Discovery and Validation of Metabolite Biomarkers in Breast Cancer Exosomes Using Liquid Chromatography-Mass Spectrometry

Rochelle D'mello

Thesis Submitted to the University of Ottawa in partial fulfillment of the requirements of the degree of

Master of Science sp. Chemical and Environmental Toxicology

Department of Chemistry and Biomolecular Sciences
Faculty of Science
University of Ottawa

Abstract

Breast cancer (BC) is the second most diagnosed cancer in Canadian women. Early detection of this cancer is critical to improve patient survival and prognoses. Exosomes are proposed to be involved in tumor proliferation through the transfer of diverse biomolecules, including metabolites. The use of exosomes as biomarkers for early diagnosis of BC has recently garnered interest due to them having unique biomolecules in diseased cohorts. Hence, an untargeted metabolomic analysis of BC exosomes was performed using nano high-performance liquid chromatography coupled to tandem mass spectrometry (nLC-MS/MS) for BC diagnostic biomarker discovery. A total of 9 independent metabolite samples from non-tumorogenic MCF10A and highly metastatic MDA-MB-231 cell lines were analyzed. Bioinformatic analysis revealed 27 potential metabolite candidates unique to MDA-MB-231. Amongst 4 metabolites tested, one, N-Acetyl-L-Phenylalanine, was successfully validated. Overall, this study reveals that exosomes possess metabolites that can be candidates for early BC diagnosis.

Keywords: Extracellular Vesicles, exosomes, cancer, biomarker discovery, diagnosis

Acknowledgments

I would like to take this time to acknowledge the many influential people in my life that have in one way, or another contributed to this thesis, whether directly or indirectly. I would first like to acknowledge and thank Dr. Maxim Berezovski for his guidance throughout these past two years and for accepting me into his lab for the duration of my MSc degree. I would like to also thank Dr. Zoran Minic, the facility manager of the John.L.Holmes Mass Spectrometry Facility's biological branch, for his mentorship and support throughout my project as well as for analyzing my samples using the facility's Liquid Chromatography Orbitrap Mass Spectrometer. I would also like to acknowledge Emil Zaripov for his training and guidance on sample preparation and data preprocessing and Nico Huttman for sharing his bioinformatic knowledge with me and showing me how to use software such as MzMine3 and R programming. Additionally, I would like to acknowledge my family and all the other members of the Berezovski lab for their continued emotional support as well. I could not have reached this milestone without you and I am very grateful for that.

List of Abbreviations

BC= Breast Cancer
nLC-MS/MS= nano liquid chromatography tandem mass spectrometry
ER= Estrogen Receptor
PR= Progesterone Receptor
HER2= Human Epidermal Growth Factor 2
TNBC= Triple Negative Breast Cancer
EVs= Extracellular Vesicles
MVs= Microvesicles
VEGF= Vascular Endothelial Growth Factor
MVB= Multi-Vesicle Body
ILV= Intra Luminal Vesicles
MS= Mass Spectrometry
LC= Liquid Chromatography
GC= Gas Chromatography
CE= Capillary Electrophoresis
MS/MS= Tandem Mass Spectrometry
Q-TOF= Quadrupole Time of Flight
QQQ= Triple Quadrupole MS
RT= Retention Time
UC= Ultracentrifugation
UF= Ultrafiltration
PEG= Polyethylene Glycol
NTA= Nanoparticle Tracking Analysis
HRP= Horse Radish Peroxidase
CO_2= Carbon Dioxide
CID= Collision Induced Dissociation
DDA= Data Dependent Acquisition
DIA= Data Independent Acquisition
ESI= Electrospray Ionization
FC= Fold Change
MF= Molecular Formula

List of Figures

List of Tables

Table of Contents

1. Introduction

1.1 Introduction to Breast Cancer

The National Institute of Health's biological sciences curriculum defines cancer as the uncontrollable proliferation of cells, resulting in the formation of a tumor.[1] Within the last two centuries, science has made remarkable strides in cancer research.[2] Although there is currently no cure, mortality rates of some cancers have steadily decreased due to advances in medical research.[2–4] For example, in the United States of America, it was reported that from 1989-2020, the mortality rate for breast cancer (BC) decreased by a total of 43%.[4] However, BC is still prevalent in today's society. Currently, it is the most commonly diagnosed cancer worldwide, with an 86% incidence rate. Consequently, it is also reported as the second leading cause of cancer-related death in North America, with 25% of all cancer cases being attributed to BC.[5] As of 2022, it was also found that 1 in 8 women will be diagnosed with BC in their lifetime, making BC a highly prevalent disease that demands more research.[6]

Clinically, BC is classified depending on its location, with cancer cells most commonly originating in the ducts and lobules of the breast.[7] The malignant cell proliferation is enhanced by receptors on the cell surface that are involved in signaling pathways, which facilitate tumor progression.[8] The presence or absence of these receptors, which includes the estrogen receptor (ER), the progesterone receptor (PR) and the human epidermal growth factor 2 receptor (HER2), is used to denote different subtypes of BC as well. Currently, 4 subtypes of BC are recognized: Luminal A (ER positive, PR positive and HER2 negative), Luminal B (ER positive, PR negative and HER2 positive), HER2 subtype (ER negative, PR negative, HER2 positive) and Triple Negative (ER negative, PR negative and HER2 negative).[9]

Triple negative breast cancer (TNBC) is one of the most aggressive subtypes of BC. It's lack of ER,PR and HER2 receptors makes TNBC difficult to treat with traditional hormone therapy, resulting in a 5-year mortality rate of 40%.[8–10] Moreover, compared to other BC subtypes, there is a higher incidence of TNBC occurrence in pre-menopausal women, specifically under 40 years of age.[10] Subsequently, TNBC is also highly metastatic, with 46% of cases leading to metastasis and 25% of treated cases recurring after surgical interventions.[10] Usually, once TNBC recurs in an individual, their mortality rate increases to 75% on average. The voracity of this cancer emphasizes the importance of early detection, which has been shown to increase the survival rate of not only TNBC patients, but all BC patients.[11]

The current method of BC screening uses mammography, a type of x-ray instrument, to image the breast tissue.[12] Most commonly, many women start mammographic BC screening every two years from the age of 50.[13] However, mammograms are prone to false positive and false negative results, which could result in a potentially delayed diagnosis.[12] Therefore, the use of biomarkers to facilitate the diagnosis of diseases, such as BC, is of interest to the scientific community.[12] There have been many known sources of biomarkers with blood, plasma and urine being relatively common.[14–16] Recently, scientists have begun looking into extracellular vesicles (EVs) as an additional source of these diagnostic BC biomarkers.

1.2 Introduction to Extracellular Vesicles

EVs are small, lipid membrane bound organelles that exist throughout the body and can move freely either within or outside the cell.[17] They carry a variety of diverse cargo, which they can exchange in order to facilitate both inter and intracellular communication.[17] This was found to be their main role within a variety of diseases and disorders, especially BC.[17,18] Furthermore, there

has been evidence that EVs can be used as a source of diagnostic and prognostic biomarkers.[16,19,20] Recently, their interaction between cells has also pushed the development of using EVs as vehicles for various therapeutic agents.[21–24] For these reasons, there is currently a lot of interest in studying EVs, especially for biomedical applications. There are two main methods of EV classification that are used. One of these classification methods is based on size whereby small EVs (sEVs) are <200nm, and medium and large EVs are >200nm.[25] Alternatively, EVs can also be classified based on their biogenesis. In this type of classification, there are three main categories of EVs: Apoptotic bodies, which are formed during apoptotsis, microvesicles (MVs) which are formed from the cell membrane, and exosomes, which are formed in the endosomal pathway within the cell.[25]

1.2.1 Apoptotic Bodies; Biogenesis, Characterization & Purpose

Apoptotic bodies are the largest type of EV, ranging from approximately 500nm-2μm in diameter.[26] In 1972, Kerr *et. al* elucidated the full mechanism of apoptosis, programmed cell death, leading to a better understanding of apoptotic bodies formation .[27,28] They found that apoptotic bodies are formed in the first stage if apoptosis when the cells fragment into clusters of varying size and shape (Figure 1). Moreover, the contents of these clusters vary and may include DNA fragments, chromatin, degraded proteins, cytosol, and intact organelles (Figure 1).[26,28]

However, this highly dependent on the contents of the cell.[28]

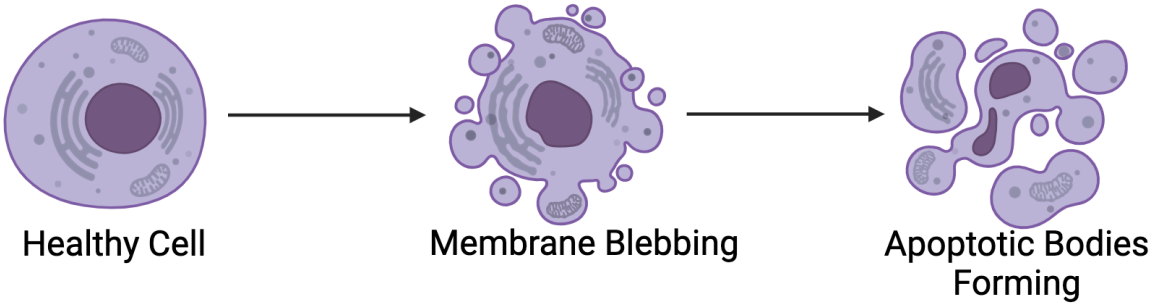

Figure 1: Formation of Apoptotic Bodies.

Illustration of the process of apoptotic formation starting with a healthy cell, which goes through membrane blebbing and finally forms apoptotic bodies consisting with DNA, whole organelles, and cytosol.[26,28] Figure created using BioRender.

After formation, apoptotic bodies are taken into other cells and destroyed via phagocytosis.[28] These cells can detect apoptotic bodies by two main mechanisms. Firstly, the formation of apoptotic bodies enables phosphotidyl-serine, a biomolecule that normally exists within the cell, to come out onto the surface of the large EV.[29] The presence of phosphotidyl serine allows other cells to recognize the presence of an apoptotic body and consume it. Additionally, proteases are that cleave various biomolecules are activated during apoptosis which results in morphological changes such as chromatin condensation, DNA fragmentation and membrane blebbing, essential for the formation of apoptotic bodies.[28,30]

1.2.2 Microvesicles: Biogenesis and Relevance in Disease Pathology and Diagnosis

MVs are smaller than apoptotic bodies, ranging from 0.2-1μm in diameter and are formed by the systematic inward pinching and budding of the cell membrane.[17,28,31] Although the mechanism of formation is not fully known, there are 3 main aspects of MV biogenesis: cell membrane

dynamic curving, protein and cargo recruitment, and vesicle fission. The step-by-step mechanism was described and illustrated in 2018 by Van Niel *et. al*.[32] Before membrane commences, lipids and proteins, like tetraspannins, are clustered together at the cell surface (Figure 2A). Other biomolecular cargo such as RNA, metabolites, or proteins, are then selectively recruited towards the surface cluster.[32] The clustering on the membrane surface creates a microdomain, a specific microenvironment, usually comprised of lipids or proteins. [32,33] Consequently, the formation of the microdomains spurs membrane budding and fission, generating MVs. However, the direct mechanism of how that occurs is not yet fully understood (Figure 2B).

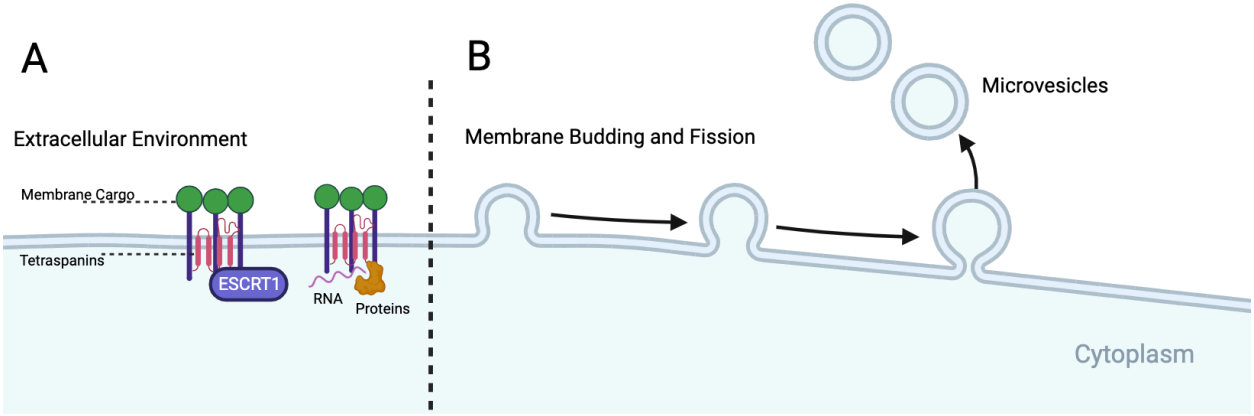

Figure 2: First and Second Stage of MV Formation

A) Membrane cargo in the form of various receptors are clustered together on the cell surface. This initial clustering promotes the selective recruitment of MV cargo which include RNA and proteins. B) Membrane budding and fission resulting in the formation of MVs. Figure created using BioRender.

MVs can play a variety of roles in disease pathology and diagnoses depending on their origin and subtype, by transferring proteins between the intra and extracellular environment.[31] For example, tumor derived MVs have been implicated in cancer disease progression via the transfer of various growth factors.[31] The presence of tumor derived MVs signals the release of vascular endothelial growth factor (VEGF) containing MVs by endothelial cells. VEGF plays an important role in tumor angiogenesis, a critical aspect of tumour survival.[19] Additionally, in a

study conducted by Choi *et. al* in colorectal cancer patients, it was reported that many proteins identified only in MVs were involved in disrupting tumour invasion, angiogenesis, migration, and tumour growth, promoting tumor progression.[19,34]

1.2.3Exosomes: Characteristics and Biogenesis of the Smallest EV

The smallest EVs are called exosomes. Exosomes typically ranging in size from approximately 30-150nm in diameter.[17] They are often characterized via the enrichment of protein markers from the tetraspanin family such as CD63, CD81, CD9, heat shock protein 70 (HSP70), Tsg1 and ALIX .[20] The biogenesis of exosomes is illustrated in Figure 3. Briefly, exosomes are born as intra luminal vesicles (ILVs), within the cell.[35] ILVs exist in larger vacuoles known as the endosome, which, during the late endosomal period, are referred to as the multi-vesicle body (MVB). On the surface of the MVB membrane, specific cargo-like proteins cluster, creating a microdomain.[32] This clustering can occur either in an ESCRT protein complex-dependent or independent fashion. After clustering, ESCRT III-mediated scissions of the ILVs are created and the new small EVs are generated inside the MVB.[32] When the MVB merges with the plasma lipid membrane of the cell, it secretes those small EVs via exocytosis.[35] Hence, these EVs are known as exosomes. This process was well documented using electron microscopy by Pan *et al.* in 1985.[36]

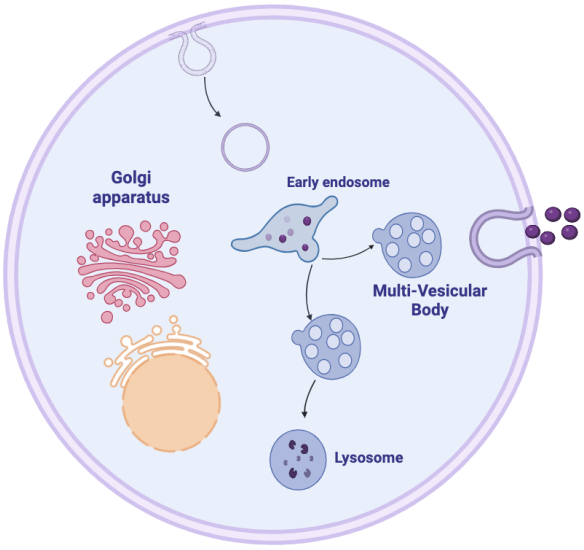

Figure 3: Formation of Exosomes. The invagination of the cell membrane forms a closed vesicle known as the endosome. The early endosome then matures and takes in vesicles from other organelles, namely the Golgi Body via endocytosis as well. The exosomes that now exist within MVBs are pushed out of the cell via exocytosis. Figure created using BioRender.

1.3 The Role of Exosomes in BC

1.3.1 How Exosome Cell-Cell Communication Affects BC Development

It is well known that generally, EVs such as exosomes, are involved in the transportation of particles between cells, cellular communication, and signalling.[20] Subsequently, the inter-tumor communication that is indirectly facilitated by exosomes results in BC proliferation and metastasis.[37] Exosomes aid in cell communication by transferring specific packages, including DNA, RNA, proteins and metabolites, between cells.[17] In the context of BC, this cargo may contain signalling molecules that initiate certain pathways. This communication allows exosomes to drive BC processes such as proliferation and metastasis.[20] In the literature, it is evident that these processes are regulated by cargo such as RNA. For example, the binding of miR-105 to the ZO-1 protein results in the destruction of endothelial walls, promoting tumor cell migration in BC.[38] The binding of these two molecules is a direct result of exosomes carrying the

RNA biomolecular package from one cell to another. Another study, conducted in 2018, found that the transfer of metastasis-associated lung cancer adenoma transcript 1 (MALAT-1), a non-coding RNA, between cells, promoted cell proliferation in highly metastatic MDA-MB-231 human BC cells.[39] Interestingly, RNA is also associated with both BC tumor cells and the bone marrow microenvironment, implicating it in BC metastasis towards bone marrow.[40] This is supported by Guo *et. al* who found that the transfer of the micro RNA miR-20a-5p via exosomes, from BC cells to bone marrow macrophages promoted both cell migration and invasion.[40]

1.4 Using MS-Based OMICS to Study Exosomes

1.4.1 Overview of Omics

Since the late 1990's, scientists have used OMICS based experiments, studying DNA, RNA and proteins in exosomes to understand the molecular bases of diseases such as cancer.[12,20,41,42] When studying these diseases, OMICS approaches have been beneficial for disease diagnosis and can also help physicians predict therapeutic response, prognosis and even disease recurrence.[41] Moreover, the rapid evolution of OMICS technology has increased in sensitivity, resolution, and has become economical, allowing for OMICS to be more frequently used in research.[41] OMICS approaches are subdivided into four main branches of study: genomics, transcriptomics, proteomics, and metabolomics.

Genomics is the oldest OMICS platform and applies the study of DNA and the genome as a whole. It has a variety of applications in cancer research.[43] For example, as recently as 2019, genomics was used to study metastatic BC, investigating somatic alterations during the evolution of early-stage to late-stage BC.[44] Interestingly, through the application of genomics, it was found that there were 9 genes that mutated more frequently, specifically in BC subtypes that possess

ER and PR receptors, but lack HER2 receptors. In similar ways, genomics can be used to map and study genome variations and their contributions to the development and expression of BC and other diseases.[42,44]

Transcriptomics is the study of RNA and can be used either qualitatively, to identify and investigate the role of specific RNAs within a disease, or quantitatively, to quantify the expression of specific RNA.[42] Much like genomics, it has been used in the study of cancer, with applications in biomarker discovery and BC subtype diagnosis.[45,46] A recently published article found that mRNA expression of the ligand of Programmed Cell Protein 1 was a prognostic biomarker for disease free survival, which is the amount of time measured between treatment and disease recurrance.[47]

Proteomics, the study of proteins has been an intensively investigated field, especially for pathologies such as cancer. More specifically, proteins have been highly researched for biomarker discovery because they can provide insight into the cellular processes occurring in BC.[48] Posttranslational modification (PTM) is one of the approaches for large scale protein analysis. PTM, such as protein phosphorylation, can provide information on cell regulatory mechanisms, as reported in Minic *et. al.*[49] Kinases, such as CMGC, CAMKs, STE and AGC, phosphorylate proteins, which, if done in a recipient cell, can promote BC cell proliferation and metastasis.[49] The metabolic dysregulation caused by phosphorylated enzymes, such as ATP citrate lyase (ACLY), which is known to participate in lipogenesis, can directly impact cancer cell proliferation and, thus, are important to study. [49,50]

Additionally, other PTMs, such as protein acetylation, have also been investigated, in relation to BC. Similarly to phosphorylation, acetylation is a reversible PTM, carried out by various acetyltransferases. Previously, histone acetylation and its relationship to tumor formation was

investigated. It was found that the overexpression of histone deacetylase (HDAC) enzymes can lead to tumor formation.[51] This finding is important because histones are involved in processes related to cell replication such as DNA replication and repair.[51] While most acetylation research focuses on histones, non-histone acetylation has also started to be investigated with regards to cancer.[52] It is known that cancer causes dysregulation in the metabolic pathways, especially in processes such as glycolysis. It was found that the acetylation of lysine residues in key glycolytic enzymes such as aldolase, phosphoglycerate kinase, and enolase, resulted in increased enzyme activity in BC derived exosomes compared to the exosomes derived from healthy breast cell lines. As such, they have been proposed as tumor markers for BC.[53]

The proteome of BC derived exosomes has been heavily studied resulting in discovery of proteins, which can act as diagnostic and prognostic biomarkers for BC and have the ability to alter cellular processes. Therefore, the dysregulation of these proteins potentially may result in the dysregulation of the small metabolites that they produce and transform. Therefore, the next step is to investigate these small chemical compounds, and their role in BC.

1.4.2 Using LC-MS-Based Metabolomics in BC Research

BC is known to cause metabolic dysregulation, affecting pathways like glycolysis, the citric acid cycle, and lipid metabolism.[12,54] The study of the small metabolites within these pathways and their biological relevance is called metabolomics.[55] The existence of metabolites in exosomes found circulating in non-invasive biofluids allows them to easily be studied using a variety of analytical instruments.[12]

Improvement in sensitivity and resolution has allowed mass spectrometry (MS) to become a widely used platform for the metabolomic analysis of biological samples.[55] Figure 4 illustrates the standard workflow that is commonly used in metabolomics. First, the samples, which can

include biofluids like urine and blood or cell culture media, are collected and preprocessed for analysis. This often involves using methods like centrifugation or resin columns to separate metabolites from other biological components in the sample. Next, the sample is analyzed using MS, often coupled to either liquid chromatography (LC), gas chromatography (GC) or capillary electrophoresis (CE) to increase the resolution of separation and allow for better overall detection of individual metabolites (Figure 4).[56] The data is then analyzed and interpreted to draw out meaningful conclusions. Metabolomic studies that have used MS for analyte identification have found merit in tandem MS (MS/MS) which allows for the addition of one or more mass analyzers, increasing mass accuracy, precision and selectivity.

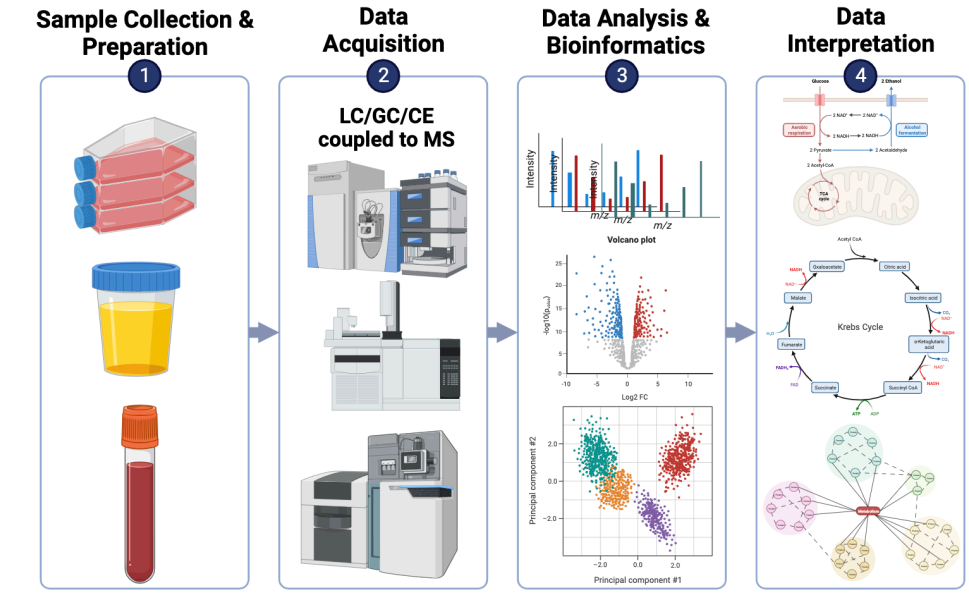

Figure 4: General MS-Based Metabolomics Workflow

Illustration of general steps that goes into the process of metabolomics including sample collection and preparation of cell culture media, urine and blood, data acquisition using separatory methods coupled to MS, data analysis and bioinformatics and finally interpretation of results. Figure created using BioRender.

