that the 66400 m/z peak had larger areas than the 33200 m/z peak, as expected since it corresponds to the single-charged BSA protein.

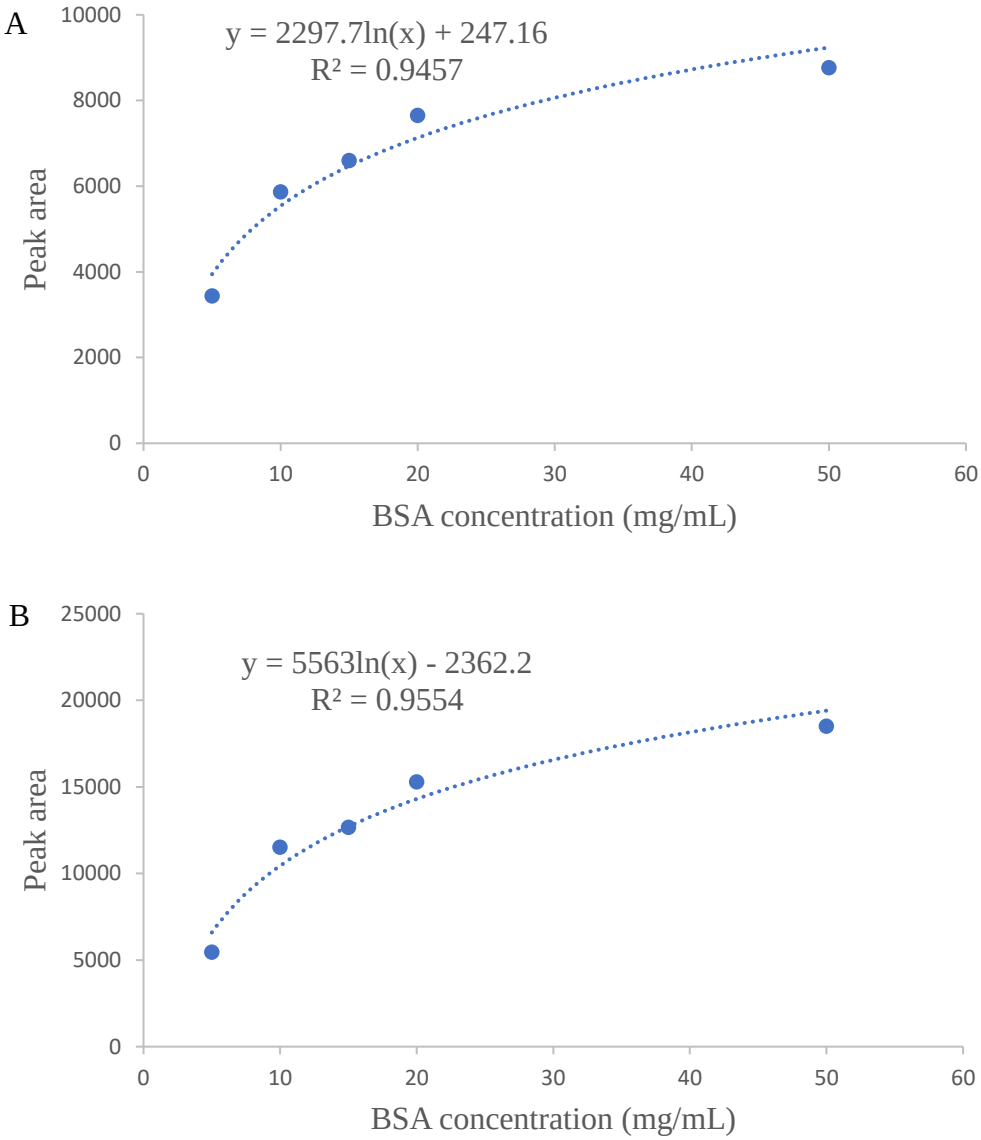


Figure 13 – Calibration curves of peak area in function of BSA concentration for 33200 (A) and 66400 (B) m/z signals.

Complementary, to prove the ability of MALDI-TOF MS for the detection of proteins/peptides in urine, using the methods outlined in the current study, the BSA

samples were submitted to the same procedure and afterwards analyzed in the low mass range (600–2000 Da) in reflector mode. According to Stensballe *et al.* [133] and Shevchenko *et al.* [134], the two peaks with m/z 847.0 and 927.0 values are considered peptide precursors of BSA. As a result, given that the digestion technique is significantly different from those described in the aforementioned research, it was considered that the two peaks with 876 and 985 m/z might have originated from BSA. This is particularly visible because there is a good correlation between the increase in BSA concentration and the area of the corresponding peaks, as shown in figures 14A and 14B.

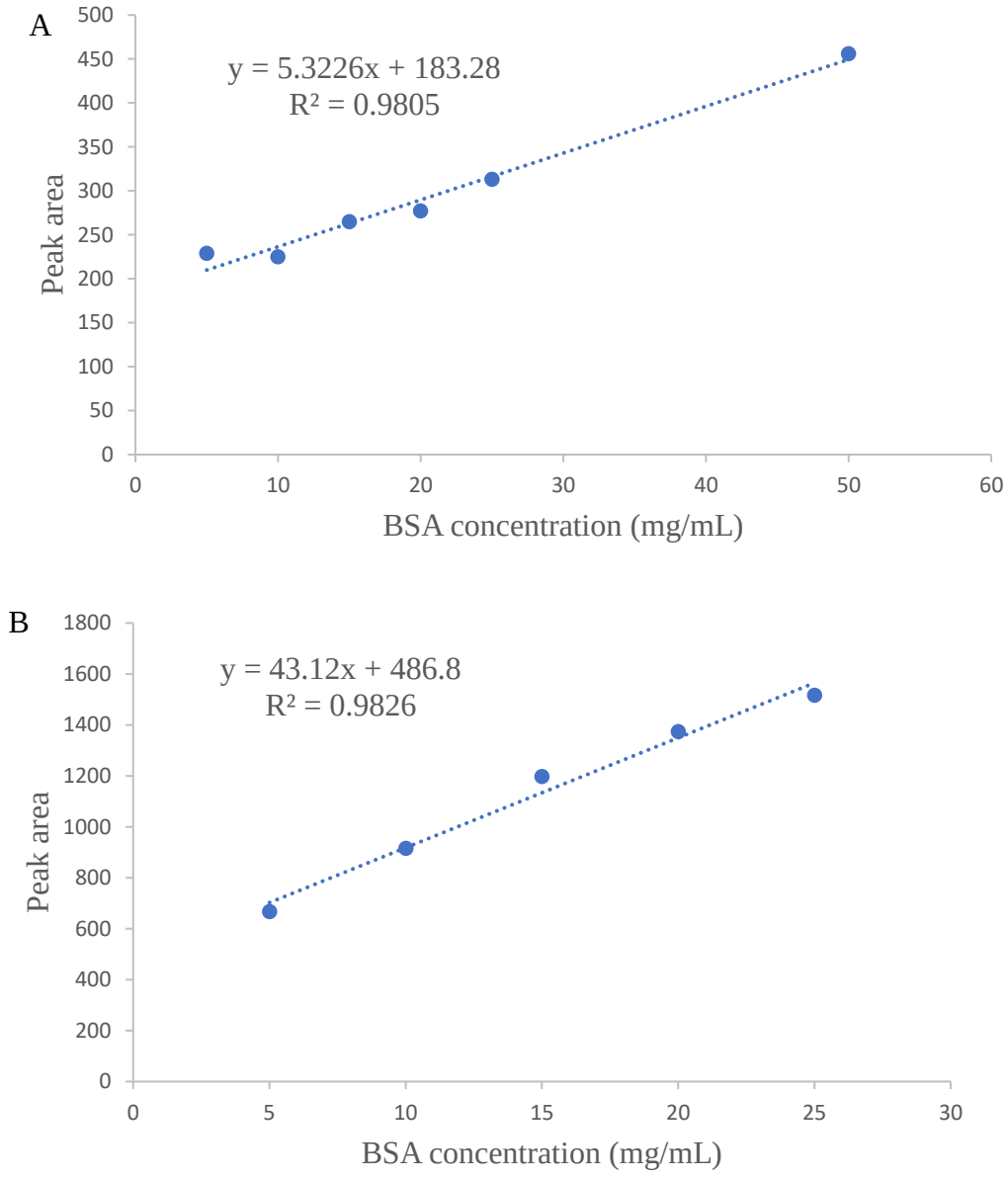


Figure 14 – Calibration curves of peak area in function of BSA concentration for (A) 847 and (B) 927 m/z peptide BSA signals.

In addition, considering that the MALDI-MS analysis of low molecular weight peptides faces some limitations because matrix clusters are quite abundant, Salum *et al.* [135] showed that cis-sinapic acid (Z-SA) is a good matrix of choice for short peptide analysis since provides spectra with higher intensity, peptide signal and fewer matrix clusters, but specifically in positive ionization mode. This is because SA belongs to the cinnamic acid derivatives class, and these currently used are E-cinnamics. Additionally, Choi *et al.* [136] proved that SA is the most suitable MALDI matrix since it gives only a few background peaks while providing high sensitivity detection for the analysis of proteins and peptides. As shown in Figure 15, in positive mode, it is possible to observe that for 1:1 urine/matrix solution spectra there are several peaks originating from analytes present in urine and not necessarily by SA.

This study used the bottom-up MS-based approach to identify candidate proteins/peptides from MALDI mass spectra between the two groups. The spectra were analyzed in the mass range of 600–3500 Da in the reflector mode, which requires using sufficient resolution. That is, resolved spectra and great sensitivity are achieved because the analytes are stable enough to withstand the energetic stress which is inherent to passing the reflector. It enables proper baseline separation of the analyzed peaks and highly accurate determination of their mass [137]. Our purpose here was to combine datasets of urines from BC patients and HCS to discover potential urinary protein/peptide biomarkers through commonly applied standard MALDI-TOF MS techniques. It was used MALDI MS software, respectively by FlexControl and FlexAnalysis version 3.4, to analyze the urinary peptidome of 110 samples and the mass spectra of samples from BC patients were compared with those from HCs.

Visual analysis of the raw data, as can be seen in Figure 15, shows that there are differences in peak intensity between urine mass spectra from BC patients and those from HCs. Additionally, by comparing the spectra of the different groups, it is possible to verify that a cluster of peaks identified in the 1000–2000 m/z range is only present in the urines of the HCs. However, not all HCs had these peaks, indicating that there was significant inter-individual variance in the HCs urines. On the other hand, although the resolution is significantly higher than linear MALDI-TOF profiles, it is not high enough to resolve the signal intensity from overlapping peptides. This is due to the fact that the resulting raw data spectrum consists of the signal distinctive to the samples as well as signals with varying levels of noise, making the recognition/detection of peaks indicative of the samples a difficult and, inevitably, multistep procedure. For this

reason, the high inter-individual variability in urinary profiles and their large complexity make any attempt at visual comparison of these spectra an unsuccessful task. Further, these peaks have similar m/z values compared to BSA, however since the cross-contamination of samples was avoided, there is another possible compound abundantly found in urine that could provide such signals, that is human serum albumin (HSA).

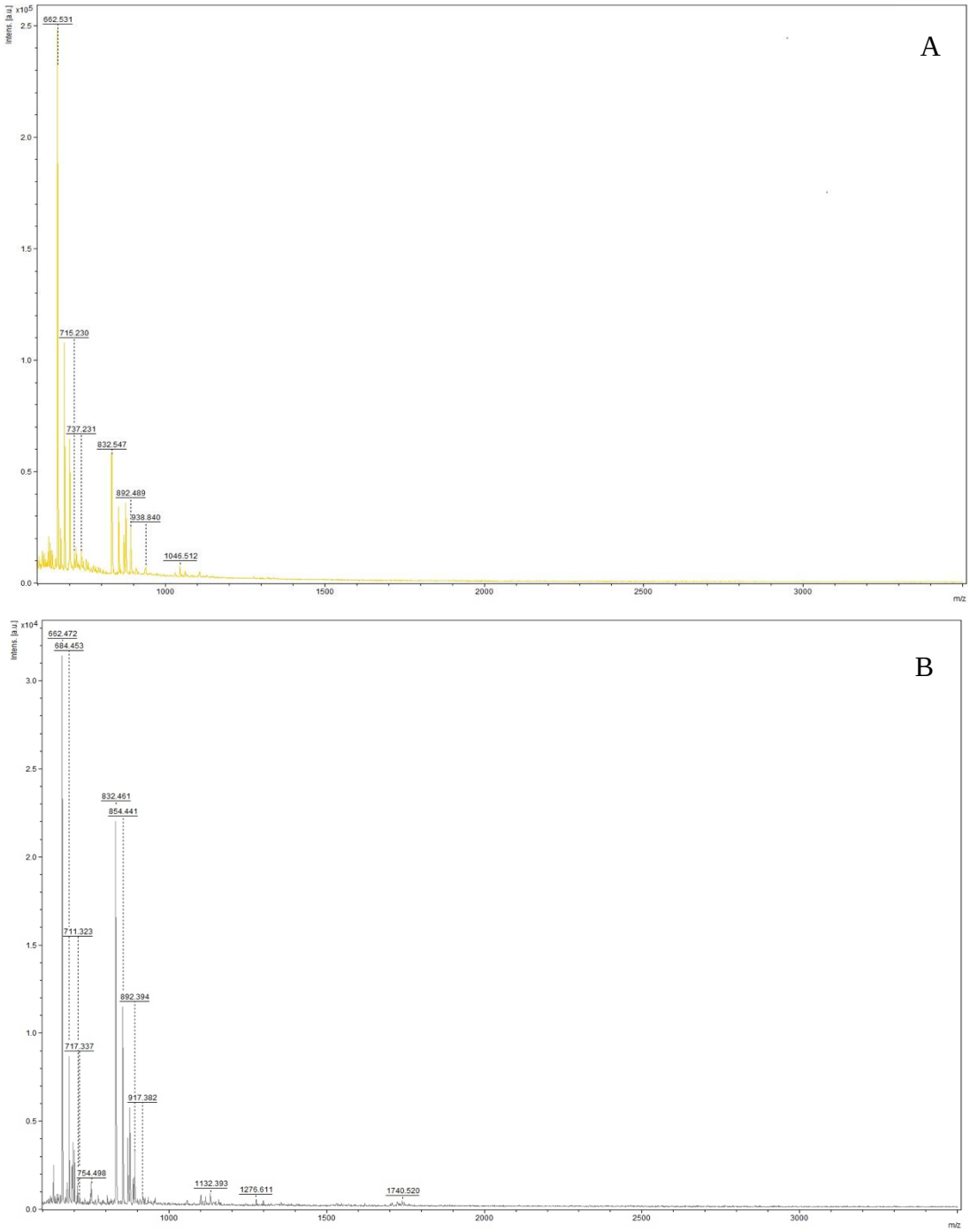


Figure 15 – MALDI-TOF MS spectra of a sample from (A) BC patient and (B) HCs.

Through the hierarchical cluster analysis (HCA), it was possible to screen the peaks considered statistically significant, with a value of $p < 0.05$, to establish an objective comparison among the proteomic/peptidomic pattern found in the urine of BC patients and HCs in terms of qualitative (identification by comparison of MS spectrum). This allowed for better analysis and interpretation of the peaks detected in the different groups. Figure 16 displays the peaks with m/z values of 1027.9, 1074.9, 1106.8, 1147.9, 1188.0, 1228.0, and 1731.6, which could discriminate between all peptide samples from BC patients compared to the HCs.

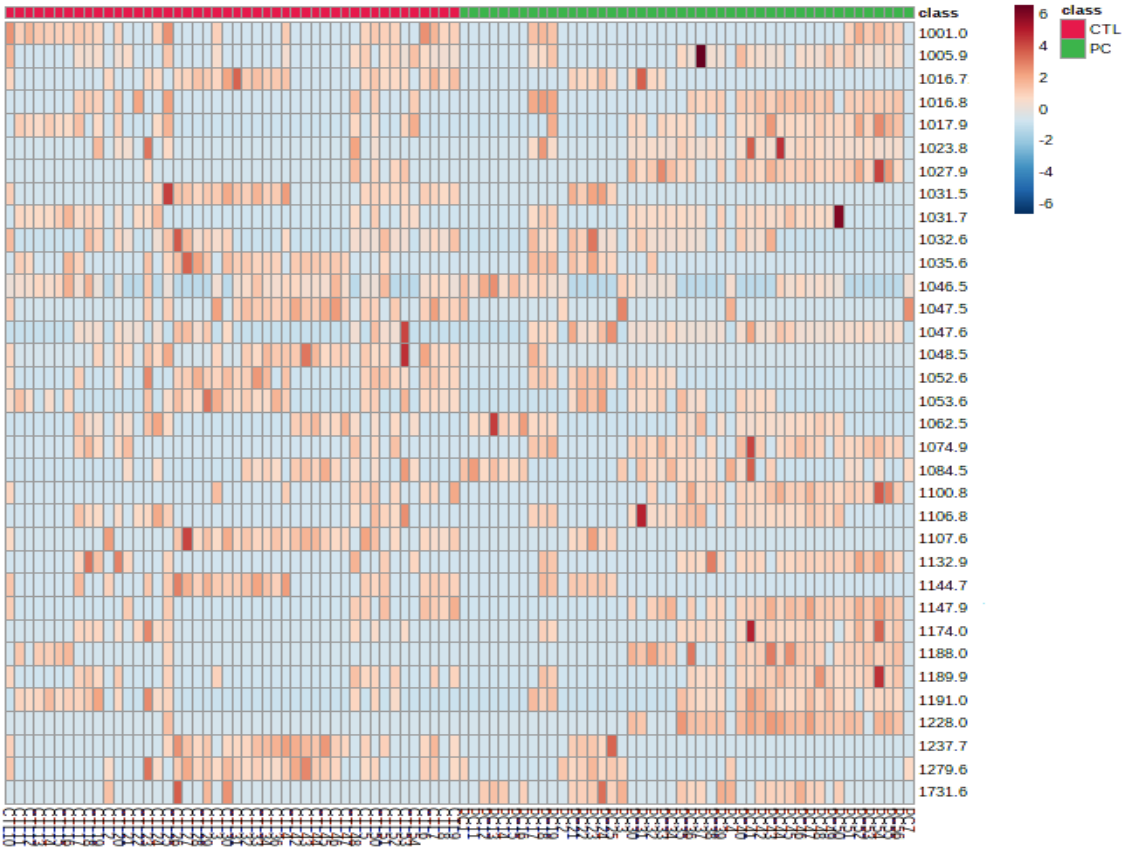


Figure 16 – Heatmap of m/z values for samples from BC patients and healthy control (CTL).

