universität wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Development of MS-based analytical tools to support discovery of metabolic biomarkers in breast cancer cell culture“

verfasst von / submitted by

Pia Schmidt

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magistra der Pharmazie (Mag.pharm.)

Wien, 2020/ Vienna, 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 449
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Diplomstudium Pharmazie
Betreut von / Supervisor:	Ao. Univ.-Prof. Mag. Dr. Franz Gabor
Mitbetreut von / Co-Supervisor:	

Declaration

The work presented in this thesis was done at the Animal Cell Technology Unit and the Mass Spectrometry Unit at iBET in Oeiras, Portugal, with the support of Inês Isidro, PhD and Patrícia Gomes Alves, PhD.

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference or acknowledgment, the work presented is entirely my own.

I have tried to trace all owners of the image rights and to obtain their consent to use the images within this work. Should a copyright infringement nevertheless become known, please do not hesitate to notify me.

Acknowledgements

First of all, I want to thank my supervisors Inês Isidro, PhD and Patrícia Gomes Alves, PhD not only for their patient guidance but also for enabling me to experience 6 months in beautiful Portugal in the Animal Cell Technology Unit and Mass Spectrometry Unit at iBET. Huge thanks to Catarina Correia from UniMS Unit of never getting bored to answer my questions regarding the LC-MS and to Rita Mendes and Maria João Sebastião, helping me a lot and making me feel very welcomed.

I really appreciate my family and my boyfriend not doubting my decision for going abroad and supporting me in every possible way, even in the distance. I would have never thought that I meet so many special people who became like a family for me and made me call Oeiras a second home. I'm very thankful to Dr. Franz Gabor for making this journey possible.

Abstract

Metabolomics, in contrast to its relative's genomics, transcriptomics and proteomics, is one of the newcomers of omics technologies trying to target the connection between genetic processes and external factors directly linked to a cell's phenotype. Not all genes are translated into active products, but metabolite concentrations are instantaneously influenced by intra- and extracellular alterations allowing a valuable snapshot of the healthy or pathogenic phenotype.

The ultimate hallmark of cancer is a rewiring in metabolism to foster proliferation. More and more evidence lead to the assumption that considering cancer solely as genetic disease is an oversimplification but approaching cancer in the context of its altered metabolic pattern allows a deeper understanding of the disease onset and progress and may help finding new therapeutic targets and biomarkers.

Since the metabolome is a complex mixture of compounds with very diverse physicochemical properties, the biggest challenge is to implement a method that covers a high number of different metabolites. The combination of liquid chromatography and high-resolution mass spectrometry (LC-MS) enables a highly sensitive measurement of a broad spectrum of metabolites involved in cancer-specific pathological processes.

In this thesis, the aim was to establish a method that allows to detect a broad number of metabolites in an untargeted approach, using ultra-high-performance liquid chromatography (UHPLC) separation coupled to a high-resolution mass spectrometer, the Q-Exactive™ hybrid quadrupole Orbitrap™ (Thermo Scientific). This method demonstrates that LC-MS metabolomics can be applied in revealing cancer specific metabolite patterns by identifying pathways and metabolite classes and further use this information in biomarker discovery.

In this work, first the best conditions regarding sample preparation, choice of column and MS conditions were evaluated by profiling cell culture media. The focus was to acquire the maximum identifications as well as the best peak shapes and separations in the chromatogram. When precipitating the sample with methanol (MeOH), lyophilizating and resuspending it in water and using two types of columns, C18 and amide, to cover polar and apolar metabolites, the highest amount of identifications could be detected. Using the optimal settings, relevant samples of breast cancer cells

belonging to different subtypes were analysed Apart from the well-recognized shift from aerobic to anaerobic glycosylation, metabolic pathways involved in lipid metabolism show alterations in order to support carcinogenesis. In conclusion, LC-MS metabolomics turns out to be a potent tool to observe differences in metabolic signatures triggered by the cell's pathological state. Developing new biomarkers, diagnostic approaches or treatment targets grounding on a deeper understanding of molecular pathways involved in disease progression, can be achieved.

Zusammenfassung

Metabolimik gehört im Gegensatz zu seinen Verwandten Genomik, Transktiptomik und Proteomik zu den Newcomern der Omics-Technologien, die versuchen genetische Prozesse mit externen Faktoren zu korrelieren, um den Phänotyp einer Zelle zu erforschen. Nicht alle Gene werden in aktive Produkte übersetzt, aber interne und externe Veränderungen haben einen sofortigen Einfluss auf die Metabolitkonzentrationen einer Zelle und repräsentieren eine wertvolle Momentaufnahme des gesunden oder pathogenen Phänotyps.

Das ultimative Kennzeichen von Krebs ist eine Neuverkabelung des Stoffwechsels, um dessen Proliferation voranzutreiben. Mehr und mehr Beweise führen zu der Annahme, dass die Betrachtung von Krebs als rein genetische Erkrankung zu sehr vereinfacht ist. Die Annäherung von Krebsentstehung in Zusammenhang mit des veränderten Stoffwechselmusters, ermöglicht ein tieferes Verständnis des Krankheitsbeginns und -fortschritts und kann zu neuen therapeutischen Angriffspunkten und Entdeckung krebsspezifischer Biomarker führen. Da das Metabolom ein komplexes Gemisch von Stoffklassen mit sehr unterschiedlichen physikalisch-chemischen Eigenschaften ist, besteht die größte Herausforderung darin, eine Methode zu implementieren, die ein breites Spektrum abdecken kann. Die Kombination von Flüssigkeitschromatographie und hochauflösender Massenspektrometrie ermöglicht eine hoch empfindliche Messung eins breiten Spektrums an Stoffwechselprodukten, welche in Krebsspezifische pathologische Prozessen involviert sind.

Das Ziel dieser Arbeit bestand in der Etablierung einer Methode, die eine breite Detektion von Metaboliten im Zuge eines “untargeted metabolomic” Ansatzes ermöglicht. Unter Verwendung von ultra-hochleistungs Flüssigkeitschromatographie (UHPLC) und hochauflösendem Q-Exactive™ Massenspektrometer (Thermo Scientific) konnte gezeigt werden, dass LC-MS ein vielversprechender Weg ist mit der Aufdeckung krebsspezifischer Metabolitmuster und deren Korrelation mit den aktivierten Stoffwechselwegen, die Biomarker Forschung voranzutreiben.

Zunächst wurden die optimalen Bedingungen bezüglich Probenvorbereitung, Säulenwahl des LC- Systems, sowie MS- Bedingungen durch Analyse von Zellkulturmediums bewertet. Der Schwerpunkt lag auf der Erfassung der maximalen

Identifikationen sowie der besten Peakformen und -trennungen im Chromatogramm. Die Ergebnisse zeigen, dass mit der Vorbereitung der Probe mit Methanol (MeOH), Gefriertrocknung und Re-Suspendierung in Wasser und der Verwendung von zwei Säulentypen, C18 sowie der Amidsäule, um polare und apolare Metabolite abzudecken, die höchste Anzahl von Identifikationen erreicht wird. Unter Verwendung der optimalen Einstellungen wurden relevante Proben von Brustkrebszellen, die zu verschiedenen Subtypen gehören, analysiert. Abgesehen von der allgemein anerkannten Verlagerung von aerober zu anaerober Glykosylierung zeigen die am Lipidstoffwechsel beteiligten Stoffwechselwege Veränderungen, um die Kanzerogenese voranzutreiben. Zusammenfassend lässt sich sagen, dass sich die LC-MS-Metabolomik als ein wirksames Instrument zur Beobachtung von Unterschieden in den Stoffwechselsignaturen erweist, die durch den pathologischen Zustand der Zelle ausgelöst werden. Das tiefere Verständnis von molekularen Pfaden, die an der der Krankheitsprogression beteiligt sind, kann die Entwicklung neuer Biomarker, Diagnoseansätze oder therapeutische Angriffspunkte fördern.

Table of Contents

List of abbreviations:

AA	amino acid
ACN	acetonitrile
ACN/H2O	acetonitrile in water
APCI	atmospheric pressure chemical ionization
APPI	atmospheric pressure photo ionization
ATP	adenosine triphosphate
BC	breast cancer
CH	carbohydrate
CV	coefficient of variation
EI	electron ionization
ER	estrogenic receptor
ESI	electrospray ionization
FA	fatty acid
FASN	fatty acid synthase
FH	fumarate hydrase
FT-ICR- MS	Fourier-Transform-Ion-Cyclotron-resonance Mass Spectrometry
GC	gas chromatography
Her2	human epidermal growth factor
HILIC	hydrophilic interaction chromatography
HMDB	human metabolome database
IDH	isocitrate dehydrogenase
LC-MS	liquid chromatography mass spectrometry
m/z	mass to charge ratio
MeOH	methanol
MeOH/H2O	methanol in water
mRNA	messenger ribonucleic acid
MS	mass spectrometry
NA	nucleic acid
NMR	nuclear magnetic resonance

PC	phosphatidylcholine
PCA	principal component analysis
PCho	phosphocholine
PE	phosphatidylethanolamine
PET	positron emission tomography
PPP	pentose phosphate pathway
ppm	parts per million
PR	progesterone receptor
PSA	prostate specific antigen
RT	retention time
SDH	succinyl dehydrogenase
TCA	tricarboxylic acid cycle
tCho	total choline
TNBC	triple negative breast cancer
TOF-MS	time-of-flight mass spectrometer
UHPLC, HPLC	ultra-high-performance liquid chromatography, high performance liquid chromatography

1 Introduction

Metabolomics, compared to its predecessor's genomics, transcriptomics and proteomics, gives the ultimate approach to connect the genetic processes and dynamic impact of external factors represented in an altered metabolic state, thus allowing the best image of its pathogenic phenotype.

1.1 Metabolomics as a diagnostic approach in cancer research

The ultimate hallmark of cancer is a rewiring in metabolism to foster proliferation. Approaching cancer in the context of its altered metabolic pattern allows a deeper understanding of the disease and progress and may help finding new therapeutic targets and biomarkers [1–5].

1.1.1 The metabolome

Since the breakthrough of the human genome sequencing [6], and the knowledge of the flow of information described as the central dogma, our understanding of cellular processes has increased drastically during the last century. The main conclusion obtained, was that our hereditary instruction is captured in the DNA, the information flows from genes to mRNA to proteins and accounts for every biological organism [7]. All cells have a fixed instruction of what can happen, captured as complete set of DNAs in the genome. The information that is actually active is transcribed into mRNA transcripts, which all combined represent the transcriptome. The mRNA transcripts serve as messenger communicating the cells work plan that is carried out by translation into proteins which catalyse biochemical reactions of the final end products, known as metabolites. Hence, the central dogma misses to consider that the actual objective of genes, transcripts and proteins is to regulate these small molecules which are major influencers of cell function and dysfunction themselves [8–10]. The term metabolome was first introduced by Stephen G. Olivier [11] and defines the complete set of all small molecular substrates less than 1500 Da, including nucleotides, amino acids, carbohydrates or lipids[12,13]. By granting biochemical feedback metabolomics represents the superglue across all omics layers, instead of being solely an information sink. Because it is considered as the endpoint of the omics cascade, it is the closest to the cell´s phenotype and the ultimate effector of the cellular machinery [14].

Metabolomics tries to identify the complete metabolome in cells and tissues to gain a comprehensive understanding of the interaction of genomics, transcriptomics and proteomics with high throughput technologies like nuclear magnetic resonance spectroscopy (NMR) or mass spectrometry (MS) [13,15].

In the genomics era the spotlight was focused on using genetic mutations for understanding diseases and individualizing therapy. Unfortunately, directly linking the genotype to the phenotype is missing important aspects because a disease is rather a result of an interplay of multiple genetic mutations and external factors like environment or lifestyle, rather than the result of only a single mutation. Many steps separate a gene from its ultimate outcome but because metabolites are seen as the downstream result of genetic variation, transcriptional changes, post-translational modifications of proteins and the influence of external factors, they capture best what is actually happening in the body and are the most representable reporters of any disease status [9,15]. Furthermore, metabolites are not only downstream products of the gene expression chain but also required building blocks for each step and disease drivers by influencing gene expression, signal transduction or substrates and products of enzymatic reactions [1,16–20].

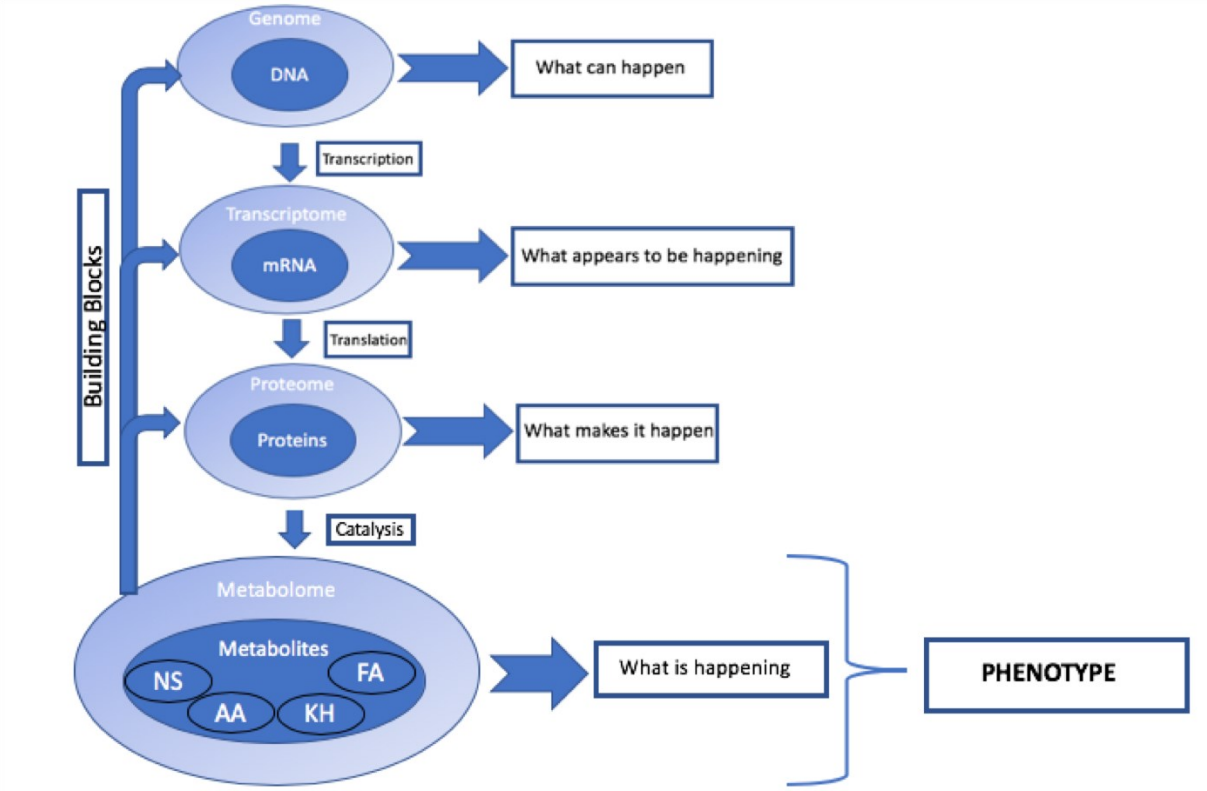

Figure 1- metabolome downstream [13,17]
The figure demonstrates how genome, transcriptome and proteome combine in the end point of the cascade, the metabolome. The metabolites, like nucleic acids (NA), amino acids (AA), carbohydrates (CH) or fatty acids (FA) are themselves influencing the whole cascade as important building blocks and representatives of the phenotype.

Figure 1 demonstrates how on one hand the metabolome symbolizes the endpoint of the omics cascade, subsequent to the central dogma framework, and thus the closest representative of its phenotype. On the other hand metabolites can loop back into steps of the chain as they are required building blocks and influencers of the genome, transcriptome and proteome [21]. Furthermore, it stresses as mentioned above that the genome gives the primary instructions to what can happen to the cell, the workplan is passed on with the produced transcripts and carried out by the proteins which are the catalysts of all biochemical reactions resulting in metabolites, like nucleic acids (NA), amino acids (AA), carbohydrates (CH) or fatty acids (FA) [13].

Alteration in the metabolome through disease or therapy lead to specific metabolic patterns which metabolomics seeks to reveal through identifying and quantifying the metabolites involved. In conclusion, uncovering the metabolome gaps the bridge of the genomic, transcriptomic and proteomic level and the macroscopic phenotype [10,15].

The main objective is to reveal more sensitive biomarkers at the metabolic level where alterations of metabolite concentrations can provide early evidence of disease outbreak, among others in cancer [22,23] diabetes and obesity [10,24] or cardiovascular diseases [25]. By targeting the individual causes leading to the pathological state of a patient, personalized diagnosis and therapy can be progressed and better disease outcomes can be achieved [26].

1.1.2 Breast cancer

Cancer is the major death cause of the 21st century, rising to 18.1 million new cases and 9.6 million deaths in 2018 worldwide [27]. Among women, the most common types of cancer are breast, lung and colorectal, from which breast cancer accounts for 30% of all female cancers (figure 2) [28].

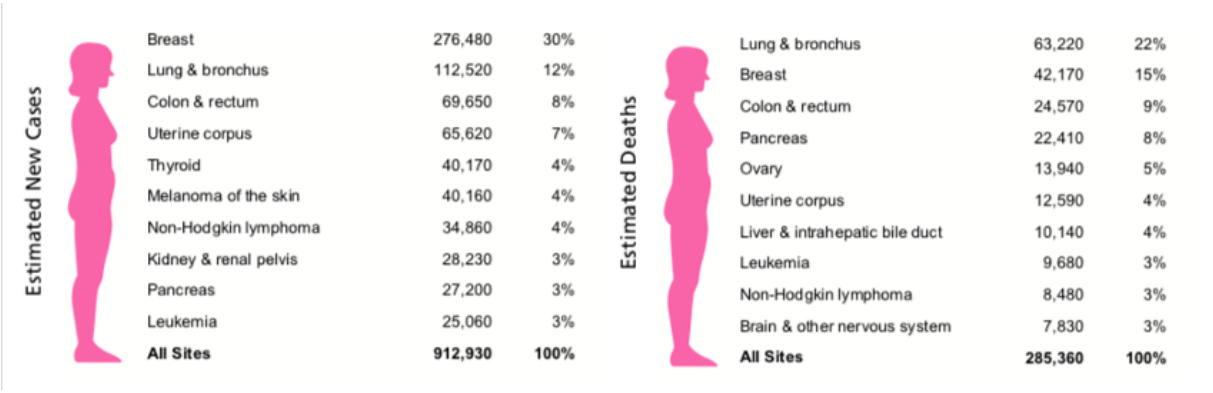

Figure 2- world cancer statistics for women population in the USA,2020.
The number of estimated new cases on the left and the number of estimated deaths on the right [28]

Risk factors like pre-existing malignancies, such as obesity, environmental factors, hormonal factors, like post menopause state, or inherited genetic variations, are main drivers of breast cancer [29]. On the molecular level, cancer cells uncouple from normal cell control mechanism and rewire their metabolism leading to uncontrolled cell division and formation of abnormal cell spreading and invasion of the surrounding area. An interplay of inherited or environmentally caused mutations in the genetic code, activation of oncogenes, or loss of tumour suppressor genes, sets into motion the alteration of several metabolic and regulatory pathways to develop different strategies to support cell proliferation and escaping the control mechanisms for apoptosis.