1.4.4 Principles of LC

For metabolomic analysis, separatory techniques, such as LC, are often coupled with MS. LC-based separation is based on multiple factors. First, separation is based on the strength difference between the interaction equilibrium between the analyte with the stationary phase, and the analyte with the mobile phase. The strength of these interactions can change based on the chemical structure of the analyte and the solvent used for the mobile phase.[57] Generally, if the analyte is more attracted to the stationary phase, it will stay in the column longer and have a higher retention time (RT) compared to an analyte that has a weaker attraction to the stationary phase.[57]

Another key factor that is important to consider for LC-based separation, especially in complex matrices, is the band broadening that can occur, which can overall lead to poorly resolved and wide chromatographic peaks.[58] This broadening is a result of diffusion occurring as the components of a sample travels through a column and is a measure of the columns efficiency. Column efficiency is often indicated by the number of theoretical plates that a column has and can also be represented by the Van Deemter equation, equation 1, where A describes Eddy diffusion whereby an analyte can take multiple different paths through the column, which will affect the RT and result in band broadening.[58] B describes longitudinal diffusion where an analyte travelling through the column in a band will diffuse out from the center to the edges. C describes the resistance to mass transfer where molecules of the sample can transfer to the stationary phase, and u represents velocity of the mobile phase. Decreasing the overall theoretical plate height (H) will allow for more plate to be present in a column and increase the overall column separation efficiency.[58]

(1)

$$H = A + \frac{B}{u} + Cu$$

In addition to factoring in column efficiency and separation, the type of stationary phase is also critical and can determine the type of experiment that can take place. One of the most common types of of materials used for the stationary phase is silica beads because of its strength and versatility.[59] Silica beads can be chemically modified, attaching a variety of different adducts that can be used for multiple types of separation. Once of the most common types of chromatography used in untargeted metabolomics is reverse phase chromatography whereby a C_{18} chain is chemically added to the silica beads. This modification allows the originally polar silica, to be used to separate non-polar analytes, which is what was used for this work.

In addition to using LC for separation, this work also employed the use of the nanoLC, a cutting-edge miniaturization of traditional LC. nLC allows for decreased injection volumes and decreased sample waste. Additionally, the decreased solvent required for nLC prevents sample dilution when sprayed into the MS increasing overall sensitivity, which is important for untargeted metabolomics.[60]

1.4.5 Approaches to Metabolomics: Untargeted vs Targeted

There are two approaches to metabolomic analysis: targeted and untargeted. Untargeted analysis involves analyzing all metabolites in a sample and is often used to understand the molecular basis of diseases.[61] During untargeted analysis, emphasis is put on the identification of new biomarkers rather than absolute quantification.[61] The discovery and identification of these novel biomarkers is facilitated through multiple opensource metabolomic databases. However, the limited capacity of these databases may result in unidentified metabolites, which poses a major

challenge in untargeted metabolomics.[61] Additionally, due to the large range of data acquired during instrumental analysis, data processing, statistic, and bioinformatic analysis is intensive and more complex than targeted metabolomics approaches. However, this approach maximizes the number of identified metabolites and aids in the discovery of novel biomarkers.[61] Alternatively, targeted metabolomics is hypothesis driven and involves analyzing a set of known metabolites, often from literature or previous studies.[61] This approach usually follows untargeted analysis and is used to obtain the absolute quantification of metabolites using commercially available standards and calibration curves. In contrast to the untargeted approach, targeted metabolomics may result in some bias due to prior knowledge of an expected m/z ratio.[62] However, this knowledge allows for optimization in the experimental conditions, including aspects like sample preparation, instrumental parameters, data preprocessing and statistical analysis.[62] As the field of metabolomics is continuously evolving, both of these approaches could be used to discover non-invasive BC biomarkers.

1.5 Project Goals

The main goal of this work is to identify potential exosomal metabolite biomarkers using nLC-MS/MS, that could be used for the early diagnosis of BC via an untargeted approach. Once the biomarker is identified using a combination of database spectral matching and manual spectral matching, it will then be validated using commercially available standards.

The early detection of BC, especially TNBC, is extremely important in healthcare. Not only does early detection improve patient survival and outcomes, but it may also increase the efficacy of the treatment and reduce treatment side effects.[63] Currently, traditional BC screening using mammography in post-menopausal women produces inaccurate results.[7,12] As such, there is a

need for a more precise and reliable method for BC screening. Therefore, the use of MS-based metabolomics to analyze BC exosomal metabolites can help address this gap in healthcare.

2. Exosome Isolation and Characterization

2.1 Introduction

In order to study exosome metabolites, the exosomes must first be isolated and characterized to verify their presence. Recently, there have been many methods developed to isolate EVs based on unique properties such as size, charge, or affinity.[64] These methods include differential ultracentrifugation (UC) as well as other methods such as ultrafiltration (UF), and precipitation based EV separation using commercially available kits such as ExoQuick.[65] Choosing an appropriate isolation method is dependent on many variables including sample type, time and labour cost, exosome purity, recovery rate, and amount of crude sample required for each isolation method.

In the literature, the yield and specificity of all three isolation methods have been compared by Risha *et. al*.[66] Through this comparison, it was found that, while ExoQuick produced the highest concentration of vesicles, there were many vesicles that were >200nm in diameter which is outside the size range for exosomes. On the other hand, UF, produced the lowest concentration of the three methods, with an average vesicle diameter of 135nm.[66] Compared to these two methods, differential UC produced the second-highest concentration of vesicles and the average vesicle size of 105nm. Additionally, MS analysis of exosomal proteins showed that the UC isolation yielded 986 exosomal proteins, higher than both UF and ExoQuick methods. These results indicated that UC was the optimal isolation method for exosomal isolation.[66] Therefore, based on this evidence, UC was chosen as the primary isolation method for this work.

2.1.1Principles of UC

Currently, UC is one of the most commonly used methods for EV isolation.[67] It involves the settling of sediment or biomolecules in a solution from the force generated by spinning as fast as 60,000 rpm.[68] There are 3 aspects to consider during centrifugation experiments: sedimentation, velocity and sedimentation equilibrium.

Sedimentation is the act of particles in solution settling at the bottom of the container. The rate at which this occurs is determined by the sedimentation coefficient (s), reported in Svedbergs, which is described by equation 2 where $\omega^2 x$, the centrifugal field is represented by the ω the angular velocity, and v represents the speed at which a biological compound will sediment.[68] Additionally, sedimentation is also determined by the density and radius of the biological particles as well as the surrounding medium.

(2)

$$s = \frac{v}{\omega^2 x}$$

Relative centrifugal force (RCF), also known as "xg" is more common unit used to express the centrifugal field. It is mathematically described in equation 3 as the ratio of centrifugal acceleration at a certain radius to the acceleration due to gravity (-9.8 m/s^2). In equation 3, rpm representes the rotor speed, r is the distance of the sample from the axis of rotation and $4\pi^2 rad^2$ is the square of a 360° rotation around the axis and g is acceleration due to the force of gravity.[69]

(3)

$$\frac{4\pi^2 rad^2 \times rpm^2 \times r}{g} = RCF$$

2.1.2Differential UC as a Method for EV Isolation

Differential UC is an application of traditional UC that utilizes the different size and density characteristics of biological structures existing in a sample to separate them.[70] Biostructures that are larger or denser, will sediment while spinning at lower speeds compared to biostructures that are smaller or less dense. This approach, therefore, allows for the separation of multiple types of EVs, such as exosomes and MVs, into different fractions based on size (Figure 5).

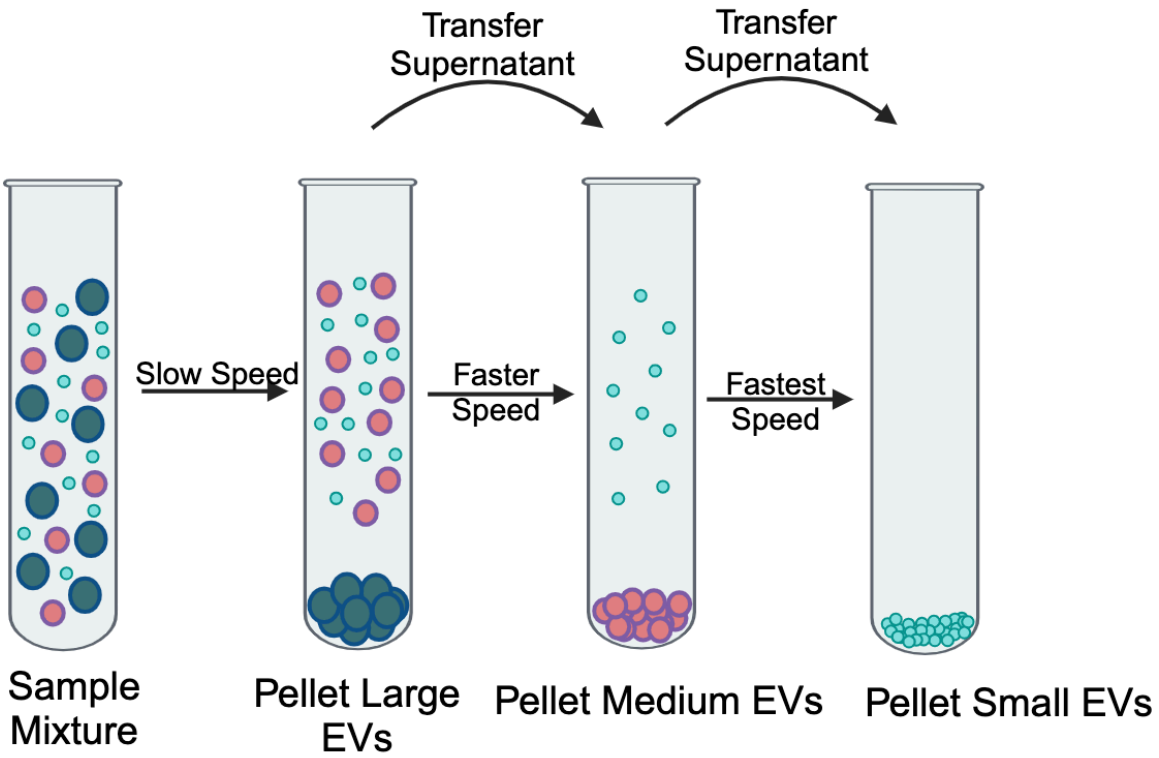

Figure 5:Schematic showing the separation of differently sized EVs in relation to speed in differential UC. As the speed of the centrifugation increases, smaller EVs can be pelleted and isolated. This figure

also illustrates the capability of this UC technique to completely isolate different fractions UC and transferring the supernatant to continue the next UC step. Figure created using BioRender.

2.1.3Characterizing EVs using Nanoparticle Tracking Analysis

After isolation, EVs can be characterized using nanoparticle tracking analysis (NTA). This is a relatively newer approach, being used since 2006, that stems from dynamic light scattering.[71] NTA combines laser light scattering microscopy with cameras to not only visualize biomolecules, but as also to detect their size. According to dynamic light scattering theory, scattered light intensity is proportional to the diameter of a particle to the power of 6.[72] This is due to fluctuation in the scattered light caused by Brownian motion, the random movement of particles suspended in solution.[71] Hence, overall speed is faster when particle radius is smaller, as seen in equation 4 where (x, y) is the average speed of the particle, K_B is the Boltzmann constant, T is temperature, η is the viscosity of the medium and r_h is the hydrodynamic radius of the particle.[71] Furthermore, since radius and particle speed are inversely proportional to each other, smaller particles will have faster Brownian motion than larger particles.

(4)

$$(x, y)^2 = \frac{2K_B T}{3r_h \pi \eta}$$

2.1.4 Characterizing EVs using Exocheck Exosome Array

Another method for characterizing exosomes is called the ExoCheck Antibody Array (ExoCheck). Contrarily to NTA, which is used to detect particle size, the ExoCheck is an immunoblotting technique, similar to western blotting, that analyzes and identifies protein

markers enriched specifically in exosomes compared to other EVs.[73] Unlike more traditional characterization, such as western blot, that may only be able to detect one or a few protein markers at a time, ExoCheck can be used to detect 8 protein markers simultaneously, making it more time efficient.

Typically, with antibody arrays, whole exosomes are lysed, and the protein content is added to a mixture of antibodies. Exocheck commonly contains antibodies for CD81, CD9, CD63, ALIX, ANAX5 and TSG101 protein markers enriched in exosomes. The primary antibody binds to the proteins. The secondary antibody is then added to the testing kit and binds to the primary antibody. An example of this process for the detection of the CD81 protein marker is illustrated in Figure 6.

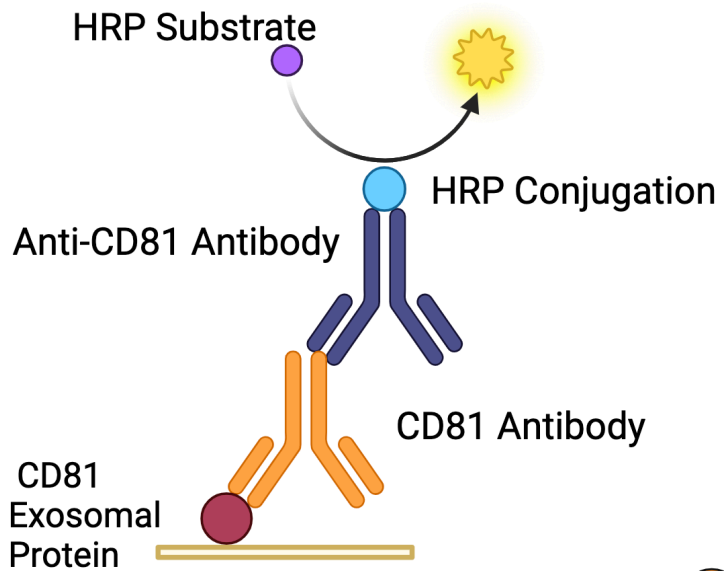

Figure 6: General Principle of ExoCheck Array

The primary antibody binds to the exosomal protein marker. A horse radish peroxidase (HRP) conjugated secondary antibody then binds to the primary antibody. A colour change then occurs when the HRP reacts with the substrate. Figure created using BioRender.

2.2 Methods

2.2.1 Preparing EV Free Serums

Horse Serum (HS) (Sigma Aldrich, cat# H1138) or Fetal Bovine Serum (FBS)(Gibco, REF 12484-028) was aliquoted into a 1:1 ratio of serum to Gibco's DMEM/F12 (1:1) supplemented with L-Glutamine and 15mM HEPES (REF 11330-032), into a 50mL falcon tube. The serum:DMEM/F12 mixtures were ultracentrifuged using a Beckman Coulter UC with a SW28 rotor (BeckmanCoulter,CatNo.342204) at 27,500 rpm for 20 hours to pellet EVs. Once the spinning was complete, the supernatant was transferred to clean 50mL falcon tubes and filtered using MilliporeSigma™ Steriflip™ Sterile Disposable Vacuum Filter Units (Cat No. SCGP00525), to ensure sterility appropriate for cell culture use, and stored at -20℃.

2.2.2 Culturing MDA-MB-231 and MCF10A Cells

MDA-MB-231(ATCC®, Cat No.HTB-26) and MCF10A (ATCC®CRL-10317™) cell lines were cultured for this project. Both cell lines were grown using Gibco's DMEM/F12 (1:1) supplemented with L-Glutamine and 15mM HEPES (REF 11330-032) growth media. Media recipes for each cell line can be found in Table 1. All cells were incubated at 37℃ in 5% carbon dioxide (CO_2). Once cells grew to 80% confluency, they were sub-cultured, a process otherwise known as passaging. Cell viability was determined using the NanoEntek EVE™ cell counter and cells that were 91-98% viable were then plated onto sterile VWR 182.5cm^2 surface treated tissue culture flasks (REF 734-2315) with a seeding density of approximately 10 million cells per plate. Once the cells grown in the 182.5cm^2 flask reached confluency, the cells were incubated with 40mL of growth media for 24 hours at 37℃ in 5% CO_2. This media, now containing EVs, was collected into 50mL falcon tubes. The tubes were centrifuged at 300xg for 5 minutes and at

2,000xg for 20 minutes to remove cell debris and large biomolecules. The supernatant was then transferred to a clean 50mL falcon tube and stored at -80℃ until isolation.

Table 1:Media Recipes of MCF10A and MDA-MB-231 Cell Lines. *Table Created in Microsoft Word*

Cell Line	Growth Media	Final Concentration of Supplements
MCF-10A	DMEM/F12	5% HS 1% Antibiotic/Antimitotic 100ng/mL cholera toxin 20ng/mL EGF 10μg/mL Insulin 1mg/mL Hydrocortisone
MDA-MB-231	DMEM/F12	10% FBS 1% Antibiotic/Antimitotic

Additional Information:

1. HS (Sigma Aldrich cat# H1138, 500mL bottle) = Horse Serum
2. Antibiotic/Antimitotic (100x solution, Gibco REF 15240-062) = Penicillin/ Streptomycin
3. Cholera Toxin (Sigma cat# C-8052, 1mg vials): Resuspended in 1mg/mL of sterile dH_2O. Aliquots stored in 4℃
4. EGF (Peprotech, 1mg) = Epithelial Growth Factor: Resuspended at 100μg/mL in sterile dH_2O. Aliquots stored at -20℃
5. Insulin (Sigma Aldrich cat# I1882, 100mg vials): Resuspended at 10mg/mL in sterile dH_2O. Aliquots stored at -20℃
6. Hydrocortisone (Sigma Aldrich cat# H0888, 1g bottle): Resuspended at 1mg/mL in ethanol. Aliquots stored at -20℃
7. FBS (Gibco REF 12484-028, 500mL bottle) = Fetal Bovine Serum

2.2.3 Exosome Isolation Using UC

Exosome isolation was done using differential UC as illustrated in Figure 7. The harvested cell culture media supernatant was removed from the -80℃ storage and thawed overnight at 4℃. The supernatant was then spun in the UC sw28rotor at 16,500xg for 1hr at 4℃ to remove larger vesicles such as apoptotic bodies and microvesicles. The pellet was discarded and the supernatant was transferred into new clean tubes and spun again using the same rotor at 100,000xg for 3hrs to pellet exosomes. The supernatant was then discarded and the exosome pellets were vigorously resuspended in 150mM ammonium acetate and spun at 100,000xg for 3 hours to wash the exosomes of any contaminants remaining from the cell culture media. Finally, the supernatant was discarded and any remaining supernatant was removed from the tubes by placing them upside down on a kimwipe. Metabolites were either immediately extracted from the newly isolated exosome pellets or they were resuspended in 150mM ammonium acetate and placed in the -80℃ storage container for later.

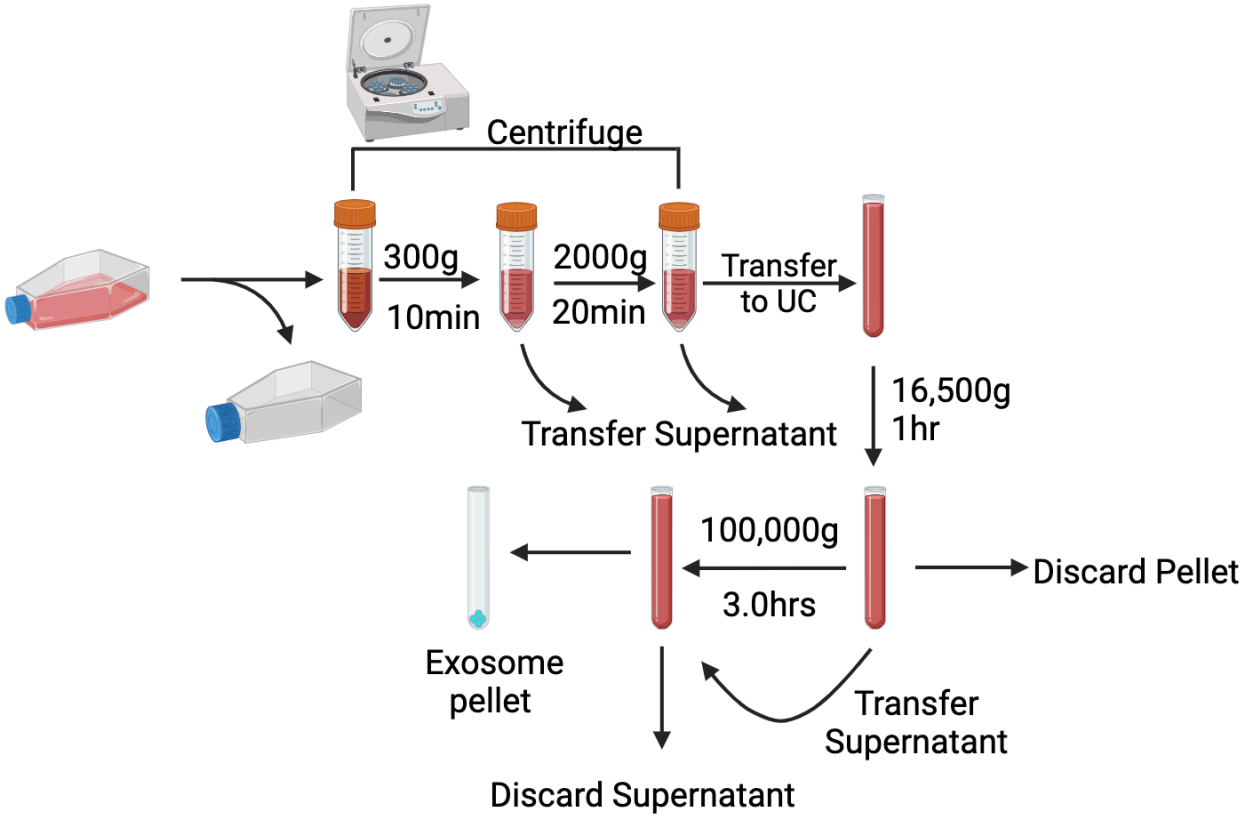

Figure 7:Exosome Isolation Protocol Using Differential UC. Whole cell culture media is centrifuged. The supernatant is then ultracentrifuged as described, allowing for the isolation of exosomes. Figure created using BioRender

2.2.4 Methanol Extraction of Metabolites

The exosome pellet was resuspended vigorously in 1mL of 100% MS-grade methanol (MeOH) for 5 minutes per sample before being transferred to a 1.5mL Eppendorf tube. The solvent was then evaporated using the Thermo Savant SPD111V SpeedVac Concentrator and stored at -80°C until needed.

2.2.5 Nanoparticle Tracking Analysis (NTA)

For NTA based experiments, exosome pellets were resuspended in 150µL of 150mM ammonium acetate of pH 6.9. All samples were diluted in a series of 10-1000 fold using a serial dilution technique before injecting into the ZetaView NTA microscope PMX-110 (Particle Matrix) to

find the concentration in the effective range of 10^8-10^9particles/mL. Instrument settings were as follows: Shutter speed of 40 and 85, minimum brightness of 15, size range of 10-500nm. Before the samples were run, 100nm polystyrene beads (Microtrac, Cat# 900383) were injected into the ZetaView to calibrate the instrument and focus the camera.

2.2.6 ExoCheck Analysis to Characterize Isolated Exosomes based on Protein Biomarkers

Isolated Exosomes were characterized using a commercially purchased kit, ExoCheck (System Biosciences, Palo Alto, CA, USA). Characterization was carried out in accordance with the manufacturers protocol. Briefly, a lysis buffer was added to 200μL of isolated exosome sample in order lyse the exosomes and expose the proteins. 1 μL of labelling reagent was then added and the samples were incubated for 30min with constant shaking. The samples were then run through the provided column in order to remove any excess labelling reagent. The exosome lysate was then combined with the provided blocking buffer. This mixture was then added to the provided membrane and incubated at 4℃ for 24hrs. Following the 24hr incubation, the membrane was then washed with the provided washing buffer, and the membrane was imaged.

2.2 Results

2.3.1 NTA Characterization of Particle Size of Isolated Exosomes

NTA is capable of providing information regarding particle size and distribution, taking into account any dilution of the samples. The particles in the 100-fold diluted sample of MCF10A EV isolate was found to have a mean diameter of 116nm (Figure 8A). The particles in the 10-fold diluted sample of MDA-MB-231 EV isolate was found to have a mean diameter of 124nm

approximately (Figure 8B). Additionally, the MV isolated fraction was also run for a size comparison. It was found that the 100-fold dilution of MCF10A MV had a mean diameter of 181 nm, larger than the 116nm exosome fraction.