However, they cannot be used as possible BC biomarkers because not all BC patients' urine mass spectra have these peaks. In addition, due to a lack of specificity, more samples will need to be collected in future studies to create a urine proteomic signature that is more accurate for detecting BC.

3.6. Pre-treatment of the MALDI-TOF spectra

Correspondingly, the primary objectives of preprocessing are the determination of the m/z and along with correctly accurate intensity, and the designation of peaks with normalized intensities below a threshold as noise with consideration of any of the m/z values of interest [116, 138].

An example of the spectrums obtained from the BC patients and HCs after pre-treatment and peak alignment procedures following the Mass-UP options is presented in Figures 17 and 18, respectively. To determine reliably detectable peaks, a combination of parameter values was applied, to clean the large data set from signal noise and identify true signals, following the software workflow [139]. With this pre-treatment it was reduced the m/z range, reducing the number of peaks to be analyzed from the 200000 in the raw spectra to 1000 – 1500 of the pre-processed data. Therefore, based on this obtained profile, a cluster of signals in the 1000 – 3500 m/z regions were chosen as the specific signature capable of differentiating BC patients from the HCs out of the MS features with the highest discriminatory value, statistical rank, and relative m/z peak intensity.

With a view to minimizing systematic variance and to improve the performance for downstream statistical analysis, normalization and transformation of the intensity are performed. According to Meuleman *et al.* [140], the spectra normalization is a crucial step in pre-processing and that, despite its simplicity, total ion current (TIC) is the best option in profile experiments, with the assumption that the quantity of proteins/peptides with variable expression is significantly less than the quantity of total proteins/peptides in the sample [141]. Since the raw data are counts of ionized molecules with intensity values roughly following a Poisson distribution [142], a square root transformation can be used to convert the Poisson distributed data to approximately normal data, with constant variance independent of mean, which is an important requirement for many statistical tests [143]. Subsequently, the transformed spectral data were smoothed by Savitzky Golay algorithm, which is based on polynomial regressions in a moving window [144-146]. Moreover, the statistics-sensitive non-linear iterative peak-clipping algorithm (SNIP) algorithm, an interactive algorithm that computes the baseline by considering the local minima and local mean intensities in windows of increasing size, was used to remove the background effects to reduce their influence in the quantification of the peak intensities [147]. As stated by Bromba *et al.* [148], the impact

of half window size (HWS) values should be chosen to be smaller than twice the full width at half-maximum (FWHM) of the peaks.

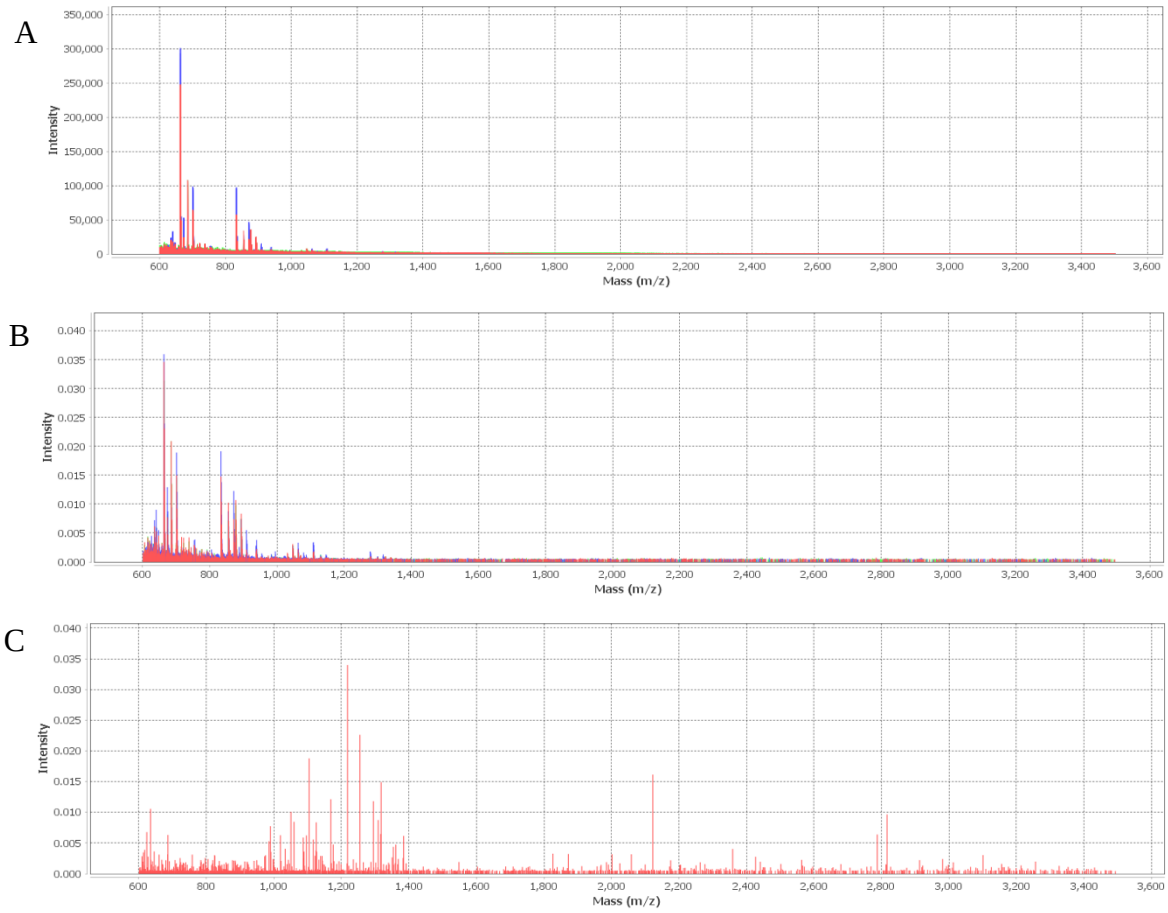
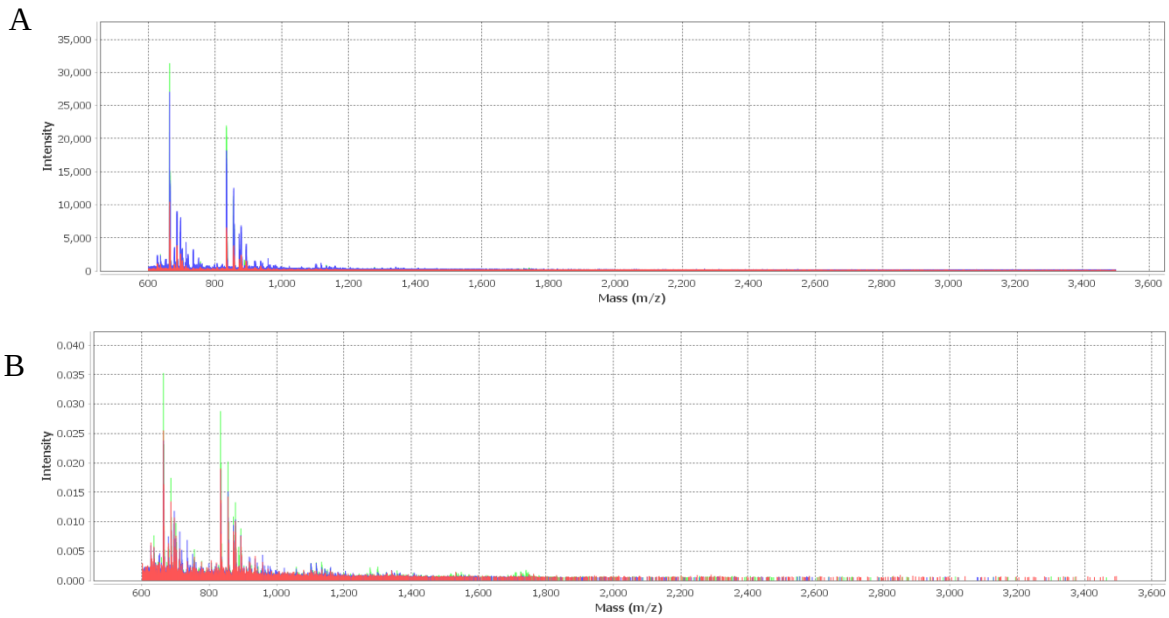


Figure 17 – Example of MALDI-TOF MS spectra from a BC patient on the Mass-UP software. (A) Raw spectra; (B) Preprocess and peak detection; (C) Peak alignment.



C

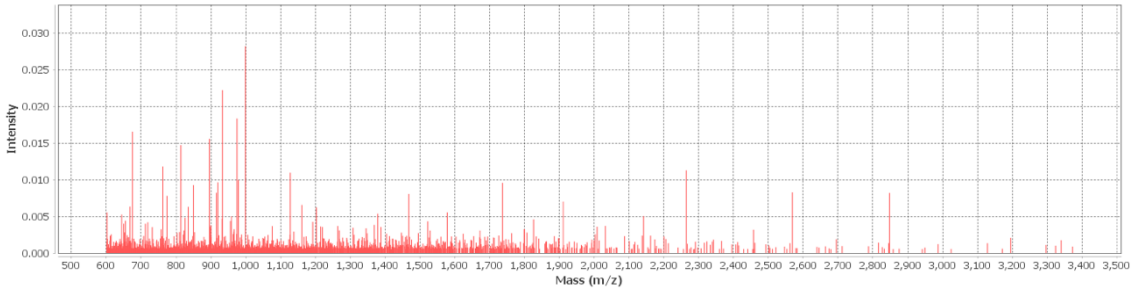


Figure 18 – Example of MALDI-TOF MS spectra from a HC on the Mass-UP software. (A) Raw spectra; (B) Preprocess and peak detection; (C) Peak alignment.

However, as the protein/peptide composition in urine is complex and variable, which is greatly affected by diet and exercise, the statistical results was performed on 26 healthy control subjects and 46 BC patients, suggesting that the remainder are putative outliers. After filtering the noise and clustering of the 72 spectra obtained by combining the list of peaks of the spectra in triplicate of each sample, a total of 74391 peaks were identified, with an average of 1033 peaks per sample.

3.7. Multivariate analysis

The statistical analysis was performed using the Metaboanalyst 5.0 [149] web server as described in the material and methods section. To study the principal sources of variation among results, detect sample clustering and possible outliers, and to establish if the cancer patients could be distinguished from healthy individuals, exploratory PCA was applied to each mass spectra obtained for the urinary samples of both groups. Only the potential BC biomarkers with confidence regions higher than 95% were considered for the statistical analysis, where peaks with a similar m/z value across all spectra (within 0.025%) were considered the same.

Initially, data were transformed by cube root transformation (to compare the magnitudes of each feature to one another) and mean centering approaches, before being subjected to multivariate statistical analysis. In a univariate statistical analysis, a total of 6 distinct peaks were identified between the two sets of samples ranked by *p* values from the *t-test* and four of them were significantly expressed. Of those, four peaks were up-regulated while two peaks were down-regulated in BC patients compared with the control subjects.

To gain a better perception between the similarities and differences among the groups, an unsupervised multivariate approach PCA was used. To obtain better visualization between the correlations, the first (PC1) and second (PC2) principal components, accounting for 59.4% of the total variability, were used to create a 2D representation (Figure 19) that illustrates the PCA results for BC patients and HCs. For both groups, the projection of the sample forms distinct clusters. The group located in PC1 positive comprises the samples corresponding to the BC patients, while the group located in PC1 negative corresponds to the HCs. These results showed that general similarities and differences of the samples under study can be visualized and, therefore, the set of cases can be divided into two different groups according to the clinical condition of the subjects. The most influential features with a VIP value between 0.8 – 1.2 are represented in Figure 20.

Figure 19 – Loading score plots of principal component analysis, PCA (59.4% of total variance).

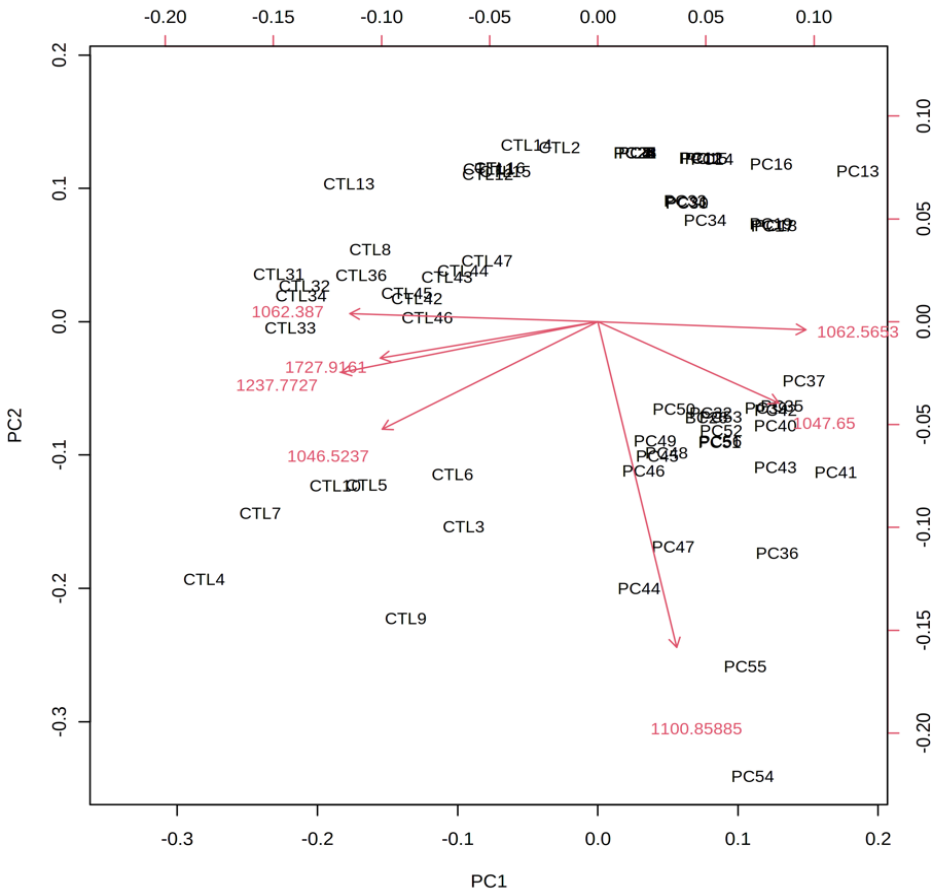


Figure 20 – Biplot representation on the factor-plane (PC1 vs PC2).

Moreover, the orthogonal projection to latent structures discriminant analysis (OPLS-DA) was applied to the urine proteomic/peptidomic profile dataset to maximize the separation of BC and healthy control groups. Additionally, this type of statistical analysis allows verifying the existence of possible outliers through the discrimination among the two groups. Thus, significant group separation was observed in the OPLS-DA score plot between BC patients and healthy control group indicating intrinsic proteomic alterations in each group (Figure 21).

The VIP value obtained by OPLS-DA is an important measure of each independent variable. Higher VIP values are considered more relevant in the ranking. Hence, the top four features (m/z 1046.5, 1062.4, 1237.7 and 1727.9) with the highest contribution to group discrimination were selected with variable importance in projection (VIP > 1) – Figure 22.

To attest to robustness of the model, a random permutation test with 1000 permutations was performed with OPLS-DA (Figure 23). The permutation test yielded R^2 (represents goodness of fit) as 0.821 and Q^2 (represents predictive ability) as 0.777 indicating that the model is not over-fitted and has a relatively good predictive ability to distinguish between study groups.

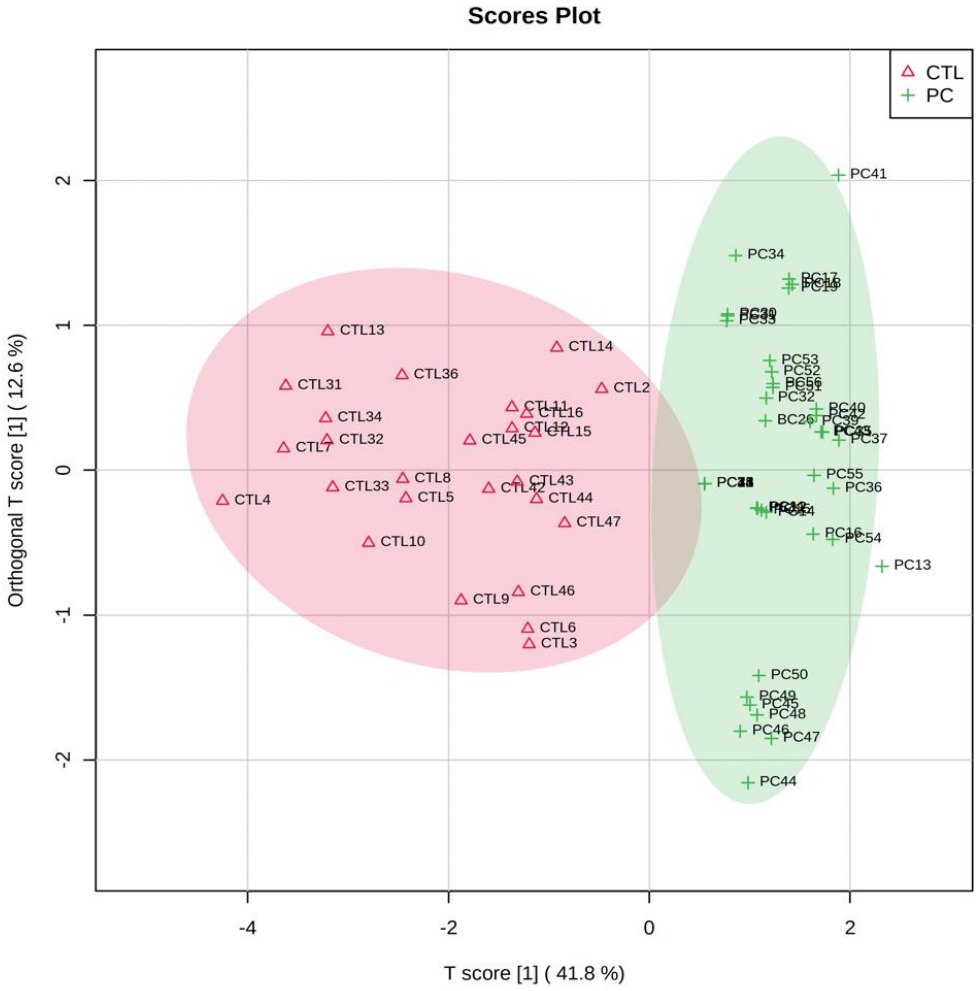


Figure 21 – Loading score plots of orthogonal projection to latent structure discriminant analysis (OPLS-DA).