This makes cancer the perfect target for metabolomic studies. In this thesis, the main focus is directed to breast cancer, which symbolizes a perfect representation of the heterogeneity of cancer because of different subtypes with distinct metabolic alterations [30]. The breast is composed of different cell types and tissues with different characteristics and thus cannot be simply summarized to one disease with one treatment option. Trying to target breast cancer as a collection of diseases, it can be divided into subtypes based on morphological examination and more precisely through molecular subtyping which can be distinguished by its receptor expression including human epidermal growth factor receptor 2 (HER2) and estrogen receptor (ER) (table 1) [31–34].

Table 1- molecular breast cancer subtyping based on the expressed surface receptor
Only the luminal subtypes express estrogen receptor (ER) and additionally expression of progesterone receptor (PR). Her2 is expressed on Luminal B and Her2 overexpressed breast cancers, while triple negative subtype lacks any receptor expression.

	ER	PR	HER2
LUMINAL A	+	+/-	-
LUMINAL B	+	+/-	+
HER2 OVEREXPRESSED	-	-	+
TRIPLE NEGATIVE	-	-	-

The characteristics of HER2 and ER can be used to roughly divide breast cancer into four major molecular subtypes, including Luminal A (HER2 negative and ER positive), Luminal B (HER2 positive and ER positive), HER2-enriched (HER2 positive and ER negative), and triple negative breast cancer (TNBC) (Her2 negative, ER negative) (figure 3) [33] . Compared to the ER+ subtypes,TNBC and Her2 overexpressed is associated with poor prognosis and shorter survival rate, mainly because they cannot be targeted by estrogen receptors [30].

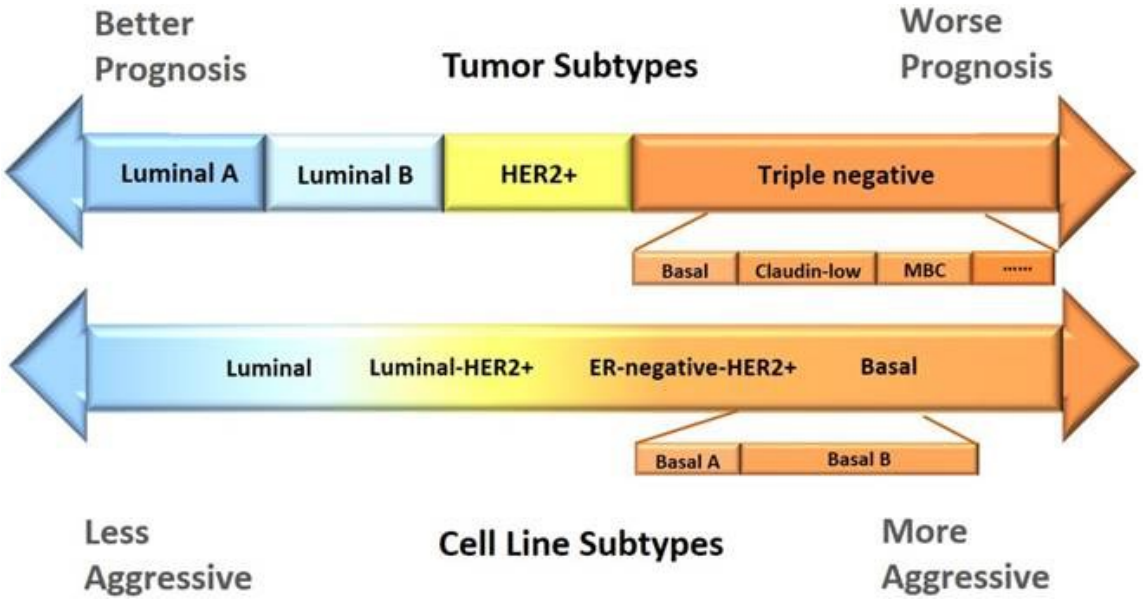

Figure 3- comparison of Breast Cancer subtypes [33]
The molecular subtyping based on the receptor status ER, PR, HER2 can be observed. They can be divided into four main groups, luminal A and B, HER2 positive and TNB which differ in prognosis and aggressiveness.

Despite immense progress in cancer research breast cancer still leads to high death rates among women, due to the lack of adequate systemic therapies and the ability of early diagnosis before metastasis [32,35,36]. Mammography, ultrasonography or invasive biopsies are applied for diagnosing symptomatic breast cancer as well as several antigens used as serum markers. Average survival can be increased dramatically by early stage diagnosis and treatment before metastasis, therefore non-invasive detection methods with high sensitivity and specificity are urgently needed [37].

The heterogeneity of breast cancer due to different genetic expressions of receptors is reflected in the clinical behaviour and used as treatment guidance. However, subtyping breast cancer at the molecular level still only leads to assertion of breast cancer pathogenesis and disease progression. Metabolomics offers a deeper understanding and approach of metabolic mechanisms and exploited pathways causing tumour proliferation, in order to improve BC classification and early diagnosing [36].

Many studies focused on revealing reliable metabolic signatures, but so far clinical application has not been achieved. Existing studies show high heterogeneity in identified metabolic patterns and lack validation. Some reasons may be pointed to these various outcomes, the use of different detection methods, sample preparation, collection (biofluid, plasma, tissue, urine) and storage [32].

Metabolomics as a new expanding field which analyses cellular metabolism through profiling biofluids like urine or blood, could deepen our understanding of the dynamic

interaction between tumour and host and help in non-invasive early disease detection, prediction of therapy response or cancer reoccurrence and personalized therapy [36,38].

1.1.3 Alterations in cancer cell metabolism

Cancer cells uncouple from normal cell control mechanism and rewire their metabolism leading to uncontrolled cell division. Formation of abnormal cell spreading and invading the surrounding area represent the primary cause of mortality in cancer patients. For a long time, cancer has primarily been viewed as a genetic disease driven by somatic mutations in oncogenes or tumour suppressor genes. In 1920 Otto Warburg was the first to observe that even though different cancer types vary a lot in their genetic mutations, all show a common hallmark in their metabolism, namely shifting aerobic respiration to fermentation and largely relying on glucose for energy. Further research suggests that gene mutations can be considered a downstream epiphenomenon caused by metabolic rewiring [39–41]. This makes cancer metabolism the ultimate target for metabolomics research and by deepening our knowledge of cancer metabolism our understanding of cancer pathophysiology will be progressed and support clinical oncology.

But how to increase biomass so a new cell can be formed although using the same metabolic pathways? A multi- step process occurs throughout transforming a healthy cell to a highly proliferative cancer cell [42].
All cells have to take up nutrients from the environment and feed it into metabolic pathways to maintain homeostasis between anabolism and catabolism and to sustain alive. Thermodynamically unfavourable reactions can only occur when coupled to adenosine triphosphate (ATP) hydrolysis. Hence it is not surprising that healthy cells dedicate their metabolism to efficient ATP production. Cancer cells aim to proliferate as much as possible, therefore their metabolism focuses on increased nutrient uptake to generate the required biomass. Unrestrained proliferation can be achieved by a tight interplay of nutrient availability, genetic alteration like loss of tumour suppressors, mutations in oncogenes integrated and metabolic rewiring controlling activity of oncogenic signalling cascades, demonstrated in figure 4 [43].

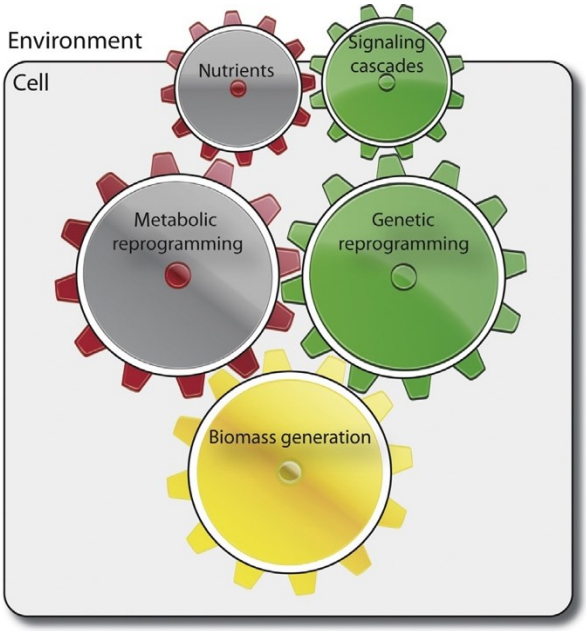

Figure 4- interplay of metabolic rewiring [43]
The figure shows that metabolism is a tight interplay of nutrients, signaling cascades, metabolic and genetic reprogramming as well as biomass generation.

The most inevitable step as described before, is the increase in nutrient uptake and especially glucose and glutamine addiction. This phenomenon is known as the Warburg effect and recognized as key hallmark of cancer. This glucose dependency is clinically applied for diagnosing and staging through positron emission tomography (PET) scanning with radiolabelled 2-deoxyglucose and 18-fludeoxyglucose. Tumour cells tremendously increase glucose uptake and instead of fully oxidizing it through the oxidative phosphorylation chain, it is fermented to pyruvate and further to lactate, although generating less ATP. Since healthy cells only rely on this pathway in the absence of oxygen, Warburg considered this mechanism the result of dysregulation in the mitochondria [44].

This rewiring seems to be a suspect matter because of much less efficient energy outcome but looking closer it turns out to be a survival advantage. The cells profit from oxygen independency, accumulation of more building blocks through shifting glycolytic intermediates to different pathways like tricarboxylic acid cycle (TCA) and pentose phosphate pathway (PPP), generation of lactate to acidify the cells environment and thus supressing immune response, the establishment of epigenetic regulations of specific metabolites (oncometabolites) and enabling an overall faster ATP production. Glucose and the produced intermediates as well as glutamine are the most important precursors serving as nitrogen source for ribose sugars, nucleotides and non-essential amino acids, generating glycerol and citrate for lipids as well as supply the reducing

agent nicotinamide adenine dinucleotide phosphate (NADPH) and nicotinamide adenine dinucleotide (NADH) through PPP. As Warburg discovered, mitochondria represent the metabolic core to meet the anabolic requirements of cancer cells and are crucial in cell proliferation. They are responsible for the synthesis of lipids, proteins, nucleic acids and maintain redox homeostasis by supplying the reducing agents NADH and NADPH. To perfectly match the metabolic needs, mitochondria are tightly controlled by oncogenes and tumour suppressor genes. Genetic mutations in metabolic enzymes, like isocitrate dehydrogenase (IDH), succinate dehydrogenase (SDH) or fumarate hydratase (FH) lead to accumulation of the oncometabolites (D)-2-hydroxyglutarate, succinate and fumarate. Those metabolites establish epigenetic alterations that directly contribute to cell transformation and demonstrate the interface between metabolic rewiring and altered gene expression in cancer [42–46].

1.1.4 Application of metabolomics: biomarker discovery

While traditional biomarkers often consist of single molecules, metabolomics tries to uncover metabolic patterns serving as biomarker signatures enabling a more sensitive and specific detection of the pathophysiological state for diagnosing, predicting, staging cancer or evaluating treatment response, drug discovery and personalized medicine [47]. An ideal biomarker is an objectively, quantifiable and reproducibly measured indicator of a patient's medical state and should be present in easily, non-invasively accessible body fluids (e.g. blood, urine, plasma) to simplify clinical application. It should be highly specific and sensitive for achieving early and accurate diagnosis, should alter instantly due to treatment and disease progression and offer a comprehensive knowledge about disease mechanism [10,48].

Metabolomics offers many benefits to satisfy the high requirements of biomarker discovery. First it allows non-invasive analysis due to metabolites being secreted, in comparison to proteomic or genomic biomarkers. Changes at the metabolite level are natural outcomes of the onset and prognosis of many diseases and such changes often appear in biological fluids before the appearance of clinical symptoms [15]. The metabolome is considered much smaller than genes, transcripts and proteins and subtle changes result in substantial changes in the metabolome representing an instantaneous snapshot of the biological state [10,49,50]. Due to metabolites being the

amplified products of gene, transcript and protein activity, the required sensitivity of a metabolic biomarker is achieved, which is particularly important for the evaluation of rapid and progressive diseases. Another advantage occurs due to the fact that metabolites are not species dependant compared to some genes and proteins, hence one method is suitable to study different species and because of highly conserved metabolic pathways developing disease models could be achieved. [10]. Furthermore, the metabolomics approach enables the analysis of multiple sample types, such as urine, blood, tissue, or cerebrospinal fluid, among which blood and urine have the greatest potential for biomarker identification [13].

Despite all this promising features metabolomic biomarker discovery still needs to overcome various limitations. Compared to the genome, which has been fully, and definitely sequenced, metabolome databases are still lacking information and furthermore have to be integrated with systems biology in order to place metabolites into biological context. Despite the progresses in database compilation and analytic detection, identifying the entire metabolome is presently not possible due to the great variety in chemical compound classes and incomplete understanding of cellular function. The proteome can be mapped based on the genome, but the size of the metabolome still remains unknown and does not consist of a limited number of building blocks like nucleotides building the genome or a set of amino acids building the proteome [17].

The advantage of the highly sensitive metabolome also results in a downside because of many factors influencing the metabolic profile that have to be considered including internal impacts such as age, sex and genetics and external like lifestyle, diet, analytical techniques and drug treatments and need also be taken into account and controlled when designing the metabolic experiment. Therefore it must be distinguished if the metabolic alteration results from the pathological state or from confounding factors [9]. Furthermore, the metabolism is very dynamic, many metabolites are produced and degraded in a fast turnover. Samples should be delicately treated and considered that sample storing and temperature (ideally -20° to -80°) impact the stability of the metabolome [10,51,52]. To minimize these challenges and bias, a well-designed metabolic experiment with large number of replicates and carefully selected patients will be necessary to accurately link the biological status with the metabotype [10].

Apparently, the best application of metabolomics so far has been in breast cancer revealing an increase in choline containing compounds, a decrease in glycerophosphocholine as well as glucose for early detection and reoccurrence [49,53–58]. Kidney cancer shows altered concentrations of acylcarnitine and metabolites associated with tryptophan metabolism [49,59,60]. Prostate cancer develop different levels of citrate concentration which appears to be a better indicator than the currently used PSA biomarker[49,61,62], as well as total choline (tCho) and phosphocholine concentrations, along with an increase in the glycolytic products lactate and alanine[13,49,63,64]. Furthermore the sum of concentration of choline and creatine rationed to citrate show aggressiveness as well as sarcosine and N-methyl glycine[65,66]. In ovarian cancer metabolites involved in purine, pyrimidine and glycerolipid metabolism are altered [49,67]. Lung cancer is significant for lower levels of hippurate and trigonelline, and elevated 3-hydroxyisovalerate, α-hydroxyisobutyrate and N- acetylglutamine [49,68]. In colon cancer, lactate, pyruvate, malic acid and long-chain polyunsaturated fatty acid are recognized disease indicators [49,69–71].

1.2 Metabolomics techniques

1.2.1 Analytical methods for metabolomics

The metabolome consists of compounds with high differences in chemical properties, size and concentrations leading to one of the bottlenecks of metabolomics, which is implementing an analytical technique that is able to cover the whole metabolite spectrum [72]. Currently no instrument is capable to target the entire metabolome, therefore the choice of the analytical approach adds a bias in which metabolites will be preferably captured. Achieving the goal to identify and quantify as many metabolites as possible, a good method to differentiate metabolites is required and among all analytical techniques to choose from, challenges, advantages and differences in sensitivity, reproducibility and equipment cost need to be faced. The selection should be based on the purpose of analysis and the nature of samples and compounds of interest. The most commonly used tools are NMR and MS. NMR differentiates metabolites based on chemical shifts. It is highly reproducible and non-destructive to the sample plus requires little sample preparation. MS analyses metabolites based on their mass to charge ratio, it is considered destructive to the sample due to the required

ionization step and demands more extensive sample preparation compared to NMR. Nevertheless, MS is often the first choice for metabolomics due to higher sensitivity. This enables to detect lower abundance metabolites which is beneficial to increase the number of detections in a highly dynamic metabolism. Another advantage is the possibility to couple MS to an a priori separation method like gas chromatography (GC) or liquid chromatography (LC). With this separation step, an additional dimension with the retention time is gained and a decrease in sample complexity is achieved. Further it improves sensitivity and facilitates identification by limiting ion suppression due to allowing compounds to enter ion source at different times instead of simultaneously by direct injection. The majority of metabolites in a cell are polar and involatile demanding a derivatization method for GC separation. In comparison, LC allows the separation of metabolites in solution and of polar and non-polar metabolites depending on the column used and convinces to be the method of choice due to high versatility. [49,52,73].The analytical differences of NMR and MS are summarised in Table 2.

Table 2-comparison NMR and MS [13,73]

CHARACTERIZATION	NMR	MS
SAMPLE		
VOLUME	large sample (500µl)	small sample(1-19µl)
INTERVENTION	non destructive	destructive
PREPARATION	simple	extensive
REPRODUCIBILITY	very reproducible	possible variation due to sample preparation
SENSITIVITY	low	high
STRUCTURAL INFORMATION	less	high
PER SAMPLE COST	low	high

1.2.2 LC-MS based metabolomics

This thesis will focus on the application of LC-MS methodology, as it represents a very flexible and promising approach for metabolite analysis due to the broadness of LC columns chemistries and mass spectrometers available providing the researcher with enormous options. Enabling the detection of many metabolites, LC-MS has become the best choice for metabolite profiling. [74,75]

To start, the most suitable chromatography for metabolite separation prior to MS detection, must be determined, since it is the most limiting part in metabolome coverage [75]. To overcome the challenge of covering a vast range of compound classes, it can be chosen between different types of columns, like reversed-phase (RP) or hydrophilic interaction chromatography (HILIC), to favour physicochemical properties and concentration ranges in the same sample [52,75]. The sample is injected and transported in the mobile phase through a column by applying high pressure. The properties of the stationary phase and mobile phase determine how metabolites are separated in time by offering different interactions. A higher retention occurs if compounds have higher affinity to the stationary phase. The time measured from sample injection to elution is described as retention time. The base of the stationary phase is made of silica and can be modified by covalently bonding either amino groups for hydrophilic interaction and retaining polar substances, known as (HILIC) columns, or with octadecyl carbon chains for apolar molecules, known as (RP) columns. The mobile phase is mainly injected in gradient elution meaning that the composition changes throughout the separation to elute also the most retained metabolites. Hence narrow peaks and thus better and faster separation can be achieved. In HILIC columns the initial mobile phase condition consists of organic solvents like acetonitrile (ACN) or methanol (MeOH) and gradually increasing the aqueous amount and the opposite in reversed phase chromatography [76]. RP columns are more reliable and robust stationary phases compared to HILIC columns but HILIC on the other hand fulfils retention of many primary metabolites (including numerous amino acids, amines, organic acids, sugars and carbohydrates), that play a key role in biochemical processes and can be analysed effectively. Furthermore, the organic solvents increase the sensitivity of the MS as consequence to better ionization but demands to consider low content of water in the sample injected. Additionally, the use of shallow gradients and allowing sufficient re-equilibration time is required to reach acceptable analytical

reproducibility to apply HILIC as complementary technique parallel to RP for metabolic profiling studies [75]. LC-MS based metabolomics uses UHPL rather than HPLC, to achieve most efficient analysis due to smaller particle size of the stationary phase which improves metabolite separation[15,77].