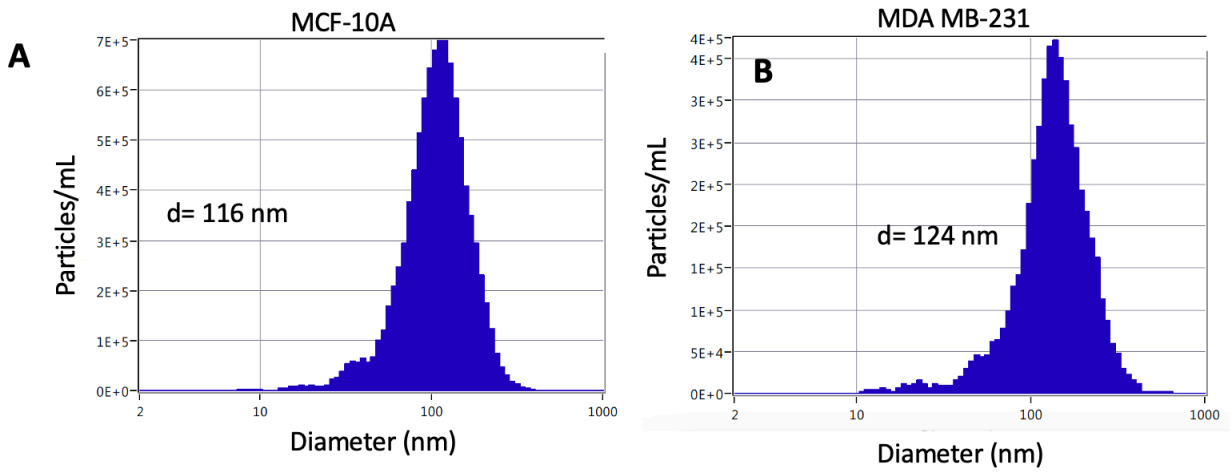

Figure 8: Particle Size Distribution of MCF10A and MDA-MB-231 Derived EVs. **A)**Particle distribution of MCF10A derived EVs diluted 100 fold in 150mM Ammonium Acetate buffer solution created by dissolving solid ammonium acetate in millicule water. B) Particle distribution of MDA MB-231 diluted 10-fold in 150mM ammonium acetate buffer created by dissolving solid ammonium acetate in millicule water. Figure created using ZetaView Data and formatted using Microsoft PowerPoint.

2.3.2 Protein Marker Characterization of EV isolate using ExoCheck Exosome Array

In addition to using a particle size-based method for characterization, EVs were also characterized by observing the presence of protein markers specifically enriched in exosomes. The membrane displayed the faint, but visible presence of all membrane and internal protein markers in MCF10A-derived exosomes and most in MDA-MB-231-derived exosomes. Additionally, the presence of a dark line in the positive control and the complete absence of a like in the blank spot was observed (Figure 9A and Figure 9B).

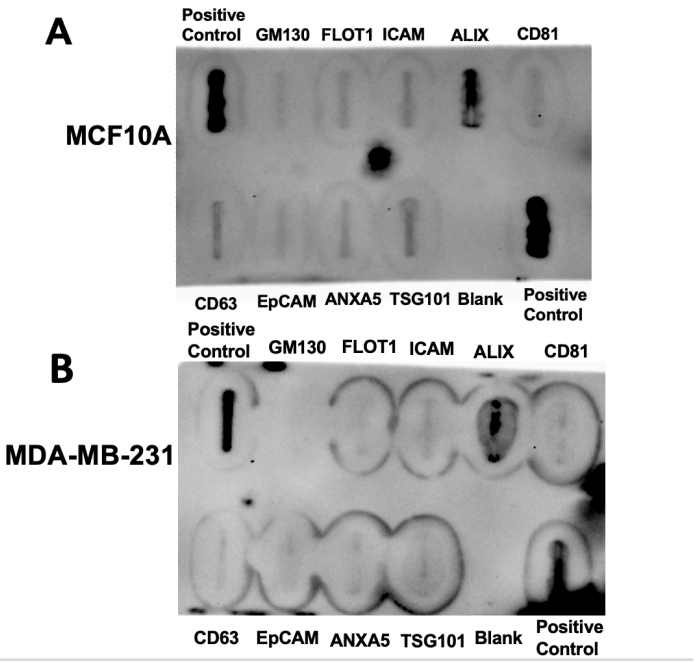

Figure 9: ExoCheck Array Membrane Images of MCF10A and MDA-MB-231 derived EV isolate. A) ExoCheck Array membrane image of MCF10A EV isolate. B) ExoCheck Array membrane image of MDA-MB-231 EV isolate. Figure created using Microsoft Powerpoint.

2.4 Discussion

2.4.1 Discussion of NTA EV Characterization

Characterization of EV isolates from MDA-MB-231 and MCF10A cell lines resulted in mean diameters that were both less than 150nm in diameter. This result indicates that most of the particles isolated are likely exosomes as they are in the range of 30-150nm.[17] While the average diameter of the exosome fraction was smaller than the MV fraction, which had an average diameter of 180nm, there were still particles detected in both cell lines with a diameter of >200nm as evidenced by the peak width seen in Figures 8A and 8B. This indicates that, other, larger EVs, such as MVs may have also been isolated along with the exosomes. MVs and

exosomes share some similarities. Not only do they share an overlap in diameter range, but they also share a number of proteins in common, as found in Risha *et. al.*[74] While some of the larger particle sizes may be eliminated by the use of subsequent purification methods such as size exclusion chromatography, those MVs, which have been found to be as small as 100nm, may not be able to be fully separated from exosomes using preestablished methods.[65]

The ZetaView instrument can accurately detect based on a certain range of particles. As such, the sample had to be made into many dilutions because the optimal concentration for diameter detection was unknown and may vary between sample types. For this reason, the MCF10A exosome sample was detected using the 100-fold dilution and the MDA exosome sample was detected using the 10-fold dilution. These were the dilutions that had the correct range of particles for the instrument to accurately function. It is important to note that the variation in the diameters from the exosomes of these two cell lines may be inter-day variation caused by running the samples on separate days. In order to minimize the instrumental variance caused by changes in condition, the diameter data for both was normalized to the measurement of the 100nm blank control polystyrene beads. When running MDA-MB-231 cell line, the ZetaView read the 100nm beads as 109.1nm and when running the MCF-10A cell line, the instrument read the 100nm beads as 90.4nm. Thus, when the measurements were taken, they were adjusted by the difference between the bead reading and the true bead diameter.

2.4.2 Characterization of Exosome Enriched Proteins Via ExoCheck Array

Further confirmation of exosome isolation can be obtained by the characterization of proteins that are specifically enriched in exosomes, compared to other types of EVs. These include surface proteins such as tetraspanins (CD63, CD81), other integrin proteins (ICAM, EpCAM)

and proteins found on the interior of the exosome (TSG101, ALIX, FLOT1, GM130).[75] As seen in Figure 9, both MCF10A and MDA-MB-231 exosomes have visible bands where the antibodies for the labeled exosome markers are. This confirms that, although faint, the exosome markers are present in the EV isolate, thereby confirming that exosomes were isolated using differential UC. The lines may be faint for a variety of reasons. One of the main reasons could be the sensitivity of the assay as it requires more protein than can be obtained just by pooling the exosome isolate obtained from 480mL of cell culture media. Overall, the presence of the spots indicates the presence of the specific exosome biomarker. That, along with the measured size falling into the range of exosome diameter indicates that exosomes were indeed isolated exosomal metabolite analysis can continue as described in Chapter 3.

3. Untargeted Metabolomic Analysis Using nLC-MS/MS

3.1 Introduction

3.1.1 Untargeted Metabolite Analysis Using Quadrupole-Orbitrap MS/MS

MS as a method of detection has become a powerful and versatile tool, capable of being applied to a variety of different research fields including metabolomics. For metabolomic studies specifically, the use of MS is often favoured because of the increased sensitivity and analysis speed that it can provide to the user.[76] However, only utilizing MS^1 spectra, whereby the whole ionized m/z of the metabolite is measured, is not sufficient to predict chemical structures for untargeted analyses. Therefore, the addition of the MS^2 spectra was necessary to help facilitate the prediction and identification of new metabolites solely based on chemical structure.

MS^2 shows metabolite fragment ions which would fragment off the parent structure after being bombarded by neutral molecules, commonly inert gas molecules.[77]

This work utilized Thermo Fisher's Q-Exactive Plus MS to generate both MS^1 and MS^2 spectra. A general schematic of the instrumental components used, after ionization, is illustrated in Figure 10. Briefly, after ionization, an ion beam is sent through the system.[78,79] During this time, neutral molecules are filtered out. From there, the ions travel through the quadrupole mass filter. Through a combination of alternating and direct current, the quadrupole can either preselect ions in chosen mass ranges or select single mass ion. The ions are then trapped in the C-Trap, before being sent up for MS^1 scanning in the orbitrap mass analyzer. In the orbitrap mass analyzer, the ions ocillate in an electric field at a frequency that is inversely dependent to their m/z. Then each ion identified from full scan is selected by quadrupole and sent to the higher energy C-Trap dissociation (HCD) cell for fragmention. The fragments are analyed once again by the orbitrap mass analyzer to produce the MS^2 spectra. Overall, this instrument provides a resolution of 240,000. Furthermore, it has <1ppm mass accuracy in full scan mode and femtogram level sensitivity, which is beneficial for untargeted metabolomics.[80]

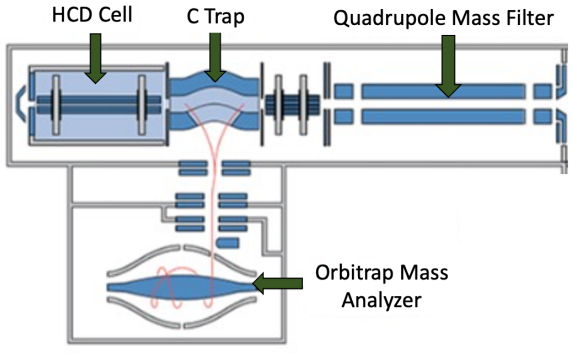

Figure 10: Schematic of Thermo Q-Exactiver Plus MS. This figure illustrates components of Thermo's Q-Exactive Plus MS instrument used for mass analysis after ionization. Figure modified from the Thermo Fisher Scientific product specifications sheet.[79]

3.1.2 Using Data Dependent Acquisition for Untargeted Metabolomics

This work uses Data Dependent Acquisition (DDA), a common MS mode for data acquisition in untargeted metabolomics. DDA allows for a full scan of m/z values and automatically fragments and detects the most abundant m/z ions in the MS^1 spectra (Figure 11).[81] Additionally, the data acquired from DDA can easily be processed by a variety of opensource software, making it a more accessible.[81]

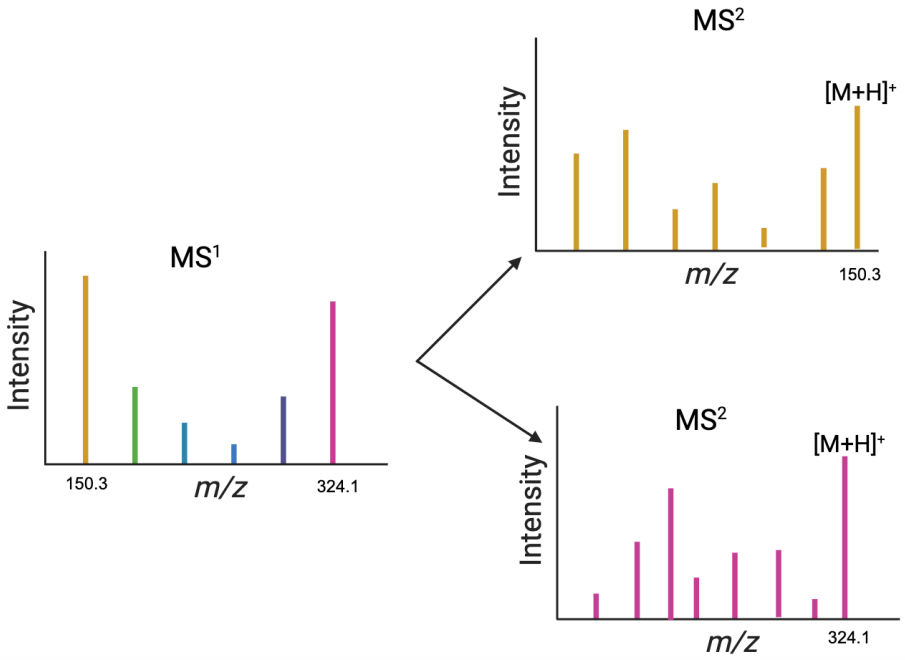

Figure 11:Descriptive Image of the General Principle of DDA. The MS does a full scan, generating an MS^1 of different precursor ions shown in different colours. It then automatically fragments the most abundant precursor ions (yellow and pink in MS^1) and generates the MS^2 spectra for those specific ions. Figure created using BioRender.

3.1.3 Untargeted Metabolomics Analysis for Biomarker Discovery of BC

The primary objective of this work is to discover early diagnostic metabolite biomarkers for BC. Since an untargeted analysis detects all m/z within a certain range, the data requires preprocessing, often using software such as MzMine3 or MSDial, so that true peaks can be identified. This can be done by specifying parameters including the mass detection range and

peak duration. Furthermore, all m/z found in the blank are removed from the samples data. The goal of preprocessing is to obtain a list of all peaks detected for further analysis, along with their m/z and RT. However, to analyze the data and make meaningful conclusions, the data needs to be filtered, to increase the quality.

Data filtering is a necessary step to data analysis, especially during an untargeted approach. Since untargeted analyses are often conducted without knowing the total amount of metabolites that should be present and the identity of those metabolites, data filtering can help remove duplicate metabolites or metabolites that were detected as false positive hits during the preprocessing stage.

Once a list of potential metabolite features has been created, the metabolites are then identified. Currently, this can be done using different identification software such as MS-Finder, MetFrag, and Sirius CSI:FingerID (Sirius) .[82] These *in silico* identification software rely on matching fragment peaks with the MS^2 housed in different databases in order to find the molecular structure of the compound and name it. Depending on the complexity of the fragmentation spectra, identification may also be done manually using the monoisotopic mass and known open-access databases such as the PubMed, Human Metabolite Database (HMDB) or Lipid Maps.[83]

Overall, this chapter will cover the analysis of metabolites using nLC-MS/MS, the data preprocessing using MzMine3, and the statistic and bioinformatic analysis often done in an untargeted analysis using R programming language.

3.2 Methods

3.2.1 Pre-Injection Metabolite Sample Preparation

9 Extracted metabolite samples were taken out of the -80°C freezer and thawed at room temperature for 10 minutes. The samples were then resuspended in 50μL of resuspension solvent comprised of 50% acetonitrile (ACN) and 50% MS-grade water with 1μg/mL of reserpine used as an internal standard. After samples were resuspended, they were then centrifuged at 21,000xg for 10 mins and filtered using 0.2-micron filter tips. The samples were then injected into the Thermo Fisher nano liquid chromatography (nLC) system coupled to the Thermo Q-Exactive quadrupole Orbitrap instrument.

3.2.2 LC-MS/MS Method

Metabolites were separated using an Easy Spray^{TM} C18 column (Thermo Fisher Scientific, Cat no. ES900) with dimensions of 150 mm x 75 μm. The separation was done using a gradient elution with a water and acetonitrile. Samples were loaded onto the column for 50 min at a flow rate of 250nL/min. Metabolites were separated using a gradient that starts out with 10% acetonitrile and 90% water. The gradient gradually increases in % acetonitrile and decreases in % water until 36minutes where it reaches 100% acetonitrile and 0% water. Separation continues using this ratio of solvents for 5 minutes until it gradually decreases to 0% acetonitrile and 100% water at 45 minutes before ending at 10% acetonitrile and 90% water to 50 minutes.

Eluted metabolites were directly sprayed into a mass spectrometer using positive electrospray ionization (ESI) at an ion source temperature of 250℃ and an ion spray voltage of 2.1 kV. The Full-scan MS spectra (m/z 80-1200) were acquired at a resolution of 70,000. Precursor ions were filtered according to the monoisotopic precursor selection, with charge state of +1 and +2, and

dynamic exclusion of 10s. The automatic gain control settings were $3x10^6$ for full FTMS scans and $1x10^5$ for MS/MS scans. Precursors were isolated using a 2 m/z isolation window and fragmented with a normalized high energy collision of 35%.

3.2.3 Data Pre-Processing Using MzMine

Each data file was uploaded onto the opensource MzMine3 software for preprocessing. All masses for each sample were detected using a 5.0E5 noise threshold limit for both MS^1 and MS^2 spectra. ADAP Chromatogram Builder, a function of MzMine3 was then used to peak pick based on the following settings: minimum highest intensity 5.0E5 and noise of 5.0E5 with 6 data points per peak. Once the chromatograms were created, the resolving function was then used to clean up the features and link the MS^1 and MS^2 spectra using "MS/MS scan pairing". The chromatogram threshold was set to 90% with a minimum search range RT of 0.03 and a minimum relative height of 50%. The ratio was 1.2 and the peak duration was set to 1 minute. Deisotoping was then performed where the retention time tolerance was set to 0.5 mins and the m/z tolerance was set to 5ppm. The samples were then aligned using a retention time of 0.5min and a m/z tolerance of 10ppm. The peaks that were in the blank were also subtracted from all samples using the default setting for data points and 150% for the fold increase meaning that true peaks were identified if the area was 150% more than the noise level .

3.2.4 Data Analysis Using R Programming

All data filtering and analysis was done using R programming, the pOmics package created by Nico Hüttman (https://github.com/nicohuttmann/pOmics) and the BiocManager package (https://bioconductor.github.io/BiocManager/). The .csv file containing all features obtained from MzMine3 was uploaded into RStudio. Thedata was normalized by taking the log2 of all

peak areas and then it was transposed and subsetted such that it contained only feature areas and feature IDs of all samples. The data was cleaned and separated into three separate dataframes consisting of candidates that contained 75% true values in MDA-MB-231 samples, candidates that contained 75% true values in MCF10A and candidates that contained 75% true values in all samples.

To determine the number of unique metabolites in MDA-MB-231, features had to satisfy conditions of being present in 75% of MDA-MB-231 samples and absent in 80% of MCF10A samples. This threshold provided the maximum amount of features for each cell line without compromising the quality of the data.To determine the number of unique features in MCF10A, the same conditions apply such that features that were present in 75% of MCF10A samples and absent in 80% of MDA-MB-231 samples were selected. This data was used to create a venn diagram, volcano plots and a PCA plot using R programming language and the aforementioned package functions. In addition to normalizing the data by scaling, the p-values were also corrected when constructing the Volcano Plot using the Benjimini Hotchberg method in order to lower the false discovery rate.

3.3 Results

3.3.1 Comparison of MDA-MB-231and MCF10A Cell Lines

Overall, upon comparing the exosome metabolites of highly metastatic MDA-MB-231 and non-metastatic MCF10A cell lines, it was found that while the cell lines share some similarities, there is an overall difference between them (Figure 12). This can be seen in the PCA plot, which is meant to visually illustrated the varience of the samples. In Figure 12 where every dot is representative of 1 metabolite sample, either in the MDA-MB-231 or MCF10A cell line.

Interestingly, upon further inspection of the metabolite areas, 132 metabolites were found to be common between the two cell lines (Figure 13). From those common metabolites, a total of 27 metabolites were found to be significantly dysregulated ($p<0.05$) (Figure 14). Furthermore, it was found that 10 features were significantly upregulated in MDA-MB-231 with a fold change (FC) ≥ 1.5 and 6 features being significantly downregulated in MDA-MB-231 with a FC$\leq$ -1.5. Additionally, there were 15 and 27 unique metabolites in MCF10A and MDA-MB-231, respectively (Figure 13). Lists of the unique metabolites' m/z and RT for each cell line can be found in Tables 2 and 3.

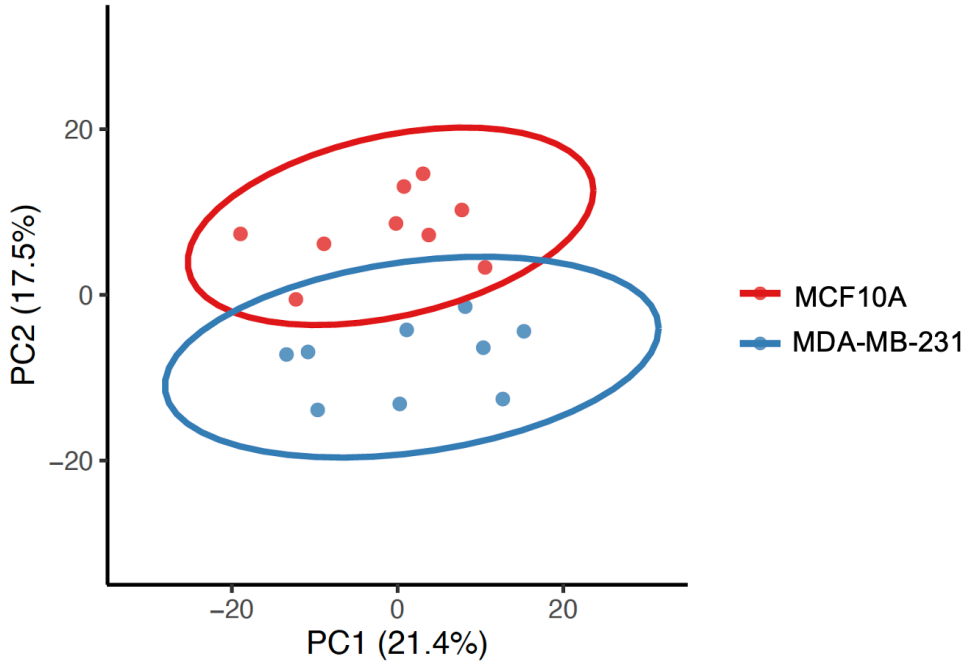

Figure 12: PCA Plot Comparing MCF10A and MDA-MB-231 Cell Lines. Each dot is representative of one sample where the red dots represent samples from the MCF10A cell line and the blue dots represent samples from the MDA-MB-231 cell line. Figure created using R programming.

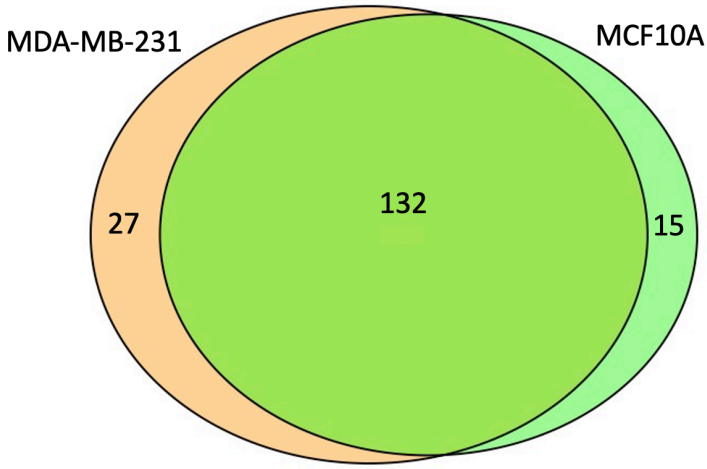

Figure 13. Venn diagram of unique and common metabolites between MDA-MB-231 and MCF10A cell lines. Figure created using RStudio.

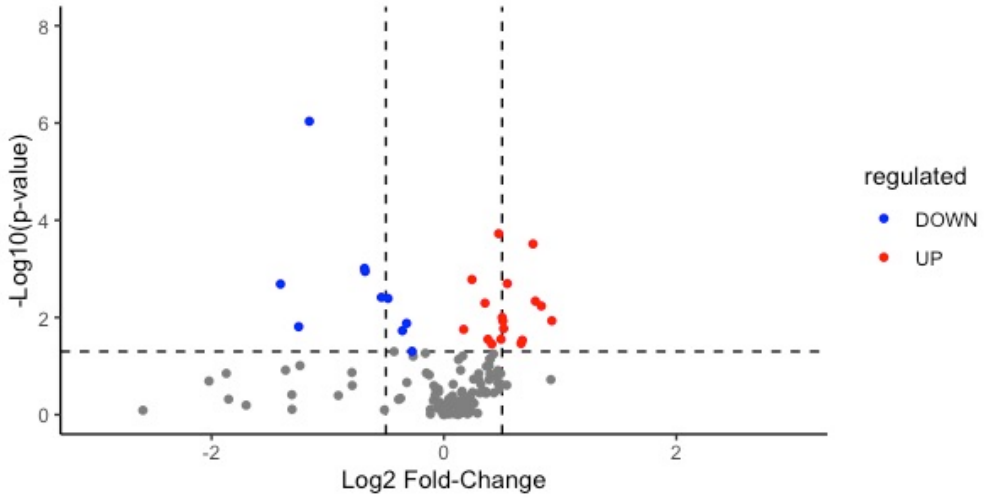

Figure 14: Volcano plot comparison of common features between MDA-MB-231 and MCF10A cell lines. Each dot represents a feature. Blue dots depict the features significantly downregulated in MDA MB-231 and red dots depict the features significantly upregulated in MDA-MB-231.X-axis represents the $\log_2$ FC between MDA-MB-231 and MCF10A and the Y-axis represents the $-\log_{10}$ of the p-value obtained from conducting student's t-tests. The dashed horizontal line represents a p-value of 0.05, with all data points above that line having $p<0.05$. The vertical dashed lines represent the threshold of a FC of 1.5. The diagram was made using R programming language and RStudio and size of each section was created in proportion to the size of each element.