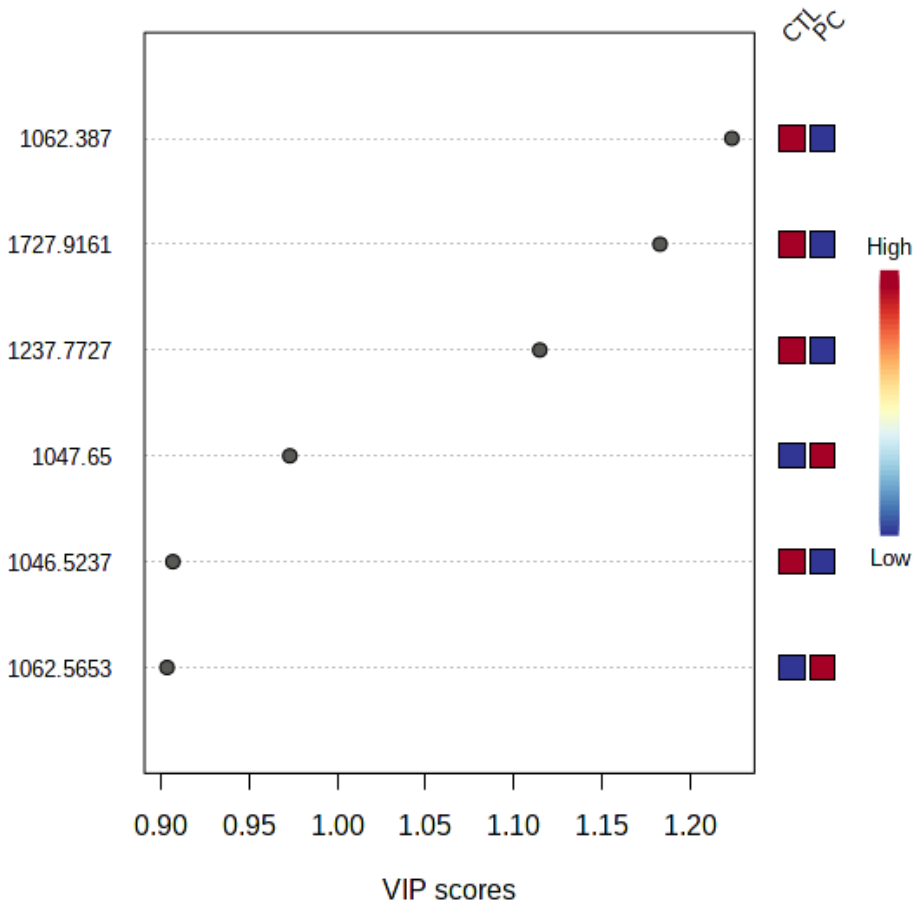


Figure 22 – Main significant characteristics based on VIP values of the component under study. The colored boxes on the right indicate the relative concentration of proteins/peptides in each sample group.

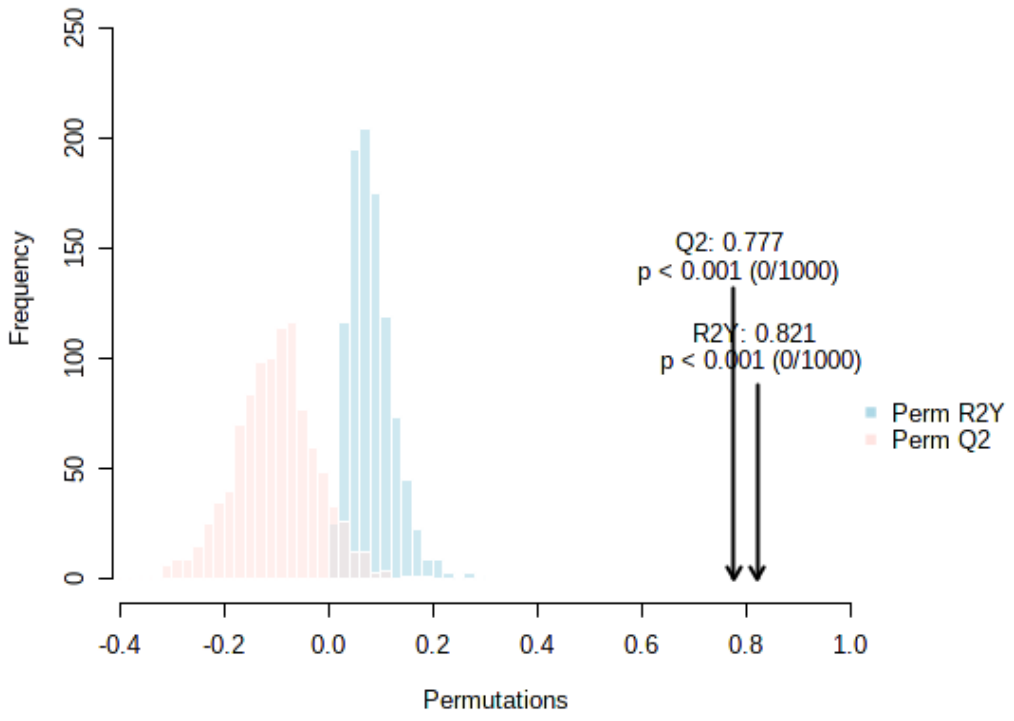


Figure 23 – Model validation by the permutation test based on 1000 permutations of proteins/peptides obtained by MALDI-TOF MS of urines from the 2 groups under study.

To further evaluate the predictive value of the peaks to discriminate between BC patients and HCs, a ROC analysis was generated using the top four features identified by VIP values (Figure 24). 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) [150, 151]. According to Xia *et al.* [152], 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 = 0.996), representing good accuracy in discriminating the two groups.

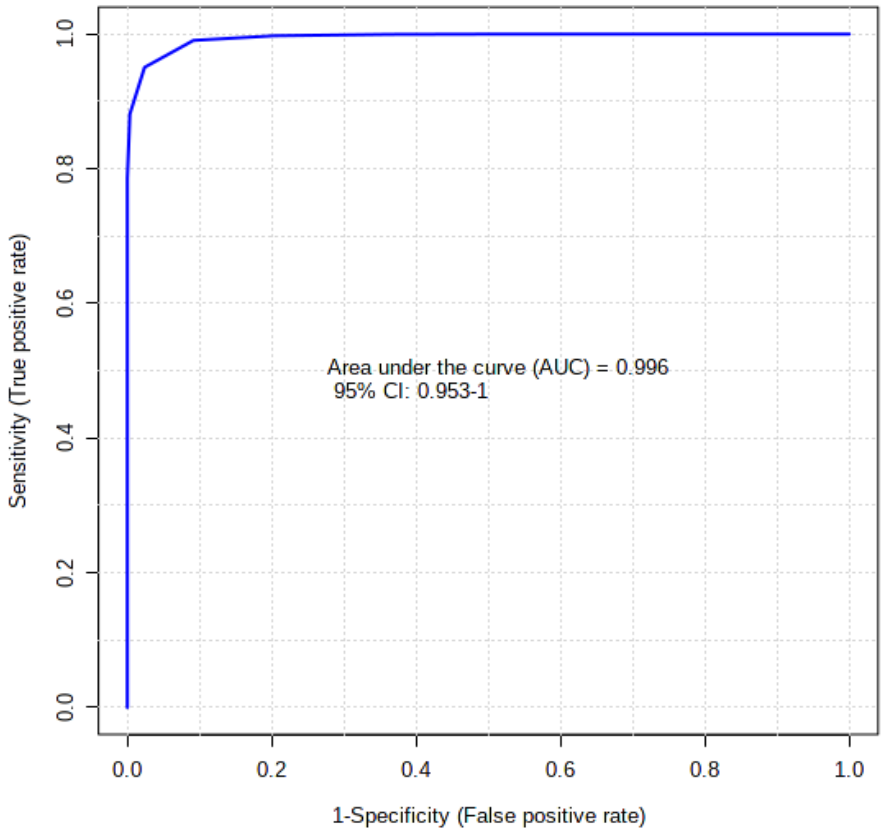


Figure 24 –ROC curve for the predictive model. ROC curve for the top four features (m/z 1046.5, 1062.4, 1237.7 and 1727.9) with the highest ability to discriminate BC patients against the HCs.

For HCA analysis the data set that was uploaded and the reduced discriminant peak list was used to restrain the analysis only to those peaks. Thus, the heat map was constructed with selected proteins/peptides by VIP > 1, using Euclidean's correlation, to provide an intuitive visualization of the data set and the correlations between samples of the two groups (Figure 25). Similar samples are clustered hierarchically. It is possible to see that, for both groups, the features m/z 1046.5 and 1237.7 and the m/z 1062.4 and 1727.9, form clusters with a closer association between them, confirming that they were significant peaks for the HCA discrimination. However, confirmation of such relations requires bigger data set for pattern verification.

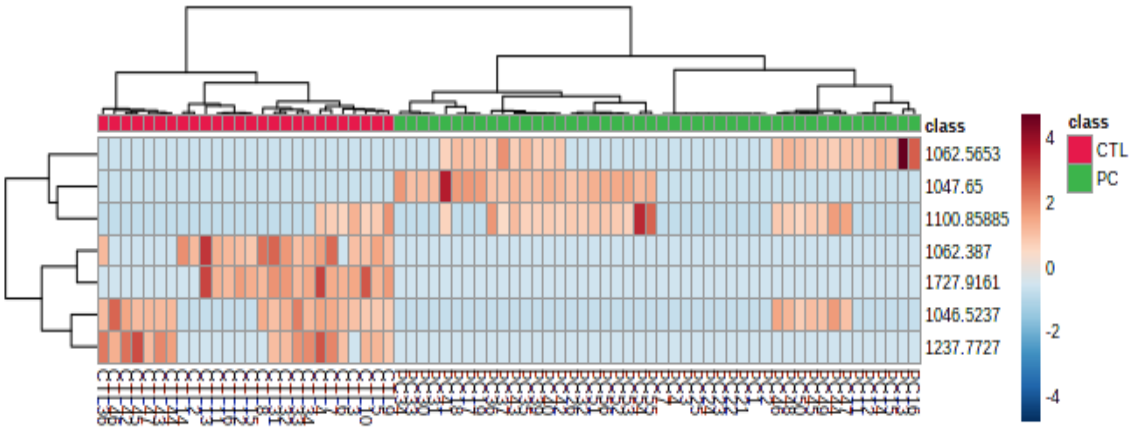


Figure 25 – Hierarchical cluster analysis: Heatmap of proteins/peptides present in two distinct groups of urine samples, BC patients *versus* healthy control (CTL). Columns correspond to each sample group, while rows correspond to the proteins/peptides colored from minimum (-2, dark blue) to maximum (2, dark red).

Therefore, based on the results obtained, the final set of discriminative peaks of urines was successful, using MALDI-TOF MS, showing that the urine proteomic/peptidomic signature of BC can be a useful approach to identify potential BC biomarkers. Obviously, future studies on a larger cohort of control subjects and patients will be needed to confirm our preliminary findings and assess the clinical relevance of our approach.

4. CONCLUSION

The identification of potential BC biomarkers in urines is a strategy of growing importance for the prevention and monitoring of disease progression, whose biological research, through analysis by MALDI-TOF MS, aims to differentiate HCs from BC patients. From this perspective, 1D SDS-PAGE combined with MALDI-TOF MS and statistical analysis could offer significant potential for the clinical routine and for customized treatment of breast cancer, through the establishment of urine proteomic/peptidomic biosignature. The Lowry's assay was suitable for the total protein dosage in urine. Regarding the protein extraction method, to further optimize the yield of the precipitated samples, the preliminary analysis of urines merits further investigation, to produce reliable results and, consequently, better validate this conclusion. 1D SDS-PAGE, a simple and speedy separation method, allowed identifying the different protein patterns between the two groups.

This study demonstrated that analysis of the urinary proteome from BC patients and HCs by 1D gel electrophoresis and MALDI-TOF MS is suitable for the establishment of specific proteomic/peptidomic patterns useful on BC diagnostics. According to multivariate statistical analysis, there were four features out of a total of 7 identified signals that were shown to be statistically different between the BC patients and the HCs. Of the features identified, m/z 1046.5, 1062.4, 1237.7, and 1727.9, showed the highest sensitivity and specificity (99.6%) to discriminate BC patients from the HCs.

The obtained results shows the potential of the used methodology establish BC urinary proteomic patterns as a promising strategy for the screening, detection, and treatment follow-up of the disease, which may make up for the lack of biomarkers for BC.

In future work, the two-dimensional gel electrophoresis technique should be used since it allows, more effectively, the separation of complex protein mixtures in proteomic studies. Due to its coupling with immobilized pH gradients, IPG-Dalt, it provides better reproducibility in addition to higher resolution and load capacity for preparative purposes. That is, 2-DE can achieve the separation of several thousand different proteins in a gel. Also, to simplify the sample and improve peak intensities, the digested peptides present in urines should be separated by HPLC before MALDI-TOF MS analysis.

5. REFERENCES

[1] Bray, F., Laversanne, M., Weiderpass, E., Soerjomantaram, I. The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer*, 2021; 127(16): 3029. DOI: 10.1002/cncr.33587.

[2] Eggert, J. Precision medicine: biologics and targeted therapies. In Eggert, J. (Ed), Cancer Basics, 2nd Edition, Oncology Nursing Society, Pittsburgh, PA, 2017: 1.

[3] World Cancer Research Fund International (WCRF). Worldwide cancer data, 2022. Accessed December 11, 2021. URL: https://www.wcrf.org/cancer-trends/worldwide-cancer-data/.

[4] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjimantarm, I., Jemal, A., Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *A Cancer Journal for Clinicians*, 2021; 71(3): 209. DOI: https://doi.org/10.3322/caac.21660.

[5] Yabroff, K.R., Wu, X.C., Negoita, S., Stevens, J., Coyle, L., Zhao, J., Mumphrey, B. J., Jemal, A., Ward, K.C. Association of the COVID-19 pandemic with patterns of statewide cancer services. *Journal of the National Cancer Institute*, 2022, 114(6): 907. DOI: 10.1093/jnci/djab122.

[6] Sharma, G.N., Dave, R., Sanadya, J., Sharma, P., Sharma, K.K. Various types and management of breast cancer: An Overview. *Journal of Advanced Pharmaceutical Technology & Research*, 2010; 1(2): 109.

[7] Yalaza, M., Inan, A., Bozer, M. Male breast cancer. *The Journal of Breast Health*, 2016; 12(1): 1. DOI: 10.5152/tjbh.2015.2711.

[8] Jain, S., Gradishar, W.J. Male breast cancer. In The Breast, comprehensive management of benign and malignant diseases, 5th edition, 2018: 1. DOI: https://doi.org/10.1016/C2014-0-01946-6.

[9] Irwig, L. Evaluating new screening tests for breast cancer. *British Medical Journal*; 2006, 332(7543): 678. DOI: 10.1136/bmj.332.7543.678.

[10] Liang, W., Lawrence, W.F., Burnett, C.B., Hwang, Y., Freedman, M., Trock, B.J., Mandelblatt, J.S., Lippman, M.E. Acceptability of diagnostic tests for breast cancer. *Breast Cancer Research and Treatment;* 2003, 79: 199. DOI: 10.1023/a:1023914612152.

[11] World Health Organization (WHO). WHO Position Paper on Mammography Screening WHO; 2014. Accessed December 14, 2021. URL: paho.org/hq/dmdocuments/2015/WHO-ENG-Mammography-Factsheet.pdf.

[12] Banach, P., Derezinski, P., Matuszewska, E., Matysiak, J., Bochynski, H., Kokot, Z.J., Nowak-Markwitz, E. MALDI-TOF-MS analysis in the identification of urine proteomic patterns of gestational trophoblastic disease. *Metabolites*; 2019, 9(2): 1. DOI: 10.3390/metabo9020030.

[13] Pin E, Fredolini C, Petricoin EF, 3rd. The role of proteomics in prostate cancer research: biomarker discovery and validation. *Clinical Biochemistry,* 2013; 46(6): 524. DOI: 10.1016/j. clinbiochem.2012.12.012.

[14] Tanase, C. P., Codrici, E., Popescu, I. D., Mihai, S., Enciu, A.M., Necula, L.G., Preda, A., Ismail, G, Albulesco, R. Prostate cancer proteomics: Current trends and future perspectives for biomarker discovery. *Oncotarget,* 2017; 8(11): 18497. DOI: 10.18632/oncotarget.14501.

[15] Monti, C., Zilocchi, M., Colugnat, I., Alberio, T. Proteomics turns functional. *Journal of Proteomics*; 2019, 198(1): 36. DOI: https://doi.org/10.1016/j.jprot.2018.12.012).

[16] Njoku, K., Chiasserini, P., Whetton, A.D., Crosbie, E.J. Proteomic biomarkers for the dectetion of endometrial cancer. *Cancers,* 2019; 11(10): 1572. DOI: https://doi.org/10.3390/cancers11101572.

[17] Zhang, S-Q., Pan, S-M., Liang, S-X., Han, Y-S., Chen, H-B., Li, J-C. Research status and prospects of biomarkers for nasopharyngeal carcinoma in the era of high-throughput omics (Review). *International Journal of Oncology,* 2021; 58(4): 1. DOI: https://doi.org/10.3892/ijo.2021.5188.

[18] Berger, M.F., Mardis, E.R. The emerging clinical relevance of genomics in cancer medicine. *Nature Reviews Clinical Oncology*, 2018; 15(6): 353. DOI: 10.1038/s41571-018-0002-6.