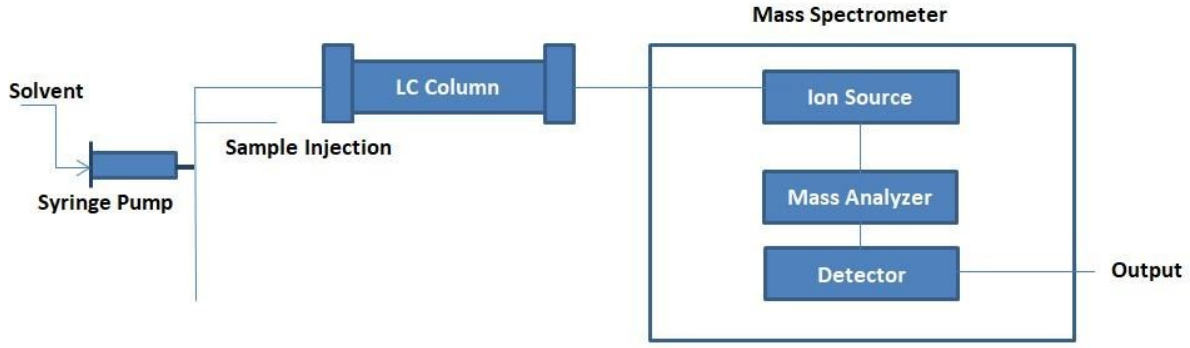

Figure 5- schematic of LC-MS system [78]
The sample is separated by liquid chromatography and then analysed with a mass spectrometer consisting of an ion source, a mass analyser and a detector.

After the chromatography, the sample is forwarded into the MS which separates ions under high vacuum in gas phase, based on their mass to charge (m/z) ratio using an electric or magnetic field. Ions are formed with either electrospray Ionization (ESI), atmosphere pressure chemical ionization (APCI) or atmospheric pressure photoionization (APPI) (figure 5). Which one to choose will have a great impact on the resulting metabolic profiles. Some metabolites prefer a specific ionisation or polarity mode but among all analysing in positive and negative mode with ESI is the favourite approach because it is considered a soft ionization due to forming intact ions (figure 6). ESI uses electrical energy to transform ions into gas phase by forming charged droplets. [52,77,79–81]

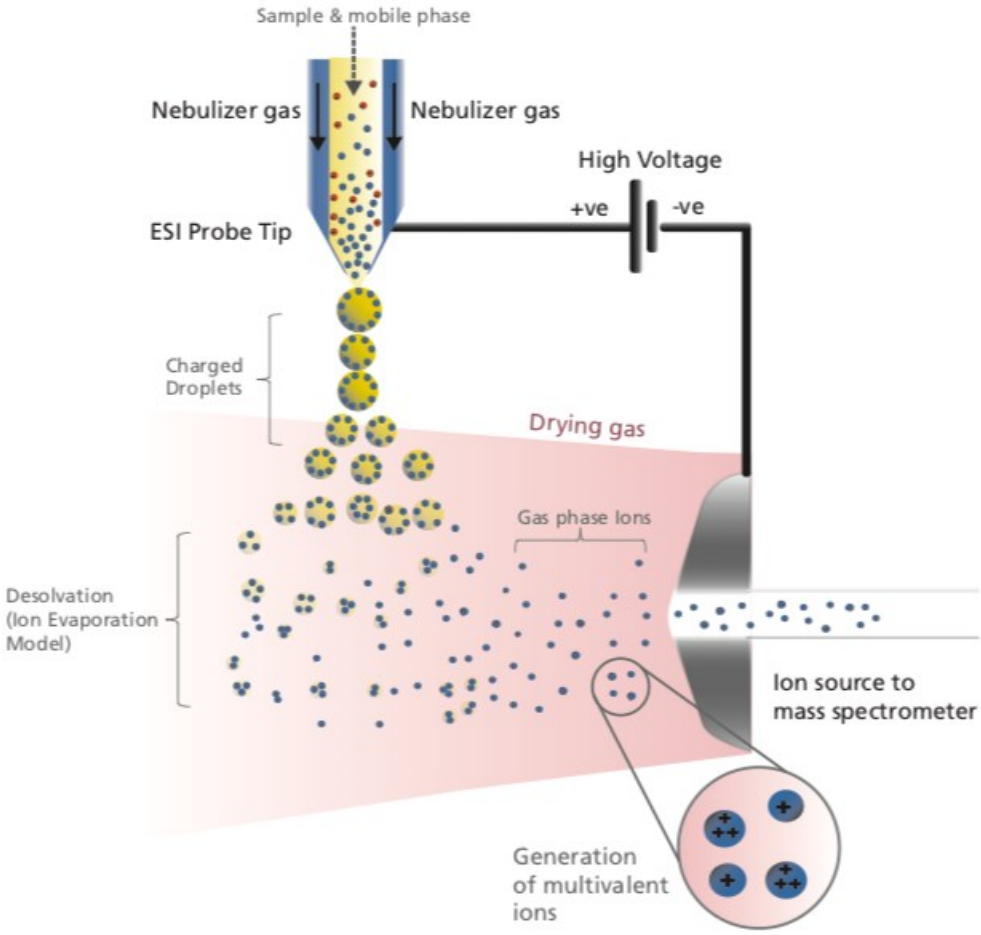

Figure 6- schematic o electrospray ionization[82]
High voltage creates a fine mist of droplets and evaporation processes remove the solvent until ions can enter the MS

Then ions are focused on a mass analyser and separated depending on their mass to charge ratio. The detector plots the m/z values against the signal intensity creating a mass spectrum. Metabolomic studies have great interest in the use of high resolution mass spectrometers such as orbitrap, time-of-flight mass spectrometer (TOF-MS) and Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR- MS). The high resolution allows a higher mass accuracy and decreased mass overlap, compared to single quadrupole and ion-trap mass analysers [10].

The most important parameters that determine the efficiency of MS are high resolution which describes the ability to ability distinguish two ions with similar m/z values (figure 7) and high accuracy which represents the difference between the measured accurate mass and calculated exact mass expressed in mass error in parts per million (ppm)[76]. Since the experimentally determined mass and the calculated mass are not equal, mass deviation is used for an accurate assignment of a measured m/z value to a specific molecular formula [83]. Mass accuracy restricts the number of possible

molecular formulae for a specific m/z value ergo the lower ppm the less possible formulas. The main objective is to ensure that only one type of ion contributes to a particular measurement [84]

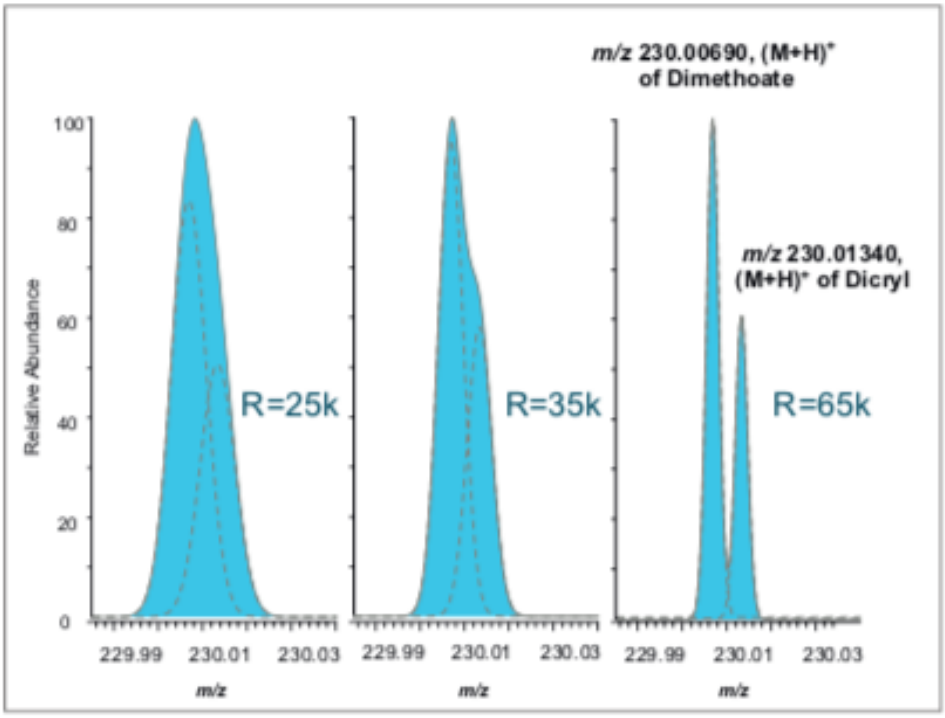

Figure 7- two compounds (m/z 230) detected with different resolutions
With lower resolution the two compounds with similar monoisotopic masses appear in one peak [84]

In metabolomics, where complex biological samples are analysed and cause background ions from the matrix, high resolution is from high importance. Therefore, this thesis uses Q-Exactive™ Focus hybrid quadrupole Orbitrap™ mass spectrometry (Thermo Scientific), which is widely used in untargeted metabolomics due to high resolution (R=70 0000) and mass accuracy.

In a Q-Exactive™ ions and neutrals enter the mass spectrometer from the ion source. A bent flatpole prevents neutrals to enter the quadrupole where m/z values can be preselected and only the corresponding ions can pass and enter the C-trap which shoots the ions into the orbitrap mass analyser or send them to the collision cell for fragmentation and MS/MS analysis. The fragments are sent back to the C-trap and further to the orbitrap mass analyser for detection [76]. In an orbitrap mass analyser, ions oscillate around a spindle in trapped orbits causing an ion separation depending on its individual frequency. An image current of these oscillation frequencies can be measured and transformed into a mass spectrum by fast Fourier transform of the current. [76,85,86] (figure 8). To accurately identify a compound and gain structural information, producing fragmentation patterns with MS/MS on top of received retention

time and mass to charge ratio, is advised to narrow down the list of potential metabolite identifications [75]. The fragmentation patterns of the tandem MS can be compared to standards in spectral databases to confirm the identification [87].

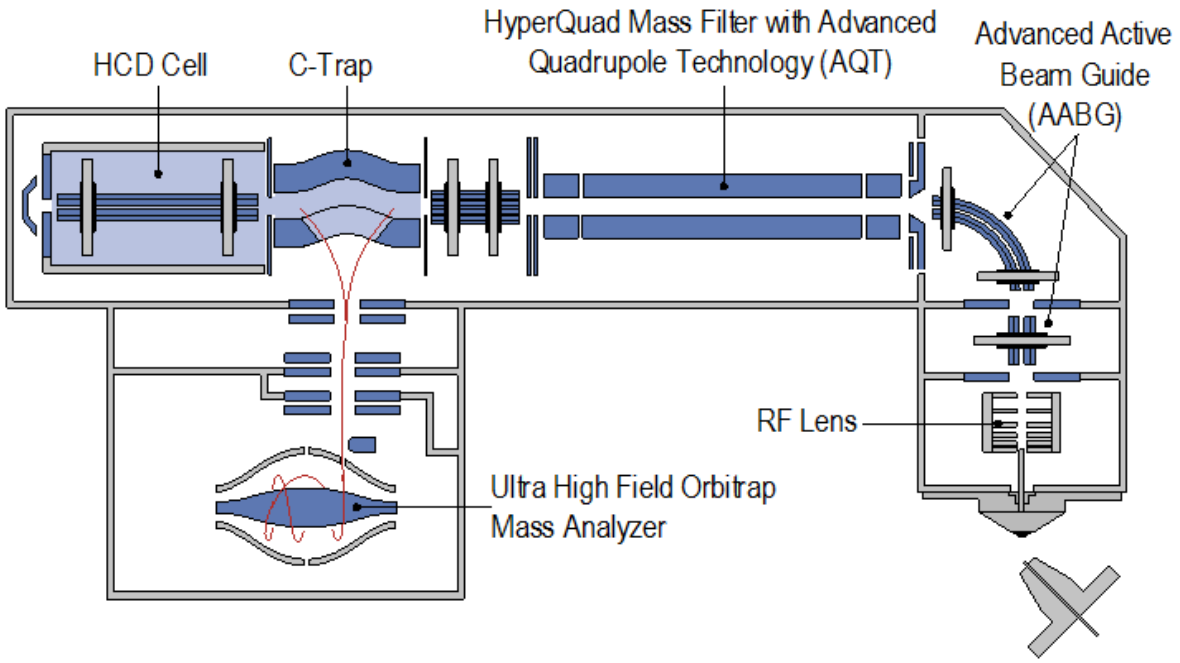

Figure 8- schematic of the Thermo Scientific™ Q-Exactive™ mass spectrometer
A bent flatpole, a quadropole mass filter, a C-trap a higher collision cell induced dissociation cell (HCD) and an orbitrap mass analyser are the important elements of a Q-Exactive™ mass analyser

1.2.3 Metabolomic workflow: targeted and untargeted

The main steps for a metabolomics approach in biomarker discovery can be summarized in three parts: data collection, data processing and data analysis (figure 9).

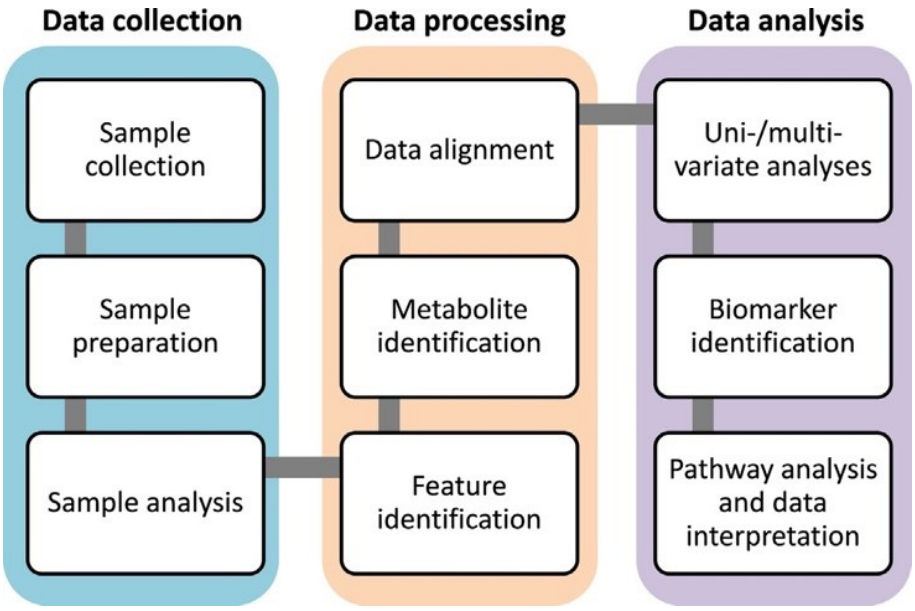

Figure 9- steps for biomarker discovery in metabolomics [88]
The backbone of biomarker discovery by metabolomic is built by collection, processing and analyzing of huge amount of data.

Samples can be collected from several biological fluids, like tissue, blood, serum, plasma, urine, cero spinal fluid or saliva [89]. Sample preparation depends on the sample type and analytical tool applied but should be as simple and non-destructive as possible to avoid metabolite degradation and allow fast analysis [52,90]. Essential steps include a non-selective, wide metabolite range and reproducibility in measurements should be strived for, as the sample preparation impacts the observed metabolite profile as well as the quality of the obtained data [90].Then the samples can be analysed with LC-MS, the raw signal obtained is called a feature and graphed as MS1 in a three-dimension plot with m/z, retention time and intensity on the axes. An ion of interest is isolated and further fragmentated by collision energy resulting in a pattern and used for compound identification (Figure 10)[17].

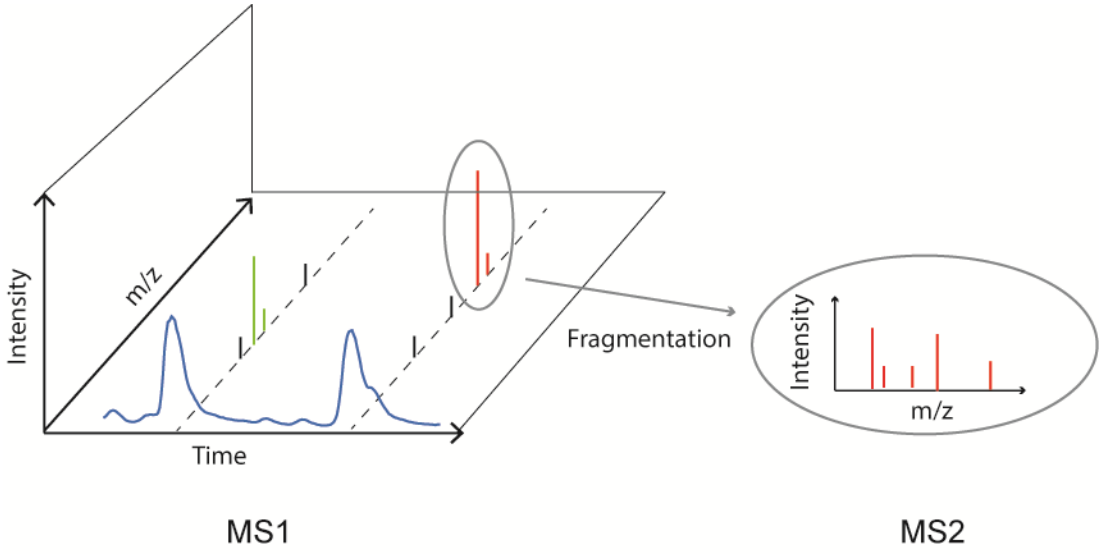

Figure 10- symbolic illustration of LC-MS data [17]
The detected compouds are scaled with their mass to charge ratio plotted against the intensity of the peak. To obtain an MS2 spectrum singular peaks of MS1 can be further fragmented.

Since high resolution MS operate in a scan mode of 50-1000 amu to cover mass ranges of small metabolites, a big set of mass spectra is generated and needs to be mined by noise filtering, peak picking, peak alignment and peak integration, to compress the detection into one feature. This is usually carried out by a MS manufacturer software which also carries out feature identification with web based resources [52] Compounds are identified by comparing the accurate mass , retention time and MS2 fragmentation patterns with metabolite libraries and databases like HMDB (https://hmdb.ca), MzCloud (https://www.mzcloud.org) or Chemspider (http://www.chemspider.com) [52,91]. A major bottleneck is that the human metabolome is by far not completely identified and unknown metabolites cannot be matched. Due to incompleteness of databases the accuracy of identification is limited and therefore recommended to validate with standards because if a substance is not registered in a database it is matched with the next most similar component and incorrectly identified [92]. Although confident identification remains a challenge, the most reliable way is the comparison to authentic standards, identification on m/z ratios, retention time and MS2 pattern. This grants important guidance for metabolite verification by reducing the number of possible identifications. [93]

Metabolomics can be accessed targeted or untargeted. Targeted metabolomics is considered as hypothesis driven rather than hypothesis generating because it focuses on specific pathways that play an important role for the function of a cell and quantifies preselected metabolites. This way only a selected amount of the whole metabolite spectrum is considered but with the aim to analyse those in a more sensitive, accurate and quantitative approach. Absolute identification is still a bottleneck in metabolomics, especially in an untargeted approach, in targeted metabolomics isotope labelled internal standards can be incorporated to quantify. Untargeted metabolomics on the other hand strives to cover the full spectrum of a cell's metabolism including known and unknown compounds. It focuses on relative quantifications due to comparing peak intensities as concentration indicator trying to detect unforeseen changes in metabolite concentrations [17,94]. While in targeted approach internal standards and calibration curves with authentic standards can be used for absolute quantification[95],in untargeted analysis it is impossible to have internal standards and calibration curves for all the detected unknown metabolites. Therefore targeted metabolomics is better in terms of quantification, whereas untargeted metabolomics strives for broader coverage [76].