Table 2:List of 27 identified metabolites with corresponding m/z and RT unique to the MDA-MB-231 cell line. MzMine3 preprocessing output provided each metabolite with a corresponding identification which is outlined in the table as the "Feature ID".

Feature ID	m/z	RT(min)
9297	182.1909	14.43
11677	353.2232	17.66
12415	367.239	18.59
12960	352.1335	19.23
13003	233.139	19.27
14252	208.0975	20.98
14307	247.1547	21.04
14348	220.1338	21.09
14399	453.7476	21.16
14615	453.7475	21.47
14785	220.1338	21.72
15454	234.1496	22.52
15731	218.1183	22.81
16029	192.1023	23.12
16226	234.1496	23.34
16926	232.1338	24.23
17635	204.1389	24.91
18907	361.2229	26.18
22001	493.3022	30.33

22334	373.1866	30.74
23050	381.2178	31.75
23446	285.0951	32.7
23516	526.2946	32.89
24092	285.0952	34.2
24162	413.2181	34.36
25006	409.2384	35.98
25229	409.2384	36.29

Table 3:List of 15 m/z and RT of features unique to the MCF10A cell line. MzMine3 preprocessing output provided each metabolite with a corresponding identification which is outlined in the table as the "Feature ID". Table was created using Microsoft Word.

Feature ID	m/z	Retention Time (min)
16248	274.2015	23.37
16478	257.2227	23.64
16635	218.1755	23.85
17607	272.1861	24.88
17739	249.1103	25.04
17781	244.155	25.08
19656	445.1865	27.03
20902	240.1962	28.63
21616	284.2225	29.75

21721	431.0958	29.95
22097	431.2072	30.45
22806	259.2009	31.35
23047	431.2073	31.74
23343	281.1389	32.44
26005	235.1697	37.62

3.3.2 Identification of Potential Biomarkers of BC

There were multiple rounds of metabolite identification throughout this project using two different methods: Sirius and Manual Database Searches. All identification using Sirius was conducted using the default databases available on the software, excluding any databases that pertained to plant metabolites. Since Sirius identification produces a list of matches for each m/z identified, each identification candidate in each list was additionally screened manually for biological relevance in humans. Overall, identification of MDA-MB-231 unique features using Sirius resulted in the identification of 10 metabolites out of 27 (Table 4). The identification of the other cancer cell specific features yielded only partial identification of the chemical structure, but they were not named and thus were not considered for further validation.

Table 4:Metabolite Identification of MDA-MB-231 specific features using Sirius. MF denotes the molecular formula of the biomarker candidate.

m/z	MF	Putative ID
353.2232	$C_{22}H_{28}N_2O_2$	**N,N''-bis(3-phenylpropyl) succinamide**
367.239	$C_{23}H_{30}N_2O_2$	**N,N" Dibenzyl-N,N"-diisopropyl-malonamide**

220.1338	$C_{13}H_{17}NO_2$	**1-Amino-4-phenylcyclohexanecarboxylic acid**
220.1338	$C_{13}H_{17}NO_2$	**Cyclohexyl 2-amino benzoate**
373.1866	$C_{18}H_{28}O8$	**(1S,2S, 3R, 4R, 5S, 6S)- 3,4 di(butanoxy)-2,5,6-trihydroxycyclohexyl butanoate**
285.0951	$C_{13}H_{16}O_7$	**p-Cresol Glucuronide**
285.0951	$C_{13}H_{16}O_7$	**6-O-Benzoyl D-Mannose**
413.2181	$C_{22}H_{28}N_4O_4$	**6-bis 2 anilo 2 oxoethyl amino-N-hydroxyhexanamide**
208.0975	$C_{11}H_{13}NO_3$	**Phenylpropionylglycine**
381.2178	$C_{23}H_{28}N_2O_3$	**Atalaphylline**

Metabolite identification was also done sans software using a combination of manual structural elucidation of the fragmentation mass spectra as well as a comparison of identification results obtained from the HMDB with the monoisotopic mass of a feature. Using this combinatory method, 3 potential biomarkers of BC, previously found in human samples, were identified. N-Acetyl-L-Phenylalanine was identified as the potential biomarker corresponding to m/z 208.0975, nona-decanol was identified as the potential biomarker corresponding to m/z 285.0941 and 6-O-Benzoyl-alpha-D-glucose was identified as the potential biomarker corresponding to m/z 285.0951(Figure 15).

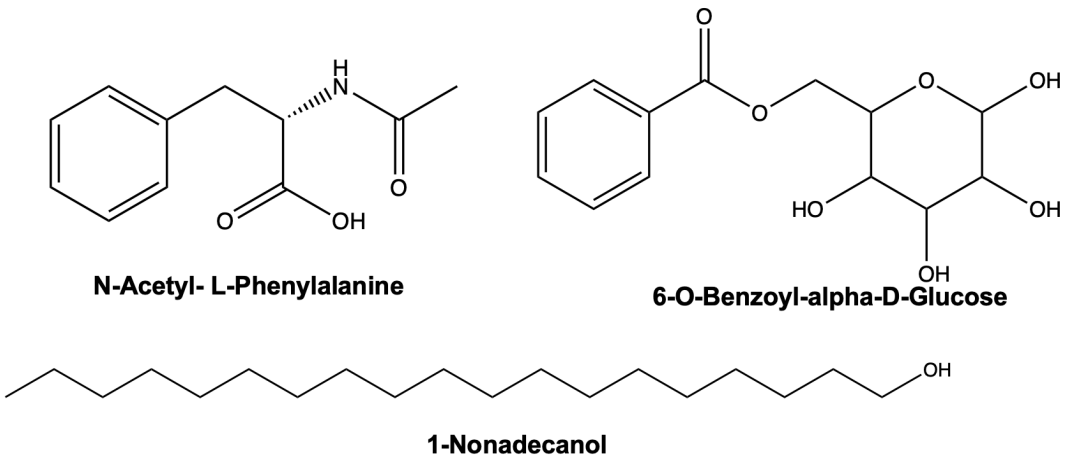

Figure 15:Chemical Structures of Potential Metabolite Biomarkers for BC.

Structures obtained using the HMDB database and manual comparison of the MS2 spectra and their biological relevance found in literature. Structures made using ChemDraw 20.0 software and Microsoft PowerPoint.

3.3.3 Identification of MCF10A specific metabolites.

MCF10A-specific metabolites were identified using both Sirius and manual analysis. 7 out of 15 metabolites were identified using software identification (Table 5). All metabolites identified using the Sirius were $[M+H]^+$ adducts except for the $[M+K]^+$ adduct of D-phenylalanyl-D-alanyl-D-valyl-D-alanine and the $[M+Na]^+$ adduct of 9-tridecanoic acid. Alternatively, manual identification using the HMDB resulted in 6 out of 15 metabolites identified with the majority being $[M+H]^+$ adducts (Table 5).

Table 5:Identification of MCF10A unique metabolites using Sirius and manual HMDB search. Table includes m/z, molecular formula (MF), putative ID, the adduct type and the identification method used. Table created using Microsoft Word.

m/z	MF	Putitive ID	Adduct	Identification Method
235.1697	C13H24O2 C15H22O2	9-tridecanoic acid 2-hydroxy-3-octyl-benzaldehyde	[M+Na] [M+H]	Sirius
240.1962	C14H25NO2	9-tetradecenamide-5-one	[M+H]	Sirius

244.155	C12H21NO4	2-Ethylacryloylcarnitine Tigylylcarnitine	[M+H]	HMDB
249.1103	C12H18O4	2,5,8,11-Dodecanetetrone	[M+H]	Sirius
249.1103	C12H18O4 C9H16N2O6	4-Phenylbutanoylcarnitine Glutamylthreonine	[M+H]	HMDB
257.2227	C12H26NO2	Heptanoylcholine	[M+ACN+H]	HMDB
259.2009	C13H26N2O3	3-[[(aminoacetyl)amino)methyl]-4,5-dimethyl octanoic acid	[M+H]	Sirius
272.1861	C14H25NO4	Heptenoylcarnitine	[M+H]	HMDB
274.2015	C14H27NO4	Heptanoylcarnitine	[M+H]	HMDB
274.2015	C14H27NO4	14-nitrotetradecanoic acid	[M+H]	Sirius
281.1389	C15H20O5	13-Hydroxyabscisic Acid	[M+H]	HMDB
284.2225	C16H29NO3	N-Dodecanoyl-DL-homoserine lactone	[M+H]	HMDB
284.2225	C16H29NO3	4-[methyl(undec-10-enoyl)amino]butanoate	[M+H]	Sirius
445.1865	C20H30N4O5	D-phenylalanyl-D-alanyl-D-valyl-D-alanine	[M+K]	Sirius

3.4Discussion

3.4.1 Discussion of MzMine3 Results as Viewed by Bioinformatic Analysis

Bioinformatic analysis with RStudio resulted in 27 features unique to MDA-MB-231, 15 features unique to MCF10A and 132 common features found in 75% of both cell lines (Figure 13). This result indicates that while there may be potential metabolites that are common with all exosomes, there are certain metabolite candidates that are unique to each cell line. Since the focus of this thesis was to determine novel biomarkers of BC, during data preprocessing, the parameters were refined by systematically changing each parameter until the maximum number of observations

were obtained. During the refinement process, the quality of the observations was also visually inspected. The parameter refinement was done until there was enough MDA-MB-231 unique features that could be used for identification. Additional optimization could increase the number of unique metabolites detected in both cell lines. However, it is important to note that compared to more established fields like proteomics or genomics, untargeted metabolomic data analysis is inherently subjective, with an overall lack of standardization in data analyses and metabolite annotation practices.[84]

Additionally, there were 10 features that were seen as significantly upregulated in MDA-MB-231 (Figure 14). For this analysis, a feature was deemed as being 'significant' when the p-value of < 0.05 was obtained with a student's t-test. Upregulation and downregulation was determined relative to MDA-MB-231 whereby if the feature exhibited a FC increase of more than 1.5 times, it was categorized as significantly upregulated and if it exhibited a FC decrease of more than 1.5 times it was categorized as significantly downregulated. The plot indicates that there may be 10 potential metabolites that are important in the regulation of metabolic processes in cancer cells. However, identification and validation of these metabolites are necessary in order to better understand the potential roles they may have in contributing to cancer metastasis and progression.

3.4.2 Comparison Between Software and Manual Identification

Throughout the metabolite identification process there were many discrepancies between software-assisted and manual metabolite identification. Sirius4 is an opensource specialized analytical tool that was created with the purpose of identifying compounds using its ability to elucidate chemical structure and molecular formulas.[85] In 2015, the creators of Sirius added the CSI:FingerID feature which utilizes machine learning and the presence of hypothetical

fragmentation patterns to analyze large sets of MS/MS data and obtain identifications efficiently. Using this analytical software, 10 metabolites out of a possible 27 were able to be identified with outputs consisting of MF, structure, and name. These three parameters were required to be identified for a complete identification. Overall, the structures of the software-identified metabolites were complex and for many of the biomarker candidates listed in Table 4, such as N,N''-bis(3-phenylpropyl) succinimide, conducting a literature search for instances of the biological presence did not yield many results. The lack of accurate identifications could potentially be attributed to human based metabolite databases being incomplete and continually updated. Additionally, updates include metabolites from a variety of sources including food sources and newly approved drugs and their metabolites.[86] Consequently, the large scope of these databases may increase the margin of error during the in-silico identification of software like Sirius. This would account for the many structurally complex metabolites that were identified during this analysis.

In addition to incorrect identifications, there were also many incomplete identifications whereby the software was unable to produce any result. This may be a result of the incomplete nature of the databases. In the future, with the discovery of new metabolites and the completion of the databases, identification may become easier and more accurate. However, currently, this remains a major pitfall in the field. One merit of using the software, however, was the efficiency that results could be obtained. With Sirius, the whole dataset was able to be identified in under an hour, with individual identifications able to be made in seconds.

In comparison, attempting to structurally elucidate the biomarker candidates and identify them using the opensource HMDB was much more time consuming, taking approximately 3-4 days to find a complete structure. However, the HMDB, while still incomplete, produced more viable

results, with smaller simpler compounds, such as N-acetyl-L-phenylalanine, that were found to be endogenous in humans during a literature search. The manual database searches required a comparison of the sample MS^2 to fragmentation spectra housed within the database. While there were many spectra obtained from true experiments, the fragmentation energy and mass analyzers that were used were different compared to the Q-Exactive Plus, resulting in multiple spectra of the same compound appearing very different. This was an additional challenge during identification. Furthermore, it is known that each metabolite has an optimum fragmentation energy. However, during untargeted MS-metabolomics, the optimum fragmentation energy requirement for each metabolite cannot be met, and thus some metabolites had spectra that were more informative than others, posing another challenge during metabolite identification.[87]

4 Metabolite Biomarker Validation

4.1 Introduction

The last chapter of this work focuses on validating the potential BC biomarkers that were found in Chapter 3. Since the primary focus of this project was to identify biomarkers unique to BC, only features that were found solely in the MDA-MB-231 cell line exosomes were examined for validation. In the field of metabolomics, the term validation means verifying that the metabolite that was identified is not a false positive result. This is commonly done by comparing the RT, and both MS^1 and MS^2 with those of a commercially available standard. Validation is an extremely important step because software that may help with metabolite identification can identify incorrectly. If this occurs, especially in the context of publication, it may lead to the spread of misinformation amongst the scientific community and potentially a decrease in

credibility. Therefore, validation is necessary to prevent this. However, validation may not be successful every time. This chapter will also cover the failed attempts at validation and furthermore, will also cover the targeted analysis of the successfully validated compound for the purposes of quantification.

4.2 Methods

4.2.1 Sample Preparation for Commercially Available Standards

All standards were obtained and prepared in the same manor, except for 1-nonadecanol, whose dilutions were created using MS-grade ACN instead of 1:1 ACN: H_2O. approximately 1mg of each standard, weighed out on an analytical balance was placed into an Eppendorf tube and solvated in 1mL of 1:1 ACN: H_2O. The initial molarity of the 'stock' was then found using the calculation present in equation 5 where C represents stock molarity in mol/L, m represents mass of the standard, M represents the molar mass of the compound in g/mol and V represents the volume in L. The stock was then diluted to a final concentration of 1µM.

(5)

$$C = \frac{(\frac{m}{M})}{V}$$

4.2.2 Validating Unique Features with m/z and Retention Time

The validation workflow was conducted according to Figure 16. To validate the true presence of the unique feature found using MzMine3 and RStudio, as described in Chapter 3, the m/z values and the RT of each feature, unique to MDA-MB-231 cell line was extracted from the total ion chromatogram using the XCalibur Software (Thermo Fisher Scientific) at a mass accuracy of

10ppm. The presence of the m/z and the accuracy of the RT was evaluated visually. Then a visual comparison was made between the MCF10A and MDA-MB-231 metabolite samples by comparing the extracted ion chromatograms (EIC) of each feature. A unique MDA-MB-231 feature was confirmed by the presence of a peak in the MDA-MB-231 sample and the absence of a peak in the MCF10A samples.

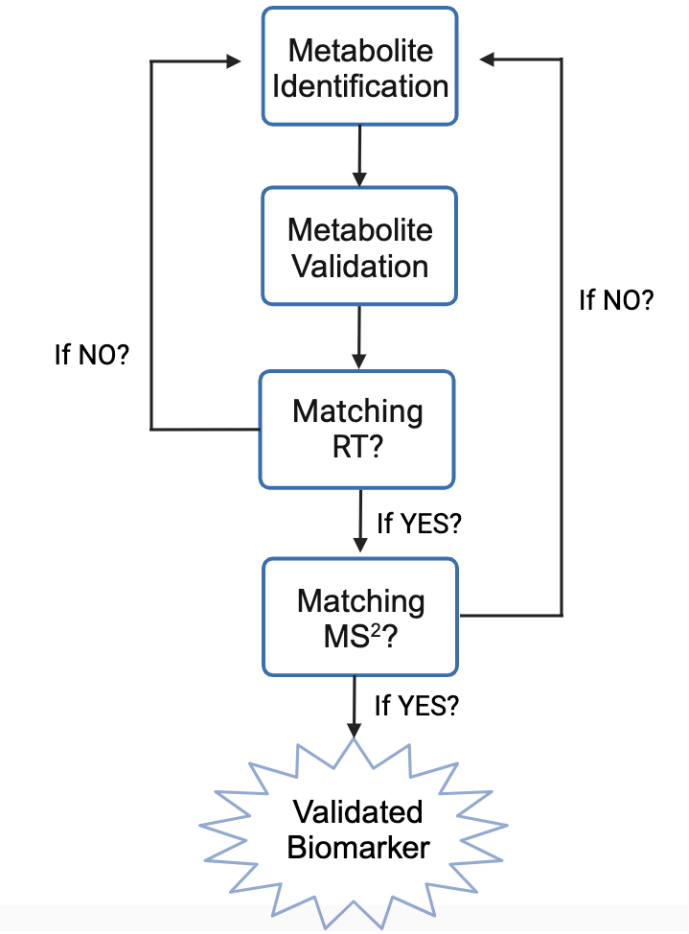

Figure 16:BC Biomarker Validation Workflow. Flowchart illustrating the steps that went into the metabolite biomarker validation. Figure created using BioRender.

4.2.3 Validating Identified Metabolites by Retention Time, m/z, and Fragmentation Pattern

1µM of each commercially available standard was analyzed along with an MCF10A and MDA-MB-231 metabolite sample. The biological metabolite samples were prepared as described in 2.2.3, and 3.2.1-3.2.2. All samples were analyzed as described in 3.2.2. After MS analysis was completed, the EIC and the fragmentation spectra were visually compared between the commercial standard and the two biological samples in order to validate identification.

4.3 Results

4.3.1 Validation of MDA Unique Metabolites Based on RT and m/z

To validate the existence and the uniqueness of the 27 MDA-MB-231 unique metabolites, the presence of these metabolites was checked using XCalibur (Thermo-Fisher Scientific). The EICs of these metabolites were extracted within a mass accuracy of 10 ppm. Upon the comparison of each EIC with m/z corresponding to Table 2 it was found that all 26 metabolites were present with RT corresponding to those presented in Table 2 and 25 of those were unique to the MDA-MB-231 cell line (Supplementary Information). Figure 16 depicts a representative EIC comparison of m/z 413.2181 at RT 34.36. It plots time in minutes on the x-axis and relative abundance of the ion on the y-axis. A clear peak is seen at 34.36minutes. This is followed by the absence of a peak in the MCF10A cell line exosome metabolite sample. This comparison was done for all MDA-MB-231 unique metabolites.

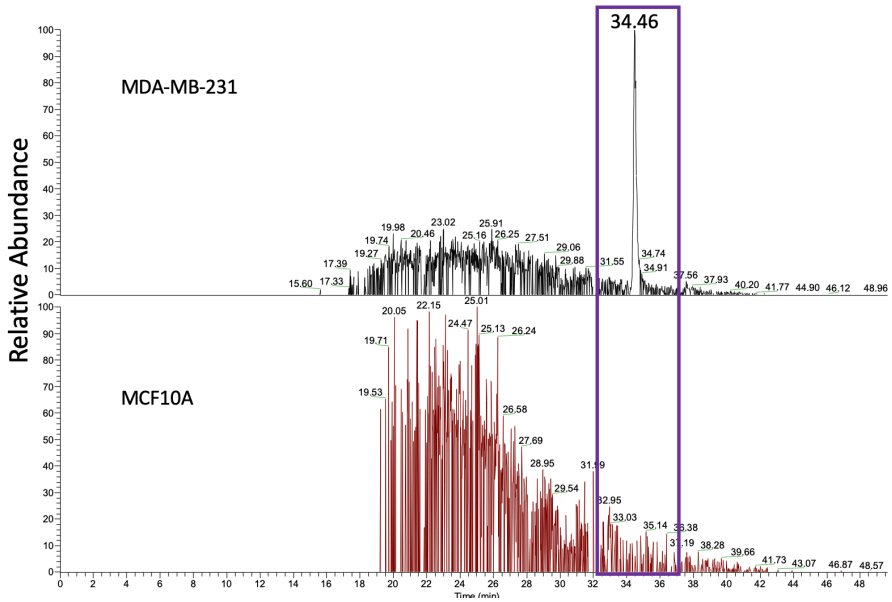

Figure 17:Representative EIC of m/z 413.2181. This figure illustrates the presence of m/z **413.2181** in MDA-MB-231 and its absence in MCF10A as a visual verification of the uniqueness of the BC biomarker candidates.

4.3.2 Validation of Select Sirius4 Identified Metabolites with nLC-MS/MS

4.3.2.1 First Round of Validation.

During the first round of validation, 4 metabolites were analyzed against a commercially available standard for validation. The first was N,N''-bis(3-phenylpropyl) succinimide with m/z 353.2232 and RT 17.66min. The second was N,N" Dibenzyl-N,N"-diisopropyl-malonamide with m/z 367.239 at RT 18.59min, the third was 1-Amino-4-phenylcyclohexanecarboxylic acid with m/z 220.1338 at RT 21.09 and the fourth was p-Cresol Glucuronide with m/z 285.0951 and RT 32.7 min.

When comparing RT of 1µM of N, N''-bis(3-phenylpropyl) succinimide, it was found that the RT of the standard compound was 29.81 min, 12 minutes longer compared to the RT of the identified metabolite (Figure 17). This discrepancy in RT indicates that the identification of m/z 353.2232 and RT 17.66min was incorrect.

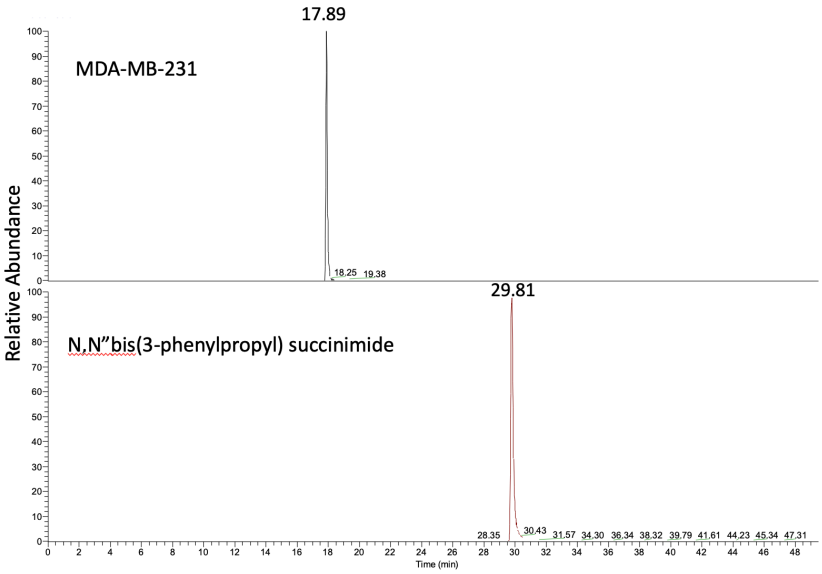

Figure 18:Stacked EIC of N, N''-bis(3-phenylpropyl) succinimide and MDA-MB-231samples with m/z 353.2232.EICs obtained from XCalibur Software and formatted using Microsoft Powerpoint.

When comparing RT of 1μM of N,N" Dibenzyl-N,N"-diisopropyl-malonamide, it was found that the RT of the standard compound was 33.35 min, 15 minutes longer compared to MDA-MB-231 sample(Figure 18). This discrepancy in RT indicates that the identification of m/z 367.239 at RT 18.59min was incorrect.

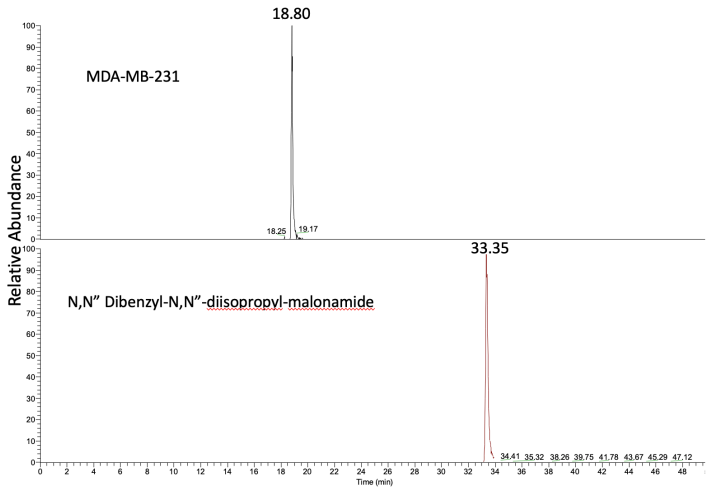

Figure 19:Stacked EIC of N,N" Dibenzyl-N,N"-diisopropyl-malonamide and MDA-MB-231 samples with m/z 353.2232.EICs obtained from XCalibur Software and formatted using Microsoft Powerpoint.