[19] Shruthi, B.S., Vinodhkumar, P., Selvamani. Proteomics: A new perspective for cancer. *Advanced Biomedical Research,* 2016; 5: 1. DOI: 10.4103/2277-9175.180636.

[20] Cieślik, M., Chinnaiyan, A.M. Cancer transcriptome profiling at the juncture of clinical translation. *Nature Reviews Genetics,* 2018; 19(2): 93. DOI: 10.1038/nrg.2017.96.

[21] Moor, A.E., Itzkovitz, S. Spatial transcriptomics: Paving the way for tissue-level systems biology. *Current Opinion in Biotechnology*, 2017; 46(1): 126. DOI: http://dx.doi.org/10.1016/j.copbio.2017.02.004.

[22] Caudle, W.M., Bammler, T.K., Lin, Y., Pan, S., Zhang, J. Using "omics" to define pathogenesis and biomarkers of Parkinson's disease. *Expert Review Neurotherapeutics,* 2010; 10 (6): 925. DOI: 10.1586/ern.10.54.

[23] Rinschen, M.M., Ivanisevic, J., Giera, M., Siuzdak, G. Identification of bioactive metabolites using activity metabolomics. *Nature Reviews Molecular Cell Biology*, 2019; 20(6): 353. DOI: 10.1038/s41580-019-0108-4.

[24] Guijas, C., Montenegro-Burke, J.R., Warth, B., Spilker, M.E., Siuzdak, G. Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nature Biotechnology,* 2018, 36(4): 316. DOI: 10.1038/nbt.4101.

[25] Zhou, Y., Song, R., Ma, C., Zhou, L., Liu, X., Yin, P., Zhang, Z., Sun, Y., Xu, C., Lu, X., Xu, G. Discovery and validation of potential urinary biomarkers for bladder cancer diagnosis using a pseudotargeted GC-MS metabolomics method. *Oncotarget*, 2017; 8(13): 20719. DOI: 10.18632/oncotarget.14988.

[26] Park, J., Shin, Y., Kim, T.H., Kim, D-H., Lee, A. Plasma metabolites as possible biomarkers for diagnosis of breast cancer. *PLOS ONE,* 2019; 14(12): e0225129. DOI: https://doi.org/10.1371/journal.pone.0225129.

[27] Xu, J., Li, J., Zhang, R., He, J., Chen, Y., Bi, N., Song, Y., Wang, L., Zhan, Q., Abliz, Z. Development of a metabolic pathway-based pseudo-targeted metabolomics method using liquid chromatography coupled with mass spectrometry. *Talanta,* 2018; 192(1): 160. DOI: 10.1016/j.talanta.2018.09.021.

[28] Li, X., Wenes, M., Romero, P., Huang, S.C., Fendt, S.M., Ho, P.C. Navigating metabolic pathways to enhance antitumour immunity and immunotherapy. *Nature Reviews Clinical Oncology,* 2019; 16(7): 425. DOI: 10.1038/s41571-019-0203-7.

[29] Yang, M., Soga, T., Pollard, P.J. Oncometabolites: Linking altered metabolism with cancer. J*ournal of Clinical Investigation*, 2013; 123(9): 3652. DOI: 10.1172/JCI67228.

[30] Li, S., Fu, L., Tian, T., Deng, L., Li, H., Xia, W., Gong, Q. Disrupting SOD1 activity inhibits cell growth and enhances lipid accumulation in nasopharyngeal carcinoma. *Cell Communication and Signaling,* 2018; 16(1): 1. DOI: 10.1186/s12964-018-0240-3.

[31] Tan, Z., Xiao, L., Tang, M., Bai, F., Li, J., Li, L., Shi, F., Li, N., Li, Y., Du, Q., Lu, J., Weng, X., Yi, W., Zhang, H., Fan, J., Zhou, J., Gao, Q., Onuchic, J.N., Bode, A.N., Luo, X., Cao, Y. Targeting CPT1A-mediated fatty acid oxidation sensitizes nasopharyngeal carcinoma to radiation therapy. *Theranostics*, 2018; 8(9): 2329. DOI: 10.7150/thno.21451.

[32] Xiao, Q., Zhang, F., Xu, L., Yue, L., Kon, O.L., Zhu, Y., Guo, T. High-throughput proteomics and AI for cancer biomarker discovery. *Advanced Drug Delivery Reviews*, 2021; 176(1): 113844. DOI: https://doi.org/10.1016/j.addr.2021.113844.

[33] Huang, R., Chen, Z., He, L., He, N., Xi, Z., Li, Z., Deng, Y., Zeng, X. Mass spectrometry-assisted gel-based proteomics in cancer biomarker discovery: approaches and application. *Theranostics*, 2017; 7(14): 3559. DOI: 10.7150/thno.20797.

[34] Sallam, R.M. Proteomics in cancer biomarkers discovery: challenges and applications. *Disease markers*, 2015; 2015(1): 321370. DOI: https://doi.org/10.1155/2015/321370.

[35] Hudler, P., Kocevar, N., Komel, R. Proteomic approaches in biomarker discovery: new perspectives in cancer diagnostics. *The Scientific World Journal*, 2014; 2014(1): 1. DOI: https://doi.org/10.1155/2014/260348.

[36] Section 7.8 – Systems Biology. In Kyriakopoulou, C., From Fundamental Genomics to Systems Biology: Understanding the Book of Life, Luxembourg: Office for Official Publications of the European Communities, 2008: 58. DOI: 10.2777/49314.

[37] Chapter 8. Manipulating Proteins, DNA and RNA; Analyzing Protein Structure and Function. In Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. Molecular Biology of the Cell, 4th edition, New York: Garland Science, 2002: 1. Available from URL: https://www.ncbi.nlm.nih.gov/books/NBK26820/.

[38] Wilkins, M.R., Sanchez, J.-C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F., Williams, K.L. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology and Genetic Engineering Reviews*, 1996; 13(1): 19. DOI: https://doi.org/10.1080/02648725.1996.10647923.

[39] Bendixen, E. Understanding the Proteome. In Toldrá, F., Nollet, L. (eds), Proteomics in Foods. Food Microbiology and Food Safety, vol. 2, Springer, Boston, MA, 2013: 3. DOI: https://doi.org/10.1007/978-1-4614-5626-1_1.

[40] Gamble, P., Jaroensri, R., Wang, H., Tan, F., Moran, M., Brown, T., Flament-Auvigne, I., Rakha, E.A., Toss, M., Dabbs, D.J., Regitnig, P., Olson, N., Wren, J.H., Robinson, C., Corrado, G.S., Peng, L.H., Liu, Y., Mermel, C.H., Steiner, D.F., Chen, P-H.C. Determining breast cancer biomarker status and associated morphological features using deep learning. *Communications Medicine;* 2021, 1(14): 1. DOI: https://doi.org/10.1038/s43856-021-00013-3.

[41] Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P. Ruddy, K., Tsang, J., Cardoso, F. Breast Cancer. *Nature Reviews Disease Primers*, 2019; 5(1): 1. DOI: 10.1038/s41572-019-0111-2.

[42] Loibl, S., Poortmans, P., Morrow, M., Denkert, C., Curigliano, G. Breast Cancer. *The Lancet*, 2021; 397(10286): 1750. DOI: https://doi.org/10.1016/S0140-6736(20)32381-3.

[43] Nicolò, E., Zagami, P., Curigliano, G. Antibody-drug conjugates in breast cancer: the chemotherapy of the future? *Current Opinion in Oncology*, 2020; 32(5): 494. DOI: 10.1097/CCO.0000000000000656.

[44] Li, H., Fu, Y. Assessing Clinical Significance of Serum CA15-3 and Carcinoembryonic Antigen (CEA) Levels in Breast Cancer Patients: A Meta-Analysis. *Medical Science Monitor*, 2016; 22(1): 3154. DOI: 10.12659/MSM.896563.

[45] Gangadharan, A., Nyirenda, T., Patel, K., Delgadillo, J-. N., Coletta, D., Tanaka, T., Walland, A.C., Jameel, Z., Vedantam, S.,Tang, S., Mannion, C., Lee, G.Y., Goy, A., Pecora, A., Suh, K.S. Prolactin induced protein (PIP) is a potential biomarker for early stage and malignant breast cancer. *The Breast*, 2018; 39(1): 101. DOI: https://doi.org/10.1016/j.breast.2018.03.015.

[46] Cano, I.G-, Constâncio, V., Artigues, A.A-, Lameirinhas, A., Simón, S., Ortega, B., Martínez, M.T., Hernando, C., Bermejo, B., Lluch, A., Lopes, P., Henrique, R., Jerónimo, C., Cejalvo, J.M., Eroles, P. Circulating miR-99ª-5p Expression in Plasma: A Potential Biomarker for Early Diagnosis of Breast Cancer. *International Journal of Molecular Sciences*, 2020; 21(19): 7427. DOI: https://doi.org/10.3390/ijms21197427.

[47] Artigues, A.A., Cano, G., Simón, S., Ortega, B., Moragón, S., Lameirinhas, A., Constâncio, V., Salta, S., Burgués, O., Bermejo, B., Henrique, R., Lluch, A., Jerónimo, C., Eroles, P., Cejalvo, J.M. Circulating miR-30b-5p levels in plasma as a novel potential biomarker for early detection of breast cancer. *European Society*

for Medical Oncology´s Open, 2021; 6(1): 100039. DOI: https://doi.org/10.1016/j.esmoop.2020.100039.

[48] Palma, F.D.E., Salvatore, F., Pol, J.G., Kroemer, G., Maiuri, M.C. Circular RNAs as potential biomarkers in breast cancer. *Biomedicines*, 2022; 10(3): 725. DOI: https://doi.org/10.3390/biomedicines10030725.

[49] Anderson, L. Candidate-based proteomics in the search for biomarkers of cardiovascular disease. *The Journal of Physiology*, 2005; 563(Pt.1): 23. DOI: 10.1113/jphysiol.2004.080473.

[50] Duong, V-A., Park, J-M., Lim, H-J., Lee, H. Proteomics in forensic analysis: applications for human samples. *Applied Sciences*, 2021; 11(1): 3393. DOI: https://doi.org/10.3390/app11083393.

[51] Tainsky, M. A. Genomic and proteomic biomarkers for cancer: a multitude of opportunities. *Biochimica et biophysica acta*, 2009; *1796*(2): 176. DOI: https://doi.org/10.1016/j.bbcan.2009.04.004.

[52] Hsiao, Y-C., Chu, L.J., Chen, J-T., Yeh, T-S., Yu, J-S. Proteomic profiling of the cancer cell secretome: informing clinical research. *Expert Review of Proteomics*, 2017; 14(9): 1. DOI: 10.1080/14789450.2017.1353913.

[53] Wu, J., Gao, Y. Physiological conditions can be reflected in human urine proteome and metabolome. *Expert Review Proteomics;* 2015, 12(6): 623. DOI: 10.1586/14789450.2015.1094380.

[54] Harpole, M., Davis, J., Espina, V. Current state of the art for enhancing urine biomarkers discovery. *Expert Review Proteomics*, 2016; 13(6): 609. DOI: 10.1080/14789450.2016.1190651

[55] Candiano, G., Santucci, L., Petretto, A., Lavarello, C., Inglese, E., Bruschi, M., Ghiggeri, G.M., Boschetti, E., Righetti, P.G. Widening and diversifying the proteome capture by combinatorial peptide ligand libraries via alcian Blue dye binding. *Analytical Chemistry*, 2015; 87(9): 4814. DOI: https://doi.org/10.1021/acs.analchem.5b00218.

[56] Santucci, L., Candiano, G., Petretto, A., Bruschi, M., Lavarello, C., Inglese, E., Righetti, P.G., Ghiggeri, G.M. From hundreds to thousands: widening the normal human urinome (1). *Journal of Proteomics*, 2015; 112(1): 53. DOI: 10.1016/j.jprot.2014.07.021.

[57] Li, M., Xu, Y., Xu, M., Ma, L., Wang, T., Liu, Y., Dai, M., Chen, Y., Lu, J., Liu, J., Bi, Y., Ning, G. Association between nonalcoholic fatty liver disease (NAFLD)

and osteoporotic fracture in middle-aged and elderly Chinese. *The Journal of Clinical Endocrinology and Metabolism*, 2012; 97(6): 2033. DOI: 10.1210/jc.2011-3010.

[58] Jonscher, K.R., Osypuk, A.A., Bokhoven, A.V., Lucia, M.S. Evaluation of urinary protein precipitation protocols for the multidisciplinary approach to the study of chronic pelvic pain research network. *Journal of Biomolecular Techniques,* 2014; 25(4): 118. DOI: 10.7171/jbt.14-2504-004.

[59] Silva, C.L., Perestrelo, R., Sousa-Ferreira, I., Capelinha, F., Câmara, J.S., Petkovic, M. Lipid biosignature of breast cancer tissues by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Breast Cancer Research and Treatment*; 2020, 182(1): 9. DOI: https://doi.org/10.1007/s10549-020-05672-9.

[60] Irwig, L. Evaluating new screening tests for breast cancer. *British Medical Journal*; 2006, 332(7543): 678. DOI: 10.1136/bmj.332.7543.678.

[61] Shruthi, B.S., Vinodhkumar, P., Selvamani. Proteomics: a new perspective for cancer. *Advanced Biomedical Research,* 2016; 5(1): 67. DOI: 10.4103/2277-9175.180636.

[62] Adachi, J., Kumar, C., Zhang, Y., Olsen, J. V., Mann, M. The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome biology,* 2006; 7(1): R80. DOI:10.1186/gb-2006-7-9-R80.

[63] Kentsis, A., Monigatti, F., Dorff, K., Campagne, F., Bachur, R., Steen, H. Urine proteomics for profiling of human disease using high accuracy mass spectrometry. *Proteomics Clinical Applications*, 2009; 3(1): 1052. DOI: 10.1002/prca.200900008.

[64] Marimuthu, A., O´Meally, R.N., Chaerkady, R., Subbannayya, Y., Nanjappa, V., Kumar, P., Kelkar, D.S., Pinto, S.M., Sharma, R., Renuse, S., Goel, R., Christopher, R., Delanghe, B., Cole, R.N., Harsha, H.C., Pandey, A. A comprehensive map of the human urinary proteome. *Journal of proteome research,* 2011; 10(1): 2734. DOI: 10.1021/pr2003038.

[65] Li, Q.R., Fan, K.X., Li, R-X., Dai, J., Wu, C-C., Zhao, S-L., Wu, J-R., Shieh, C-H., Zheng, R. A comprehensive and non-prefractionation on the protein level approach for the human urinary proteome: touching phosphorylation in urine. *Rapid Communications in Mass Spectrometry,* 2010; 24(1): 823. DOI: 10.1002/rcm.4441.

[66] Paul, P., Antonydhason, V., Gopal, J., Haga, S.W., Hasan, N., Oh, J-W. Bioinformatics for Renal and Urinary Proteomics: Call for Aggrandization. *International Journal of Molecular Sciences*, 2020; 21(3): 961. DOI: https://doi.org/10.3390/ijms21030961.

[67] Swensen, A.C., He, J., Fang, A.C., Ye, Y., Nicora, C.D., Shi, T., Liu, A.Y., Sigdel, T.K., Sarwal, M.M., Qian, W-J. A comprehensive urine proteome database generated from patients with various renal conditions and prostate cancer. *Frontiers in Medicine*, 2021; 8(1): 548212. DOI: https://doi.org/10.3389/fmed.2021.548212.

[68] Leng, W., Ni, X., Sun, C., Lu, T., Malovannaya, A., Jung, S.Y., Huang, Y., Qiu, Y., Sun, G., Holt, M.V., Ding, C., Sun, W., Men, X., Shi, T., Zhu, W., Wang, Y., He, F., Zhen, B., Wang, G., Qin, J. Proof-of-concept workflow for establishing reference intervals of human urine proteome for monitoring physiological and pathological changes. *EBioMedicine*, 2017; 18(1): 300. DOI: 10.1016/j.ebiom.2017.03.028.

[69] Santucci, L., Bruschi, M., Candiano, G., Lugani, F., Petretto, A., Bonanni, A., Ghiggeri, G.M. Urine proteome biomarkers in kidney diseases. I. Limits, perspectives, and first focus on normal urine. *Biomarker Insights*, 2016; 11(1): 41. DOI: https://doi.org/10.4137/BMI.S26229.

[70] Xiao, Q., Zhang, F., Xu, L., Yue, L., Kon, O.L., Zhu, Y., Guo, T. High-throughput proteomics and AI for cancer biomarker discovery. *Advanced Drug Delivery Reviews*, 2021; 176(1): 113844. DOI: 10.1016/j.addr.2021.113844.

[71] Ritchie, H., Roser, M. Causes of death. Published online at OurWorldInData.org, 2018. Available from URL: https://ourworldindata.org/causes-of-death.

[72] D'Costa, J.J., Goldsmith, J.C., Wilson, J.S., Bryan, R.T., Ward, D.G. Systematic review of the diagnostic and prognostic value of urinary protein biomarkers in urothelial bladder cancer. *Bladder Cancer*, 2016; 2(3): 301. DOI: 10.3233/BLC-160054.

[73] Masuda, N., Ogawa, O., Park, M., Liu, A.Y., Goodison, S., Dai, Y., Kozai, L., Furuya, H., Lotan, Y., Rosser, C.J., Kobayashi, T. Meta-analysis of a 10-plex urine-based biomarker assay for the detection of bladder cancer. *Oncotarget*, 2018, 9(6): 7101. DOI: 10.18632/oncotarget.23872.

[74] Cheung, C.H.Y., Juan, H.F. Quantitative proteomics in lung cancer. *Journal of Biomedical Science*, 2017; 24(37): 372017. DOI: https://doi.org/10.1186/s12929-017-0343-y.

[75] Rifai, N., Gillette, M., Carr, S. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature Biotechnology*, 2006; 24(8): 971. DOI: https://doi.org/10.1038/nbt1235.

[76] Callesen, A.K., Madsen, J.S., Vach, W., Kruse, T.A., Mogensen, O., Jensen, O.N. Serum protein profiling by solid phase extraction and mass spectrometry: A future diagnostics tool. *Proteomics*, 2009; 9(6): 1428. DOI: 10.1002/pmic.200800382.