Due to instrumental variations, such as changes in pressure, temperature or column aging, QC samples can be incorporated in the analysis to assure the stability of the analytical system. and correct for signal drifts. QC samples include a mixture of all samples that are analysed and signal drifts can be corrected [96]. Since high throughput technologies generate a large amount of data, statistical analysis, like uni- or multivariate, should be carried out. Principal component analysis (PCA)as unsupervised dimension reduction is the most applied. It reduces the number of variables in the dataset while still retaining most variation of the original data. Without prior knowledge of sample classes, differences in samples based on their metabolite composition and identification of outliers can be achieved [10]. From this, possible biomarkers are suggested and can be correlated to the underlying metabolic pathway to allow meaningful interpretation [88]. Identification of putative biomarkers is only the starting point to reach clinical diagnostic a validation is the essential following step [49].

1.3 Aim of the thesis

The main objective of this thesis is to establish a methodology for profiling cancer cell media and identifying excreted metabolites in an untargeted approach by using LC-MS metabolomics. The second objective is to evaluate metabolic patterns specific for

breast cancer cells through observing significantly changing metabolites from fresh and spent media which may support discovery of potential biomarkers and novel disease mechanisms. Resulting biomarkers can be used as early and non-invasive diagnostic tools, to individualize therapy options or as treatment monitoring in the field of precision medicine.

2 Materials and methods

2.1 Samples

2.1.1 Experiment 1: optimization of sample processing methods

Cell culture medium DMEM (# 41966-029, Invitrogen) supplemented with 10% FBS (Life Technologies) and 1% (v/v) Penicillin and streptavidin (Life Technologies). Breast cancer cell line HCC 1806 was originally obtained from ATCC (ATCC #CRL-2335). Samples of the plain media DMEM were collected and represent the sample day zero (D0). After four days the medium DMEM containing the breast cancer cell line HCC 1806 was collected and represents the sample day four (D4)

2.1.2 Experiment 2: untargeted profiling of cell culture media

Cell culture medium DMEM (# 41966-029, Invitrogen) supplemented with 10% FBS (Life Technologies) and 1% (v/v) Penicillin and streptavidin (Life Technologies). Breast cancer cell line HCC 1806 was originally obtained from ATCC (ATCC #CRL-2335). Samples of the plain media DMEM were collected and represent the sample day zero (D0). After four days the medium DMEM containing the breast cancer cell line HCC 1806 was collected and represents the sample day four (D4)

2.1.3 : Experiment 3: comparing relevant samples of breast cancer

Cell lines HCC1954 cell line was originally obtained from ATCC (ATCC #CRL-2338). The culture medium for this cell line is composed by Roswell Park Memorial Institute 1640 (RPMI1640) medium, no phenol red (#11835-063, Invitrogen) supplemented with 10% (v/v) FBS (Life Technologies), 1% (v/v) Penicillin and streptavidin (Life Technologies), 0.5mL 2-β-Mercaptoethanol (#21985-023, Life Technologies) and 3mL HEPES (Life technologies). Cells were cultured at 37 °C in a 5% CO2 atmosphere and medium were changed twice a week. BT474 cell line was also obtained from ATCC (ATCC #HTB-20) and cell culture was performed with the same method. The medium used with this cell line was RPMI 1640 medium, no phenol red (#11835-063, Invitrogen) supplemented with 10% (v/v) FBS (Life Technologies) and 1% (v/v)

Penicillin (Basal RPMI). MDA-MB-231cell line was obtained from ATCC (#HTB-26) and cell culture method was as described as before. The medium used was DMEM (# 41966-029, Invitrogen) supplemented with 10% FBS (Life Technologies) and 1% (v/v) Penicillin and streptavidin (Life Technologies).

Table 3- cell line classification
Breast cancer subtypes based on the expression of surface receptor for hormones and growth factors. Estrogen Receptor (ER) is expressed only on the luminal subtypes while the Progesterone Receptor (PR) has phased expression on the same subtypes. Her2 is expressed on Luminal B and Her2 overexpressed breast cancers.

CELL LINE	BC SUBTYPE	SOURCE	MEDIA
HCC 1806	Triple negative	ATCC CRL-2335	RPMI+FBS+Penstrep+ HEPES+2-β-Mercaptoethanol
HCC 1954	Her2 +	ATCC CRL-2338	RPMI+FBS+Penstrep+HEPES + 2-β-Mercaptoethanol
BT474	Luminal B	ATCC HTB-20	RPMI+ FBS+Penstrep
MDA-MB-231	Triple negative	ATCC HTB-26	DMEM+ FBS+ Penstrep

From three different cell lines HCC1806, HCC1954, BT474, MDA-MB-231 (table 2) fresh media (day 0) and spent media after 4 days (day 4) was collected.

2.2 Sample preparation

2.2.1 Experiment 1:

The cell culture media DMEM was analysed to develop the best precipitation mode. 100 µl DMEM were placed in four different 1.5 ml Eppendorf vials. In two 300 µl MeOH was added and in the other two 400µl ACN. Protein precipitation was performed over night by storing the vials at - 20 °C. The next day all samples were centrifuged at -4 °C 14000 RPM for 10 minutes. The supernatant was collected from one MeOH and one ACN vial for directly injecting 5µl and 10 µl into an amide column. The other MeOH and ACN vial were lyophilizated using a vacufuge (Eppendorf). The dried samples were resuspended in 200µl H2O and 5 µl injected into a reversed phase C18 column prior mass spectrometry.

2.2.2 Experiment 2 and 3:

Supernatant form cancer cell line was collected and centrifuged for 5min 300G at 4°C, again collecting the supernatant and centrifuging 5 min 1000G at the previous temperature and storing the samples at minus 80°C.
Frozen samples were thawed and 100 µl were placed in a 2ml Eppendorf vial. Protein precipitation was performed over night by using 300 µl MeOH and stored at - 20 °C. The next day sample was centrifuged at -4 °C 14000 RPM for 10 minutes. The supernatant was collected and 200 µl of the supernatant were placed into a new Eppendorf vial and the remaining 200 µl into another Eppendorf vial. One Eppendorf vial was used for direct injection (10µl) into the amide column the other one was lyophilizated using a vacufuge. The dried samples were resuspended in 200 µl H2O and injected (5µl) into a reversed phase C18 column prior mass spectrometry.

2.3 Liquid Chromatography and mass spectrometry conditions

2.3.1 Reagents

Following solvents have been used for LC-MS analysis, including 0,1% formic acid in water and 0.1% formic acid in ACN Optima LC-MS (Fisher Chemical), ACN and MeOH Optima LC-MS (Fisher Chemical), Ammonium Acetate 0.25M HPLC (Fisher Chemical).

2.3.2 C18 Column

Chromatographic analysis was performed on an UltiMate 3000 UHPLC (Thermo Scientific). The separation was performed using a Waters XBridge column C18 (2.1x150mm, 3.5 µm particle size, P/N 186003023. The mobile phase A was water with 0.1% formic acid (v/v), and mobile phase B was acetonitrile with 0.1% formic acid (v/v) (Optima LC/MS Grade, Fisher Scientific). The gradient program is listed in table 3. The column temperature was maintained at 30°C, and a flow rate of 400 µl/min was used.

Table 4- gradient program for C18 column

TIME (MIN)	MOBILE PHASE A (%)	MOBILE PHASE B (%)
1.0	99	1
13.0	1	99
15.0	1	99
16.0	99	1
20.0	99	1

The data was acquired on Q-Exactive™ Focus hybrid quadrupole Orbitrap™(Thermo Scientific) coupled to UHPLC, using Xcalibur software v.4.0.27.19 (Thermo Scientific). The method consisted of several cycles of Full MS scans (R=70000) followed by 3 ddMS2 scans (R=17500), in positive and negative mode.

The raw MS and MS/MS data was analysed using Compound Discoverer software v2.1(Thermo Scientific) with HighChem HighRes algorithm. The searches were performed against MzCloud and Chemspider database.

2.3.3 Amide Column

Chromatographic analysis was performed with an UltiMate 3000 UHPLC (Thermo Scientific). The separation was performed using a Waters Acquity UPLC BEH Amide 2.1x150mm, 2.5 µm particle size, P/N 186003023. The mobile phase A was ammonium acetate with acetic acid (pH3) and mobile phase B was acetonitrile (LC/MS Grade, Fisher Scientific). The gradient program is listed in table 4. The column temperature was maintained at 30°C, and a flow rate of 350 µl/min was used.

Table 5- gradient program for amide column

TIME	MOBILE PHASE A (%)	MOBILE PHASE B (%)
1.0	10	90
6.0	50	50
7.0	60	40
9.0	60	40
10.0	10	90
20.0	10	90

For the untargeted experiment the amide column as well as the C18 column were selected to cover polar and apolar metabolites. All the samples from the untargeted approach were analysed in positive and negative mode in the MS.

2.3.4 Workflow for Compound Discoverer software

The workflow for metabolite identification is demonstrated in figure 11.
A node-based processing workflow was custom-built in Compound Discoverer software v 2.1 (Thermo Scientific) to search and identify the metabolites in an untargeted approach.

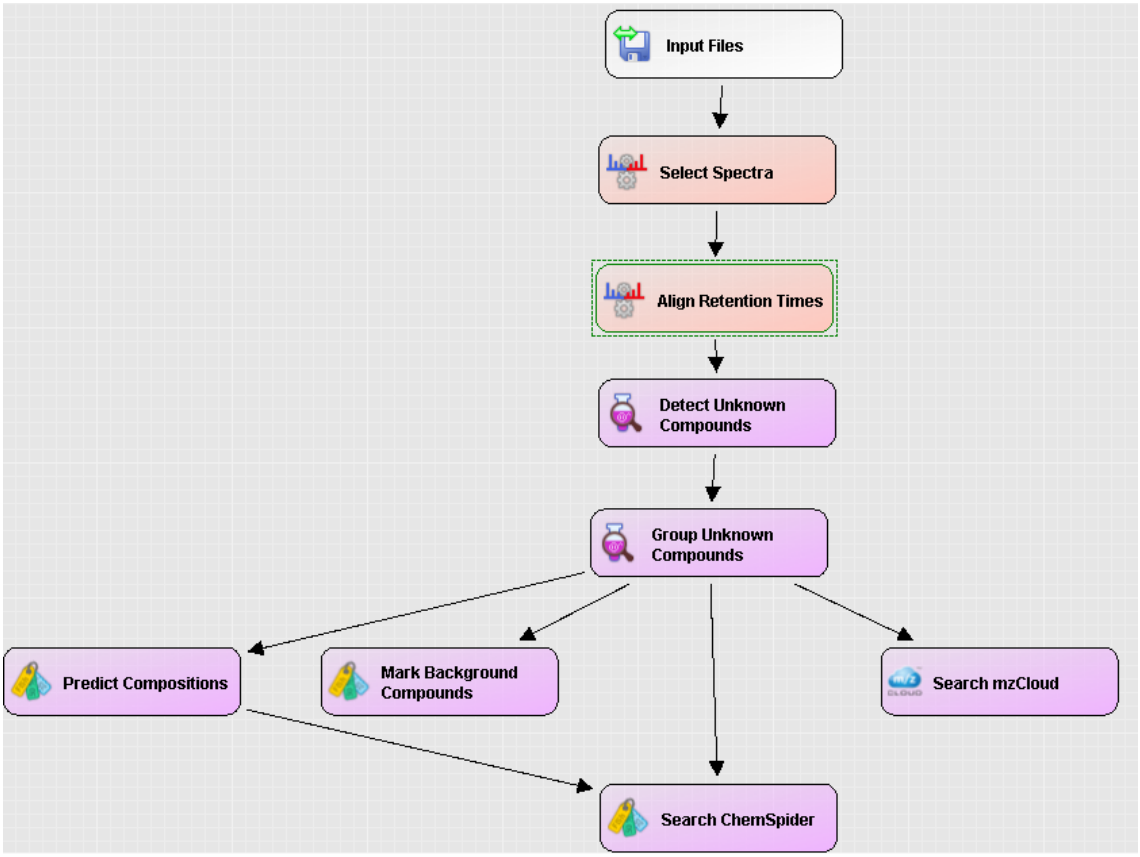

Figure 11- workflow for metabolite identification with Compound Discoverer software.
After file input, spectra selction, alignment of retention times, unknown compounds can be detected and identified with mzCloud or chemspider.

The search description included detecting and identifying all compounds in a single sample (with ddMS2), even compounds with very low abundance. The software performs unknown compound detection and predicts elemental composition for all compounds. It Identifies compounds using mzCloud (ddMsS2) and chemspider (formula or exact mass) and performs similarity search for all compounds with ddMS2 data using mzCloud. The intensity threshold for unknown compound detection was set

extremely low and peak detection filters are turned off to enable the detection of compounds with very low abundance. Important general settings for peak detection included a mass tolerance of 3 ppm, intensity tolerance of 30 and a minimum peak intensity of 1000000.

Raw data was selected and sent to “select spectra” node which forwards the data to align the retention time. Here retention times were chromatographically aligned with a specified alignment algorithm. The “detect unknown compounds” node detects unknown compounds using component elucidator algorithm. Compounds with the same molecular weight with no retention time shift higher than 0.1 minute were grouped together in the “group unknown compounds” node, parameters shown in figure 12.

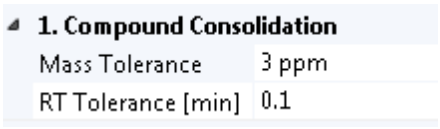

Figure 12- group unknown compounds parameters with Compound Discoverer software.
Compounds with the same molecular weight with a mass tolerance of 3 ppm and with a retention time shift lower or equal 0,1 minutes can be grouped together.

In the “predict compositions” node the elemental composition of unknown compounds was predicted by using isotope pattern matching and a set of prediction algorithms. Then predicted compositions and matching masses were searched in the chemspider database. The “mark background” node is used to eliminate compounds present in the blank samples. For the identifications in chemspider and mzCloud a filter is applied to correspond the results to mzCloud full match with scores ≥80% and full match witch chemspider database shown in figure 13. This filter is applied in experiment 3, analysing relevant samples of breast cancer.

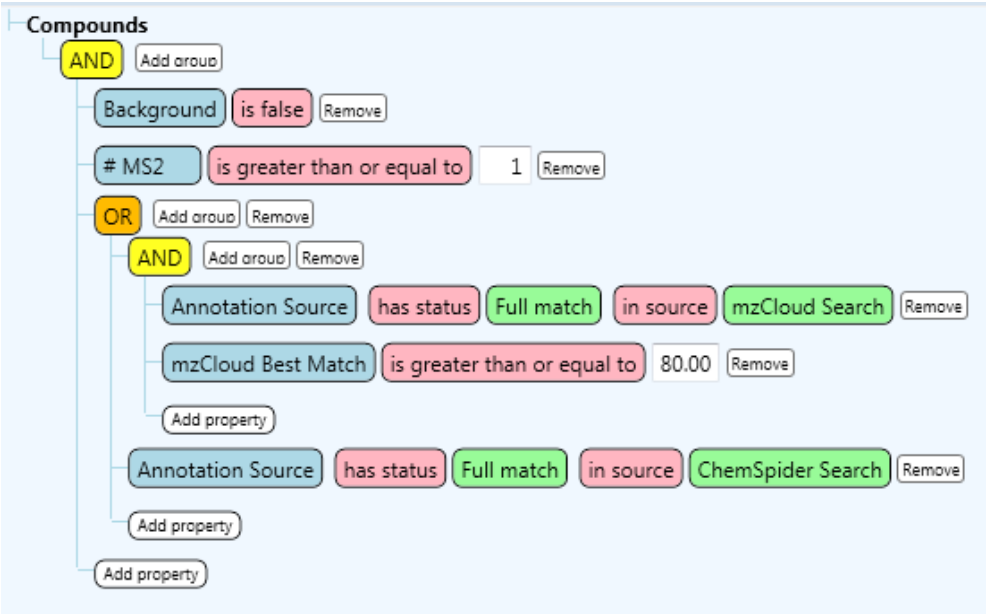

Figure 13- chemspider and mzCloud filter with Compound Discoverer software.
Peaks are identified with full match with scores ≥80% in mzCloud and chemspider.

The data was filtered by only mzCloud greater than 80% match giving more confident identifications based on the ddMS2 but eliminating a lot of identifications, demonstrated in figure 14.

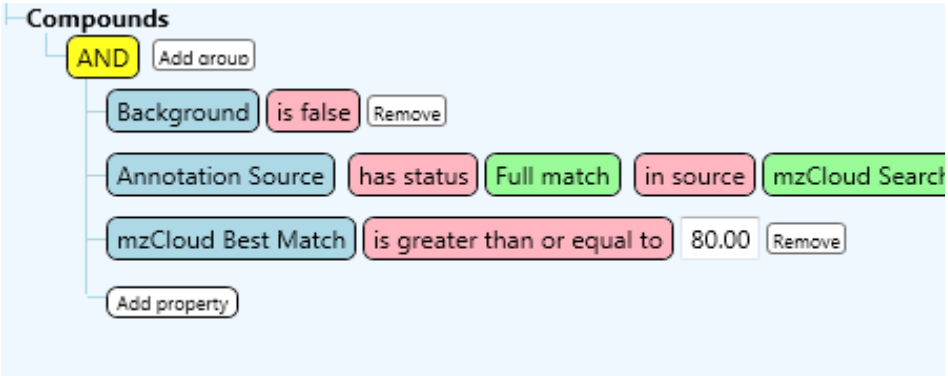

Figure 14- mzCloud filter with Compound Discoverer software.
Peaks are identified with full match with scores ≥80% in mzCloud

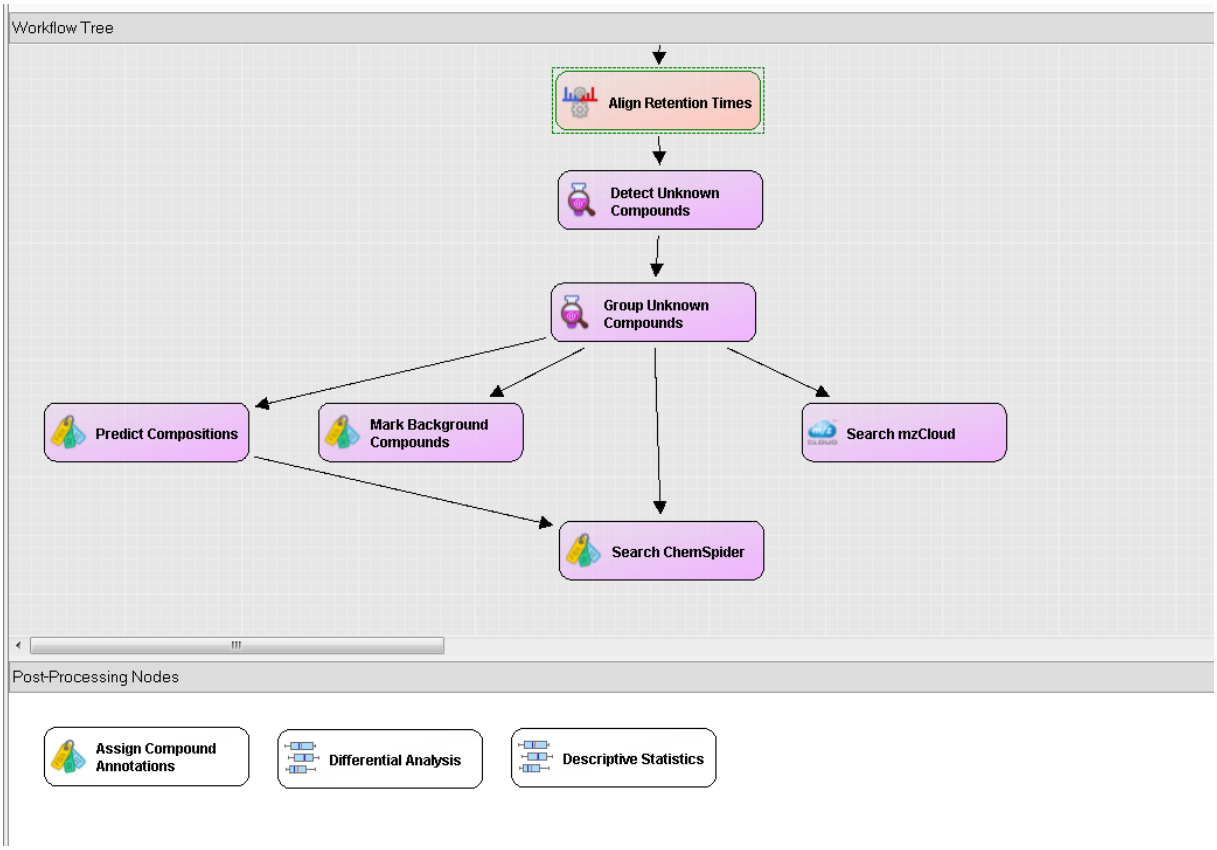

Figure 15- workflow tree and post-processing nodes with Compound Discoverer software
The workflow tree ends with assigning the name, formula and structure of the compounds as well as differential analysis and descriptive statistics.