When comparing RT of 1µM of 1-Amino-4-phenylcyclohexanecarboxylic acid, it was found that the RT of the standard compound was 16.40 min, 5 minutes shorter compared to the RT of what was identified as 1-Amino-4-phenylcyclohexanecarboxylic acid (Figure 19). This discrepancy in RT indicates that the identification of m/z 367.239 at RT 18.59min was incorrect.

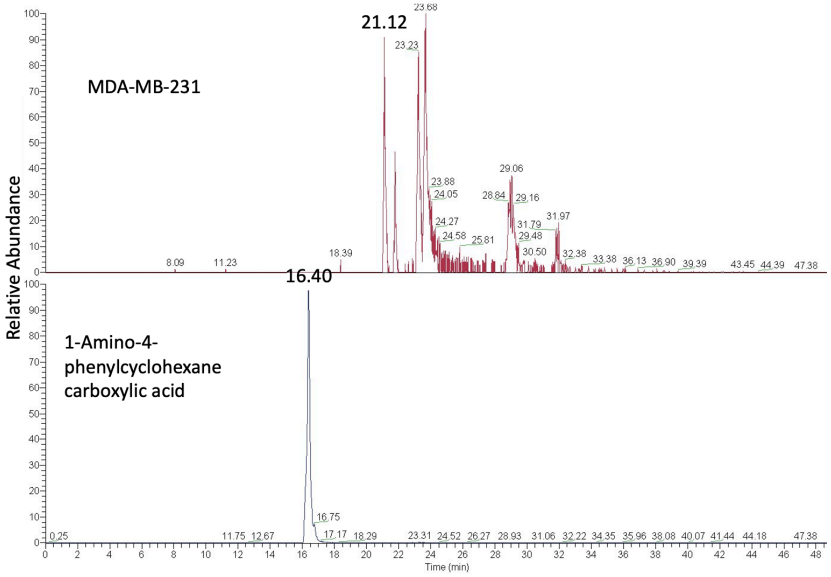

Figure 20:Stacked EIC of 1-Amino-4-phenylcyclohexanecarboxylic acid and MDA-MB-231 samples with m/z 367.239.EICs obtained from XCalibur Software and formatted using Microsoft Powerpoint.

When validating 1µM p-Cresol Glucuronide it was found that there was only a peak at m/z 285.0951 in MDA-MB-231 but not in the standard (Figure 21). As such, this compound was deemed to be incorrectly identified.

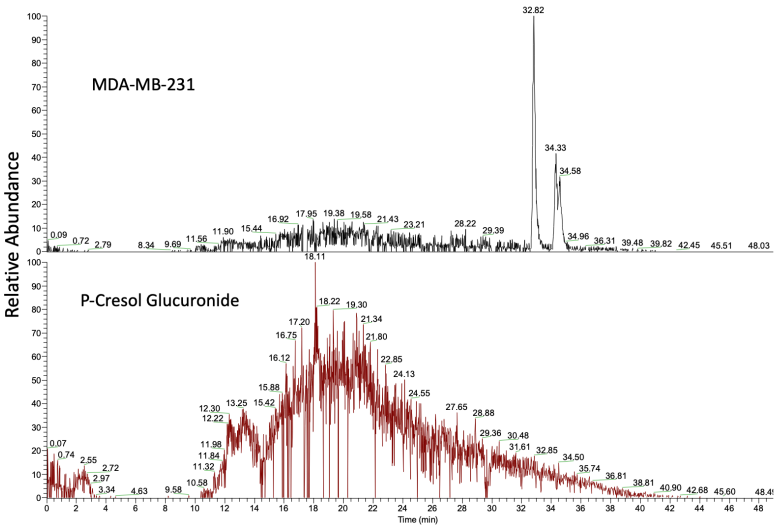

Figure 21:Stacked EIC of p-Cresol Glucuronide, MDA-MB-231 and MCF10A samples with m/z 285.0951.EICs obtained from XCalibur Software and formatted using Microsoft Powerpoint.

4.3.2.2 Second Round of Validation.

During the second round of validation, metabolites, now identified manually using database searches and manual structural elucidation from the MS data, were run against commercially obtained standards in order to confirm identification. For m/z 285.0951 and RT 32.7 min, 1-nonadecanol was identified as a possible candidate. Comparing the RT of 1-nonadecanol also yielded a negative identification result. The RT of the standard was greater than the total duration of the run time. This was concluded based on part of the peak being visible at 50 minutes but not the whole peak being visible. Due to the variation in RT between the standard and the metabolite samples, the result was a false identification.

The last compound that was validated was N-acetyl-L-Phenylalanine at m/z 208.0975. Compared with a 1μM standard, it was found that the MS^1 and the RT of the sample and the standard were comparable with the RT in MDA-MB-231 at 19.42 min and 19.40min, respectively (Figure 22A and 22B). Furthermore, since the RT were comparable, the MS^2 spectra between the standard

and the samples were also comparable as well (Figure 23). Spectral comparison shows a total of 7 matching peaks at m/z 120.08, 123.08, 151.07, 162.09, 166.08, 191.06, and 209.07 at 10ppm mass accuracy. Additionally, the figure also depicts the structure of potential fragments, obtained from the HMDB database. The compatibility of both the RT, the MS^1 and MS^2 spectra demonstrates that m/z 208.0975 is N-acetyl-L-Phenylalanine, a potential TNBC diagnostic biomarker.

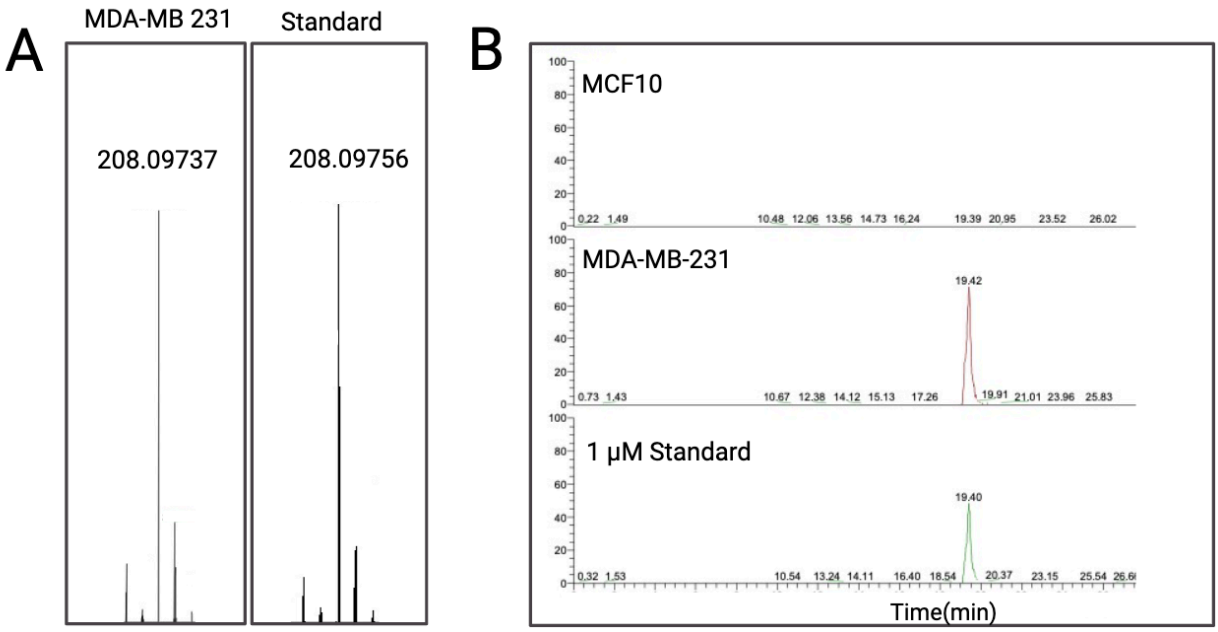

Figure 22: Validating Presence of N-Acetyl-L-Phenylalanine. A) MS^1 comparison between MDA-MB-231 derived exosome metabolite and the N-acetyl-L-Phenylalanine standard. B) Stacked EIC of N-acetyl-Phenylalanine, MDA-MB-231 and MCF10A samples with m/z 208.0975. EICs obtained from XCalibur Software and formatted using Biorender.

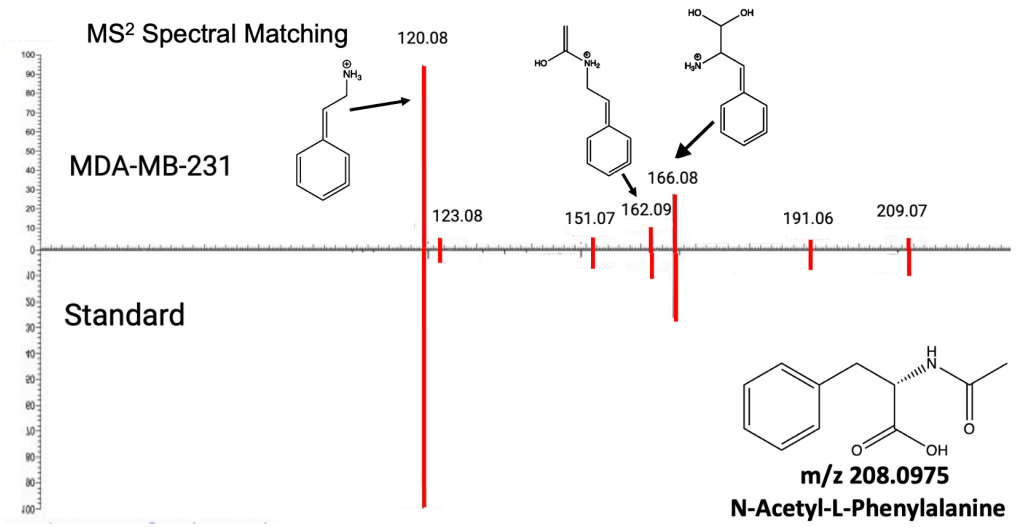

Figure 23:MS2 spectral comparison of N-Acetyl-Phenylalanine and m/z 208.0975.m/z highlighted in red signifies the matching m/z between the two spectra. Figure created using BioRender.

4.3.3 Absolute Quantification of N-Acetyl-L-Phenylalanine

N-acetyl-L-Phenylalanine is a derivative of the amino acid phenylalanine. In order to perform absolute quantification of this metabolite, a calibration curve was created based on the concentrations showed in Table 6 from 10-2000nM (Figure 24). The N-acetyl-L-Phenylalanine calibration curve showed good linearity with a R^2 value of 0.9948. Before the concentrations were calculated, peak integrations were done and are presented in Table S1with an average peak area of approx. 3.55E8 $\pm$ 3.12E7 for 9 biological replicates. Using these peak areas and the calibration curve, calculations indicated that samples contained an average of approximately 406.6 $\pm$ 21 nM. .

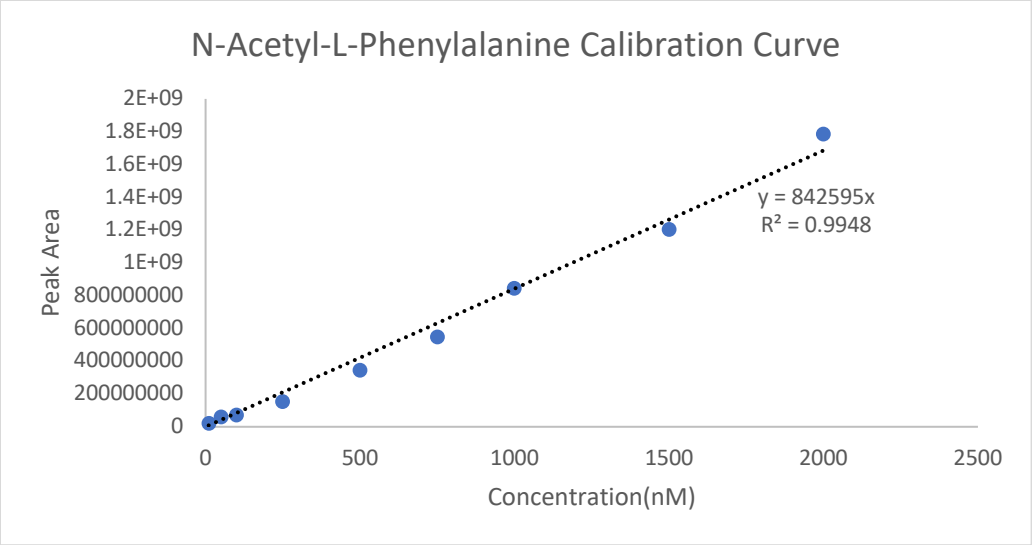

Figure 24:Calibration curve of N-Acetyl-L-Phenylalanine. Calibration curve created based on peak areas taken from concentrations of 10-2000nM with an R^2 value was 0.9948.

Table 6:Table of N-Acetyl-L-Phenylalanine Concentrations created for the calibration curve along with the corresponding peak areas and RT. Table created using Microsoft Word.

Concentration(nM)	Peak Area	Retention Time (min)
10.00	2.057E7	19.50
50.00	6.015E7	19.48
100.0	7.253E7	19.41
250.0	1.529E8	19.44
500.0	3.453E8	19.50
750.0	5.497E8	19.53
1.000E3	8.435E8	19.45
1.500E3	1.205E9	19.52
2.000E3	1.786E9	19.52

4.4 Discussion

4.4.1 Validation of MDA Unique Metabolites Based on Retention Time and m/z Values for MS^1 and MS^2 Pattern

The first step of validating a potential BC metabolite biomarker was to verify the true presence of the MDA-MB-231 unique metabolites that were found using bioinformatic analysis in RStudio. For this validation step, Thermo Fisher XCalibur software was used. This software

allows for the visualization of both chromatograms and mass spectra. As such, the EIC of each unique m/z was analyzed to account for true presence and uniqueness in MDA-MB-231 upon comparing the results of both cell lines. An example of a positive result would look like a peak at the selected m/z, solely in MDA-MB-231, and the RT, corresponding to the results obtained in RStudio and the absence of a peak in the same RT range in the corresponding MCF10A samples. The verification that all but one unique MDA-MB-231 peaks were found during this validation step indicates that both the mzMine3 data pre-processing and the RStudio data-cleanup was optimized to a point, capable of producing valid metabolites, that, when identified, could act as potential BC diagnostic biomarkers. However, it is worth noting that optimization can always be improved. It is just a matter of whether the current method produces meaningful results or not that determined, in this project, the extent of optimization.

4.4.2 Validation of Select Sirius- Identified Metabolites with nLC-MS/MS

4.4.2.1 The Significance of Biomarker validation in Metabolomics Research

The validation of metabolite identifications is a long and arduous process. However, it is required to provide the necessary rigor to the scientific experimentation and support the omics results. In the field of untargeted metabolomics, there are many publications that limit their metabolomics process to putative identifications, often done by in-house databases and software.[88–90] While this may provide some confidence in the initial result, it cannot be a substitution for validation against a commercially available standard. For example, *Zhu et. al* conducted an untargeted GC-MS analysis for biomarker discovery of early colorectal cancer.[90] While they chose metabolites based on rigorous statistical criteria, their identification was solely putative based on manual MS^2 spectral matching with NIST, Metlin, HMDB and

MetaboAnalyst. While this would be acceptable for a preliminary identification, additional validation experiments are necessary due to the possibility of incorrect identifications, as seen in this thesis. Furthermore, it would be difficult to draw meaningful conclusions of the biological role and relevance of the metabolite biomarker based on putative identification measures. This process can also be seen in BC research in a publication by *Wei et. al*. whereby for identification, the authors matched the mass of each ion against masses found on in-instrument Agilent databases.[89] Consequently, if multiple metabolites have the same or similar masses or isomeric structures, the resulting identification may be incorrect. These putative identifications are then often used for pathway analyses whereby their implications are expanded upon and published. The lack of overall metabolite identity validation in the metabolomics community is the main reason why validation was emphasized so highly in this work. As seen in sections 4.3.1 and 4.3.2, the use of software and spectral matching can result in incorrect identifications, ultimately creating a gap in knowledge.

4.4.2.2 First Round of Validation

In the first round of identification, metabolites were identified solely using Sirius.. However, after comparing both RT and MS^2 of m/z 353.2232, m/z 367.239, m/z 220.1338 and 285.0951 with the N,N’’-bis(3-phenylpropyl) succinimide, N,N” Dibenzyl-N,N”-diisopropyl-malonamide,1-Amino-4-phenylcyclohexanecarboxylic acid and p-Cresol Glucuronide standards respectively, it was found that all four of these metabolites were falsely identified. It was initially hypothesized that this may be due to having too much variety in the databases that were chosen during the identification. These databases included those such as PubChem and PubMed, which are databases that primarily cater to medicine and biochemical compounds. As a result, there

were instances where potential identifications included metabolites that were halogenated, which cannot exist endogenously in the human body. During identification, the Sirius software outputs a list of potential identifications for the features. During the first round of identification, only those suggestions that were complete, identified by a chemical structure and a name were considered. However, in retrospect, only considering 'complete' identifications may have resulted in excluding potentially correct identifications that just showed chemical structure. Additionally, the overall complexity and structural variety of metabolites poses another challenge when using identification software.[87]

4.4.2.3 Second Round of Validation.

During the second round of validation, a manual approach was utilized, mainly looking through the HMDB. This database was used specifically because, while still incomplete, it is comprised a plethora of metabolites that pertain to humans and information regarding experimental and theoretical fragmentation mass spectra which can aid in metabolite identification. By utilizing this refined software and manual search combination, the structures found during this second round of identification were smaller and simpler structures compare to those found in the first round of identification. Furthermore, there was more literature on these compounds being found in humans that justified purchasing them for the second round of identification. After the unsuccessful identification of 1-nonadecanol, N-acetyl-L-Phenylalanine, a derivative of the amino acid phenylalanine, was successfully identified as a potential BC biomarker due to the identical m/z, RT and fragmentation patterns between the standard and m/z 208.0975. While one metabolite was successfully validated, there are currently many challenges, inherent to

metabolomics that may have contributed to limiting the number of successfully validated biomarker candidates. One of the main challenges of metabolomic analysis is the limited commercially availability of standards. [91]

4.4.3 Potential Role of N-Acetyl-L-Phenylalanine in Cancer

Although few articles have been written in reference to the link between N-acetyl-L-phenylalanine and cancer, there is evidence of this compound acting as a diagnostic biomarker of oral cancers. In 2014, Wang *et. al* found that N-acetyl-L-Phenylalanine could be used to differentiate between patients with squamous cell carcinoma and healthy controls. They found that N-acetyl-L-Phenylalanine was downregulated in the cancer patient saliva compared to the healthy controls which suggested that there may be a dysregulation in phenylalanine metabolism.[92] Interestingly, however, it was found that in the serum of patients with Clear Cell Renal Cell Carcinoma, N-acetyl-L-Phenylalanine was upregulated compared to the healthy control.[93] While this finding may contrast the 2014 biomarker discovery by Wang *et. al*, it does corroborate the finding of this project somewhat.[92,93] Additionally, during quantification of N-acetyl-L-Phenylalanine, it was found that there was an average of 406.6 nm $\pm$ 21nM. In order to contextualize this finding, a comparative literature analysis was done. However, as this is currently a minimally studied metabolite, there was no absolute quantitative analysis done on N-acetyl-L-Phenylalanine in the literature, only qualitative. Hence, this is one of the first absolute concentration measures taken of the metabolite obtained from MDA-MB-231 cell line- derived exosomes.

Furthermore, the presence of N-acetyl-L-phenylalanine in TNBC cell line exosomes potentially indicates that there is either a dysregulation in Phenylalanine-N-acetyltransferase, the enzyme

responsible for acetylating phenylalanine, or a dysregulation in phenylalanine hydroxylase which is the enzyme that is responsible for transforming phenylalanine into tyrosine.

4.5 Conclusion and Future Work

In conclusion, using an untargeted metabolomic approach, we were able to discover and validate a novel diagnostic biomarker for the early detection of BC, which was the main goal of this work. This is the first time that N-acetyl-Phenylalanine has been named as a potential BC metabolic biomarker in exosomes and validated against a commercially available standard. While an absolute quantification of this metabolite was able to be done via a calibration curve, absolute quantification of related metabolites, phenylalanine, and tyrosine, the other two main metabolites in the phenylalanine metabolic pathway, was not able to be conducted potentially due to limitations of the method that was used. Both Phenylalanine and Tyrosine were too hydrophilic to be separated and detected using a C18 column. Hydrophilic Interaction Liquid Chromatography (HILIC) could be used in the future to separate and detect amino acids.[94] Being able to quantify and compare Phenylalanine, Tyrosine and N-acetyl-L-Phenylalanine could provide insight on the relationship that these three metabolites share and how cancer metabolism can affect that relationship. A more comprehensive picture can also be provided by conducting a targeted analysis of the metabolites in whole cell extracts, specifically seeing if there are any changes in metabolites involved in the phenylalanine and tyrosine metabolism pathway and comparing it to the results obtained from the analysis of the exosomal metabolites. Furthermore, mass spectrometry analyses were done using positive ion mode. To further extend these untargeted analyses, negative ion mode mass detection could be performed. This would broaden the amount of potential BC diagnostic biomarkers that may be detected. Negative mode analysis

may also provide greater insight into more lipids, which may be more easily detected due to the deprotonation of the hydroxyl group.

Additionally, in order to get a better understanding of how this work could potentially be translated to diagnostic studies, another study could be conducted using blood-based exosome metabolites. Redoing this analysis on blood-based exosome metabolites may provide additional insight, especially on the potential of N-acetyl-L-Phenylalanine to be applied as a true diagnostic biomarker for early BC detection using liquid biopsy, a less invasive technique than tumor biopsy. As seen through this work, exosomes can be a powerful source for metabolite biomarkers for BC. The discovery of reliable BC diagnostic biomarkers from exosomes, combined with the use of liquid biopsy, may increase the accuracy of early BC diagnoses during screening, improving overall patient outcomes.

References

(1) Health, N. I. of. Understanding Cancer. In *Biological Science Curriculum Study*; National Institutes of Health (US): Bethesda, MD, USA, 2007.
(2) Weinstein, I. B.; Case, K. The History of Cancer Research: Introducing an AACR Centennial Series. *Cancer Res* **2008**, *68* (17), 6861–6862. https://doi.org/10.1158/0008-5472.CAN-08-2827.
(3) Maruvka, Y. E.; Tang, M.; Michor, F. On the Validity of Using Increases in 5-Year Survival Rates to Measure Success in the Fight against Cancer. *PLoS One* **2014**, *9* (7), e83100. https://doi.org/10.1371/JOURNAL.PONE.0083100.
(4) Giaquinto, A. N.; Sung, H.; Miller, K. D.; Kramer, J. L.; Newman, L. A.; Minihan, A.; Jemal, A.; Siegel, R. L. Breast Cancer Statistics, 2022. *CA Cancer J Clin* **2022**, *72* (6), 524–541. https://doi.org/10.3322/CAAC.21754.
(5) Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **2021**, *71* (3), 209–249. https://doi.org/10.3322/CAAC.21660.
(6) Arnold, M.; Morgan, E.; Rumgay, H.; Mafra, A.; Singh, D.; Laversanne, M.; Vignat, J.; Gralow, J. R.; Cardoso, F.; Siesling, S.; Soerjomataram, I. Current and Future Burden of Breast Cancer: Global Statistics for 2020 and 2040. *The Breast* **2022**, *66*, 15–23. https://doi.org/10.1016/J.BREAST.2022.08.010.
(7) Watkins, E. J. Overview of Breast Cancer. *J Am Acad Physician Assist* **2019**, *32* (10), 13–17. https://doi.org/10.1097/01.JAA.0000580524.95733.3D.
(8) Velloso, F. J.; Bianco, A. F. R.; Farias, J. O.; Torres, N. E. C.; Ferruzo, P. Y. M.; Anschau, V.; Jesus-Ferreira, H. C.; Chang, T. H. T.; Sogayar, M. C.; Zerbini, L. F.; Correa, R. G. The Crossroads of Breast Cancer Progression: Insights into the Modulation of Major Signaling Pathways. *Onco Targets Ther* **2017**, *10*, 5491–5524. https://doi.org/10.2147/OTT.S142154.
(9) Orrantia-Borunda, E.; Anchondo-Nuñez, P.; Acuña-Aguilar, L. E.; Gómez-Valles, F. O.; Ramírez-Valdespino, C. A. Subtypes of Breast Cancer. In *Breast Cancer*; Mayrovitz, H. N., Ed.; Exon Publications: Brisbane (AU), 2022; pp 31–42. https://doi.org/10.36255/EXON-PUBLICATIONS-BREAST-CANCER-SUBTYPES.
(10) Yin, L.; Duan, J. J.; Bian, X. W.; Yu, S. C. Triple-Negative Breast Cancer Molecular Subtyping and Treatment Progress. *Breast Cancer Research* **2020**, *22* (61), 1–13. https://doi.org/10.1186/S13058-020-01296-5/TABLES/3.
(11) Sun, Y. S.; Zhao, Z.; Yang, Z. N.; Xu, F.; Lu, H. J.; Zhu, Z. Y.; Shi, W.; Jiang, J.; Yao, P. P.; Zhu, H. P. Risk Factors and Preventions of Breast Cancer. *Int J Biol Sci* **2017**, *13* (11), 1387–1397. https://doi.org/10.7150/IJBS.21635.
(12) Li, J.; Guan, X.; Fan, Z.; Ching, L. M.; Li, Y.; Wang, X.; Cao, W. M.; Liu, D. X. Non-Invasive Biomarkers for Early Detection of Breast Cancer. *Cancers (Basel)* **2020**, *12* (10), 1–28. https://doi.org/10.3390/CANCERS12102767.
(13) Care, T. C. T. F. on P. H. Recommendations on Screening for Breast Cancer in Average-Risk Women Aged 40-74 Years. *CMAJ. Canadian Medical Association Journal* **2011**, *183* (17), 1991–2001. https://doi.org/10.1503/CMAJ.110334/-/DC1.
(14) Ma, Y.; Liu, G.; Du, M.; Stayton, I. Recent Developments in the Determination of Urinary Cancer Biomarkers by Capillary Electrophoresis. *Electrophoresis* **2004**, *25* (10–11), 1473–1484. https://doi.org/10.1002/ELPS.200405895.