[77] Tan, C. S. Statistical Methods for Biomarker Discovery in Proteomics. Karolinska Institutet, 2008. Available from URL: http://hdl.handle.net/10616/37944.

[78] Bellei, E., Bergamini, S., Monari, E., Fantoni, L.I., Cuoghi, A., Ozben, T., Tomasi, A. High-abundance proteins depletion for serum proteomic analysis: concomitant removal of non-targeted proteins. *Amino Acids*, 2011; 40(1): 145. DOI: 10.1007/s00726-010-0628-x.

[79] Hajduk, J., Matysiak, J., Kokot, P., Nowicki, P., Dereziński, P., Kokot, Z.J. The application of fuzzy statistics and linear discriminant analysis as criteria for optimizing the preparation of plasma for matrix-assisted laser desorption/ionization mass spectrometry peptide profiling. *Clinica Chimica Acta*; 2015, 448(1): 174. DOI: 10.1016/j.cca.2015.06.025.

[80] Banach, P., Suchy, W., Derezinski, P., Matysiak, J., Kokot, Z.J., Nowak-Markwitz, E. Mass spectrometry as a tool for biomarkers searching in gynecological oncology. *Biomedicine & Pharmacotherapy*, 2017; 92(1): 836. DOI: 10.1016/j.biopha.2017.05.146.

[81] Lindemann, C., Thomanek, N., Hundt, F., Lerari, T., Meyer, H.E., Wolters, D., Marcus, K. Strategies in relative and absolute quantitative mass spectrometry-based proteomics. *Biological Chemistry*, 2017; 398(5-6): 687. DOI: 10.1515/hsz-2017-0104.

[82] Meyer, J.G., Schilling, B. Clinical applications of quantitative proteomics using targeted and untargeted data-independent acquisition techniques. *Expert Reviews Proteomics*, 2017; 14(5): 419. DOI: 10.1080/14789450.2017.1322904.

[83] Cai, W., Tucholski, T.M., Gregorich, Z.R., Ge, Y. Top-down proteomics: Technology advancements and applications to heart diseases. *Expert Reviews Proteomics*, 2016; 13(8): 717. DOI: 10.1080/14789450.2016.1209414.

[84] Lan, J., Nunez Galindo, A., Doecke, J., Fowler, C., Martins, R.N., Rainey-Smith, S.R., Cominetti, O., Dayon, L. Systematic evaluation of the use of human plasma

and serum for mass-spectrometry-based shotgun proteomics. *Journal of Proteome Research*, 2018; 17(4): 1426. DOI: 10.1021/acs.jproteome.7b00788.

[85] Greco, V., Piras, C., Pieroni, L., Ronci, M., Putignani, L., Roncada, P., Urbani, A. Applications of MALDI-TOF mass spectrometry in clinical proteomics. *Expert Review of Proteomics*, 2018; 15(8): 683. DOI:10.1080/14789450.2018.1505510.

[86] Percy, A.J., Parker, C.E., Borchers, C.H. Pre-analytical and analytical variability in absolute quantitative MRM-based plasma proteomic studies. *Bioanalysis*, 2013; 5(22): 2837. DOI: 10.4155/bio.13.245.

[87] Takehana, S., Yoshida, H., Ozawa, S., Yamazaki, J., Shimbo, K., Nakayama, A., Mizukoshi, T., Miyano, H. The effects of pre-analysis sample handling on human plasma amino acid concentrations. *Clinica Chimica Acta*, 2016; 455(1): 68. DOI: 10.1016/j.cca.2016.01.026.

[88] Mateos, J., Carneiro, I., Corrales, F., Elortza, F., Paradela, A., Pino, M.S.D., Iloro, I., Marcilla, M., Mora, M.I., Valero, L., Ciordia, S., Fernández, V., Fortuño, M.A., García-Sánchez, I., Martínez, R., Muñoz, M.A., Rodriguez, C., Doménech, N. Multicentric study of the effect of pre-analytical variables in the quality of plasma samples stored in biobanks using different complementary proteomic methods. *Journal of proteomics*, 2017; 150(1): 109. DOI: 10.1016/j.jprot.2016.09.003.

[89] Hassis, M.E., Niles, R.K., Braten, M.N., Abertolle, M.E., Ewa, W.H., Hubel, C.A., Fisher, S.J., Williams, K.E. Evaluating the effects of preanalytical variables on the stability of the human plasma proteome. *Analytical biochemistry*, 2015; 478(1): 14. DOI: 10.1016/j.ab.2015.03.003.

[90] Greco, V., Piras, C., Pieroni, L., Urbani, A. Direct assessment of plasma/serum sample quality for proteomics biomarker investigation. *Methods in Molecular Biology*, 2017; 1619(1): 3. DOI: 10.1007/978-1-4939-7057-5_1.

[91] Brunzel, N.A. Physical examination of urine. In Fundamentals of urine and body fluid analysis, 4th ed., St. Louis, Missouri: Elsevier, 2018.

[92] Lin, L., Yu, Q., Zheng, J., Cai, Z., Tian, R. Fast quantitative urinary proteomic profiling workflow for biomarker discovery in kidney cancer. *Clinical Proteomics*, 2018; 15(1): 42. DOI: https://doi.org/10.1186/s12014-018-9220-2.

[93] Tkáčiková, S., Talian, I., Sabo, J. Optimisation of urine sample preparation for shotgun proteomics. *Open Chemistry*; 2020, 18(1): 850. DOI: https://doi.org/10.1515/chem-2020-0150.

[94] Tang, X., Xiao, X., Sun, H., Zheng, S., Xiao, X., Guo, Z., Liu, X., Sun, W. 96DRA-Urine: A high throughput sample preparation method for urinary proteome analysis. *Journal of Proteomics,* 2022; 257(1): 104529. DOI: https://doi.org/10.1016/j.jprot.2022.104529.

[95] Shruthi, B.S., Palani Vinodhkumar, S. Proteomics: A new perspective for cancer. *Advanced Biomedical Research,* 2016, 5(1): 67. DOI: 10.4103/2277-9175.180636.

[96] Miękus, N., Olędzka, I., Plenis, A., Woźniak, Z., Lewczuk, A., Koszałka, P., Seroczyńska, B., Bączek, T. Gel electrophoretic separation of proteins from cultured neuroendocrine tumor cell lines. *Molecular Medicine Reports,* 2015; 11(2): 1407. DOI: https://doi.org/10.3892/mmr.2014.2864.

[97] Penque, D. Two-dimensional gel electrophoresis and mass spectrometry for biomarker discovery. *Proteomics Clinical Applications,* 2009; 3(2): 155. DOI: 10.1002/prca.200800025.

[98] Karas, M., Bachmann, D., Hillenkamp, F. Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Analytical Chemistry*, 1985; 57(14): 2935. DOI: https://doi.org/10.1021/ac00291a042.

[99] Karas, M., Hillenkamp, F. Laser desorption ionisation of proteins with molecular masses exceeding 10.000 daltons. *Analytical Chemistry*, 1988; 60(20): 2299. DOI: https://doi.org/10.1021/ac00171a028.

[100] Hutchens, T. W., Yip, T-T. (1993). New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Communications in Mass Spectrometry*, 1993; 7(7): 576. DOI: https://doi.org/10.1002/rcm.1290070703.

[101] Chapter 1 – Principles and mechanism of MALDI-TOF-MS analysis. In Hosseini, S., Martinez, S.O. -C. (eds), Fundamentals of MALDI-TOF-MS analysis: applications in bio-diagnosis, tissue engineering and drug delivery, SpringerBriefs: Forensic and Medical Bioinformatics, 2017: 1. DOI: https://doi.org/10.1007/978-981-10-2356-9.

[102] Hanash, S., Taguchi, A. The grand challenge to decipher the cancer proteome. *Nature Reviews Cancer*, 2010; 10(9): 652. DOI: https://doi.org/10.1038/nrc2918.

[103] Ding, D., Chen, M., Xiao, X., Cao, P., Li, S. Novel serum peptide model revealed by MALDI-TOF-MS and its diagnostic value in early bladder cancer. *The*

international Journal of Biological Markers, 2020; 35(3): 59. DOI:10.1177/1724600820935473.

[104] Kiprijanovska, S., Stavridis, S., Stankov, O., Komina, S., Petrusevska, G., Polenakovic, M., Davalieva, K. Mapping and identification of the urine proteome of prostate cancer patients by 2D PAGE/MS. *International Journal of* Proteomics, 2014; 2014(7): 1. DOI: 10.1155/2014/594761.

[105] Wang, W., Wang, S., Zhang, M. Identification of urine biomarkers associated with lung adenocarcinoma. *Oncotarget*, 2017; 8(24):38517. DOI: 10.18632/oncotarget.15870.

[106] Tantipaiboonwong, P., Sinchaikul, S., Sriyam, S., Phutrakul, S., Chen, S-T. Different techniques for urinary protein analysis of normal and lung cancer patients. *Proteomics*, 2005; 5(4): 1140. DOI: 10.1002/pmic.200401143.

[107] Jung, Y.H., Han, D., Shin, S.H., Kim, E-K., Kim, H-S. Proteomic identification of early urinary-biomarkers of acute kidney injury in preterm infants. *Scientific Reports*, 2020; 10(1): 4057. DOI: https://doi.org/10.1038/s41598-020-60890-x.

[108] Theodorescu, D., Wittke, S., Ross, M.M., Walden, M., Conaway, M., Just, I., Mischak, H., Fierson, H.F. Discovery and validation of new protein biomarkers for urothelial cancer: a prospective analysis. *The Lancet Oncology*, 2006; 7(3): 230. DOI: 10.1016/S1470- 2045(06)70584-8.

[109] Kreunin, P., Zhao, J., Rosser, C., Urquidi, V., Lubman, D.M., Goodison, S. Bladder cancer associated glycoprotein signatures revealed by urinary proteomic profiling. *Journal of Proteome Research*, 2007; 6(7): 2631. DOI: https://doi.org/10.1021/pr0700807.

[110] Zografos, E., Anagnostopoulos, A.K., Papadopoulou, A., Legaki, E., Zagouri, F., Marinos, E., Tsangaris, G.T., Gazouli, M. Serum proteomic signatures of male breast cancer. *Cancer Genomics & Proteomics*, 2019; 16 (2): 129. DOI: https://doi.org/10.21873/cgp.20118.

[111] Ding, D., Chen, M., Xiao, X., Cao, P., Li, S. Novel serum peptide model revealed by MALDI-TOF-MS and its diagnostic value in early bladder cancer. *The international Journal of Biological Markers*, 2020; 35(3): 59. DOI:10.1177/1724600820935473.

[112] Yang, X-K., Wang, N., Yang, C., Wang, Y-M., Che, T-J. Differential protein expression in patients with urosepsis. *Chinese Journal of Traumatology*, 2018; 21(6): 316. DOI: 10.1016/j.cjtee.2018.07.003.

[113] Hepburn, S., Cairns, D.A., Jackson, D., Craven, R.A., Riley, B., Hutchinson, M., Wood, S, Smith, M.W., Thompson, D., Banks, R.E. An analysis of the impact of pre-analytical factors on the urine proteome: sample processing time, temperature, and proteolysis. *Proteomics Clinical Applications*, 2015; 9(5-6): 507. DOI: 10.1002/prca.201400079.

[114] Saito, S., Hirao, Y., Quadery, A.F., Xu, B., Elguoshy, A., Fujinaka, H., Koma, S., Yamamoto, K., Yamamoto, T. The optimized workflow for sample preparation in LC-MS/MS-based urine proteomics. *Methods and Protocols*, 2019; 2, 46(1): 1. DOI: 10.3390/mps2020046.

[115] Chertov, O., Biragyn, A., Kwak, L.W., Simpson, J.T., Boronina, T., Hoang, V.M., Prieto, D.A., Conrads, T.P., Veenstra, T.D., Fisher, R.J. Organic solvent extraction of proteins and peptides from serum as an effective sample preparation for detection and identification of biomarkers by mass spectrometry. *Proteomics*, 2004; 4(4): 1195. DOI: 10.1002/pmic.200300677.

[116] Jonscher, K.R., Osypuk, A.A., Bokhoven, A.V., Lucia, M.S. Evaluation of urinary protein precipitation protocols for the multidisciplinary approach to the study of chronic pelvic pain research network. *Journal of Biomolecular Techniques*, 2014; 25(4): 118. DOI: 10.7171/jbt.14-2504-004.

[117] Arunambiga, S., Dave, G. S., Joshi, J., Goswami, N., Raval, I., Jogani, P. A. Comparative study of urinary proteins precipitation methods. *Bioscience Biotechonolgy Research Communications*, 2020; 13(3): 1. DOI: http://dx.doi.org/10.21786/bbrc/13.3/17.

[118] GelAnalyzer 19.1 (www.gelanalyzer.com) by Istvan Lazar Jr., PhD and Istvan Lazar Sr., PhD, CSc.

[119] Hajduk, J., Matysiak, J., Kokot, P., Nowicki, P., Derezinski, P., Kokot, Z.J. The application of fuzzy statistics and linear discriminant analysis as criteria for optimizing the preparation of plasma for matrix-assisted laser desorption / ionization mass spectrometry peptide profiling. *Clinica Chimica Acta*, 2015; 448(1): 174. DOI: 10.1016/j.cca.2015.06.025.

[120] López-Fernández, H., Santos, H.M., Capelo, J.L., Fdez-Riverola, F., Glez-Peña, D., Reboiro-Jato, M. Mass-Up: an all-in-one open software application for MALDI-TOF mass spectrometry knowledge discovery. *BioMed Central Bioinformatics*, 2015; 16(1): 1. DOI: 10.1186/s12859-015-0752-4.

[121] Ferrari, E., Wittig, A., Basilico, F., Rossi, R., De Palma, A., Di Silvestre, D., Sauerwein, W.A.G., Mauri, P.L. Urinary proteomics profiles are useful for detection of cancer biomarkers and changes induced by therapeutic procedures. *Molecules*, 2019; 24(4): 794. DOI: https://doi.org/10.3390/molecules24040794.

[122] Lamb, E.J., MacKenzie, F., Stevens, P.E. How should proteinuria be detected and measured? *Annals of Clinical Biochemistry: International Journal of Laboratory Medicine*, 2009; 46(3): 205. DOI: https://doi.org/10.1258/acb.2009.009007.

[123] French, D. Chapter Five – Advances in Clinical Mass Spectrometry. *Advances in Clinical Chemistry*, 2017; 79(1): 153. DOI: https://doi.org/10.1016/bs.acc.2016.09.003.

[124] Waterborg, J.H. The lowry method for protein quantitation. In Walker, J.M. (eds), The Protein Protocols Handbook, Springer Protocols Handbooks, 2009, Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-59745-198-7_2.

[125] Guo, Z., Zhang, Y., Zou, L., Wang, D., Shao, C., Wang, Y., Sun, W., Zhang, L. A proteomic analysis of individual and gender variations in Normal human urine and cerebrospinal fluid using iTRAQ quantification. *PLoS One*, 2015; 10(7): e0133270. DOI: https://doi.org/10.1371/journal.pone.0133270.

[126] Wolfinger, R.D., Beedanagari, S., Boitier, E., Chen, T., Couttet, P., Ellinger-Ziegelbauer, H., Guillemain, G., Mariet, C., Mouritzen, P., O´Lone, R., Pine, P.S., Sharapova, T., Yan, J., Yuen, P.S., Thompson, K.L. Two approaches for estimating the lower limit of quantitation (LLOQ) of microRNA levels assayed as exploratory biomarkers by RT-qPCR. *BMC Biotechnology*, 2018; 18(1): 6. DOI: 10.1186/s12896-018-0415-4.

[127] Weber, K., Osborn, M. The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *Journal of Biological Chemistry*, 1969; 244(1): 4406. DOI: https://doi.org/10.1016/S0021-9258(18)94333-4.

[128] Robati, R.Y., Ahmadi, H., Riahi, M.M., Lari, P., Aledavood, S.A., Rashedinia, M., Abnous, K., Ramezani, M. Comparative proteome analysis of human esophageal cancer and adjacent normal tissues. *Iranian Journal of Basic Medical Sciences*, 2017; 20(3): 265. DOI: 10.22038/ijbms.2017.8354.

[129] Fayazfar, S., Zali, H., Oskouie, A.A., Aghdaei, H. A., Tavirani, M.R., Mojarad, E.N. Early diagnosis of colorectal cancer via plasma proteomic analysis of CRC

(colorectal cancer) and advanced adenomatous polyp. *Gastroenterology Hepatology from Bed to Bench*., 2019; 12(4): 328. DOI: https://doi.org/10.22037/ghfbb.v12i4.1723.

[130] Noreen, S., Akhtar, S., Batool, T., Gardner, Q.A., Akhtar, M.W. Tubulin beta 2C chain (TBB2C), a potential marker of ovarian cancer, an insight from ovarian cancer proteome profile. *American Chemical Society Omega*, 2021; 6(16): 10506. DOI: https://doi.org/10.1021/acsomega.0c03262.

[131] Riley, N.M., Westphall, M.S., Coon, J.J. Sequencing larger intact proteins (30-70 kDa) with activated ion electron transfer dissociation. *Journal of the American Society for Mass Spectrometry*, 2018; 29(1): 140. DOI: https://doi.org/10.1007/s13361-017-1808-7.

[132] Arrowsmith, C.H., Lee, W., Revington, M., Yamazaki, T., Kay, L.E. Toward the solution structure of large (>30 kDa) proteins and macromolecular complexes. *Techniques in Protein Chemistry*, 1995; 6(1): 503. DOI: https://doi.org/10.1016/S1080-8914(06)80061-X.

[133] Stensballe, A., Jensen, O.N. Simplified sample preparation method for protein identification by matrix-assisted laser desorption/ionization mass spectrometry: in-gel digestion on the probe surface. *European Journal of Mass Spectrometry*, 2001; 7(1): 111. DOI: 10.1255/ejms.396.

[134] Shevchenko, A., Loboda, A., Ens, W., Schraven, B., Standing, K.G., Shevchenko, A. Archived polyacrylamide gels as a resource for proteome characterization by mass spectrometry. *Electrophoresis*, 2001; 22(6): 1194. DOI: 10.1002/1522-2683()22:6<1194::AID-ELPS1194>3.0.CO;2-A.