After the workflow tree was completed the post processing nodes followed, as described in figure 15. The “assign compound annotations” assigns the name, the formula and the structure annotations to the compounds. The “differential analysis” node created the group areas and group CV columns in the compounds table and calculated Log 2Fold change, ratio, p-Values of per group ratios by an analysis of variance (ANOVA) with tukey post hoc tests, adjusted p-values for the false discovery rate, for the differential analysis. The descriptive statistics tool calculated the mean area and median area.

2.4 Experiment 3: Comparing relevant samples of breast cancer cell lines

Fresh media (D0) and spent media (D4) of two biological replicates of breast cancer cell lines ACC 1806, ACC1904, BT474 and one sample of MD-MB231 were analysed using a C18 and an amide column (3 technical replicates positive and negative mode).

This metabolomics experiment involves four different cell lines characterizing 3 different subtypes with distinct receptor expressions, as shown in table 6.

Table 6- subtype specific receptor expression [33]

SUBTYPE	CELL LINE	BIOLOGICAL REPLICATE	ER EXPRESSION	PR EXPRESSION	HER2 EXPRESSION
LUMINAL A	BT474	2	+	+	+
HER2	ACC 1954	2	-	-	+
TNB	ACC 1806	2	-	-	-
	MD-MB231	1	-	-	-

2.4.1 Data analysis

Differential analysis of the injected samples was applied to evaluate the significance of relative metabolite concentrations between fresh and spent media with Log2Fold change ratios and p-Values (ANOVA). Untargeted metabolite measurements were applied and a comparison of ion intensities between sample D0 and D4 revealed a significant increase with a fold change FC ≥1 and p-value 0.05 or decrease with FC ≤-1 and p-value 0.05 over time.

Afterwards Perseus software was used to perform multivariate statistics in form of principal component analysis (PCA). This dimension reduction method was intended to identify sample clustering and extract main metabolites that cause variance and further reveal group specific metabolic signatures of the different cell lines. The PCA is visualised as score plot with each point representing an individual sample and marked with different colours to differentiate the cell line. When samples group together it can be assumed that a similar metabolic phenotype is the reason.

Then the loadings of the PC were exported to an excel sheet to determine the metabolites responsible for causing a separation of the samples in the PCA. A

percentile served as a threshold to select the most extreme metabolites, meaning all lower than 2,5% and higher than 97,5%.

Perseus was also used to perform cluster analysis in form of heatmaps aiming to identify important differences in metabolite concentrations between the samples. The clustering algorithms such as distances and correlation coefficients group the samples so that the similar ones appear in the same cluster.

Kegg pathways (https://www.genome.jp/kegg/pathway.html) were used to correlate identified metabolites to specific pathways.

3 Results d

3.1 Experiment 1: ptimization of sample processing and spectra acquisition

The aim was to set the conditions for being able to profile cell media and obtaining the most identifications. For this step, cell media DMEM 11039 was analysed. First decision to make was the conditions of the sample preparation. The media required a protein precipitation to minimize sample complexity. Based on the metabolite polarity and the conditions of the column, 4 different approaches were considered. First and second approach were precipitating the sample with MeOH or ACN and directly injecting the supernatant after centrifugation. In the third and fourth approach a lyophilization step after precipitation and resuspending in water was added. This step further concentrates the sample and leads to better solubilization of the mostly polar metabolites.

Experiment 1 involved analysing the cell culture medium DMEM with all four precipitation modes and a C18 column. The following chromatogram (figure 16) pictures the quality of peak separation comparing the sample either precipitated with ACN or MeOH and analysed in positive and negative ionization mode. The time in minutes is graphed on the x-axes and the intensity on the y-axes. An ideal chromatogram should contain symmetrical peaks that are sufficiently separated and have a detectable intensity, which is best achieved with MeOH in positive mode, visible in chromatogram A in figure 16.

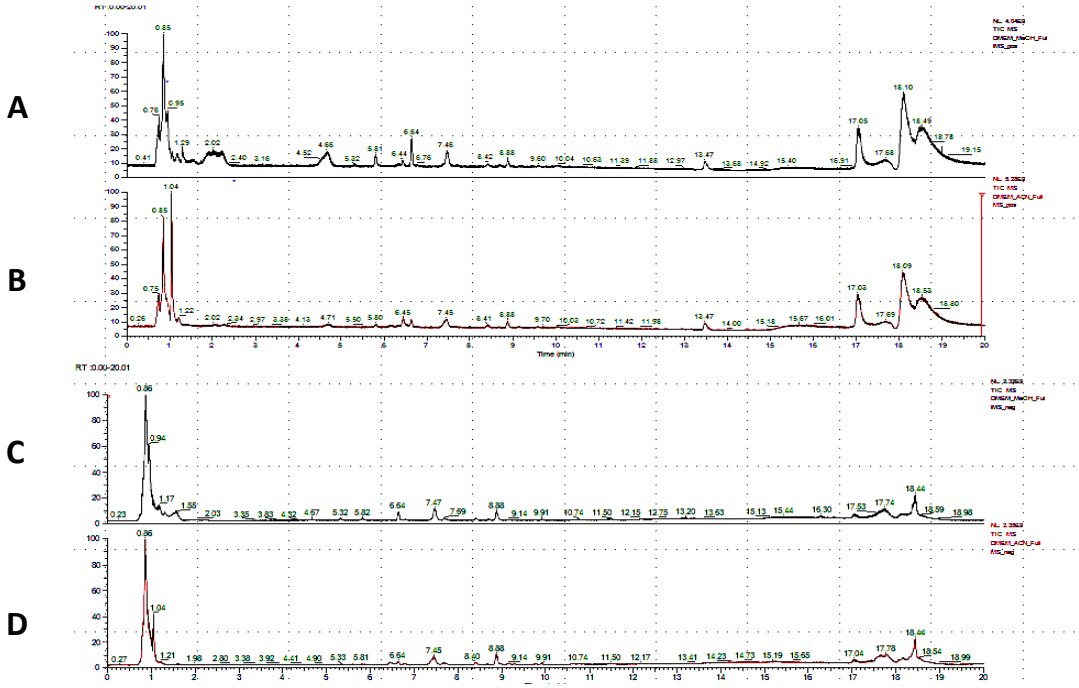

Figure 16- comparison of chromatogram with different precipitating agents (extracted from Compound Discoverer software)
Chromatogram A shows the peak shapes with MeOH in positive mode, B after precipitating with ACN in positive mode, C after precipitating with MeOH in negative mode and D after precipitating with ACN in negative mode.

The following venn diagram, figure 17, demonstrates a difference in the precipitation modes based on the obtained numbers of identifications.

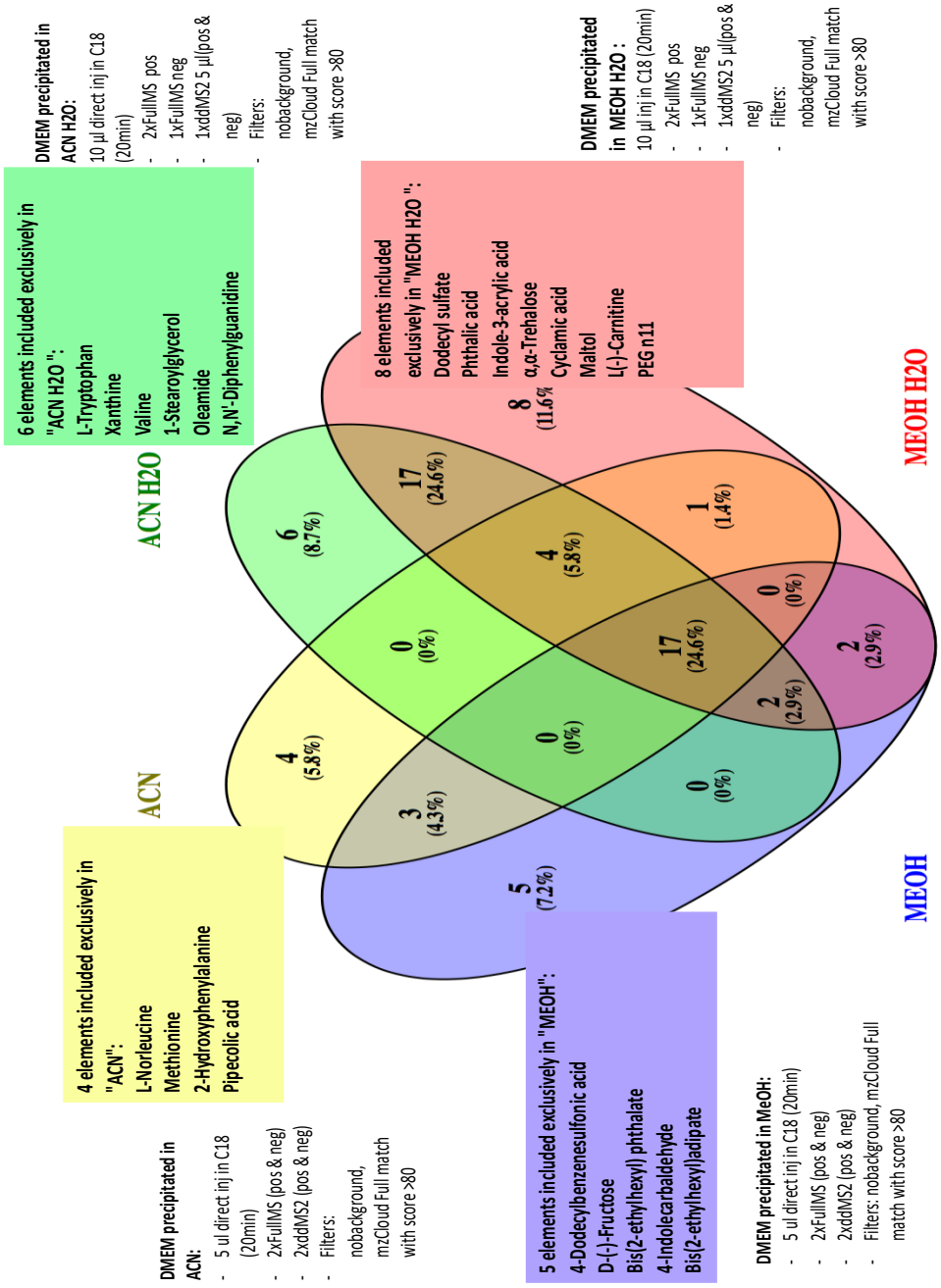

Figure 17- venn diagram- total number of identified metabolites with different precipitations using a C18 column

Injection of DMEM 5µl for MeOH and ACN (two technical replicates positive and negative mode) and 10µl for MeOH/H2O and ACN/H2O (two technical replicates positive mode and one negative mode).The yellow cluster shows 4 exclusively detected metabolites precipitated with ACN, the green cluster shows 6 exclusively detected metabolites with ACN precipitation and resuspending in water, the purple cluster shows 5 exclusively detected metabolites with precipitating with MeOH and the red cluster 8 with precipitating in MeOH and resuspending in water.

The highest amount of identifications, with 8 metabolites exclusively detected, was achieved by MEOH precipitation and resuspending in water named as MeOH/H2O.This can be explained with the polarity of the metabolites of which the majority is water soluble and therefore more peaks can be detected in the water-based precipitation mode. The sample concentration step of drying and reconstituting in 100% H2O increases the metabolome coverage resulting in a higher number of identifications [97]. Another benefit can be documented in the peak shape. Figure 18 shows the peak of the amino acid DL-tryptophan in all four different precipitation modes. The ideal peak should be symmetrical without fronting or tailing and sharp for a defined separation between multiple peaks. The peaks of the samples lyophilized and resuspended in water have higher intensities (figure 18,19). Without lyophilization tilted peaks were obtained which would make it hard to properly quantify the metabolites, because it is difficult to see the start or end of the peak (figure 20,21). The higher intensities can be explained with the higher concentration of the metabolites when lyophilized. Plus having the metabolites highly concentrated in water perfectly fit the composition of the mobile phase. Reversed phase columns, C18, are made of a lipophilic stationary phase, the mobile phase starts polar and gradually increases in organic solvents thus injecting a polar sample matches the polarity of the mobile phase.

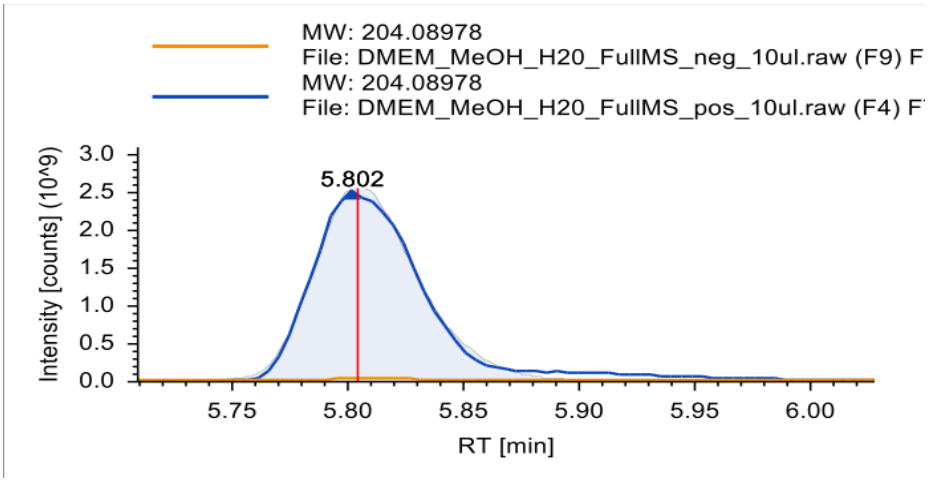

Figure 18-tryptophan peak with MeOH/H2O precipitation and C18 column.
The yellow peak is detected in negative ionization mode and the blue peak in positive ionization mode.

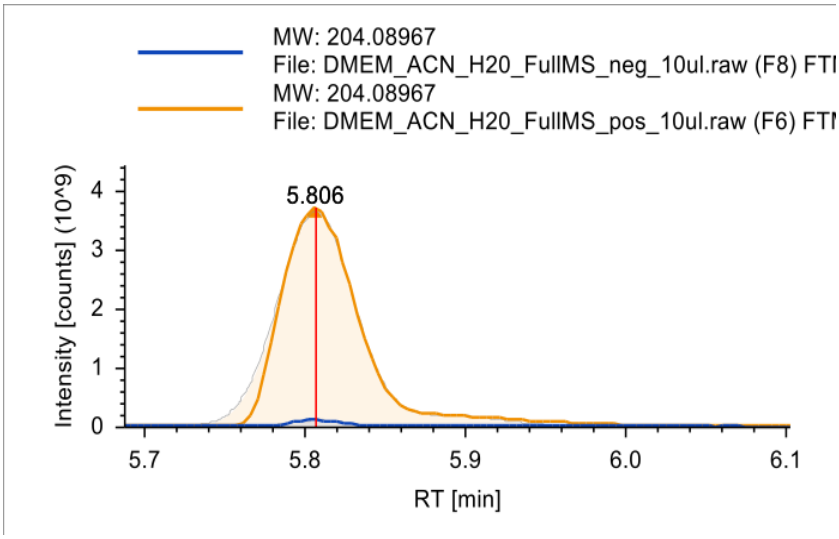

Figure 19- tryptophan peak with ACN/H2O precipitation and C18 column.
The yellow peak is detected in negative ionization mode and the blue peak in positive ionization mode.

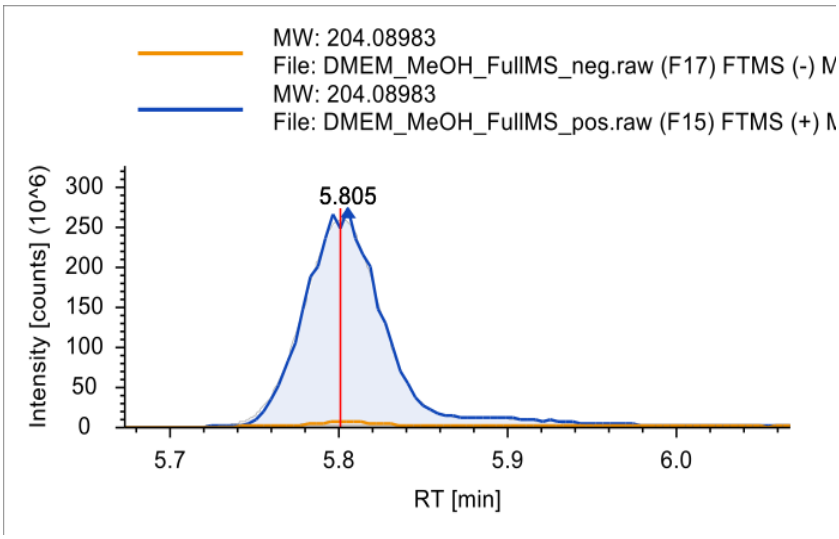

Figure 20- tryptophan peak with MeOH precipitation and C18 column.
The yellow peak is detected in negative ionization mode and the blue peak in positive ionization mode.

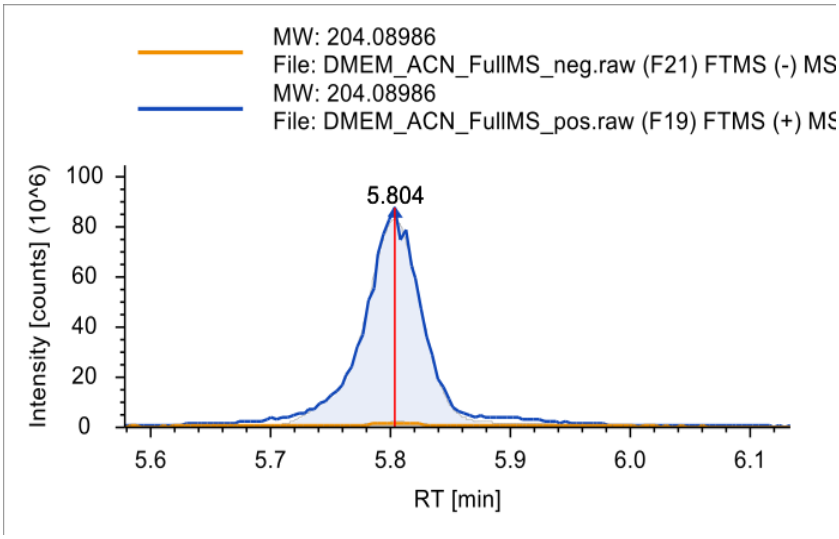

Figure 21 -tryptophan peak with ACN precipitation and C18 column.
The yellow peak is detected in negative ionization mode and the blue peak in positive ionization mode.