(15) Fujita, K.; Nonomura, N. Urinary Biomarkers of Prostate Cancer. *International Journal of Urology* **2018**, *25* (9), 770–779. https://doi.org/10.1111/IJU.13734.
(16) Marrugo-Ramírez, J.; Mir, M.; Samitier, J. Blood-Based Cancer Biomarkers in Liquid Biopsy: A Promising Non-Invasive Alternative to Tissue Biopsy. *International Journal of Molecular Sciences 2018, Vol. 19, Page 2877* **2018**, *19* (10), 2877. https://doi.org/10.3390/IJMS19102877.
(17) Manash, P. K. Introductory Chapter: Role of Extracellular Vesicles in Human Diseases and Therapy. In *Extracellular Vesicles - Role in Diseases, Pathogenesis and Therapy*; Manash, P. K., Brzozowski, T., Eds.; IntechOpen: Los Angeles (USA), 2022; pp 1–6. https://doi.org/10.5772/intechopen.94822.
(18) Gurung, S.; Perocheau, D.; Touramanidou, L.; Baruteau, J. The Exosome Journey: From Biogenesis to Uptake and Intracellular Signalling. *Cell Communication and Signaling 2021 19:1* **2021**, *19* (1), 1–19. https://doi.org/10.1186/S12964-021-00730-1.
(19) Panfoli, I.; Santucci, L.; Bruschi, M.; Petretto, A.; Calzia, D.; Ramenghi, L. A.; Ghiggeri, G.; Candiano, G. Microvesicles as Promising Biological Tools for Diagnosis and Therapy. *Expert Rev Proteomics* **2018**, *15* (10), 801–808. https://doi.org/10.1080/14789450.2018.1528149.
(20) Logozzi, M.; Mizzoni, D.; Di Raimo, R.; Fais, S. Exosomes: A Source for New and Old Biomarkers in Cancer. *Cancers (Basel)* **2020**, *12* (9), 1–18. https://doi.org/10.3390/CANCERS12092566.
(21) Zhang, M.; Xiao, B.; Wang, H.; Han, M. K.; Zhang, Z.; Viennois, E.; Xu, C.; Merlin, D. Edible Ginger-Derived Nano-Lipids Loaded with Doxorubicin as a Novel Drug-Delivery Approach for Colon Cancer Therapy. *Molecular Therapy* **2016**, *24* (10), 1783–1796. https://doi.org/10.1038/MT.2016.159.
(22) Kim, M. S.; Haney, M. J.; Zhao, Y.; Mahajan, V.; Deygen, I.; Klyachko, N. L.; Inskoe, E.; Piroyan, A.; Sokolsky, M.; Okolie, O.; Hingtgen, S. D.; Kabanov, A. V.; Batrakova, E. V. Development of Exosome-Encapsulated Paclitaxel to Overcome MDR in Cancer Cells. *Nanomedicine* **2016**, *12* (3), 655–664. https://doi.org/10.1016/J.NANO.2015.10.012.
(23) Saari, H.; Lázaro-Ibáñez, E.; Viitala, T.; Vuorimaa-Laukkanen, E.; Siljander, P.; Yliperttula, M. Microvesicle- and Exosome-Mediated Drug Delivery Enhances the Cytotoxicity of Paclitaxel in Autologous Prostate Cancer Cells. *Journal of Controlled Release* **2015**, *220*, 727–737. https://doi.org/10.1016/J.JCONREL.2015.09.031.
(24) Du, R.; Wang, C.; Zhu, L.; Yang, Y. Extracellular Vesicles as Delivery Vehicles for Therapeutic Nucleic Acids in Cancer Gene Therapy: Progress and Challenges. *Pharmaceutics* **2022**, *14* (10). https://doi.org/10.3390/PHARMACEUTICS14102236.
(25) Štok, U.; Čučnik, S.; Sodin-Šemrl, S.; Žigon, P.; Štok, U.; Čučnik, S.; Sodin-Šemrl, S.; Žigon, P. Extracellular Vesicles: Intercellular Communication Mediators in Antiphospholipid Syndrome. *Antiphospholipid Syndrome - Recent Advances in Clinical and Basic Aspects* **2021**. https://doi.org/10.5772/INTECHOPEN.97412.
(26) Battistelli, M.; Falcieri, E. Apoptotic Bodies: Particular Extracellular Vesicles Involved in Intercellular Communication. *Biology (Basel)* **2020**, *9* (1), 21. https://doi.org/10.3390/BIOLOGY9010021.
(27) Elmore, S. Apoptosis: A Review of Programmed Cell Death. *Toxicol Pathol* **2007**, *35* (4), 495–516. https://doi.org/10.1080/01926230701320337.

(28) Kerr, J. F. R.; Wyllie, A. H.; Currie, A. R. Apoptosis: A Basic Biological Phenomenon with Wide-Ranging Implications in Tissue Kinetics. *Br J Cancer* **1972**, *26* (4), 239. https://doi.org/10.1038/BJC.1972.33.
(29) Papaliagkas, V.; Anogianaki, A.; Anogianakis, G.; Ilonidis, G. The Proteins and the Mechanisms of Apoptosis: A Mini-Review of the Fundamentals. *Hippokratia* **2007**, *11* (3), 108–113.
(30) Porter, A. G.; Jänicke, R. U. Emerging Roles of Caspase-3 in Apoptosis. *Cell Death & Differentiation 1999 6:2* **1999**, *6* (2), 99–104. https://doi.org/10.1038/sj.cdd.4400476.
(31) Tricarico, C.; Clancy, J.; D'Souza-Schorey, C. Biology and Biogenesis of Shed Microvesicles. *Small GTPases* **2017**, *8* (4), 220. https://doi.org/10.1080/21541248.2016.1215283.
(32) Van Niel, G.; D'Angelo, G.; Raposo, G. Shedding Light on the Cell Biology of Extracellular Vesicles. *Nature Reviews Molecular Cell Biology 2018 19:4* **2018**, *19* (4), 213–228. https://doi.org/10.1038/nrm.2017.125.
(33) Laude, A. J.; Prior, I. A. Plasma Membrane Microdomains: Organisation, Function and Trafficking. *Mol Membr Biol* **2004**, *21* (3), 193. https://doi.org/10.1080/09687680410001700517.
(34) Choi, D. S.; Park, J. O.; Jang, S. C.; Yoon, Y. J.; Jung, J. W.; Choi, D. Y.; Kim, J. W.; Kang, J. S.; Park, J.; Hwang, D.; Lee, K. H.; Park, S. H.; Kim, Y. K.; Desiderio, D. M.; Kim, K. P.; Gho, Y. S. Proteomic Analysis of Microvesicles Derived from Human Colorectal Cancer Ascites. *Proteomics* **2011**, *11* (13), 2745–2751. https://doi.org/10.1002/PMIC.201100022.
(35) Akers, J. C.; Gonda, D.; Kim, R.; Carter, B. S.; Chen, C. C. Biogenesis of Extracellular Vesicles (EV): Exosomes, Microvesicles, Retrovirus-like Vesicles, and Apoptotic Bodies. *J Neurooncol* **2013**, *113* (1), 1–11. https://doi.org/10.1007/S11060-013-1084-8/TABLES/1.
(36) Pan, B. T.; Teng, K.; Wu, C.; Adam, M.; Johnstone, R. M. Electron Microscopic Evidence for Externalization of the Transferrin Receptor in Vesicular Form in Sheep Reticulocytes. *Journal of Cell Biology* **1985**, *101* (3), 942–948. https://doi.org/10.1083/jcb.101.3.942.
(37) Maia, J.; Caja, S.; Strano Moraes, M. C.; Couto, N.; Costa-Silva, B. Exosome-Based Cell-Cell Communication in the Tumor Microenvironment. *Front Cell Dev Biol* **2018**, *6* (FEB), 18. https://doi.org/10.3389/FCELL.2018.00018.
(38) Zhou, W.; Fong, M. Y.; Min, Y.; Somlo, G.; Liu, L.; Palomares, M. R.; Yu, Y.; Chow, A.; O'Connor, S. T. F.; Chin, A. R.; Yen, Y.; Wang, Y.; Marcusson, E. G.; Chu, P.; Wu, J.; Wu, X.; Li, A. X.; Li, Z.; Gao, H.; Ren, X.; Boldin, M. P.; Lin, P. C.; Wang, S. E. Cancer-Secreted MiR-105 Destroys Vascular Endothelial Barriers to Promote Metastasis. *Cancer Cell* **2014**, *25* (4), 501–515. https://doi.org/10.1016/J.CCR.2014.03.007.
(39) Zhang, P.; Zhou, H.; Lu, K.; Lu, Y.; Wang, Y.; Feng, T. Exosome-Mediated Delivery of MALATI Induces Cell Proliferation in Breast Cancer. *Onco Targets Ther* **2018**, *11*, 291–299. https://doi.org/10.2147/OTT.S155134.
(40) Guo, L.; Zhu, Y.; Li, L.; Zhou, S.; Yin, G.; Yu, G.; Cui, H. Breast Cancer Cell-Derived Exosomal MiR-20a-5p Promotes the Proliferation and Differentiation of Osteoclasts by Targeting SRCIN1. *Cancer Med* **2019**, *8* (12), 5687–5701. https://doi.org/10.1002/CAM4.2454.

(41) Micheel, C. M.; Nass, S. J.; Omenn, G. S.; Trials, C. on the R. of O.-B. T. for P. P. O.; Clinical, I.; Services, B. on H. C.; Policy, B. on H. S.; Medicine, I. of. Omics-Based Clinical Discovery: Science, Technology, and Applications. In *Evolution of Translational Omics: Lessons Learned and the Path Forward*; Micheel, C. M., Nass, S. J., Omenn, G. S., Eds.; National Academies Press (US): Washington, DC, 2012.

(42) Hasin, Y.; Seldin, M.; Lusis, A. Multi-Omics Approaches to Disease. *Genome Biology 2017 18:1* **2017**, *18* (1), 1–15. https://doi.org/10.1186/S13059-017-1215-1.

(43) Berger, M. F.; Mardis, E. R. The Emerging Clinical Relevance of Genomics in Cancer Medicine. *Nature Reviews Clinical Oncology 2018 15:6* **2018**, *15* (6), 353–365. https://doi.org/10.1038/s41571-018-0002-6.

(44) Bertucci, F.; Ng, C. K. Y.; Patsouris, A.; Droin, N.; Piscuoglio, S.; Carbuccia, N.; Soria, J. C.; Dien, A. T.; Adnani, Y.; Kamal, M.; Garnier, S.; Meurice, G.; Jimenez, M.; Dogan, S.; Verret, B.; Chaffanet, M.; Bachelot, T.; Campone, M.; Lefeuvre, C.; Bonnefoi, H.; Dalenc, F.; Jacquet, A.; De Filippo, M. R.; Babbar, N.; Birnbaum, D.; Filleron, T.; Le Tourneau, C.; André, F. Genomic Characterization of Metastatic Breast Cancers. *Nature 2019 569:7757* **2019**, *569* (7757), 560–564. https://doi.org/10.1038/s41586-019-1056-z.

(45) Barrón-Gallardo, C. A.; Garcia-Chagollán, M.; Morán-Mendoza, A. J.; Delgadillo-Cristerna, R.; Martínez-Silva, M. G.; Aguilar-Lemarroy, A.; Jave-Suárez, L. F. Transcriptomic Analysis of Breast Cancer Patients Sensitive and Resistant to Chemotherapy: Looking for Overall Survival and Drug Resistance Biomarkers. *Technol Cancer Res Treat* **2022**, *21*, 1–14. https://doi.org/10.1177/15330338211068965.

(46) Yoosuf, N.; Navarro, J. F.; Salmén, F.; Ståhl, P. L.; Daub, C. O. Identification and Transfer of Spatial Transcriptomics Signatures for Cancer Diagnosis. *Breast Cancer Research* **2020**, *22* (1), 1–10. https://doi.org/10.1186/S13058-019-1242-9/FIGURES/4.

(47) Matikas, A.; Zerdes, I.; Lövrot, J.; Richard, F.; Sotiriou, C.; Bergh, J.; Valachis, A.; Foukakis, T. Prognostic Implications of PD-L1 Expression in Breast Cancer: Systematic Review and Meta-Analysis of Immunohistochemistry and Pooled Analysis of Transcriptomic Data. *Clinical Cancer Research* **2019**, *25* (18), 5717–5726. https://doi.org/10.1158/1078-0432.CCR-19-1131/75408/AM/PROGNOSTIC-IMPLICATIONS-OF-PD-L1-EXPRESSION-IN.

(48) Chan, P. P.; Wasinger, V. C.; Leong, R. W. Current Application of Proteomics in Biomarker Discovery for Inflammatory Bowel Disease. *World J Gastrointest Pathophysiol* **2016**, *7* (1), 27. https://doi.org/10.4291/WJGP.V7.I1.27.

(49) Minic, Z.; Hüttmann, N.; Poolsup, S.; Li, Y.; Susevski, V.; Zaripov, E.; Berezovski, M. V. Phosphoproteomic Analysis of Breast Cancer-Derived Small Extracellular Vesicles Reveals Disease-Specific Phosphorylated Enzymes. *Biomedicines* **2022**, *10* (2), 408. https://doi.org/10.3390/BIOMEDICINES10020408/S1.

(50) Wang, D.; Yin, L.; Wei, J.; Yang, Z.; Jiang, G. ATP Citrate Lyase Is Increased in Human Breast Cancer, Depletion of Which Promotes Apoptosis. *Tumor Biology* **2017**, *39* (4). https://doi.org/10.1177/1010428317698338/ASSET/IMAGES/LARGE/10.1177_1010428317698338-FIG6.JPEG.

(51) Yang, Y.; Zhang, M.; Wang, Y. The Roles of Histone Modifications in Tumorigenesis and Associated Inhibitors in Cancer Therapy. *Journal of the National Cancer Center* **2022**, *2* (4), 277–290. https://doi.org/10.1016/J.JNCC.2022.09.002.

(52) Yang, J.; Song, C.; Zhan, X. The Role of Protein Acetylation in Carcinogenesis and Targeted Drug Discovery. *Front Endocrinol (Lausanne)* **2022**, *13*, 972312. https://doi.org/10.3389/FENDO.2022.972312/BIBTEX.
(53) Minic, Z.; Li, Y.; Hüttmann, N.; Uppal, G. K.; D'Mello, R.; Berezovski, M. V. Lysine Acetylome of Breast Cancer-Derived Small Extracellular Vesicles Reveals Specific Acetylation Patterns for Metabolic Enzymes. *Biomedicines* **2023**, *11* (4), 1076. https://doi.org/10.3390/BIOMEDICINES11041076/S1.
(54) Wang, L.; Zhang, S.; Wang, X. The Metabolic Mechanisms of Breast Cancer Metastasis. *Front Oncol* **2020**, *10*. https://doi.org/10.3389/FONC.2020.602416.
(55) Zhou, B.; Xiao, J. F.; Tuli, L.; Ressom, H. W. LC-MS-Based Metabolomics. *Mol Biosyst* **2012**, *8* (2), 470. https://doi.org/10.1039/C1MB05350G.
(56) Contreras, A. V.; Cocom-Chan, B.; Hernandez-Montes, G.; Portillo-Bobadilla, T.; Resendis-Antonio, O. Host-Microbiome Interaction and Cancer: Potential Application in Precision Medicine. *Front Physiol* **2016**, *7* (DEC). https://doi.org/10.3389/FPHYS.2016.00606.
(57) Coskun, O. Separation Techniques: Chromatography. *North Clin Istanb* **2016**, *3* (2), 156. https://doi.org/10.14744/NCI.2016.32757.
(58) Van Deemter, J. J.; Zuiderweg, F. J.; Klinkenberg, A. Longitudinal Diffusion and Resistance to Mass Transfer as Causes of Nonideality in Chromatography. *Chem Eng Sci* **1995**, *50* (24), 3869–3882. https://doi.org/10.1016/0009-2509(96)81813-6.
(59) Katz, E.; Ogan, K. L.; Scott, R. P. W. Peak Dispersion and Mobile Phase Velocity in Liquid Chromatography: The Pertinent Relationship for Porous Silica. *J Chromatogr A* **1983**, *270* (C), 51–75. https://doi.org/10.1016/S0021-9673(01)96351-4.
(60) Wilson, S. R.; Vehus, T.; Berg, H. S.; Lundanes, E. Nano-LC in Proteomics: Recent Advances and Approaches. *http://dx.doi.org.proxy.bib.uottawa.ca/10.4155/bio.15.92* **2015**, *7* (14), 1799–1815. https://doi.org/10.4155/BIO.15.92.
(61) Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *J Am Soc Mass Spectrom* **2016**, *27* (12), 1897–1905. https://doi.org/10.1007/S13361-016-1469-Y/ASSET/IMAGES/LARGE/JS8B05178_0004.JPEG.
(62) Griffiths, W. J.; Koal, T.; Wang, Y.; Kohl, M.; Enot, D. P.; Deigner, H. P. Targeted Metabolomics for Biomarker Discovery. *Angewandte Chemie - International Edition* **2010**, *49* (32), 5426–5445. https://doi.org/10.1002/anie.200905579.
(63) Fitzgerald, R. C.; Antoniou, A. C.; Fruk, L.; Rosenfeld, N. The Future of Early Cancer Detection. *Nature Medicine 2022 28:4* **2022**, *28* (4), 666–677. https://doi.org/10.1038/s41591-022-01746-x.
(64) Akbar, A.; Malekian, F.; Baghban, N.; Kodam, S. P.; Ullah, M. Methodologies to Isolate and Purify Clinical Grade Extracellular Vesicles for Medical Applications. *Cells* **2022**, *11* (2). https://doi.org/10.3390/CELLS11020186.
(65) Doyle, L. M.; Wang, Z. Overview of Extracellular Vesicles, Their Origin, Composition, Purpose, and Methods for Exosome Isolation and Analysis. *Cells* **2019**, *8* (7), 727. https://doi.org/10.3390/cells8070727.
(66) Risha, Y.; Minic, Z.; Ghobadloo, S. M.; Berezovski, M. V. The Proteomic Analysis of Breast Cell Line Exosomes Reveals Disease Patterns and Potential Biomarkers. *Scientific Reports 2020 10:1* **2020**, *10* (1), 1–12. https://doi.org/10.1038/s41598-020-70393-4.

(67) Gardiner, C.; Vizio, D. Di; Sahoo, S.; Théry, C.; Witwer, K. W.; Wauben, M.; Hill, A. F. Techniques Used for the Isolation and Characterization of Extracellular Vesicles: Results of a Worldwide Survey. *J Extracell Vesicles* **2016**, *5* (1). https://doi.org/10.3402/JEV.V5.32945.

(68) Schachman, H. *Ultracentrifugation in Biochemistry*, 3rd ed.; Academic Press: New York City, 1959.

(69) Ohlendieck, K.; Harding, S. E. Centrifugation and Ultracentrifugation. In *Wilson and Walker's Principles and Techniques of Biochemistry and Molecular Biology*; Hofmann, A., Clokie, S., Eds.; Cambridge University Press: Cambridge, 2018; pp 424–453. https://doi.org/10.1017/9781316677056.014.

(70) Jeppesen, D. K.; Hvam, M. L.; Primdahl-Bengtson, B.; Boysen, A. T.; Whitehead, B.; Dyrskjøt, L.; Ørntoft, T. F.; Howard, K. A.; Ostenfeld, M. S. Comparative Analysis of Discrete Exosome Fractions Obtained by Differential Centrifugation. *J Extracell Vesicles* **2014**, *3* (1). https://doi.org/10.3402/JEV.V3.25011/SUPPL_FILE/ZJEV_A_11815512_SM0001.PDF.

(71) Filipe, V.; Hawe, A.; Jiskoot, W. Critical Evaluation of Nanoparticle Tracking Analysis (NTA) by NanoSight for the Measurement of Nanoparticles and Protein Aggregates. *Pharm Res* **2010**, *27* (5), 796–810. https://doi.org/10.1007/S11095-010-0073-2/TABLES/4.

(72) Stetefeld, J.; McKenna, S. A.; Patel, T. R. Dynamic Light Scattering: A Practical Guide and Applications in Biomedical Sciences. *Biophys Rev* **2016**, *8* (4), 409. https://doi.org/10.1007/S12551-016-0218-6.

(73) Pang, B.; Zhu, Y.; Ni, J.; Ruan, J.; Thompson, J.; Malouf, D.; Bucci, J.; Graham, P.; Li, Y. Quality Assessment and Comparison of Plasma-Derived Extracellular Vesicles Separated by Three Commercial Kits for Prostate Cancer Diagnosis. *Int J Nanomedicine* **2020**, *15*, 10241. https://doi.org/10.2147/IJN.S283106.

(74) Risha, Y.; Susevski, V.; Hüttmann, N.; Poolsup, S.; Minic, Z.; Berezovski, M. V. Breast Cancer-Derived Microvesicles Are the Source of Functional Metabolic Enzymes as Potential Targets for Cancer Therapy. *Biomedicines 2021, Vol. 9, Page 107* **2021**, *9* (2), 107. https://doi.org/10.3390/BIOMEDICINES9020107.

(75) Dimik, M.; Abeysinghe, P.; Logan, J.; Mitchell, M. The Exosome: A Review of Current Therapeutic Roles and Capabilities in Human Reproduction. *Drug Deliv Transl Res* **2023**, *13* (2), 473–502. https://doi.org/10.1007/s13346-022-01225-3.

(76) Heiles, S. Advanced Tandem Mass Spectrometry in Metabolomics and Lipidomics—Methods and Applications. *Anal Bioanal Chem* **2021**, *413* (24), 5927. https://doi.org/10.1007/S00216-021-03425-1.

(77) Yinon, J. Tandem Mass Spectrometry (MS/MS) and Collision Induced Dissociation (CID) — an Introduction. *Chemistry and Physics of Energetic Materials* **1990**, 685–693. https://doi.org/10.1007/978-94-009-2035-4_30.

(78) *Q Exactive Plus Hybrid Quadrupole Orbitrap Mass Spectrometer - Creative Proteomics*. https://www.creative-proteomics.com/support/q-exactive-plus-hybrid-quadrupole-orbitrap-mass-spectrometer.htm (accessed 2023-09-05).

(79) *Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer*. https://www.thermofisher.com/order/catalog/product/IQLAAEGAAPFALGMBDK (accessed 2023-09-05).

(80) Michalski, A.; Damoc, E.; Hauschild, J. P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass Spectrometry-Based Proteomics Using Q Exactive, a High-Performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Molecular and Cellular Proteomics* **2011**, *10* (9), M111.011015. https://doi.org/10.1074/mcp.M111.011015.