[135] Salum, M.L., Giudicessi, S.L., León, T.S.D., Camperi, S.A., Balsells, R.E. Application of Z-sinapinic matrix in peptide MALDI-MS analysis. *Journal of Mass Spectrometry*, 2017; 52(3): 182. DOI: https://doi.org/10.1002/jms.3908.

[136] Choi, H., Lee, D., Kim, Y., Nguyen, H-Q., Han, S., Kim, J. Effects of matrices and additives on multiple charge formation of proteins in MALDI–MS analysis. *Journal of The American Society for Mass Spectrometry*, 2019; 30(1): 1174. DOI: https://doi.org/10.1007/s13361-019-02213-7.

[137] Hajduk, J, Matysiak, J, Kokot, Z.J. Challenges in biomarker discovery with MALDI-TOF MS. *Clinica Chimica Acta*, 2016; 458(1): 84. DOI: 10.1016/j.cca.2016.04.033.

[138] Eidhammer, I., Flikka, K., Martens, L., Mikalsen, S.-O. (eds), Computational Methods for Mass Spectrometry Proteomics, Wiley, Chichester, England, 2007: 81.

[139] López-Fernández, H., Santos, H.M., Capelo, J.L., Fdez-Riverola, F., Glez-Peña, D., Reboiro-Jato, M. Mass-Up: an all-in-one open software application for MALDI-TOF mass spectrometry knowledge discovery. *BioMed Central Bioinformatics*, 2015; 16 (1): 1. DOI: 10.1186/s12859-015-0752-4.

[140] Meuleman, W., Engwegen, J.Y., Gast, M.-C.W., Beijnen, J.H., Reinders, M.J., Wessels, L.F. Comparison of normalisation methods for surface-enhanced laser desorption and ionisation (SELDI) time-of-flight (TOF) mass spectrometry data. *BioMed Central Bioinformatics*, 2008; 9(1): 1. DOI: 10.1186/1471-2105-9-88.

[141] Liu, Y., Wei, F., Wang, F., Li, C., Meng, G., Duan, H., Ma, Q., Zhang, W. Serum peptidome profiling analysis for the identification of potential biomarkers in cervical intraepithelial neoplasia patients. *Biochemical and Biophysical Research Communications*, 2015; 465(3): 476. DOI: 10.1016/j.bbrc.2015.08.042.

[142] Du, P., Stolovitzky, G., Horvatovich, P., Bischoff, R., Lim, J., and Suits, F. A noise model for mass spectrometry based proteomics. *Bioinformatics*, 2008; 24(8): 1070. DOI: 10.1093/bioinformatics/btn078.

[143] Gibb, S., Strimmer, K. Chapter 11 – Mass spectrometry analysis using MALDIquant. In Datta, S., Mertens, B. (eds), Statistical analysis of proteomics, metabolomics, and lipidomics data using mass spectrometry. Frontiers in Probability and the Statistical Sciences. Springer, New York, 2016: 1.

[144] Pecks, U., Kirschner, I., Wolter, M., Schlembach, D., Koy, C., Rath, W., Glocker, M.O. Mass spectrometric profiling of cord blood serum proteomes to distinguish infants with intrauterine growth restriction from those who are small for gestational age and from control individuals. *Translational Research,* 2014; 164(1): 57. DOI: 10.1016/j.trsl.2013.12.003.

[145] Labas, V., Spina, L., Belleannee, C., Teixeira-Gomes, A-P., Gargaros, A., Dacheux, F., Dacheux, J-L. Analysis of epididymal sperm maturation by MALDI profiling and top-down mass spectrometry. *Journal of Proteome*, 2015; 113(1): 226. DOI: 10.1016/j.jprot.2014.09.031.

[146] Povey, J.F., O'Malley, C.J., Root, T., Martin, E.B., Montague, G.A., Trim, C., Feary, M., Lang, D.A., Alldread, R., Racher, A.J., Smales, C.M. Rapid high-throughput characterisation, classification and selection of recombinant mammalian cell line phenotypes using intact cell MALDI-TOF mass spectrometry

fingerprinting and PLS-DA modelling. *Journal of Biotechnology*, 2014; 184(1): 84. DOI: 10.1016/j.jbiotec.2014.04.028.

[147] Christner, M. Trusch, M., Rohde, H., Kwiatkowski, M., Schlüter, H., Wolters, M., Aepfelbacher, M., Hentschke, M. Rapid MALDI-TOF mass spectrometry strain typing during a large outbreak of Shiga-toxigenic *Escherichia coli*. *PLoS One*, 2014; 9(7): e101924. DOI: https://doi.org/10.1371/journal.pone.0101924.

[148] Bromba, M.U.A., Ziegler, H. Application hints for savitzky-golay digital smoothing filters. *Analytical Chemistry*, 1981; 53(11): 1583. DOI: https://doi.org/10.1021/ac00234a011.

[149] Pang, Z., Chong, J., Zhou, G., Morais D., Chang, L., Barrette, M., Gauthier, C., Jacques, PE., Li, S., Xia, J. (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research*, 2021; 49(1): W388. DOI: 10.1093/nar/gkab382.

[150] Baker, S.G. The central role of receiver operating characteristic (ROC) curves in evaluating tests for the early detection of cancer. *Journal of the National Cancer Institute*, 2003; 95(7): 511. DOI: 10.1093/jnci/95.7.511.

[151] Hajian-Tilaki, K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Metabolomics*, 2013; 4(2): 627. DOI: https://doi.org/10.1161/CIRCULATIONAHA.105.594929.

[152] Xia, J., Broadhurst, D.I., Wilson, M., Wishart, D.S. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics*, 2013; 9(2): 280. DOI: 10.1007/s11306-012-0482-9.

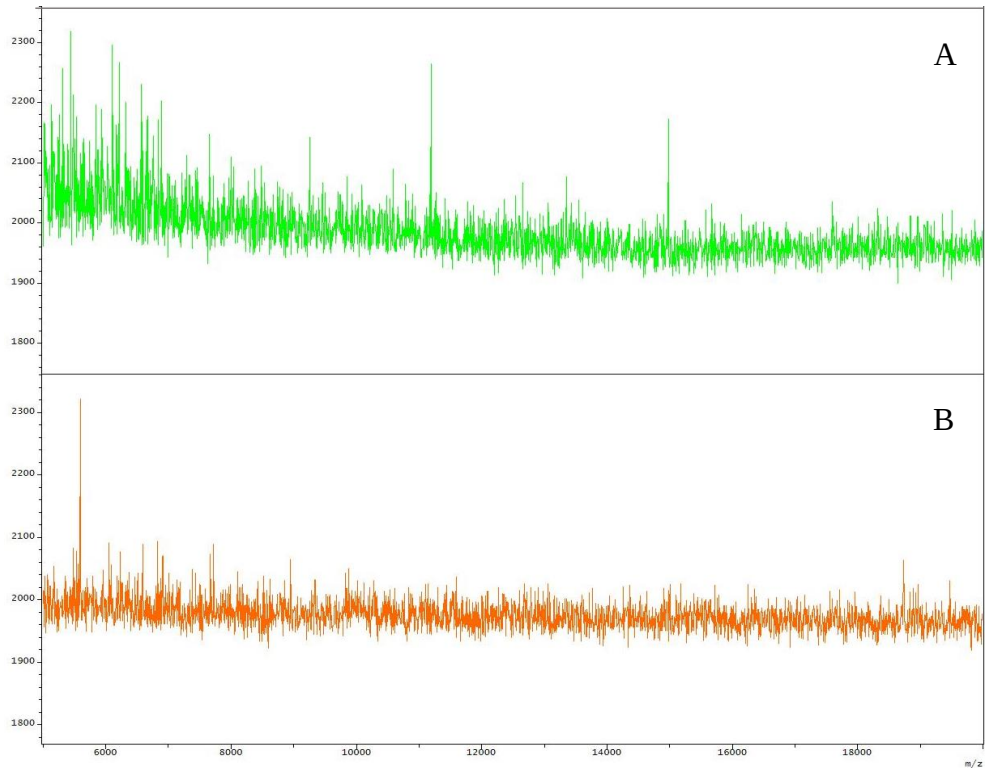


Figure A1 – MALDI-TOF MS spectra in linear positive-ion mode across the m/z range of 5 to 20 kDa of samples from (A) HC and (B) BC patient using ACN/TFA precipitation.

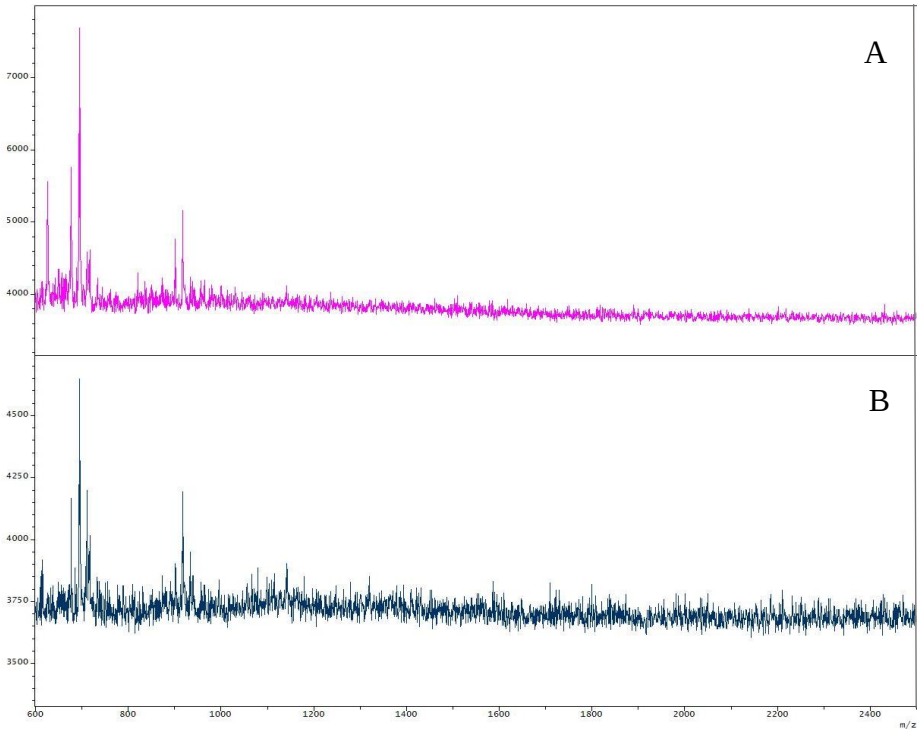


Figure A2 – MALDI-TOF MS spectra in linear positive-ion mode across the m/z range of 600 to 2500 Da of samples from (A) HC and (B) BC patient using ACN/TFA precipitation.

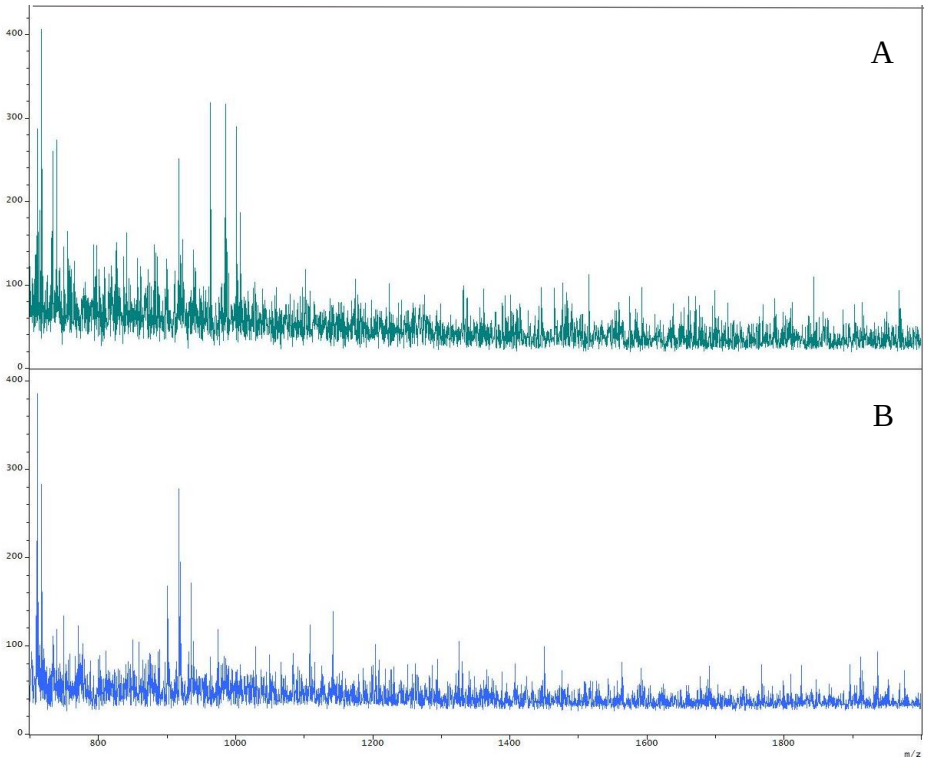


Figure A3 – MALDI-TOF MS spectra in reflector positive-ion mode across the m/z range of 600 to 2500 Da of samples from (A) HC and (B) BC patient using ACN/TFA precipitation.

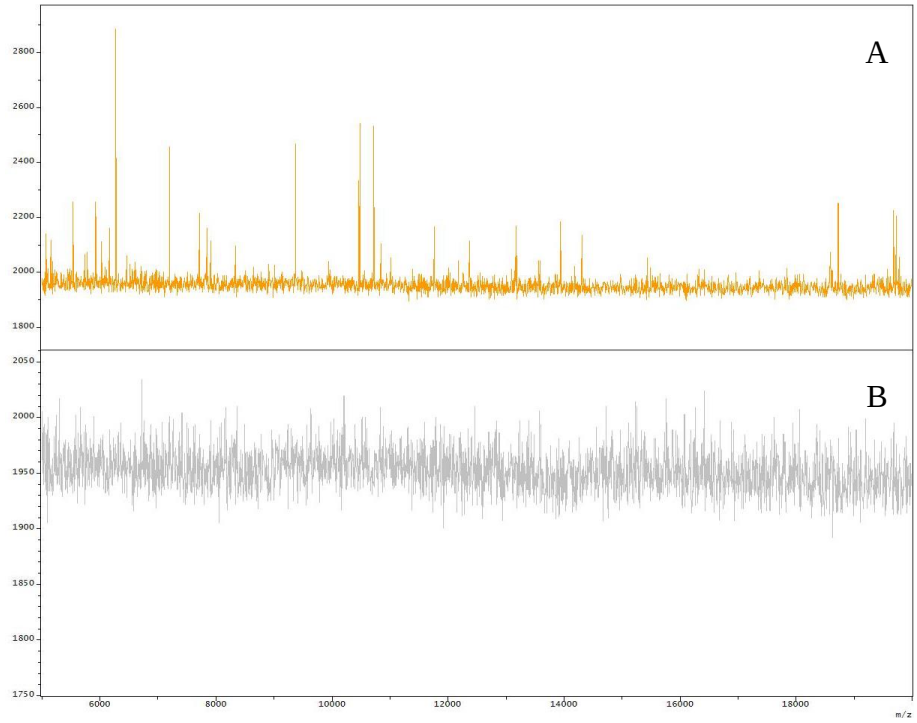


Figure A4 – MALDI-TOF MS spectra in linear positive-ion mode across the m/z range of 5 to 20 kDa of samples from (A) HC and (B) BC patient using TCA/acetone precipitation.

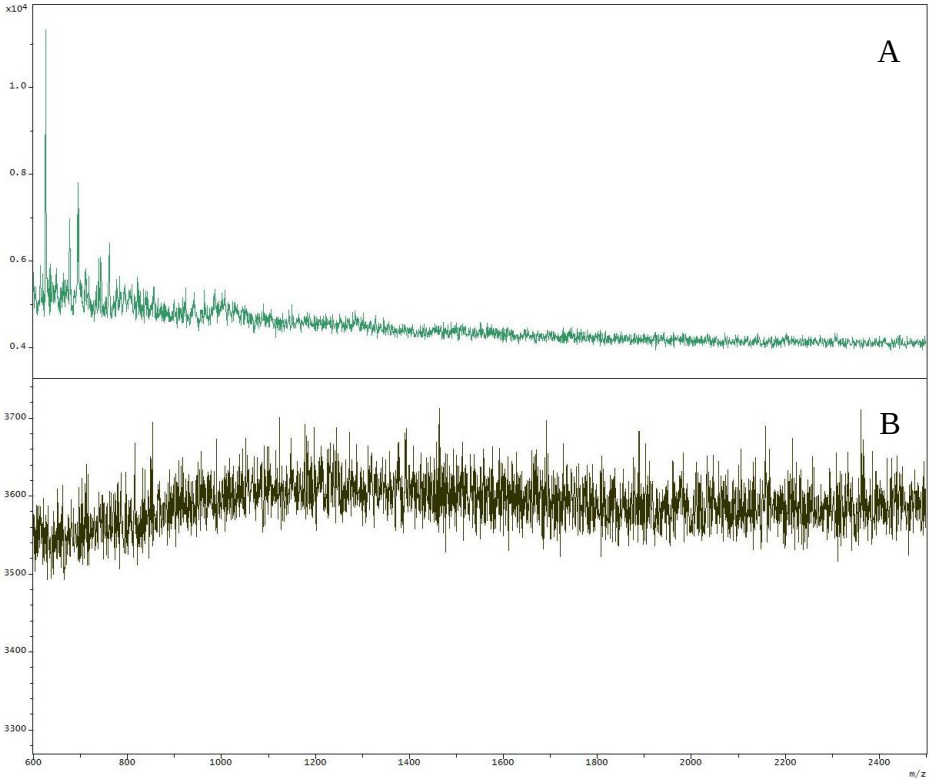


Figure A5 – MALDI-TOF MS spectra in linear positive-ion mode across the m/z range of 600 to 2500 Da of samples from (A) HC and (B) BC patient using TCA/acetone precipitation.