Based on these results, the best separation with a C18 column, the sharpest peak shapes and the most identifications were obtained with MeOH precipitation and resuspending in water.

Next the DMEM sample was analysed in all four precipitation modes in the amide column demonstrated in the following venn diagram in figure 22.

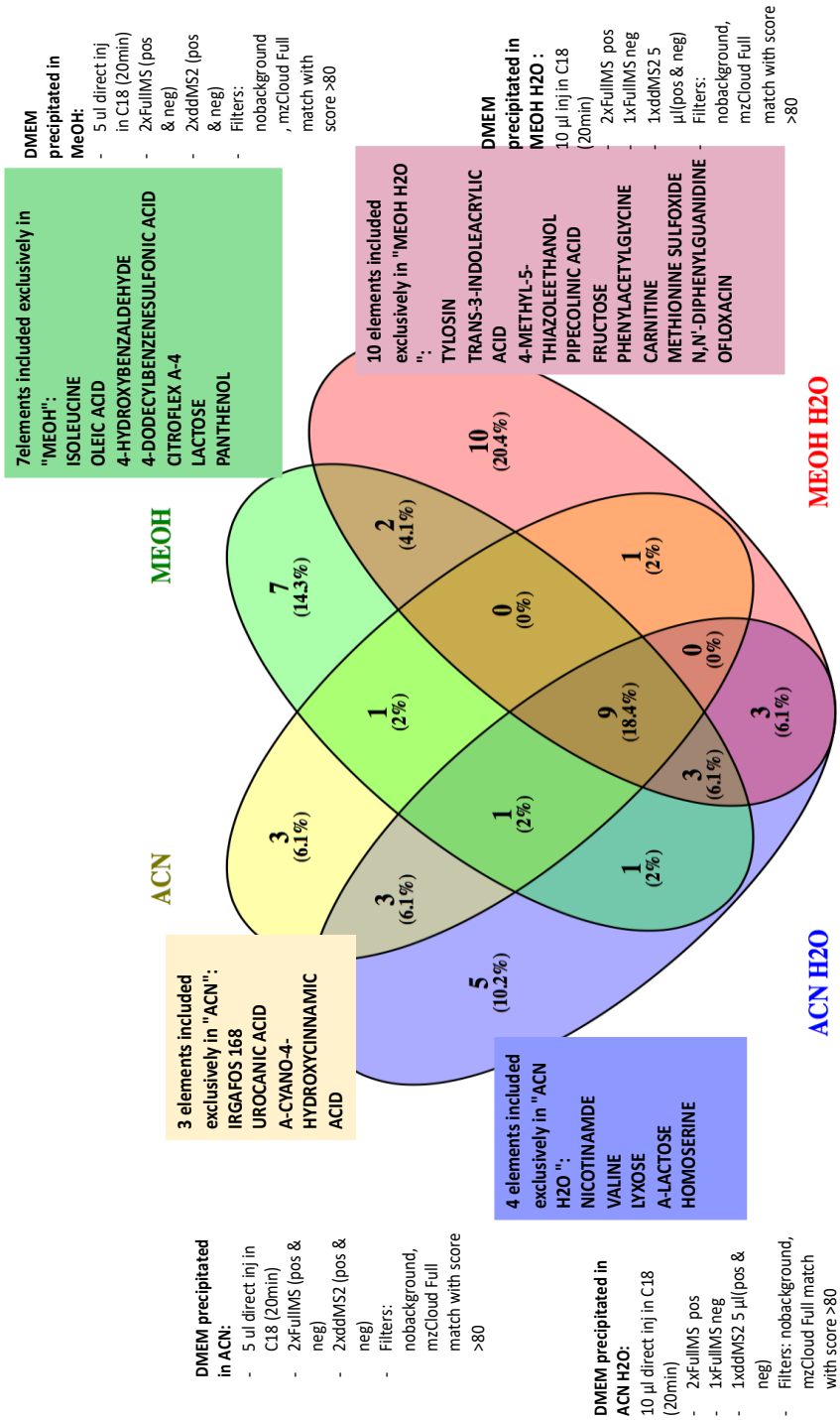

Figure 22- metabolite identifications amide column.
Injection of DMEM 5µl for MeOH and ACN (two technical replicates positive and negative mode) and 10µl for MeOH/H2O and ACN/H2O (two technical replicates positive mode and one negative mode).
The yellow cluster shows 3 exclusively detected metabolites with ACN precipitation, the purple cluster 4 exclusively detected metabolites with MeOH precipitation, the purple cluster 4 exclusively detected metabolites with ACN precipitation and resuspending in water and the red cluster 10 exclusively detected metabolites with MeOH precipitation and water resuspension.

The comparison gives us the information that MeOH/H2O is the best choice regarding to the higher number of identifications with 10 elements exclusively detected. On the

other hand, the detected peak shapes turned out to be very broadened (figure 23) compared to the symmetrical peaks after precipitation with ACN (figure 24) and MeOH (figure 25).

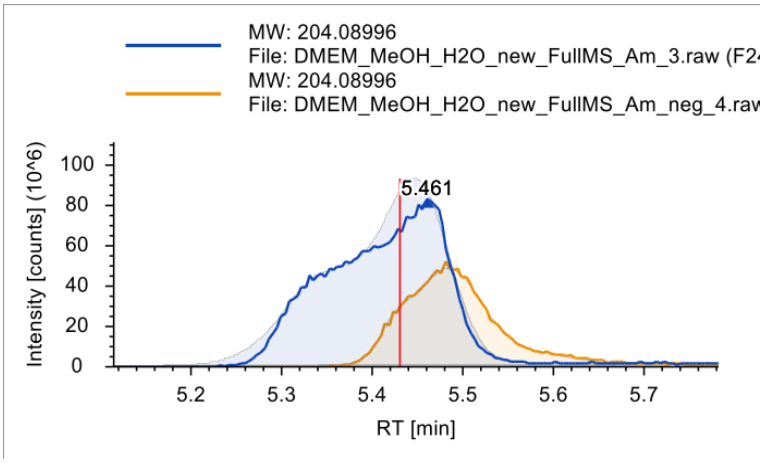

Figure 23- tryptophan peak with MeOH/H2O precipitation and amide column.
The yellow peak is detected in negative ionization mode and the blue peak in positive ionization mode.

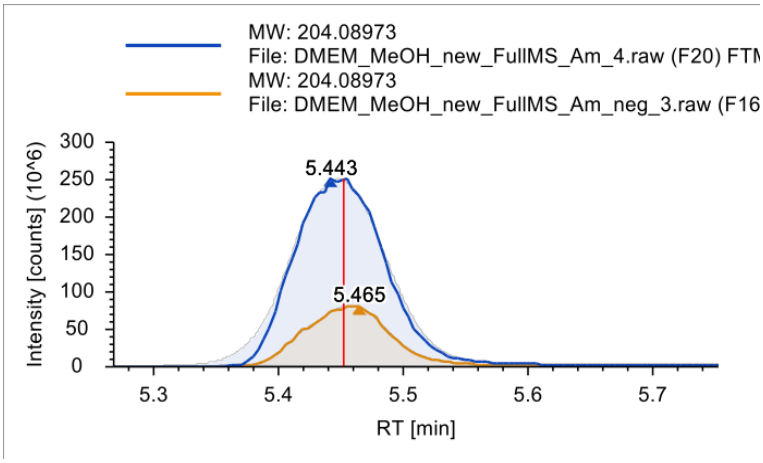

Figure 24- tryptophan peak with ACN precipitation and amide column
The yellow peak is detected in negative ionization mode and the blue peak in positive ionization mode.

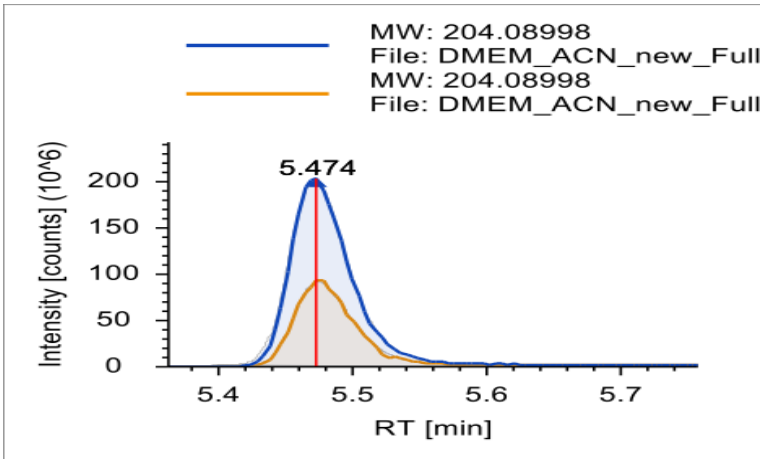

Figure 25- tryptophan peak with MeOH precipitation and amide column.
The yellow peak is detected in negative ionization mode and the blue peak in positive ionization mode.

In the amide column the conditions are reversed, consisting of a polar stationary phase and an organic mobile phase gradually increasing in polar solvents, therefore apolar substances elute before polar. Resuspending the supernatant in water lead to broad and non-quantifiable peak shapes. One reason could be the interference of the water in the sample with the composition of the mobile phase that should increase in polar components instead of starting with polar solvents. Therefore, the better results for the amide column were obtained with precipitating with MeOH and after centrifugation directly injecting the supernatant

Also, in the chromatogram in figure 26 we can observe better peak separation in MeOH/H2O and higher intensities with increasing injection volume.

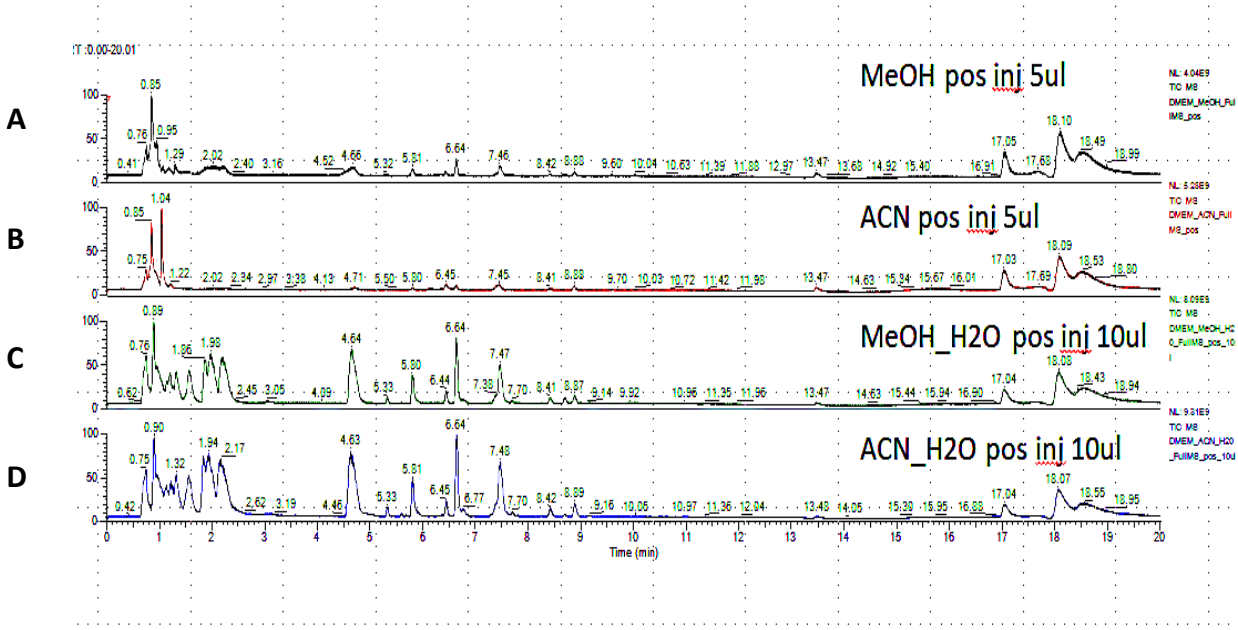

Figure 26- chromatogram comparison precipitation and solubilization.
Chromatogram A is detected with MeOH precipitation and 5 µl injection volume, B with ACN and 5 µl injection volume, C with MeOH/H2O and 10 µl injection volume and D with ACN/H2O and 10 µl injection volume.

The high-resolution Q-Exactive™ allows positive and negative ionization in separate injections, as well as fast polarity switching between positive and negative mode in the same injection allowing faster metabolite screening. Comparing DMEM media based on the acquired identifications, injecting the sample twice and ionizing separately in positive and negative mode gives more identifications, 10 exclusively, as with fast polarity switching in the same run with only 6 exclusive identifications, as demonstrated in figure 27.

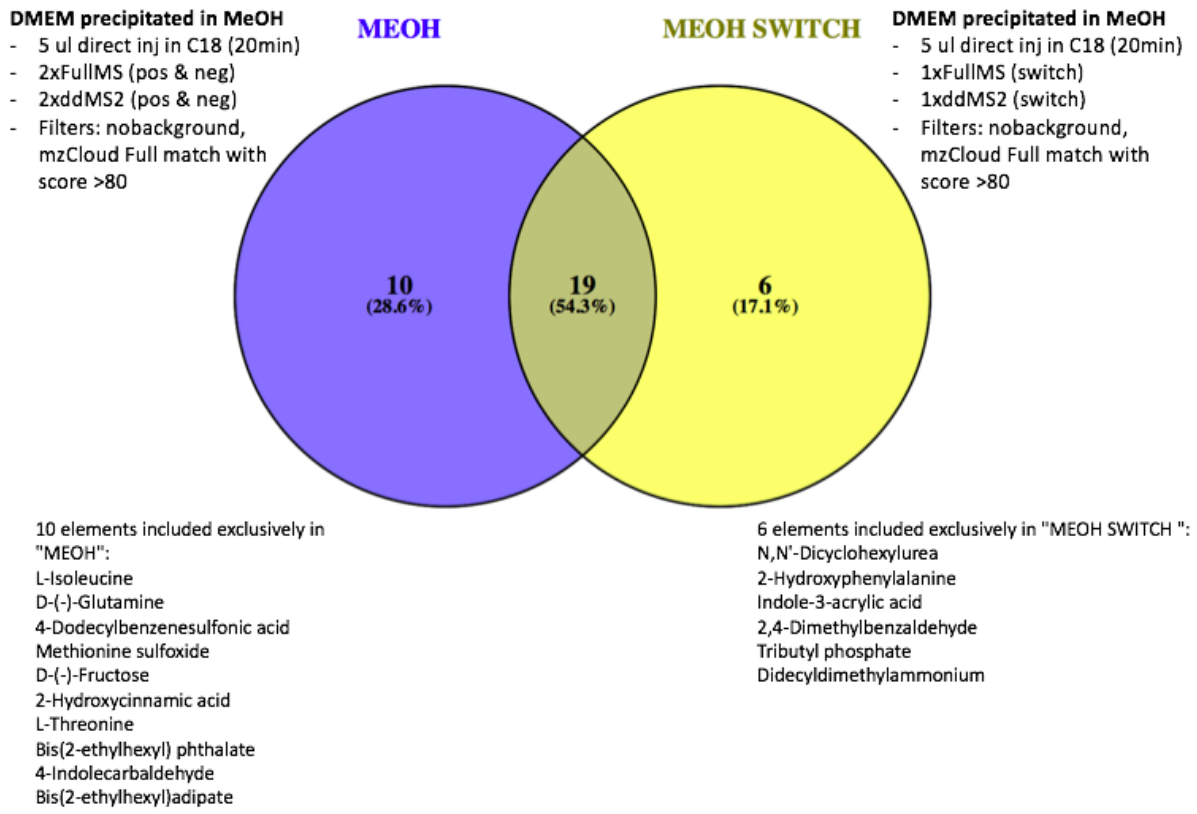

Figure 27- comparison of polarity switch and individual positive and negative mode.
Injection of two technical replicates of DMEM 5µl for MeOH in C18 column for positive and negative mode compared to 5µl MeOH in polarity switch mode. 6 exclusively detected metabolites analyzing the sample with polarity switch and 10 exclusively detected metabolites in separate analysis of negative and positive ionization mode, could be achieved.

The following experiments 2 and 3 were based on these results. A multidimensional LC approach combining RP and HILIC separation, sample preparation with MeOH/H2O for C18 and MeOH for the amide column and high performance MS operating in positive and negative ionization mode, are used to increase peak capacity and metabolite identification in complex samples.

3.2 Experiment 2: Untargeted profiling of cell culture media

A sample of fresh (day0) and spent media (day4) was assayed in LC-MS to observe changes in the metabolism over time. For this analysis a list of metabolites (table 7) involved in central metabolism, including important amino acids, organic acids, vitamins, sugars as well as oncometabolites was implemented and compared with the resulting identifications gathered from LC-MS. Untargeted metabolite measurements were applied and a comparison of ion intensities between sample D0 and D4 revealed

a significant increase with a fold change FC ≥1 and p-value 0.05 or decrease with FC ≤-1 and p-value 0.05 over time.

Table 7- list of relevant metabolites

Breast cancer			
Sugars	Glucose	Organic acids	2-Oxoisocaproate
	myo-Inositol		3-Methyl-2-oxovalerate
Amino Acids	Alanine		Acetate
	Arginine		Choline
	Asparagine		Formate
	Aspartate		Lactate
	Cystine		Nicotinurate
	Glutamate		Pyruvate
	Glutamine	Vitamins	Niacinamide
	Glycine		Pantothenate
	Histidine		Pyridoxine
	Isoleucine	Purinderivates	Hypoxanthin
	Leucine		
	Lysine		
	Methionine		
	Phenylalanine		
	Proline		
	Serine		
	Threonine		
	Tryptophan		
	Tyrosine		
	Valine		
	Pyroglutamate		

The next diagram shows how many identifications matched the relevant metabolites by running the samples in amide and C18 column in positive and negative ionization mode (figure 28).