(81) Guo, J.; Huan, T. Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography-Mass Spectrometry Based Untargeted Metabolomics. *Anal Chem* **2020**, *92* (12), 8072–8080. https://doi.org/10.1021/ACS.ANALCHEM.9B05135/ASSET/IMAGES/LARGE/AC9B05135_0006.JPEG.

(82) Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites 2018, Vol. 8, Page 31* **2018**, *8* (2), 31. https://doi.org/10.3390/METABO8020031.

(83) Díaz-Beltrán, L.; González-Olmedo, C.; Luque-Caro, N.; Díaz, C.; Martín-Blázquez, A.; Fernández-Navarro, M.; Ortega-Granados, A. L.; Gálvez-Montosa, F.; Vicente, F.; Del Palacio, J. P.; Sánchez-Rovira, P. Human Plasma Metabolomics for Biomarker Discovery: Targeting the Molecular Subtypes in Breast Cancer. *Cancers (Basel)* **2021**, *13* (1), 1–18. https://doi.org/10.3390/cancers13010147.

(84) Kumar, A.; Misra, B. B. Challenges and Opportunities in Cancer Metabolomics. *Proteomics* **2019**, *19* (21–22), 1900042. https://doi.org/10.1002/PMIC.201900042.

(85) Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. SIRIUS 4: A Rapid Tool for Turning Tandem Mass Spectra into Metabolite Structure Information. *Nature Methods 2019 16:4* **2019**, *16* (4), 299–302. https://doi.org/10.1038/s41592-019-0344-8.

(86) Wishart, D. S.; Guo, A. C.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B. L.; Berjanskii, M.; Mah, R.; Yamamoto, M.; Jovel, J.; Torres-Calzada, C.; Hiebert-Giesbrecht, M.; Lui, V. W.; Varshavi, D.; Varshavi, D.; Allen, D.; Arndt, D.; Khetarpal, N.; Sivakumaran, A.; Harford, K.; Sanford, S.; Yee, K.; Cao, X.; Budinski, Z.; Liigand, J.; Zhang, L.; Zheng, J.; Mandal, R.; Karu, N.; Dambrova, M.; Schiöth, H. B.; Greiner, R.; Gautam, V. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res* **2022**, *50* (D1), D622–D631. https://doi.org/10.1093/NAR/GKAB1062.

(87) Chaleckis, R.; Meister, I.; Zhang, P.; Wheelock, C. E. Challenges, Progress and Promises of Metabolite Annotation for LC–MS-Based Metabolomics. *Curr Opin Biotechnol* **2019**, *55*, 44–50. https://doi.org/10.1016/J.COPBIO.2018.07.010.

(88) Yang, T.; Hui, R.; Nouws, J.; Sauler, M.; Zeng, T.; Wu, Q. Untargeted Metabolomics Analysis of Esophageal Squamous Cell Cancer Progression. *J Transl Med* **2022**, *20* (1), 1–11. https://doi.org/10.1186/S12967-022-03311-Z/FIGURES/4.

(89) Wei, Y.; Jasbi, P.; Shi, X.; Turner, C.; Hrovat, J.; Liu, L.; Rabena, Y.; Porter, P.; Gu, H. Early Breast Cancer Detection Using Untargeted and Targeted Metabolomics. *J Proteome Res* **2021**, *20* (6), 3124–3133. https://doi.org/10.1021/ACS.JPROTEOME.1C00019/ASSET/IMAGES/LARGE/PR1C00019_0005.JPEG.

(90) Zhu, G.; Wang, Y.; Wang, W.; Shang, F.; Pei, B.; Zhao, Y.; Kong, D.; Fan, Z. Untargeted GC-MS-Based Metabolomics for Early Detection of Colorectal Cancer. *Front Oncol* **2021**, *11*, 729512. https://doi.org/10.3389/FONC.2021.729512/BIBTEX.

(91) Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *J Am Soc Mass Spectrom* **2016**, *27* (12), 1897–1905. https://doi.org/10.1007/S13361-016-1469-Y/ASSET/IMAGES/LARGE/JS8B05178_0004.JPEG.

(92) Wang, Q.; Gao, P.; Wang, X.; Duan, Y. The Early Diagnosis and Monitoring of Squamous Cell Carcinoma via Saliva Metabolomics. *Scientific Reports 2014 4:1* **2014**, *4* (1), 1–9. https://doi.org/10.1038/srep06802.

(93) Knott, M. E.; Manzi, M.; Zabalegui, N.; Salazar, M. O.; Puricelli, L. I.; Monge, M. E. Metabolic Footprinting of a Clear Cell Renal Cell Carcinoma in Vitro Model for Human Kidney Cancer Detection. *J Proteome Res* **2018**, *17* (11), 3877–3888. https://doi.org/10.1021/ACS.JPROTEOME.8B00538/SUPPL_FILE/PR8B00538_SI_002.XLSX.

(94) Tsochatzis, E.; Papageorgiou, M.; Kalogiannis, S. Validation of a HILIC UHPLC-MS/MS Method for Amino Acid Profiling in Triticum Species Wheat Flours. *Foods* **2019**, *8* (10). https://doi.org/10.3390/FOODS8100514.

Supplementary Information

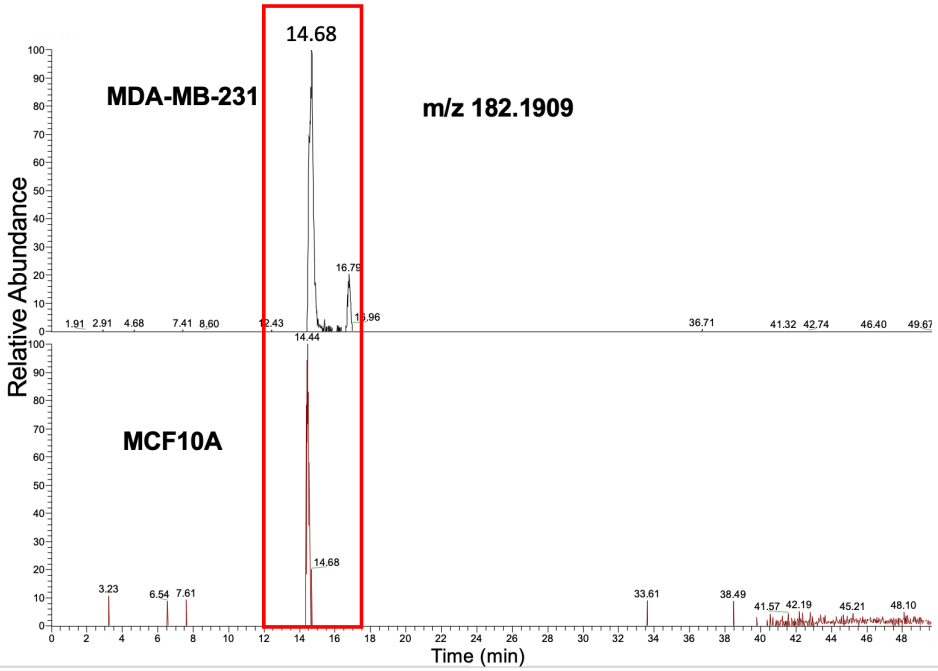

Figure S1: m/z 182.1909 Presence Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite sample verifying the presence of m/z 182.1909 at RT 14.68 obtained from statistical and bioinformatic analysis using RStudio.

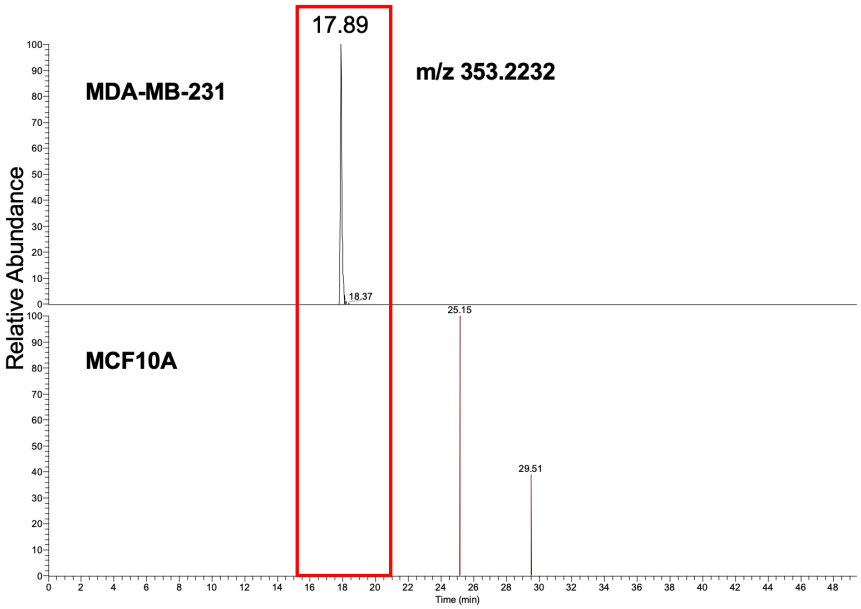

Figure S2:m/z 353.2232 Presence and Uniqueness Verification.EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 353.2232 and RT 17.89min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

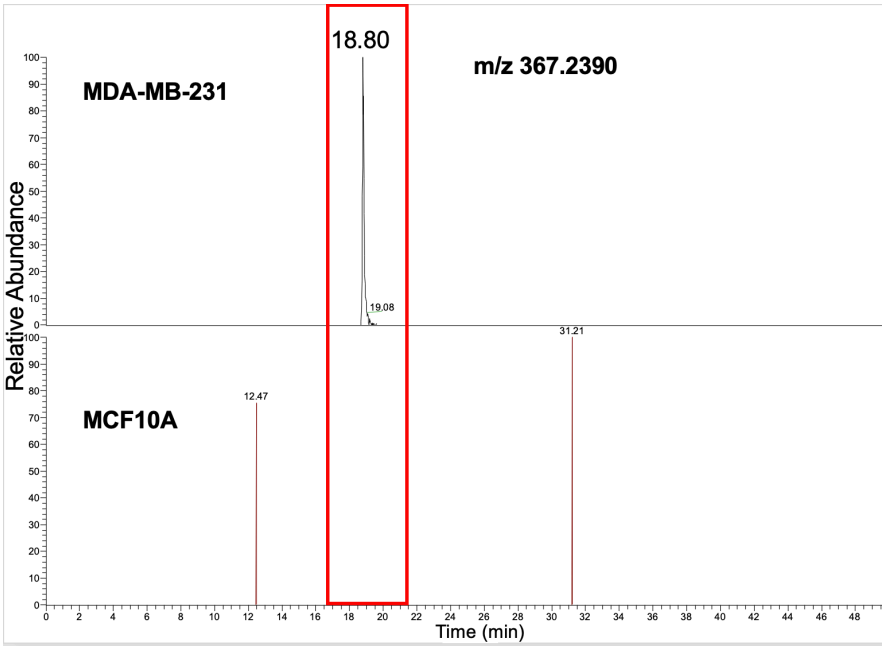

Figure S3:m/z 367.2390 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 367.2390 and RT 18.80min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line. This figure verifies the presence of m/z 353.2232 obtained from statistical and bioinformatic analysis using RStudio.

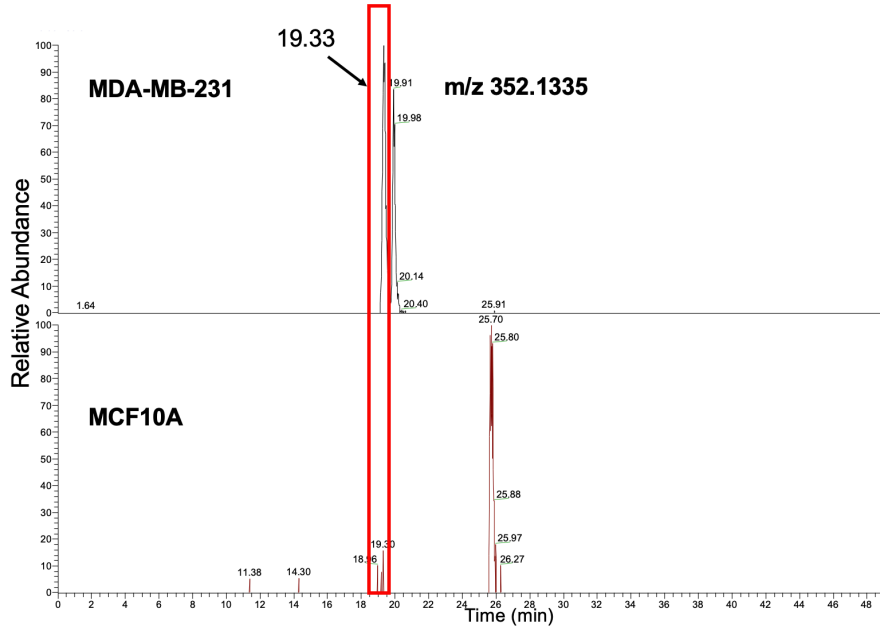

Figure S4: m/z 353.1335 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 352.1335 and RT 19.33min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line. This figure verifies the presence of m/z 352.1335 obtained from statistical and bioinformatic analysis using RStudio.

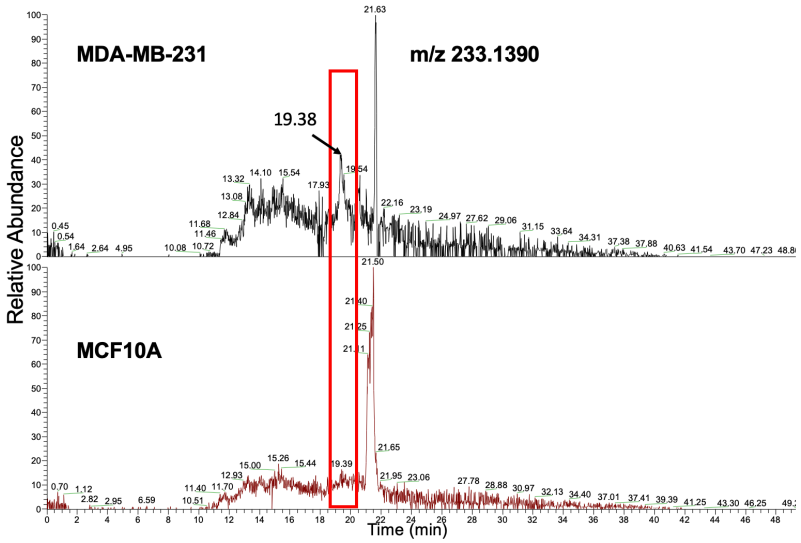

Figure S5:m/z 233.1390 Presence and Uniqueness. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 233.1390 and RT 19.38 min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

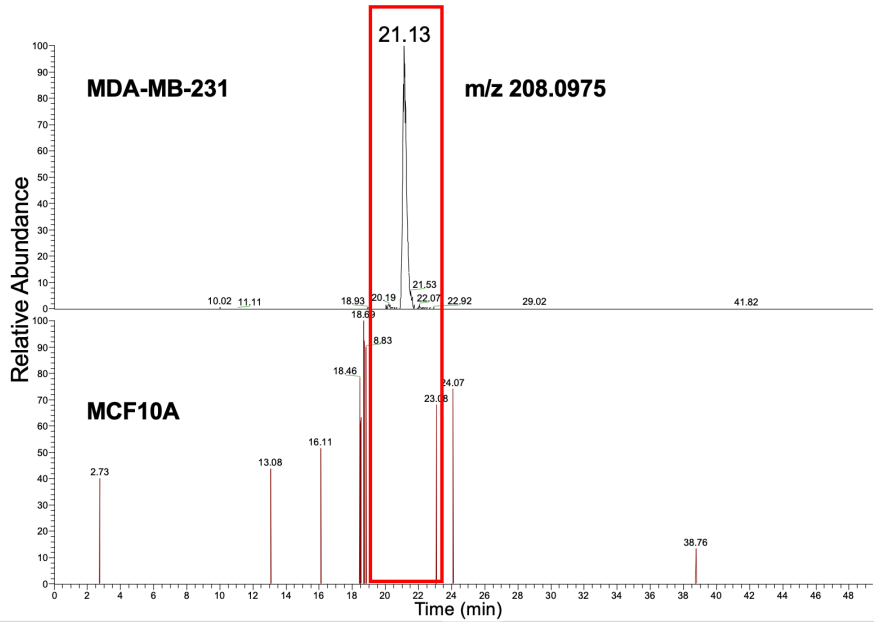

Figure S6:m/z 208.0975 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 208.0975 is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line. This figure verifies the presence of m/z 208.0975 obtained from statistical and bioinformatic analysis using RStudio.

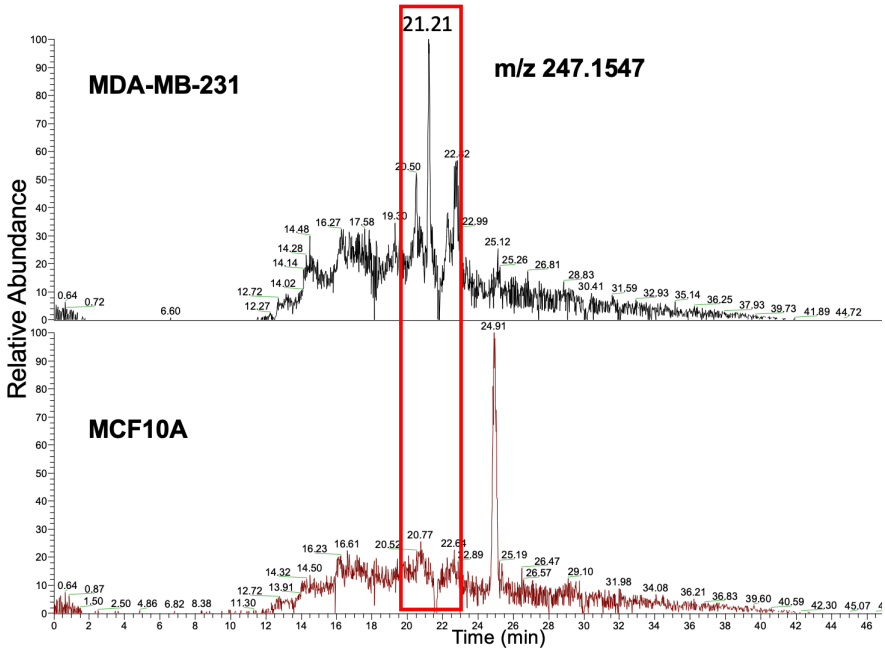

Figure S7:m/z 247.1547 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 247.1547 is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

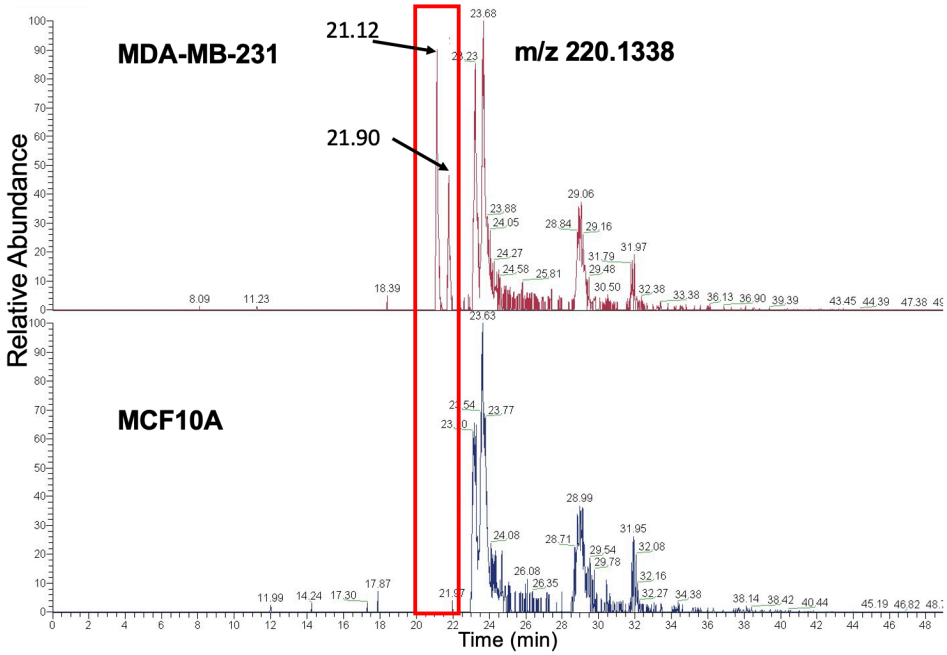

Figure S8:m/z 220.1338 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peaks at m/z 220.1338 RT 21.12 and 21.90 min are only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

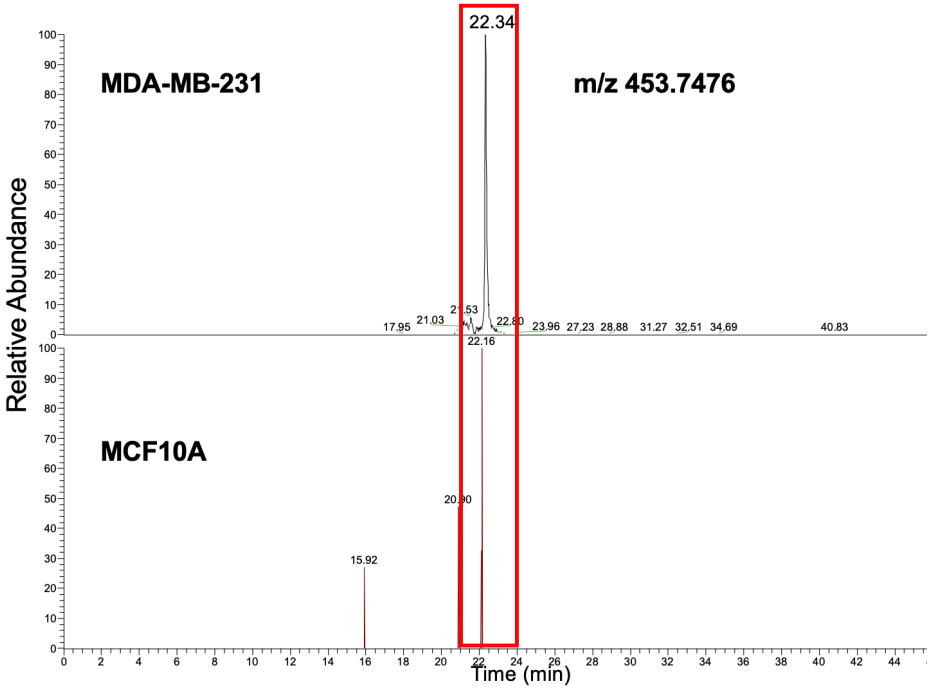

Figure S9:m/z 453.7476 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 453.7476 RT 22.34 min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

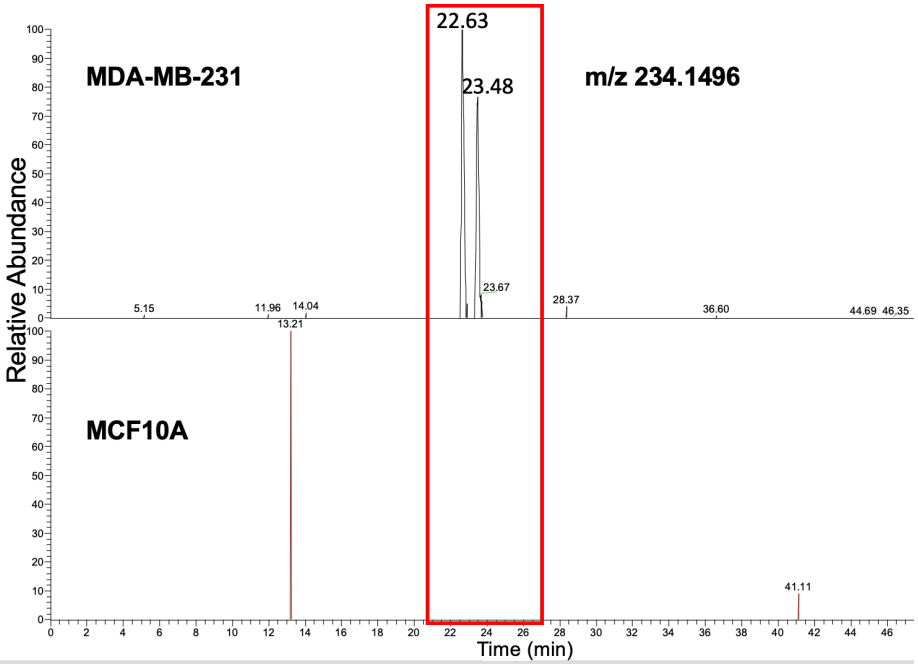

Figure S10:m/z 234.1496 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peaks at m/z 234.1496 RT 22.63 and 23.48 min are only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