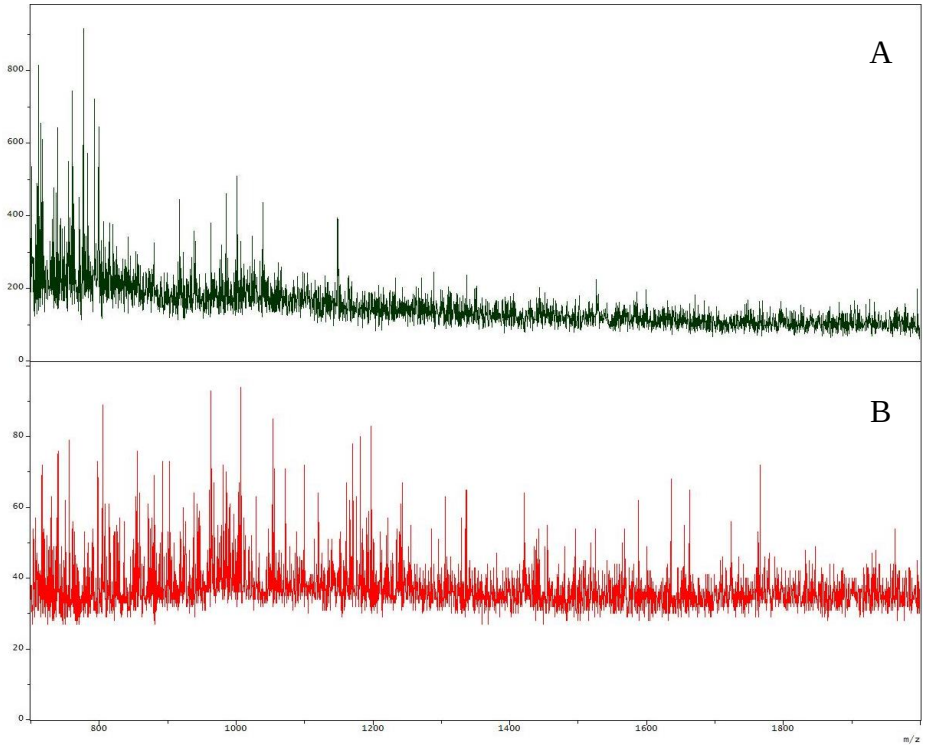


Figure A6 – MALDI-TOF MS spectra in reflector positive-ion mode across the m/z range of 600 to 2500 Da of samples from (A) HC and (B) BC patient using TCA/acetone precipitation.

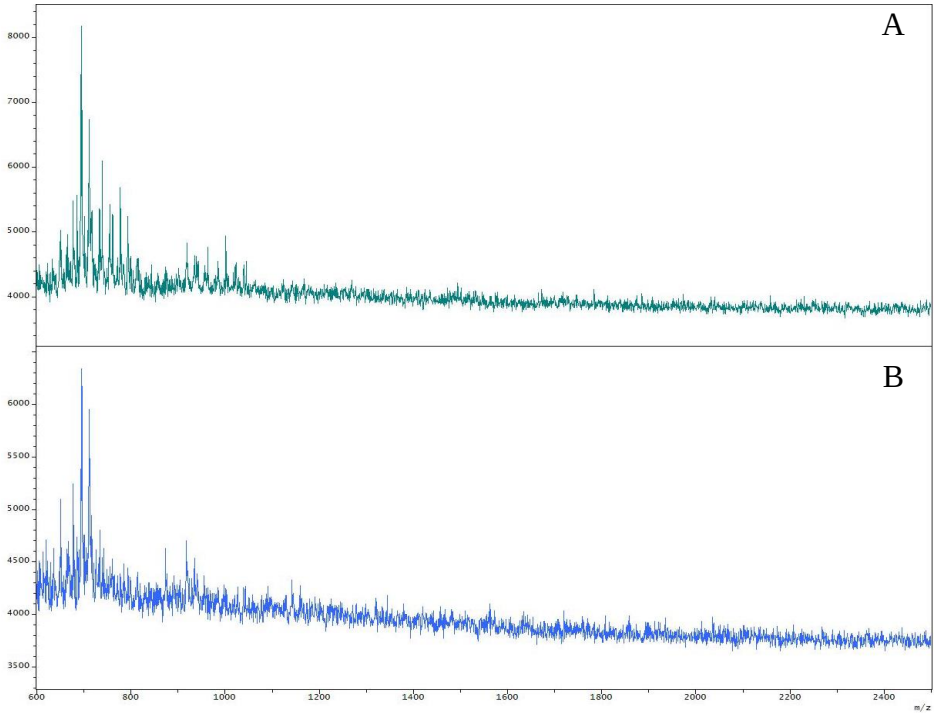
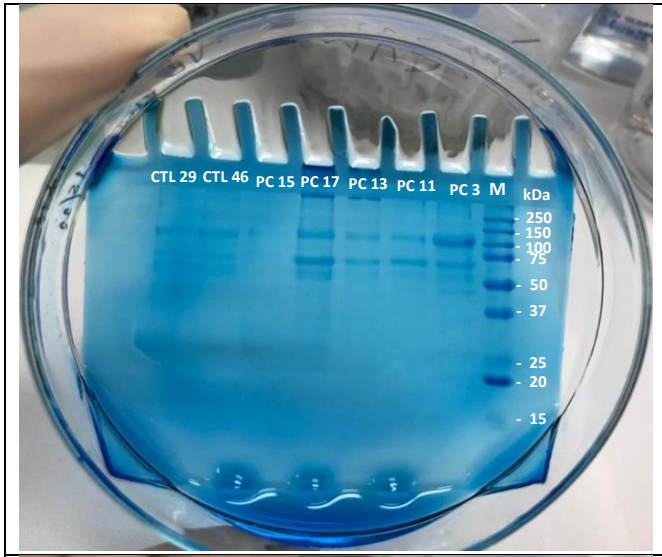
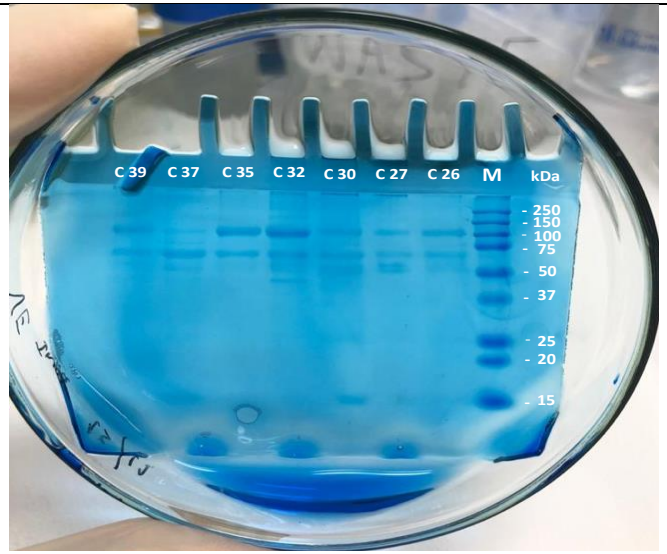
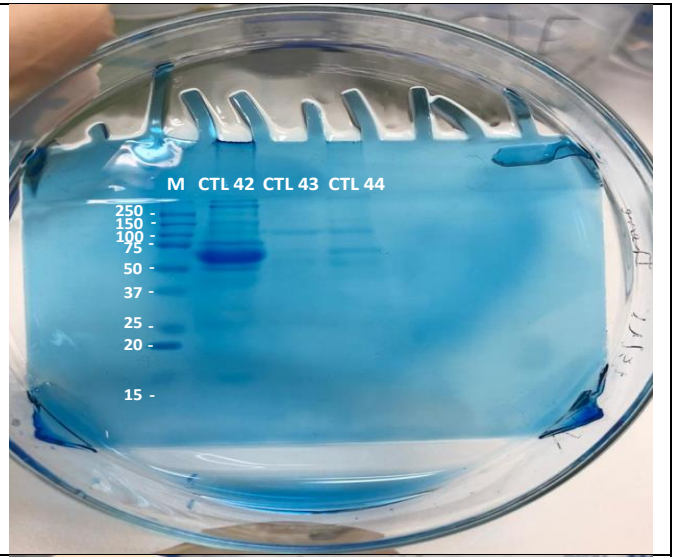
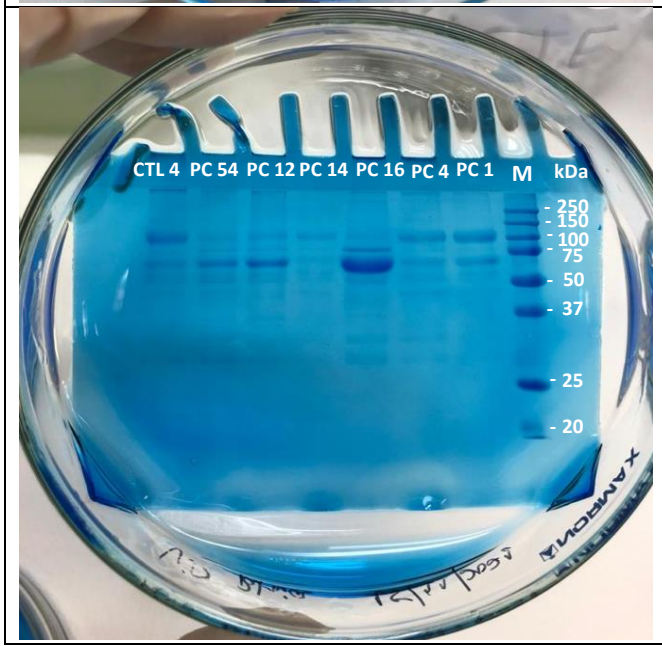


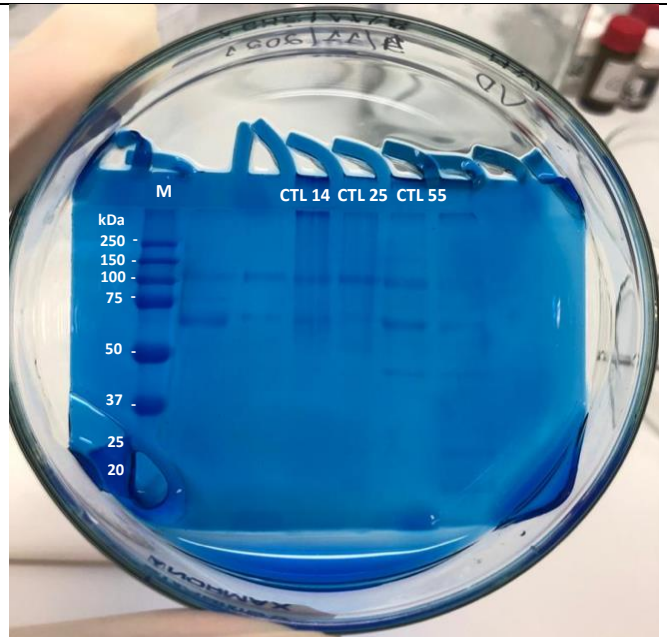
Figure A7 – MALDI-TOF MS spectra in linear positive-ion mode across the m/z range of 600 to 2500 Da of samples from (A) HC and (B) BC patient using chloroform/methanol precipitation.

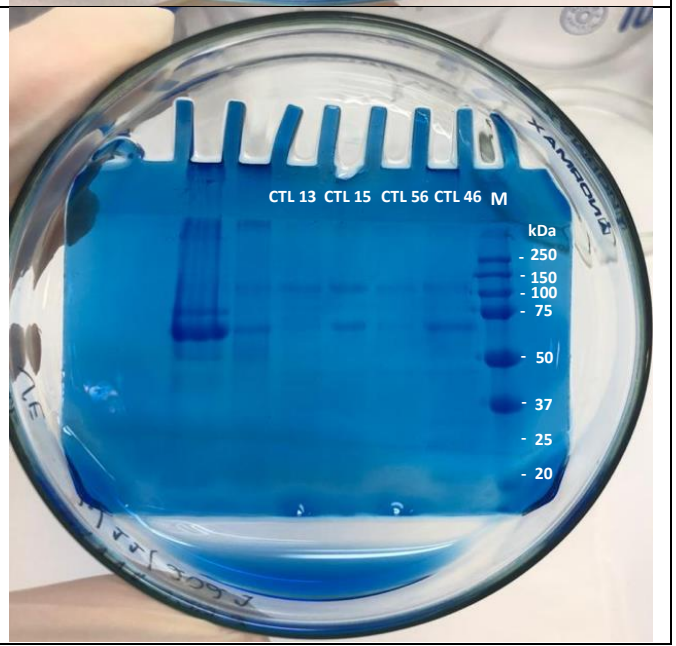

CTL 29 CTL 46 PC 15 PC 17 PC 13 PC 11 PC 3 M kDa
250
150
100
75
50
37
25
20
15

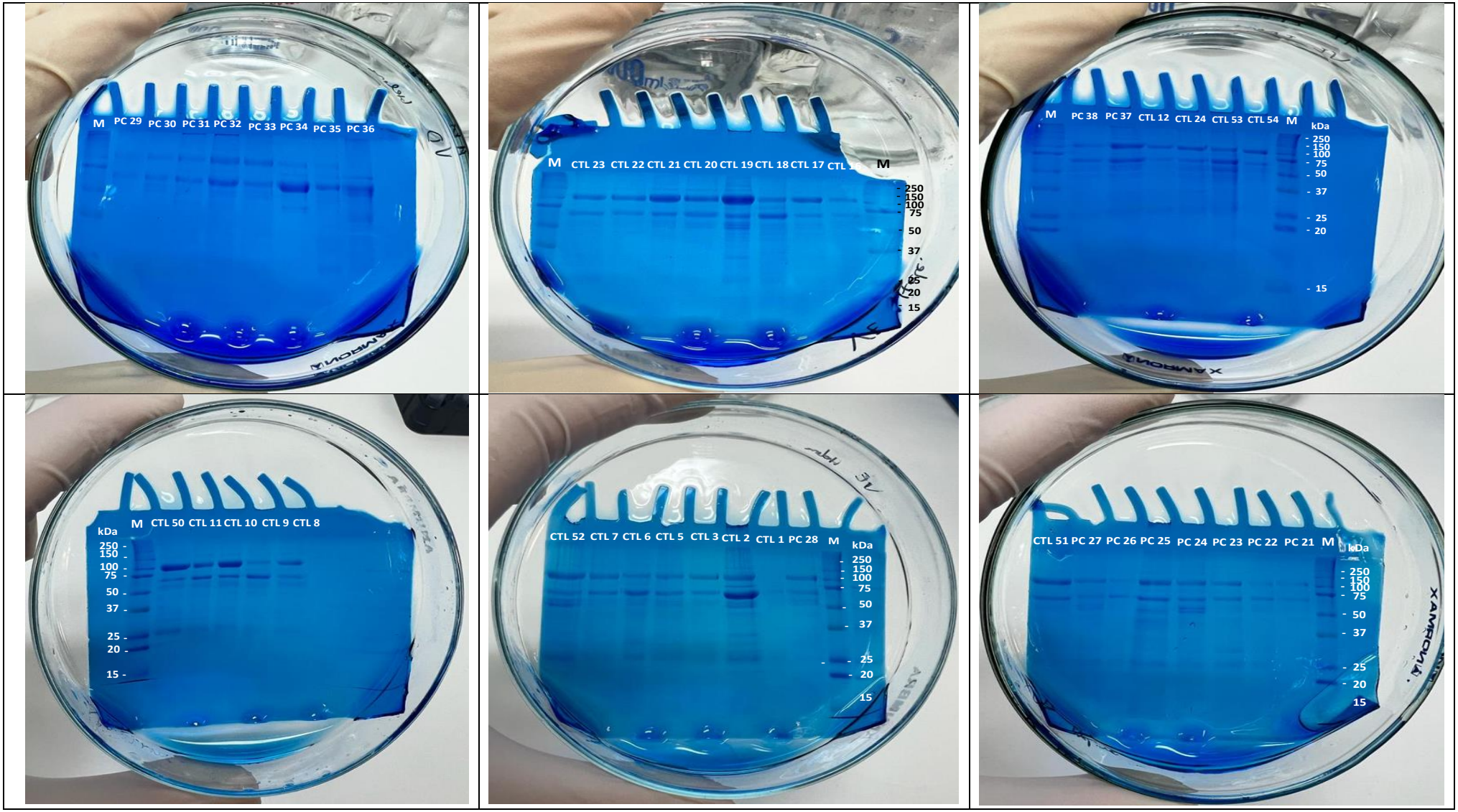

C 39 C 37 C 35 C 32 C 30 C 27 C 26 M kDa
250
150
100
75
50
37
25
20
15


M CTL 42 CTL 43 CTL 44
250
150
100
75
50
37
25
20
15


CTL 4 PC 54 PC 12 PC 14 PC 16 PC 4 PC 1 M kDa
250
150
100
75
50
37
25
20


M
CTL 14 CTL 25 CTL 55
kDa
250
150
100
75
50
37
25
20


CTL 13 CTL 15 CTL 56 CTL 46 M
kDa
250
150
100
75
50
37
25
20


M PC 29 PC 30 PC 31 PC 32 PC 33 PC 34 PC 35 PC 36
M CTL 23 CTL 22 CTL 21 CTL 20 CTL 19 CTL 18 CTL 17 M
250
150
100
75
50
37
25
20
15
M PC 38 PC 37 CTL 12 CTL 24 CTL 53 CTL 54 M
kDa
250
150
100
75
50
37
25
20
15
M CTL 50 CTL 11 CTL 10 CTL 9 CTL 8
kDa
250
150
100
75
50
37
25
20
15
CTL 52 CTL 7 CTL 6 CTL 5 CTL 3 CTL 2 CTL 1 PC 28 M
kDa
250
150
100
75
50
37
25
20
15
CTL 51 PC 27 PC 26 PC 25 PC 24 PC 23 PC 22 PC 21 M
kDa
250
150
100
75
50
37
25
20
15