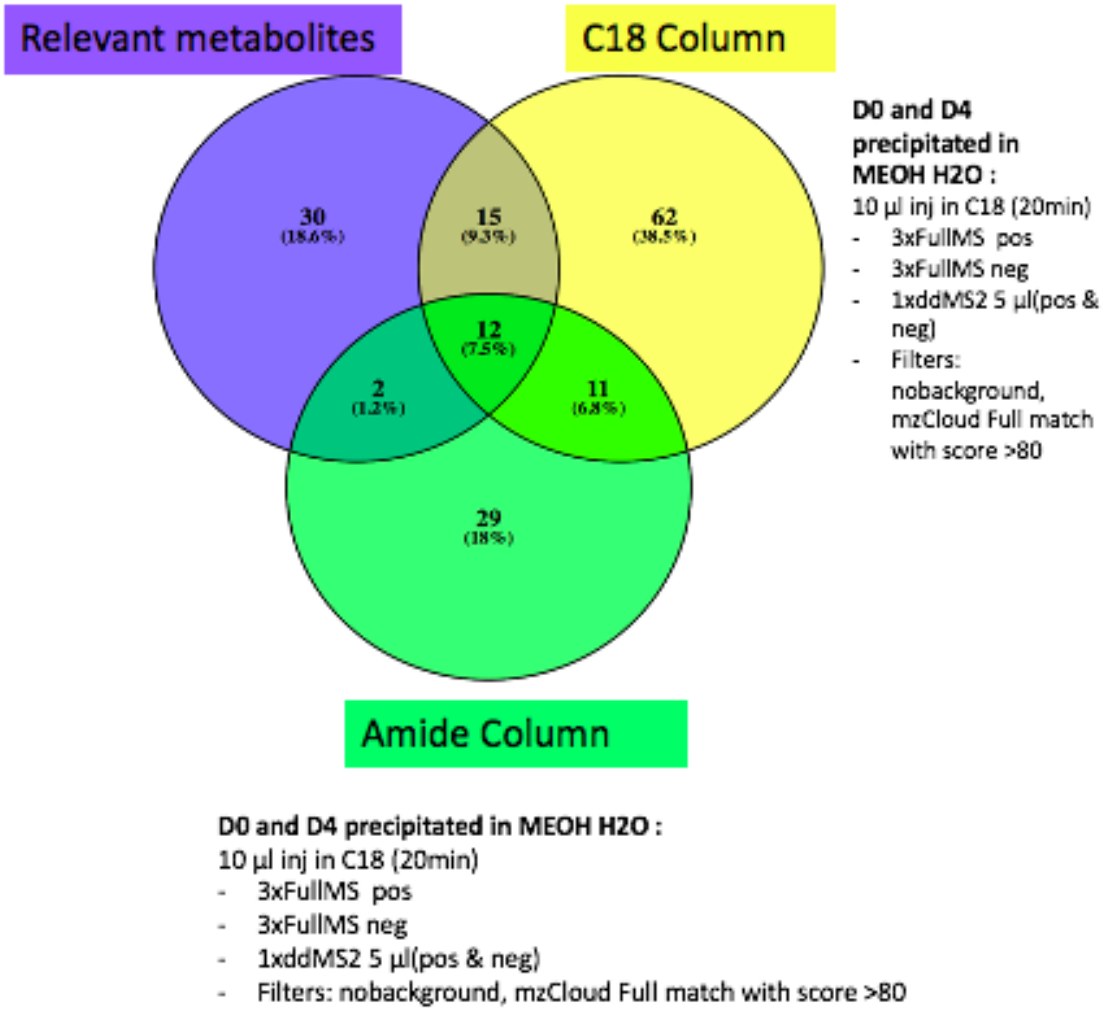

Figure 28- venn diagram showing metabolite identifications
Injection of 10µl MeOH/H2O precipitation (3 technical replicates positive and negative mode) each sample D0 and D4. In total 29 relevant metabolites could be identified in the samples, 15 with the C18 column, 2 with the amide column and 12 with both columns.

A lot of sugars were impossible to accurately identify because MS cannot differentiate between the stereoisomers with the exact same mass, in which case using standards would be a solution to overcome this problem [77,83]. Acetic acid and formic acid are included in the mobile phase and therefore considered background and do not appear as identification. Another reason for some metabolites not appearing might be due to low abundance, not being present in the media and therefore not being metabolized or they were not able be confidently identified with the m/z cloud database.

With Compound Discoverer software, descriptive analysis was used to show significant changes between fresh and spent media of the cancer cell line.

The following volcano plot (figure 29) points out the significantly increasing and decreasing metabolites comparing day 0 (D0) with day 4 (D4), analysed with C18 column.

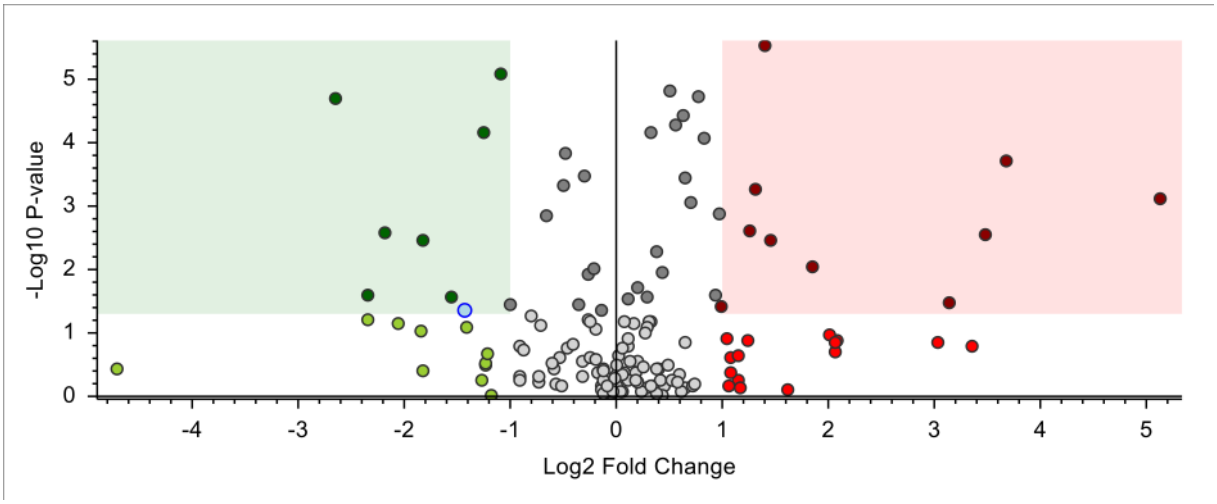

Figure 29- volcano plot positive and negative mode with C18
Injection of 10µl MeOH/H2O precipitation (3 technical replicates positive and negative mode) each sample D0 and D4.The volcano plot is generated with the log2 fold change D0 over D4, therefore the green section and negative values represent the metabolites with higher abundance in day 4 and the red section with the positive values represent the metabolites higher abundant in day 0. Only metabolites with FC≤ -1 or ≥1; p-Value ≤ 0.05 were considered significant.

From D0 to D4 following metabolites seem to be decreasing, including glutamine, and fructose, which may indicate the Warburg effect and glutamine dependency of cancer cells (figure 30 and 31). As well as indole-acetic acid, hypoxanthine and uridine are decreasing. On the other side threonic acid, pyroglutamic acid, y-L-glutamyl-glutamic acid, 7-methylguanine, thiamine and citric acid seem to be increasing after time.

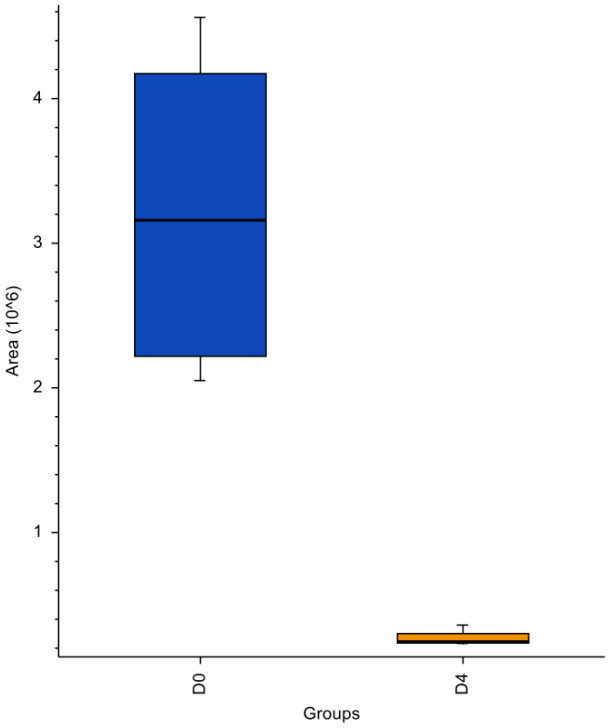

Figure 30- box plot relative concentration of glutamine.
Injection of 10µl MeOH/H2O precipitation (3 technical replicates positive and negative mode) each sample D0 and D4. A decrease of glutamine over time can be observed.

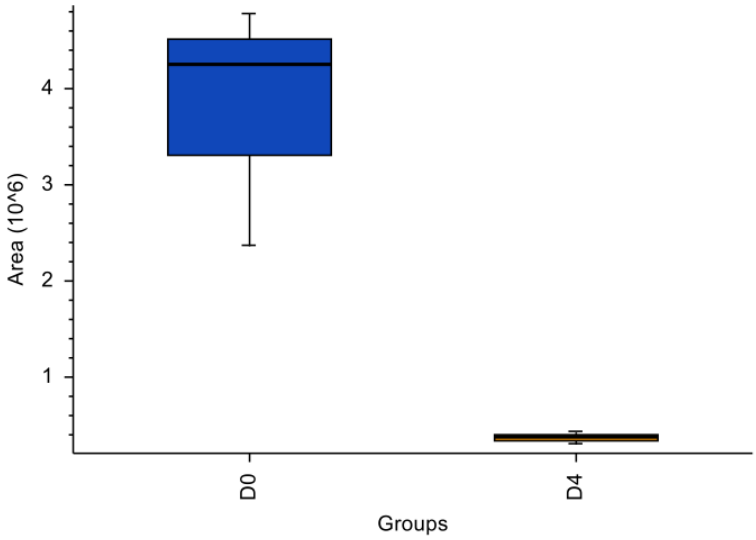

Figure 31- box plot relative concentration of fructose
Injection of 10µl MeOH/H2O precipitation (3 technical replicates positive and negative mode) each sample D0 and D4. A decrease of fructose over time can be observed.

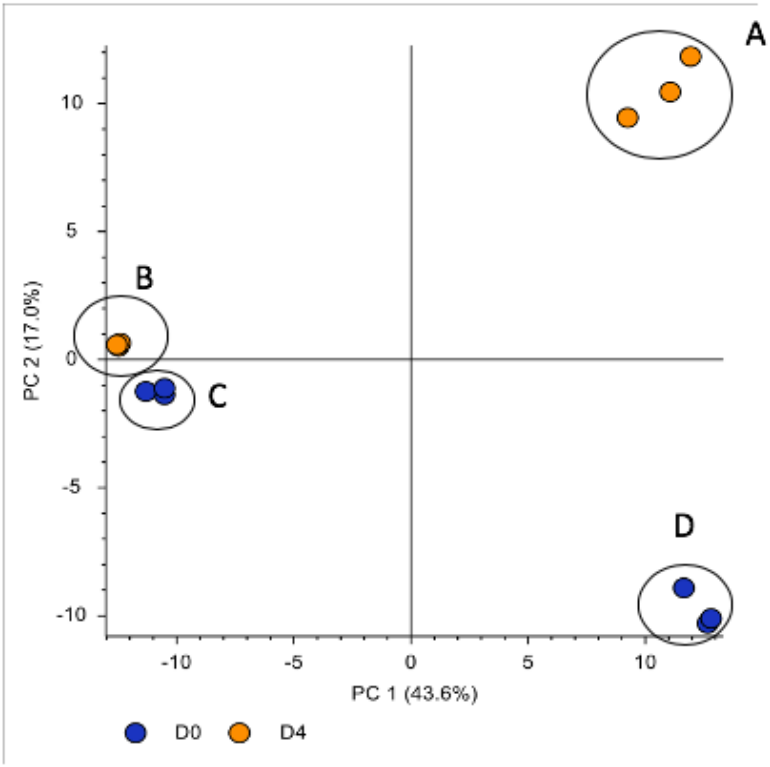

Figure 32- PCA D0 and D4 in positive and negative mode using a C18 column
Injections of 3 technical replicates from fresh media (day0) and spent media (day4) show that the 3 technical replicates in negative mode of the fresh media (C) and spent media (B) cluster together in PC2. The 3 technical replicants in positive mode of the fresh media (D) and spent media (A) can be observed in PC1.

The PCA in figure 32, shows that the 3 injections of the same sample which are considered as technical replicates cluster well together. The different modes cluster apart, samples analysed in negative mode on the left and samples analysed in positive mode on the right. In amide negative mode no significantly changing metabolites were identified. Comparing the two columns, more significantly changing metabolites could be evaluated with C18 column (figure 29), visible due to the less identifications in the volcano plot generated with amide column (figure 33). The metabolites in the green section represent those that are significantly higher concentrated in day 4 (FC ≤ -1; p-Value ≤ 0.05) and thus increasing from D0 to D4. The metabolites in the red section represent those significantly (FC ≥ 1; p-Value ≤ 0.05) higher concentrated in day 0 and thus decreasing from D0 to D4.

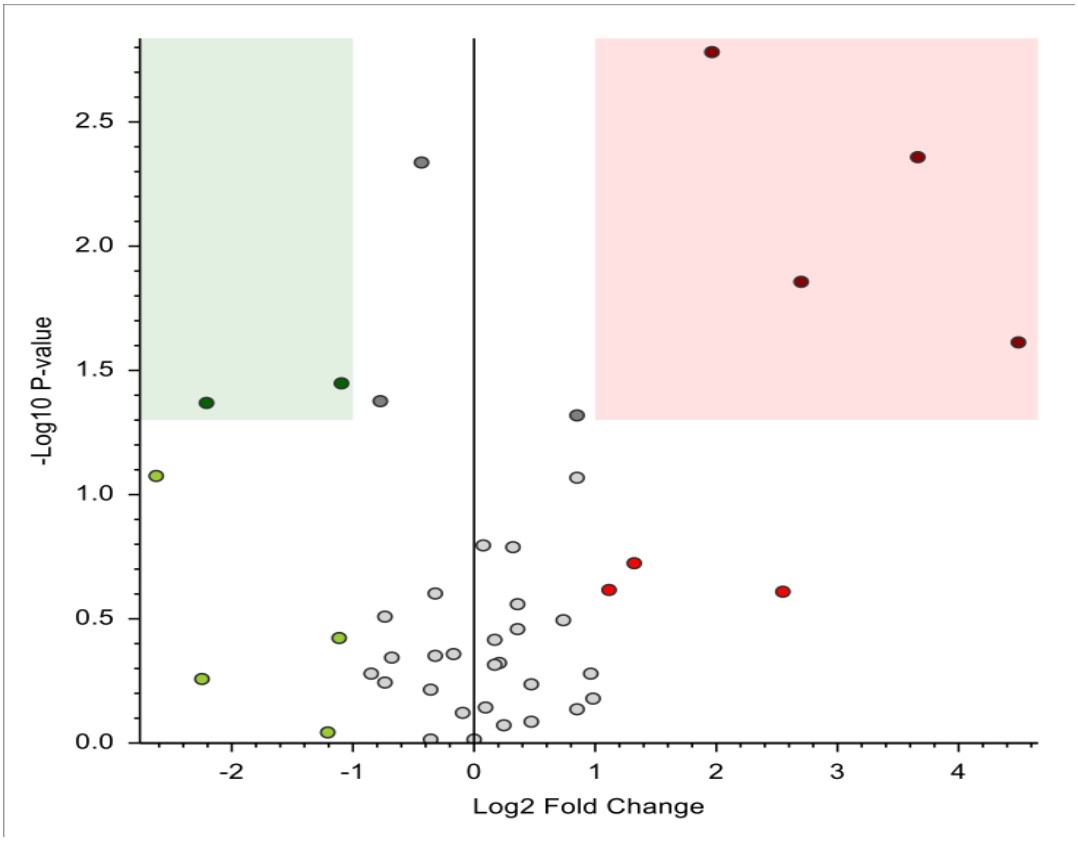

Figure 33- volcano plot showing significant metabolites in positive mode using an Amide column.
Injection of 10µl MeOH/H2O precipitation (3 technical replicates positive and negative mode) each sample D0 and D4. The volcano plot is generated with the log2 fold change D0 over D4, therefore the green section and negative values represent the metabolites with higher abundance in day 4 and the red section with the positive values represent the metabolites higher abundant in day0. Only metabolites with FC≤ -1 or ≥1; p-Value ≤ 0.05 were considered significant.

3.3 Experiment 3: comparing relevant samples of cancer cell lines

This analysis was intended to reveal the metabolic pattern and to discover variations of different breast cancer cell lines. These characteristics were attempted to be correlated with the underlying breast cancer subtype of the cell line. Based on their molecular distinction breast cancer can be divided into four main different subtypes, Luminal A, Luminal B, Her2+ and TNB. Even though, the variation of receptor expression between those subtypes already implicates the therapy approach, the full heterogeneity of this disease has not yet been revealed. Through metabolomics, pathways and up or downregulated metabolites that might be more specific for an individual subtype may be identified and used as biomarkers or treatment target.
This metabolomics experiment involves four different cell lines characterizing 3 different subtypes with distinct receptor expressions (Table 6). Fresh media (D0) and

spent media (D4) of two biological replicates of breast cancer cell lines ACC 1806, ACC1904, BT474 and one sample of MD-MB231 were analysed using a C18 and an amide column (3 technical replicates positive and negative mode).

LC-MS analysis was carried out as described in 2.3. The approach of using two different columns in positive and negative mode aimed to detect the largest possible number of metabolites. In total 3476 features could be identified using an amide column for targeting polar compound classes and a C18 column for rather lipophilic ones, both in positive and negative ionization mode. Out of this 3476 features a total number of 1952 of unique identifications was obtained. To shorten the list of identifications, descriptive analysis was carried out by Compound Discoverer software. Only significantly increasing and decreasing metabolites from fresh and spent media of the cell line supernatant (FC≤ -1.5 or ≥1.5; p-Value ≤ 0.05) were filtered out and further analysed. This reduced the list to a total of 1592 identifications (figure 34).

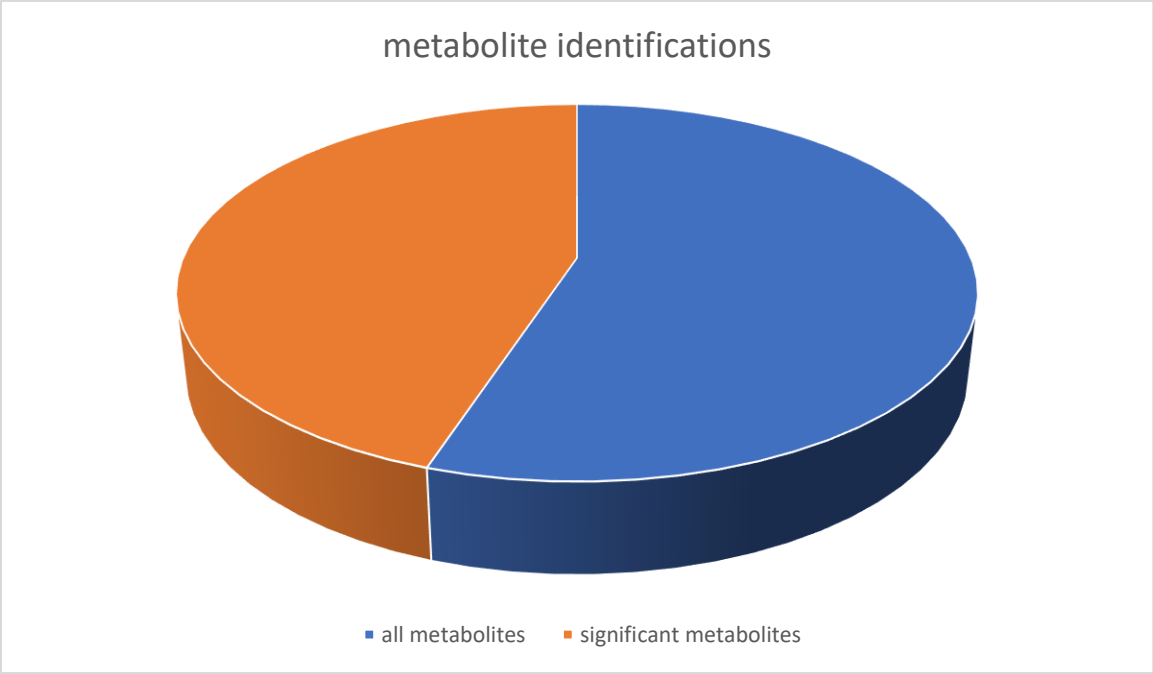

Figure 34- pie chart symbolizing the amount of significantly in or decreasing metabolites over time.
Two biological replicates of breast cancer cell lines ACC 1806, ACC1904, BT474 and one sample of MD-MB231 were analysed using a C18 and an amide column (3 technical replicates positive an
d negative mode). Only metabolites with (FC≤ -1.5 or ≥1.5; p-Value ≤ 0.05) were considered significant.