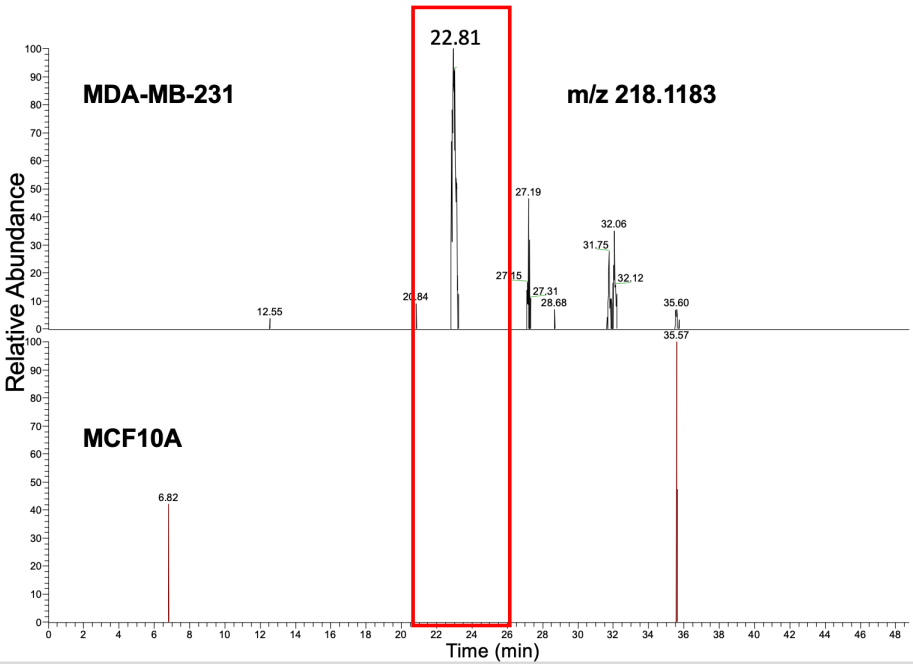

Figure S11:m/z 218.1183 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 218.1183 and RT 22.81min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

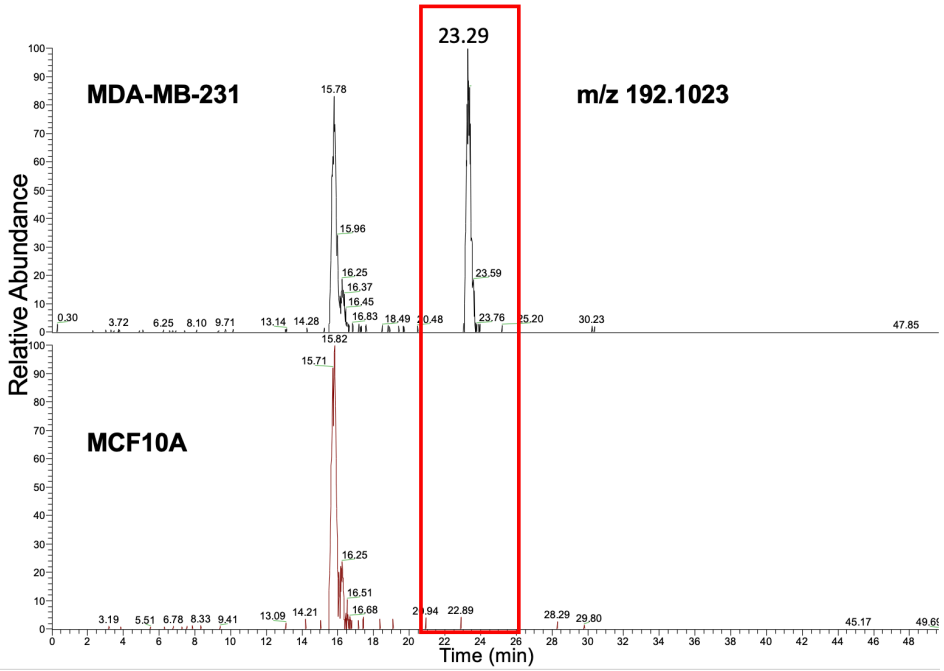

Figure S12:m/z 192.1023 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 192.1023 and Rt 23.29min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

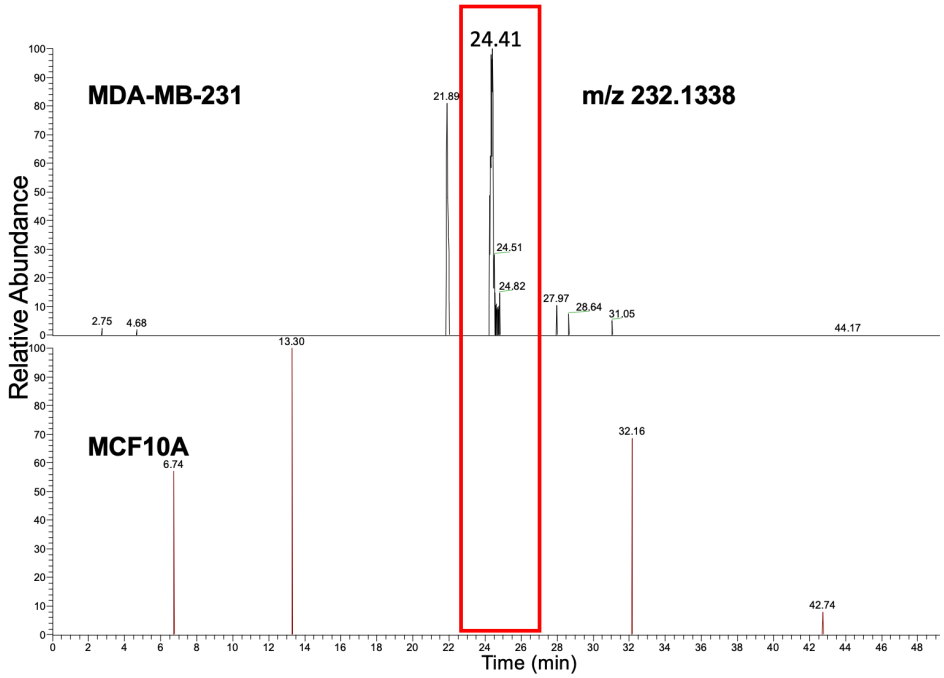

Figure 25:m/z 232.1338 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 232.1338 and RT 24.41min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

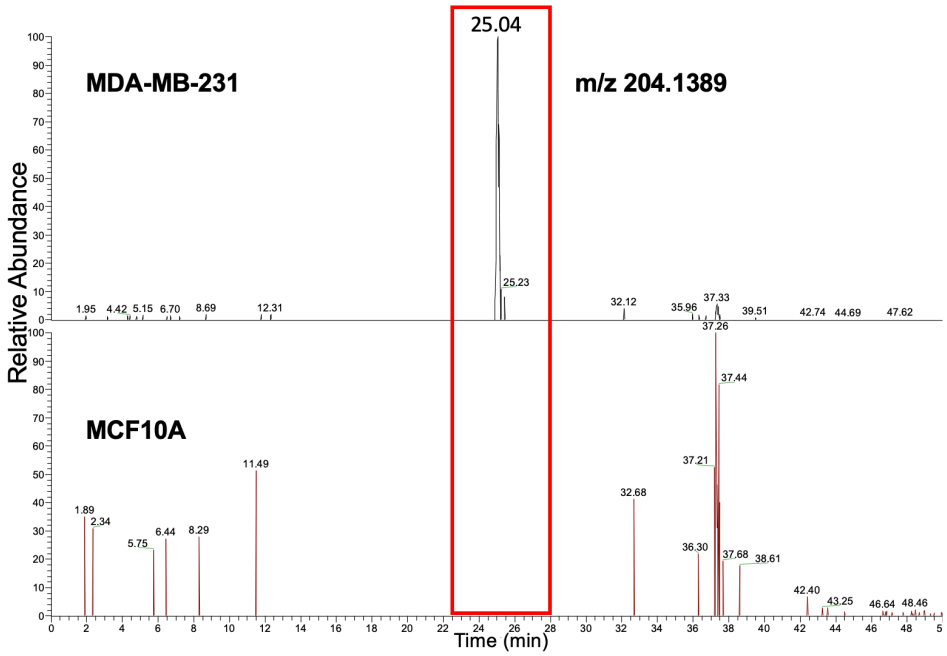

Figure S14:m/z 204.1389 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 204.1389 and Rt 25.04min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

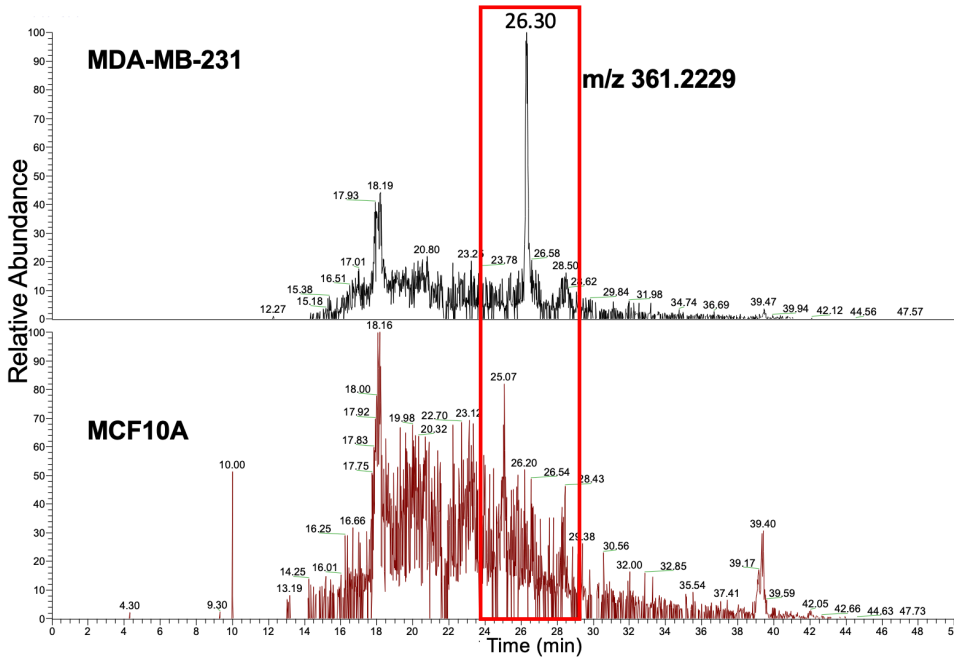

Figure 26:m/z 361.2229 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 361.2229 RT 26.30 min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

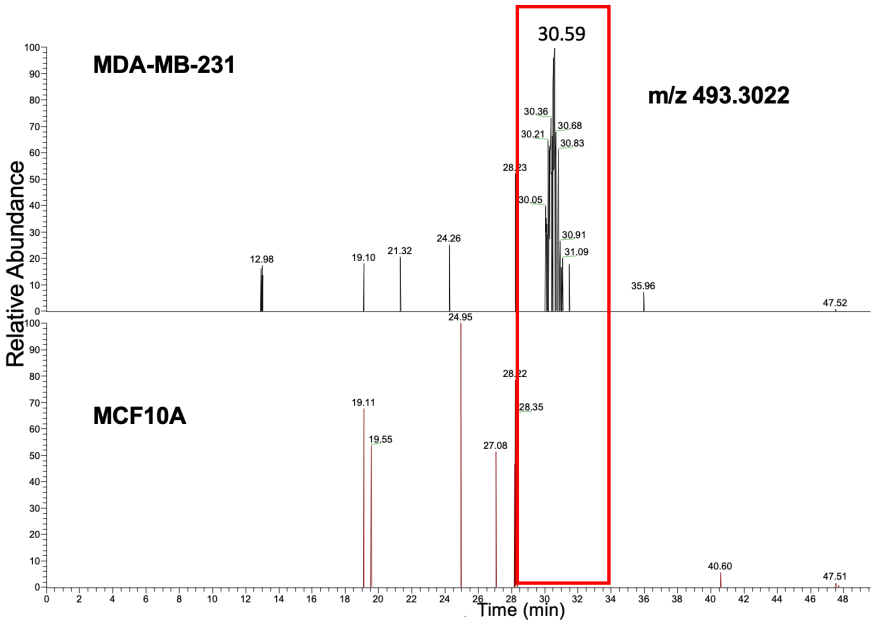

Figure S16:m/z 493.3022 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 493.3022 at RT 30.59 is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

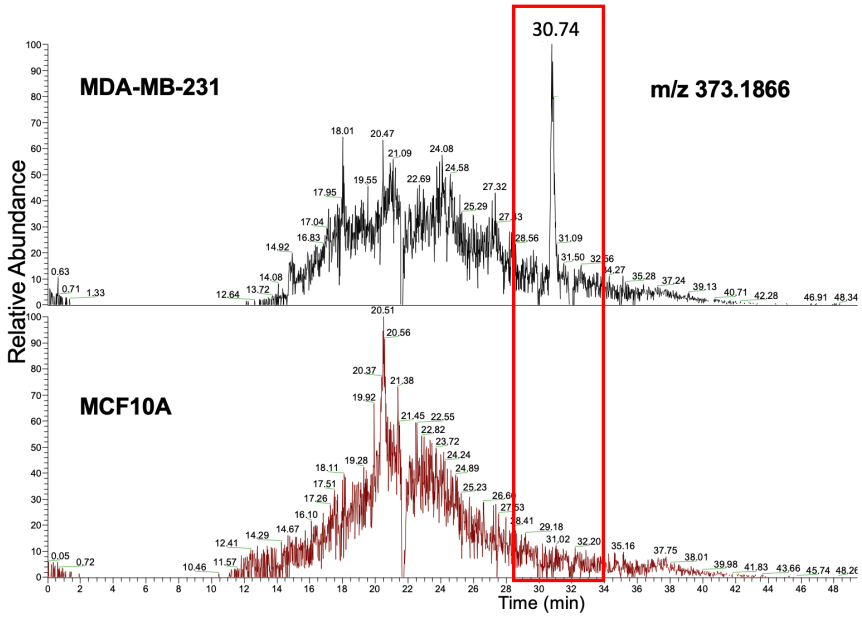

Figure 27:m/z 373.1866 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 373.1866 at RT 30.74min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

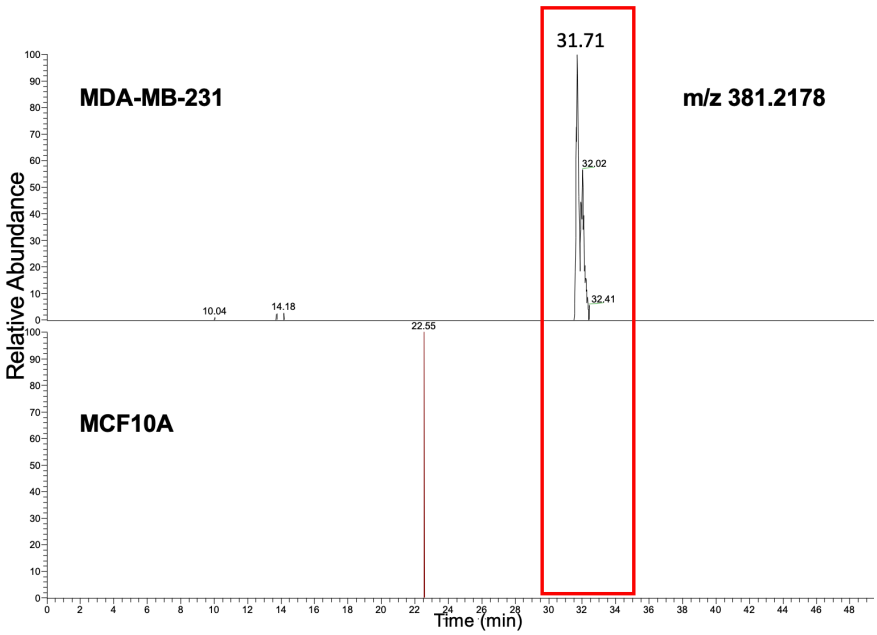

Figure S18: m/z 381.2178 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 381.2178 at RT 31.71min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

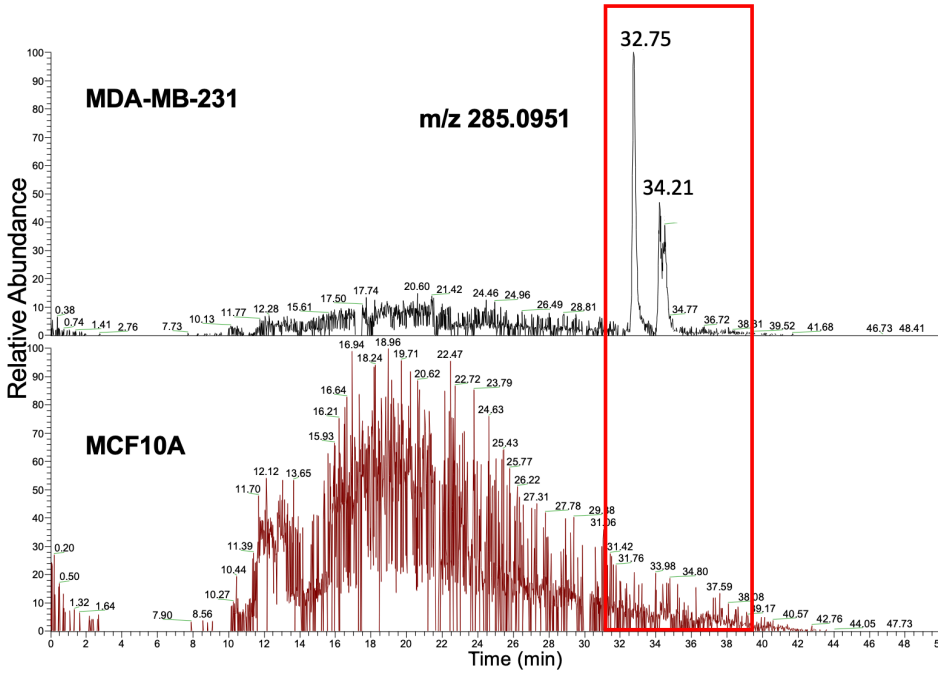

Figure S19:m/z 285.0951 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peaks at m/z 285.0951 at RT 32.75 min and 34.21 min are only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

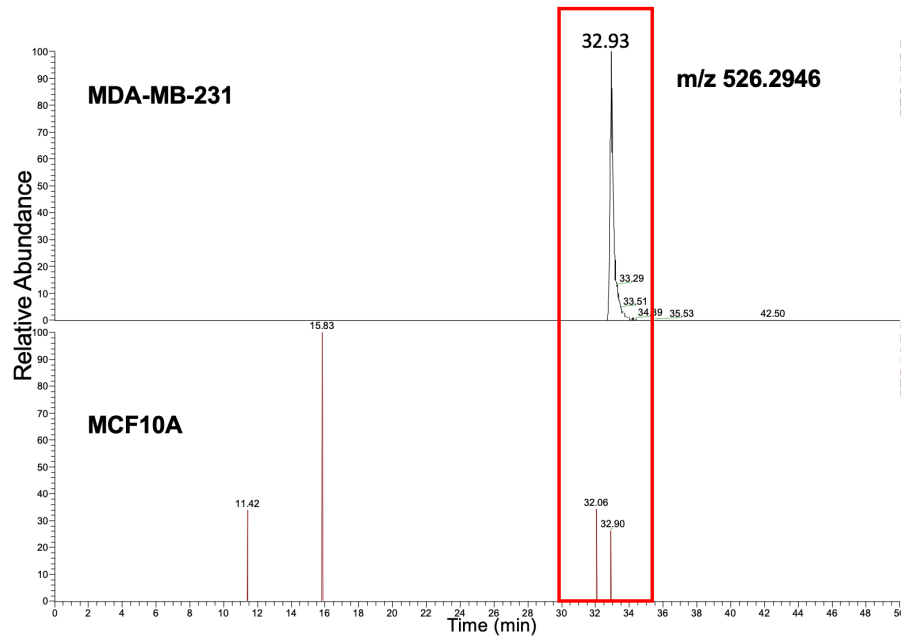

Figure S20:m/z 526.2946 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 526.2946 at RT 32.93min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line. The two small peaks seen in the red box in the MCF10A sample are just peaks from background noise.

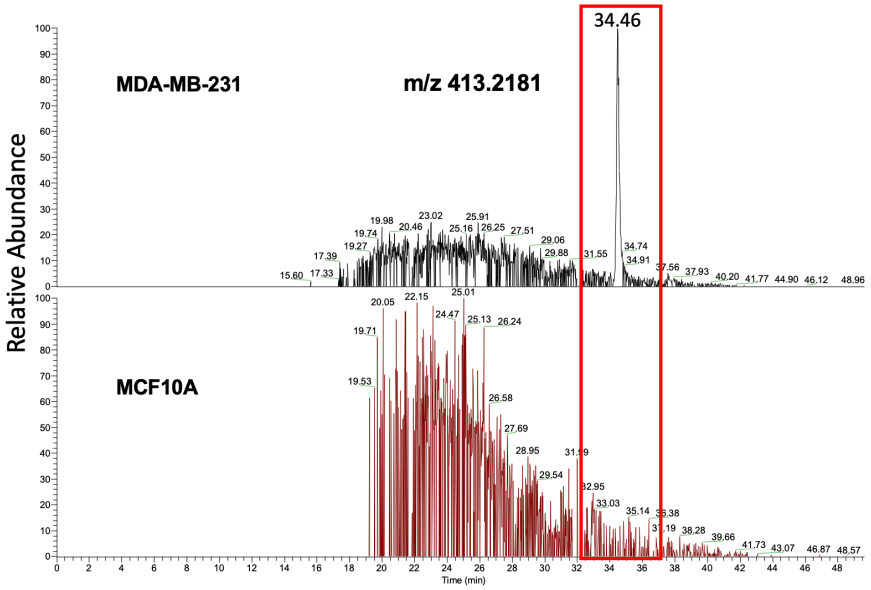

Figure S21:m/z 413.2181 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peak at m/z 413.2181 at RT 34.46min is only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

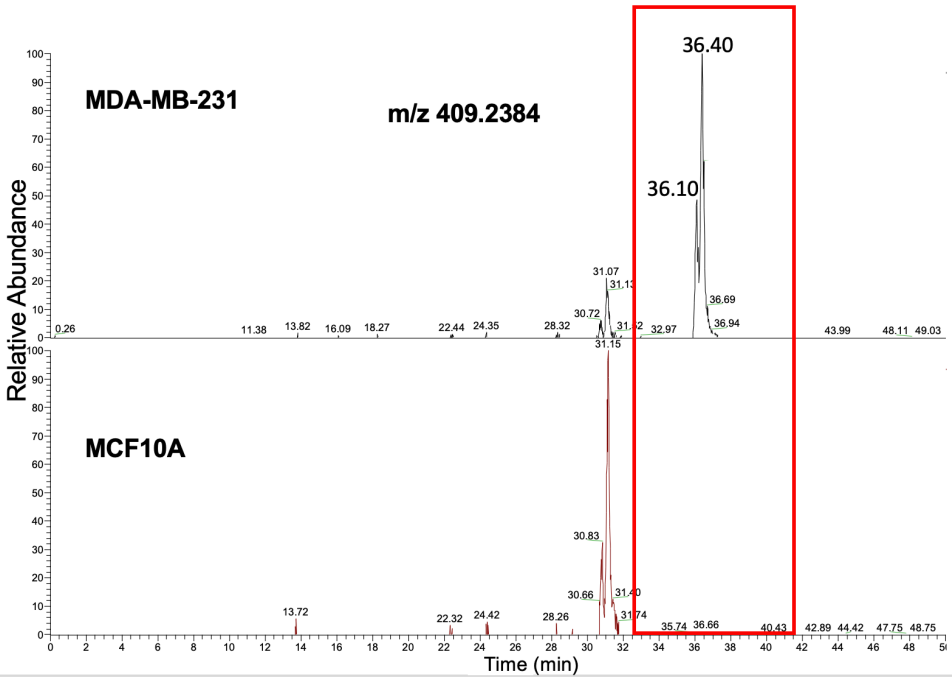

Figure S22:m/z 409.2384 Presence and Uniqueness Verification. EICs of both MDA-MB-231 metabolite sample and MCF10A metabolite samples. The peaks at m/z 409.2384 at RT 36.10 and 36.40 mins are only present in MDA-MB-231 samples, verifying its uniqueness to the cancer cell line.

Table 7: N-Acetyl-L-Phenylalanine Peak Areas and Corresponding Concentrations. Concentrations obtained using the calibration curve.

Sample Number	N-Acetyl-L-Phenylalanine Peak Area	N-Acetyl-L-Phenylalanine Concentration (nM)
1	344212197	408.514
2	354394187	420.595
3	314146848	372.833
4	337916831	401.043
5	369177595	438.144
6	316765955	375.941
7	357028407	423.725
8	342824693	406.868
9	347158807	412.011