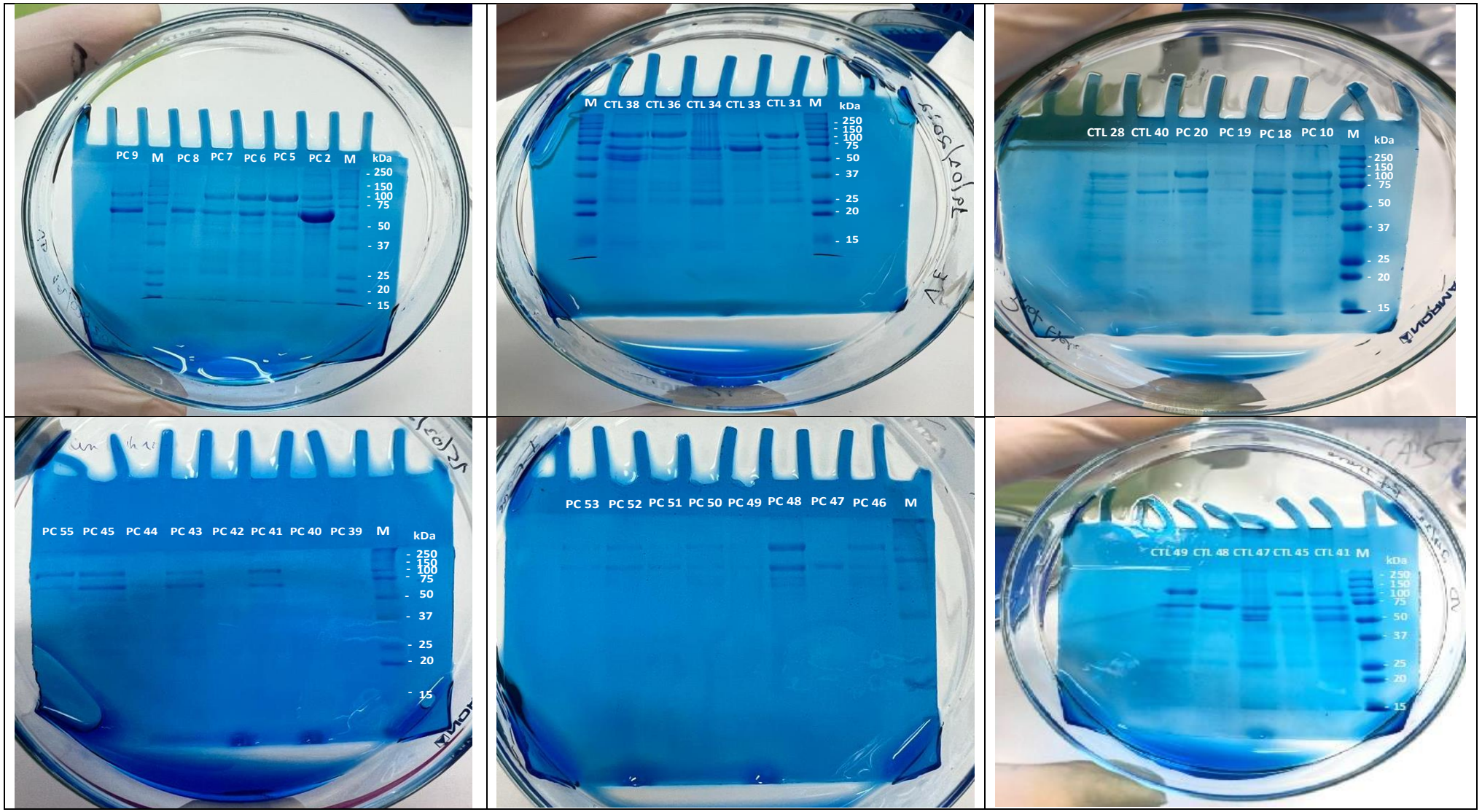


Figure A8 – 1D SDS-PAGE of urines from BC patients and HCs. M – marker (protein ladder); CTL – healthy control group; PC – BC patients.

Table A1 – The retention factor value and molecular weights of each band of all BC patients and HCs using Gel-Analyzer 19.1.

Lane	Band	Rf	MW
Ladder	1	0.103	250
	2	0.149	150
	3	0.208	100
	4	0.262	75
	5	0.387	50
	6	0.513	37
	7	0.721	25
	8	0.816	20
	9	0.926	15
CTL 26	1	0.197	112
	2	0.290	59
CTL 27	1	0.197	112
	2	0.290	59
	3	0.339	46
	4	0.369	40
CTL 30	1	0.200	109
	2	0.303	55
	3	0.960	25
CTL 32	1	0.131	194
	2	0.210	102
	3	0.298	57
	4	0.407	36
CTL35	1	0.200	109
	2	0.296	57
CTL 37	1	0.200	109
	2	0.250	76
	3	0.303	55
	4	0.369	40
CTL 39	1	0.196	113
	2	0.298	57
PC3	1	0.229	107
	2	0.319	59
	3	0.436	35
PC 11	1	0.218	117
	2	0.320	58
PC 13	1	0.222	113
	2	0.307	63
PC 17	1	0.220	115

	2	0.331	55
	3	0.377	44
	4	0.408	39
PC 15	1	0.221	114
CTL 46	1	0.234	103
	2	0.331	55
	3	0.377	44
	4	0.408	39
CTL 29	1	0.227	109
CTL 42	1	0.100	252
	2	0.148	164
	3	0.191	116
	4	0.248	76
	5	0.324	50
	6	0.382	40
CTL 43	1	0.200	108
	2	0.288	60
	3	0.342	46
CTL 44	1	0.185	121
	2	0.285	61
	3	0.352	45
PC 1	1	0.202	101
	2	0.293	56
	3	0.343	44
PC 4	1	0.132	179
	2	0.202	101
	3	0.249	73
	4	0.293	56
	5	0.355	42
	6	0.405	35
	7	0.589	27
	8	0.666	26
PC 16	1	0.129	184
	2	0.246	74
	3	0.317	49
	4	0.372	39
	5	0.504	29
	6	0.589	27
	7	0.666	26
PC 14	1	0.196	106
	2	0.238	79
	3	0.282	60

	4	0.343	44
	5	0.364	40
	6	0.405	35
PC 12	1	0.191	111
	2	0.243	76
	3	0.299	54
	4	0.340	44
	5	0.361	41
	6	0.501	29
PC 54	1	0.202	101
	2	0.246	74
	3	0.299	54
	4	0.343	44
	5	0.367	40
	6	0.408	35
CTL 4	1	0.132	179
	2	0.202	101
	3	0.305	52
	4	0.337	45
	5	0.364	40
CTL 14	1	0.321	102
	2	0.501	57
CTL 25	1	0.318	104
	2	0.483	59
CTL55	1	0.336	96
	2	0.519	55
	3	0.743	44
CTL 46	1	0.223	109
	2	0.319	73
	3	0.399	58
	4	0.612	45
CTL 56	1	0.220	111
	2	0.399	58
	3	0.612	45
CTL 15	1	0.220	111
	2	0.399	58
CTL 13	1	0.220	111
	2	0.262	91
	3	0.395	59
	4	0.532	48
PC 29	1	0.195	98
	2	0.332	58

PC 30	1	0.202	94
	2	0.351	55
	3	0.443	47
PC 31	1	0.218	87
	2	0.366	53
	3	0.534	44
PC 32	1	0.145	130
	2	0.233	81
	3	0.302	63
	4	0.393	51
	5	0.462	46
PC 33	1	0.244	77
	2	0.405	50
PC 34	1	0.262	72
	2	0.332	58
	3	0.439	48
	4	0.576	43
	5	0.603	43
PC 35	1	0.271	70
	2	0.347	55
	3	0.427	48
	4	0.523	44
PC 36	1	0.351	55
	2	0.443	47
	3	0.527	44
	4	0.653	42
CTL 23	1	0.203	90
	2	0.324	57
CTL 22	1	0.207	88
	2	0.324	57
CTL 21	1	0.137	139
	2	0.211	86
	3	0.332	56
	4	0.414	49
CTL 20	1	0.145	131
	2	0.215	84
	3	0.336	55
	4	0.422	49
CTL 19	1	0.109	173
	2	0.145	131
	3	0.223	81
	4	0.285	63

	5	0.332	56
	6	0.398	50
	7	0.414	49
	8	0.438	48
	9	0.664	45
CTL 18	1	0.219	83
	2	0.281	64
	3	0.348	54
	4	0.430	49
	5	0.523	46
	6	0.625	45
CTL 17	1	0.227	80
	2	0.344	54
	3	0.410	49
	4	0.434	48
	5	0.480	47
CTL 16	1	0.227	80
	2	0.340	55
	3	0.395	50
CTL 54	1	0.230	72
	2	0.326	46
	3	0.371	41
CTL 53	1	0.213	79
	2	0.251	64
	3	0.313	48
	4	0.380	40
	5	0.434	36
	6	0.679	31
CTL 24	1	0.143	128
	2	0.205	83
	3	0.251	64
	4	0.297	52
	5	0.367	41
	6	0.430	36
	7	0.542	32
	8	0.679	31
CTL 12	1	0.193	90
	2	0.292	53
PC 37	1	0.126	145
	2	0.188	92
	3	0.242	67
	4	0.305	50

	5	0.430	36
	6	0.617	31
	7	0.704	31
PC 38	1	0.180	98
	2	0.238	68
	3	0.301	51
	4	0.376	40
	5	0.729	31
CTL 50	1	0.205	108
	2	0.291	63
	3	0.328	52
	4	0.387	42
	5	0.502	32
	6	0.639	28
	7	0.715	28
CTL 11	1	0.194	117
	2	0.276	69
	3	0.320	54
	4	0.369	44
	5	0.506	32
CTL 10	1	0.120	210
	2	0.191	121
	3	0.268	72
	4	0.309	57
	5	0.372	44
CTL 9	1	0.176	135
	2	0.224	95
	3	0.268	72
	4	0.376	43
CTL 8	1	0.116	217
	2	0.179	131
	3	0.228	93
	4	0.268	72
	5	0.313	56
	6	0.376	43
	7	0.498	32
	8	0.628	29
PC 28	1	0.218	107
	2	0.268	79
	3	0.319	60
	4	0.373	48
	5	0.404	44

CTL 1	1	0.311	63
CTL 2	1	0.144	181
	2	0.214	110
	3	0.276	75
	4	0.346	54
	5	0.393	45
	6	0.424	41
	7	0.474	37
	8	0.630	31
	9	0.700	30
	10	0.755	30
	11	0.805	29
CTL 3	1	0.218	107
	2	0.327	58
	3	0.389	46
CTL 5	1	0.202	119
	2	0.327	58
	3	0.385	46
	4	0.404	44
	5	0.416	42
CTL 6	1	0.202	119
	2	0.256	84
	3	0.327	58
	4	0.401	44
CTL 7	1	0.210	113
	2	0.260	82
	3	0.327	58
CTL 52	1	0.210	113
	2	0.327	58
	3	0.381	47
	4	0.412	43
	5	0.712	30
	6	0.798	29
PC 21	1	0.202	114
	2	0.304	62
PC 22	1	0.198	117
	2	0.300	63
	3	0.371	46
PC 23	1	0.123	204
	2	0.198	117
	3	0.257	80
	4	0.304	62

	5	0.359	48
	6	0.659	30
	7	0.757	29
PC 24	1	0.194	120
	2	0.245	86
	3	0.308	60
	4	0.371	46
	5	0.391	43
	6	0.470	36
	7	0.572	31
	8	0.659	30
PC 25	1	0.194	120
	2	0.241	88
	3	0.308	60
	4	0.383	44
	5	0.466	36
	6	0.647	30
	7	0.749	29
PC 26	1	0.178	134
	2	0.226	97
	3	0.285	68
PC 27	1	0.182	131
	2	0.233	92
	3	0.289	67
	4	0.344	52
	5	0.367	47
	6	0.430	39
	7	0.639	30
	8	0.730	29
	9	0.749	29
CTL 51	1	0.190	123
	2	0.300	63
	3	0.359	48
	4	0.635	30
	5	0.880	29
	6	0.909	29
PC 2	1	0.184	144
	2	0.241	93
	3	0.306	62
	4	0.391	44
	5	0.444	38
	6	0.493	35

	7	0.538	33
	8	0.979	31
PC 5	1	0.253	86
	2	0.351	50
	3	0.408	42
	4	0.436	39
	5	0.489	35
PC 6	1	0.249	88
	2	0.302	64
	3	0.359	49
	4	0.412	41
	5	0.489	35
PC 7	1	0.237	96
	2	0.294	67
	3	0.347	51
	4	0.424	40
	5	0.485	35
	6	0.750	31
PC 8	1	0.212	115
	2	0.273	75
	3	0.334	54
	4	0.383	45
	5	0.412	41
PC 9	1	0.216	111
	2	0.273	75
	3	0.330	55
	4	0.412	41
	5	0.648	31
	6	0.742	31
CTL 31	1	0.177	97
	2	0.218	70
	3	0.255	54
	4	0.316	40
	5	0.361	34
	6	0.490	28
	7	0.534	27
	8	0.609	26
CTL 33	1	0.163	109
	2	0.211	74
	3	0.259	53
	4	0.296	44
	5	0.320	39

	6	0.527	27
	7	0.619	26
	8	0.701	26
CTL 34	1	0.156	116
	2	0.255	54
	3	0.299	43
	4	0.323	39
	5	0.480	28
	6	0.541	27
	7	0.619	26
CTL 36	1	0.167	106
	2	0.255	54
	3	0.299	43
	4	0.320	39
	5	0.486	28
	6	0.534	27
	7	0.609	26
CTL 38	1	0.167	106
	2	0.252	56
	3	0.310	41
	4	0.333	37
	5	0.357	35
	6	0.378	33
	7	0.473	28
	8	0.534	27
	9	0.602	26
PC 10	1	0.177	104
	2	0.275	54
	3	0.328	43
	4	0.342	40
	5	0.395	34
	6	0.435	32
	7	0.528	28
	8	0.581	27
	9	0.610	27
	10	0.663	26
PC 18	1	0.169	110
	2	0.209	82
	3	0.273	55
	4	0.342	40
	5	0.400	34
	6	0.421	32

	7	0.474	30
	8	0.496	29
	9	0.525	28
	10	0.655	27
	11	0.934	26
PC 19	1	0.159	120
	2	0.212	81
	3	0.257	60
	4	0.307	47
	5	0.328	43
PC 20	1	0.175	106
	2	0.267	57
	3	0.321	44
	4	0.336	41
CTL 40	1	0.172	108
	2	0.273	55
CTL 28	1	0.177	104
	2	0.220	76
	3	0.270	56
	4	0.342	40
	5	0.400	34
	6	0.424	32
	7	0.480	29
	8	0.583	27
	9	0.658	27
PC 39	1	0.271	65
	2	0.567	31
	3	0.667	28
PC 40	1	0.243	74
	2	0.505	34
	3	0.777	27
PC 41	1	0.185	97
	2	0.233	77
	3	0.295	59
	4	0.371	45
	5	0.777	27
PC 42	1	0.271	65
	2	0.491	34
	3	0.657	28
PC 43	1	0.200	90
	2	0.300	58
	3	0.362	47

	4	0.462	36
	5	0.500	34
	6	0.791	27
PC 44	1	0.305	57
	2	0.395	42
	3	0.676	28
	4	0.700	28
PC 45	1	0.214	84
	2	0.262	68
	3	0.319	54
	4	0.405	41
	5	0.800	27
PC 55	1	0.219	82
	2	0.314	55
PC 46	1	0.112	154
	2	0.178	109
	3	0.232	84
	4	0.290	66
	5	0.344	54
PC 47	1	0.193	101
	2	0.220	89
	3	0.290	66
PC 48	1	0.186	105
	2	0.240	81
	3	0.298	64
	4	0.352	52
	5	0.386	47
	6	0.437	41
	7	0.467	38
PC 49	1	0.092	172
	2	0.124	145
	3	0.174	111
	4	0.205	95
	5	0.228	86
	6	0.290	66
PC 50	1	0.088	176
	2	0.174	111
	3	0.290	66
	4	0.359	51
	5	0.383	47
PC 51	1	0.224	87
	2	0.294	65

PC 52	1	0.178	109
	2	0.251	77
	3	0.298	64
PC 53	1	0.112	154
	2	0.178	109
	3	0.294	65
CTL 41	1	0.185	106
	2	0.289	59
	3	0.356	45
	4	0.385	41
	5	0.448	36
	6	0.467	35
	7	0.549	31
	8	0.716	29
CTL 45	1	0.111	183
	2	0.189	104
	3	0.237	77
	4	0.289	59
	5	0.370	43
	6	0.452	35
	7	0.627	30
	8	0.712	29
CTL 47	1	0.170	118
	2	0.296	57
	3	0.352	45
	4	0.382	41
	5	0.419	38
	6	0.556	31
	7	0.630	30
	8	0.730	29
CTL 48	1	0.107	188
	2	0.170	118
	3	0.222	84
	4	0.285	60
	5	0.437	36
	6	0.701	29
CTL 49	1	0.100	200
	2	0.170	118
	3	0.226	82
	4	0.274	63
	5	0.352	45
	6	0.415	38

	7	0.545	32
	8	0.601	30
	9	0.701	29

ESTABELECIMENTO DO PERFIL PROTEÓMICO/PEPTIDÓMICO DE AMOSTRAS DE URINA BASEADA EM MALDI-TOF MS COMO PLATAFORMA INOVADORA PARA O DIAGNÓSTICO DE CANCRO DA MAMA

Projeto: Dissertação de Mestrado

Consentimento Informado Livre e Esclarecido para Participação em Investigação
de acordo com a Declaração de Helsínquia e a Convenção de Oviedo

Por favor, leia com atenção a folha de informação referente à explicação do projeto. Se achar que algo está incorreto ou que não está claro, não hesite em solicitar mais informações. Se concorda em participar de forma voluntária no estudo queira assinar este documento.

Eu, _____, (nome completo do participante do estudo), declaro que após a leitura e compreensão do documento de informação referente à pesquisa de potenciais biomarcadores com base nos perfis de proteína urinária de indivíduos com cancro da mama do projeto de dissertação de mestrado, aceito participar de livre vontade no estudo acima mencionado com a cedência de amostras de urina. Compreendi a informação que me foi dada, tive oportunidade de fazer perguntas e as minhas dúvidas foram esclarecidas.

Sei que posso recusar-me a participar ou interromper a qualquer momento a minha participação no estudo, sem nenhum tipo de penalização por este facto. Autorizo também a divulgação dos resultados obtidos no meio científico, garantindo o anonimato.

Funchal, _____ de _____ de 2022

O Declarante,

_____