Among the 1592 identifications, relevant significant metabolites correspond to the chemical classes of fatty acids, organic acids, amino acids and the rest consisting of metabolites from various pathways such as carbohydrate, vitamin, urea, or components of cell culture media or of the solvents used (figure 35).

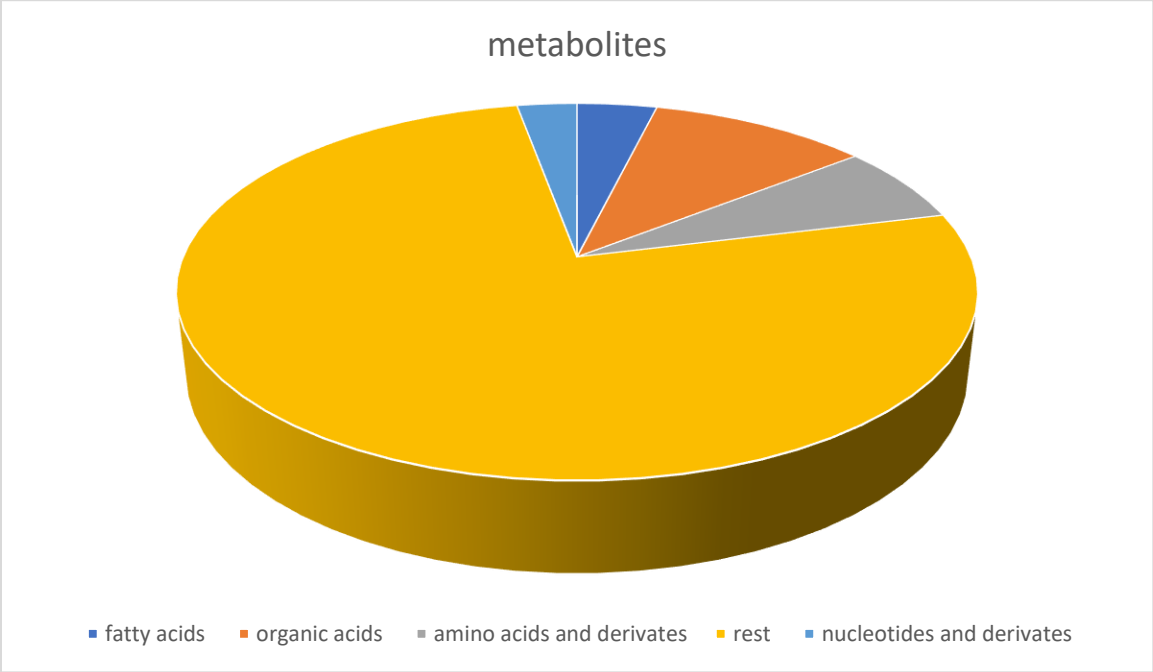

Figure 35- compound classes of identified metabolites.
4% belong to the class of fatty acids, 6% amino acids, 11% organic acids and 79% of sugars, solvents, and unknown compounds

Using the shortened list of 1592 identified metabolites that significantly change in concentration from D0 to D4 a venn diagram (figure 36) was generated to compare the metabolite variation between the relevant breast cancer samples. This Figure illustrates that different cell lines consume or produce different metabolites. A total amount of 429 metabolites could be identified in all cell lines. We can observe that TNB and Her2 have more metabolites in common (134) than TNB and Luminal B (70) and Luminal B and Her2 (75). This could be explained with the comparable aggressiveness of TNB and HER2 breast cancer subtype.

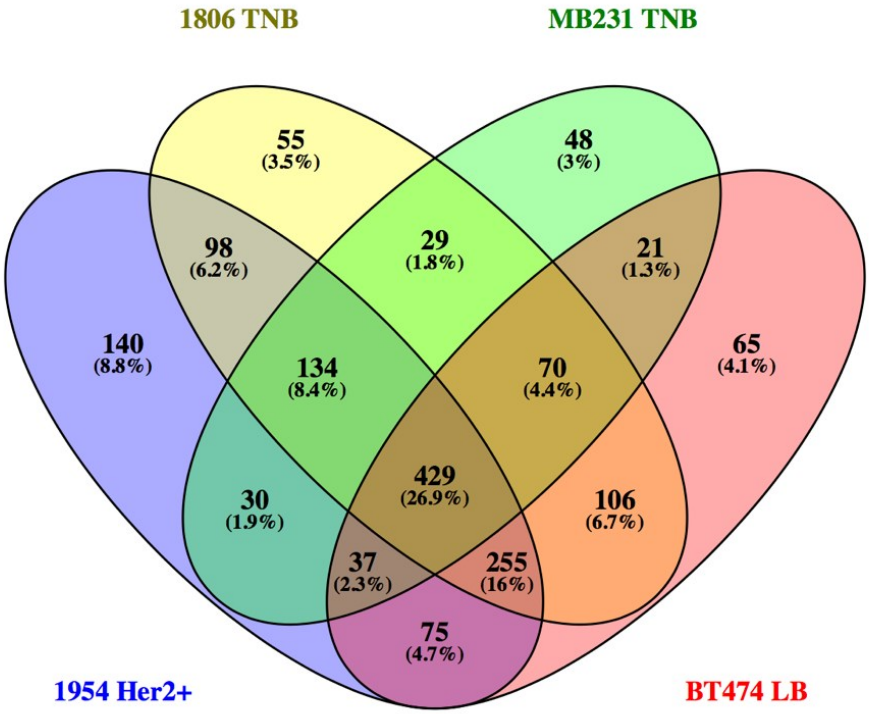

Figure 36- venn diagram showing common metabolite identifications of different breast cancer cell lines.
Fresh media (D0) and spent media (D4) of two biological replicates of breast cancer cell lines ACC 1806, ACC1904, BT474 and one sample of MD-MB231 were analysed using a C18 and an amide column(3 technical replicates positive and negative mode).Depending on the breast cancer subtype different metabolites are consumed or produced and similarities can be identified.

A PCA demonstrates which metabolites are responsible for causing a separation between the different cell lines or BC subtypes, shown in figure 37. The input for PCA was the list of all significantly changing metabolites from day 0 to day 4 (FC≤-1.5 or ≥1.5 ;p-Value ≤ 0.05) Component 1 and 2 explain 58,2% of the variation. The two biological replicates of cell line BT474 marked in green cluster well together in the negative values of component 1 and 2. The biological replicates of cell line 1954 and 1806 are more separated. This could be explained with the sample collection of biological replicates 2 two weeks after biological replicate 1 and thus causing a metabolite variance.

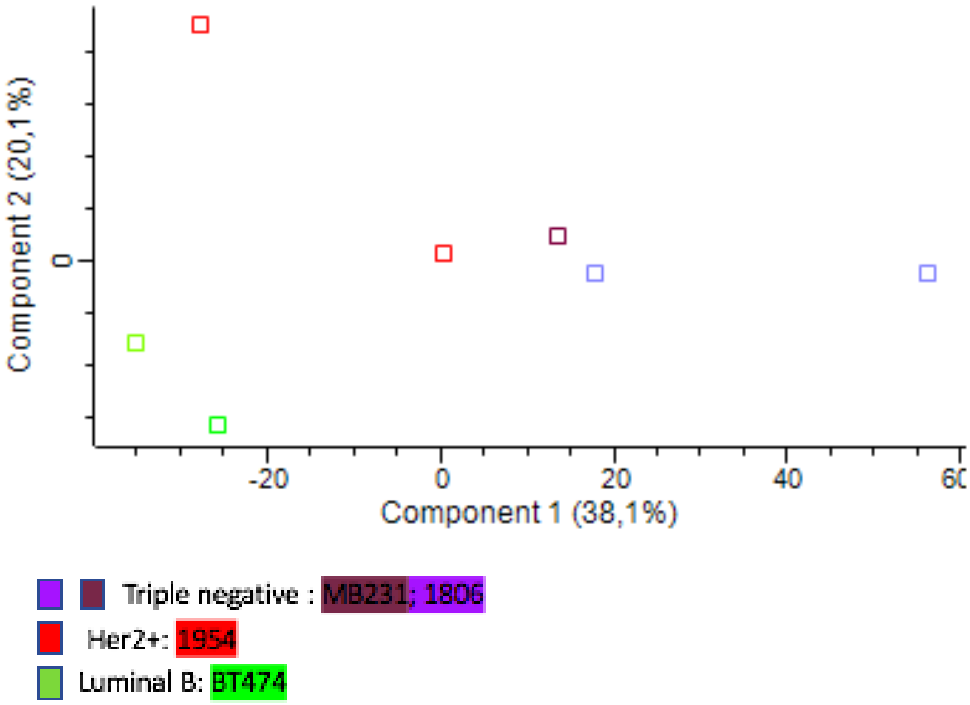

Figure 37- PCA metabolites responsible for causing a separation between the different cell lines or BC subtypes.
The biological replicates of each cell line cluster together in PC1 and a separation depending on the breast cancer subtype is visible in PC1.In PC2 BT474 is separated from the other breast cancer cell lines.

Overall the more aggressive subtypes TNB and Her2 cluster apart from the less aggressive subtype luminal B. In component 1, when leaving replicate 1954-2 out of consideration , there is a seperation between cell line BT474 (Luminal B)and MB 231, 1806 (TNB) as well as 1954 (Her2+). Component 2 seperates BT474 form 1954 replicate 2. A histogram of the biological replicates of each cell line based on the relative increase and decrease of all metabolites shows that there is no normal distribution but clear upper and lower extremes can be observed in figure 38.

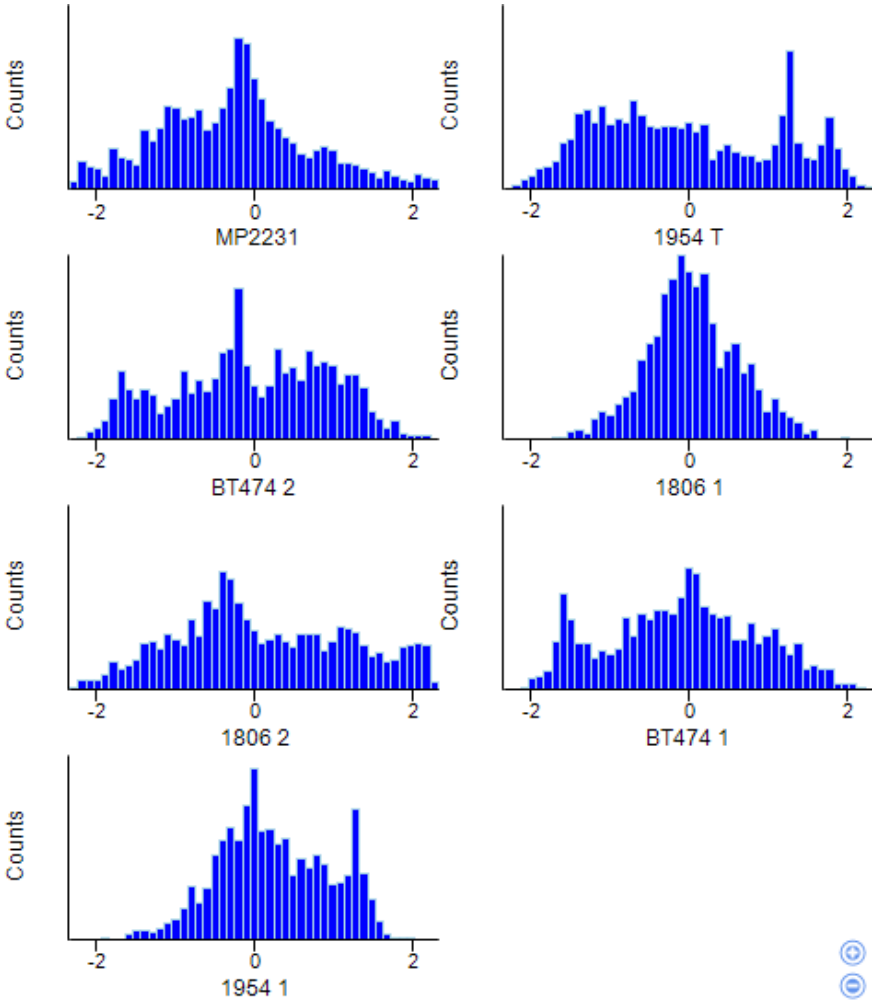

Figure 38- histogram of breast cancer subtypes.
Fresh media (D0) and spent media (D4) of two biological replicates of breast cancer cell lines ACC 1806, ACC1904, BT474 and one sample of MD-MB231 were analysed using a C18 and an amide column (3 technical replicates positive and negative mode)

Therefore, the loadings of the PC were extracted and arranged from smallest to largest. 2,5% of the most negative PC loadings and 2,5% of the most positive PC loadings were further selected and combined in a table. A heatmap generated with the fold changes shows the variation of relative increase and decrease of the metabolites (Table 7-8).The biological replicates of the cell lines are marked with different colours and the one's symbolizing the same BC subtype are grouped together.

The numbers in the columns represent the fold 2 changes from D0 to D4, with negative values showing a relative increase in concentration and positive numbers showing a relative decrease over time stressed with a colour change from green (positive) to red (negative). Observing sphingolipids, tetradecaphytosphingosine (PC1+ and PC2-) and palmiotylethanolamine (PC1+), N,N-diethylethanolamine (PC2 -) in the upper percentiles of the PC1 and PC2 loadings indicate altered sphingolipid and glycerophospholipid metabolism, shown in table 8 and 9.

Table 8- metabolites in PC1 positive

The biological replicates of the cell lines are marked with different colors and the one's symbolizing the same BC subtype are grouped together. The numbers in the columns represent the mean value of fold 2 changes from D0 to D4, of the three technical replicates of each sample, with negative values showing a relative increase in concentration and positive numbers showing a relative decrease over time stressed with a color change from green (positive) to red (negative).

						TNB			Her 2 pos		Luminal B	
2	Component 1		T: Name	T: Formula	T: Mode	MB-231	1806-1	1806-2	1954-1	1954-2	BT474-1	BT474-2
89	0,0269387	PC1 pos	tetradecaphytosphingosine	C14 H31 N O	C18+	-0,56	0,04	-2,49	-0,45	4,96	-0,12	-4,81
90	0,0269497	PC1 pos	1-[(2R,3R)-5-[(2S)-1-hydroxypropan-	C30 H36 N4	A+	1,35	-0,14	-1,32	4,41	4,34	-2,38	-0,44
91	0,0269614	PC1 pos	NI3400000	C5 H6 N2 O	A+	1,12	2,43	-0,88	1,17	4,31	0,51	0,08
92	0,0270465	PC1 pos	N,N-Bis(2-hydroxyethyl)dodecanami	C16 H33 N O	C18+	-0,43	-0,65	-1,84	-0,90	5,54	-0,89	-5,68
93	0,0270742	PC1 pos	Palmitoyl ethanolamide	C18 H37 N O	C18+	0,44	0,33	-0,51	-0,24	3,71	-0,78	-3,83

Table 9-metabolites in PC2 negative

	A	B	C	D	E	BW	BX	BY	BZ	CA	CB	CC
1						TNB			Her 2 pos		Luminal B	
2	Component 2		T: Name	T: Formula	T: Mode	MB-231	1806-1	1806-2	1954-1	1954-2	BT474-1	BT474-2
3	-5,43E+00	PC2 neg	Etoglucid	C12 H22 O6	A-	0,41	-0,88	-2,18	0,61	1,85	-3,02	-0,57
4	-5,41E+00	PC2 neg	2-HYDROXYMETHYL-2-METHYLPENTYL I	C11 H23 N O3	C18+	-0,15	-0,03	-1,86	-0,33	6,61	0,18	-6,61
5	-1,34E+00	PC2 neg	tetradecaphytosphingosine	C14 H31 N O3	A+	0,27	-0,03	-0,94	4,93	4,78	-2,96	1,87
6	-7,55E-01	PC2 neg	N,N-Diethylethanolamine	C6 H15 N O	A+	5,38	0,23	-0,09	0,27	6,72	0,07	0,23
7	-0,04157	PC2 neg	Plumbane	H4 Pb	A-	-2,34	-2,25	-5,86	-1,95	-2,01	-0,74	-1,40

Therefore, from the file including all significantly changing metabolites from day 0 to day 4, all metabolites involved in these two metabolisms based on the kegg pathway, shown in figure 39 and 40, were further extracted. The lipid identifications are summarized in table 10.

Table 10- identified lipid metabolites involved in sphingolipid and glycerophospholipid metabolism

1-(4Z,7Z,10Z,13Z,16Z,19Z-docosahexaenoyl)-sn-glycero-3-phosphocholine	N-(9Z-hexadecenoyl)sphingosine-1-phosphocholine
1-[(9Z)-hexadecenoyl]-sn-glycero-3-phosphocholine	N-(icosanoyl)ethanolamine
1-arachidonoyl-sn-glycero-3-phosphocholine	N-[(15Z)-tetracosenoyl]sphing-4-enine-1-phosphocholine
1-hexadecanoyl-2-(13Z-docosenoyl)-sn-glycero-3-phosphoethanolamine	N-C18:0 Phytoceramide
1-hexadecanoyl-2-(4Z,7Z,10Z,13Z,16Z-docosapentaenoyl)-sn-glycero-3-phosphocholine	N-Caprylyldiethanolamine
1-hexadecanoyl-2-(4Z,7Z,10Z,13Z,16Z,19Z-docosahexaenoyl)-sn-glycero-3-phosphocholine	N-dodecanoylphytosphingosine
1-Linoleoyl-sn-glycero-3-phosphocholine	N-eicosanoyl-4-hydroxysphinganine
1-Oleoyl-2-Stearoyl-sn-Glycero-3-Phosphocholine	N-hexadecanoylsphinganine
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine	N-lauroylethanolamine
1-stearoyl-sn-glycero-3-phosphoethanolamine	N-methylethanolamine phosphate
1-tetradecanoyl-2-[(9Z)-octadecenoyl]-sn-glycero-3-phosphocholine	N-octodecanoylsphinganine
3-hydroxyhexadecanoylcarnitine	N-Octylethanolamine
3-hydroxyoctadecanoylcarnitine	N-STEAROYLSPHINGOMYELIN
Caprylic diethanolamide	N-tetradecanoylsphinganine
Choline	N,N-Diethylethanolamine
Decanoylcarnitine	O-(4,8-dimethylnonanoyl)carnitine
Diethanolamine	O-3-methylglutarylcarnitine
Dihydroceramide C2	O-behenoylcarnitine
DIHYDROCERAMIDE C8	O-pimeloylcarnitine
DL-Dipalmitoylphosphatidylcholine	O-sebacoylcarnitine
HYDROXYPALMITOYL SPHINGANINE	O-suberoylcarnitine
N-(2-hydroxyhenicosanoyl)-15-methylhexadecasphing-4-enine	Palmitoyl ethanolamide
N-(2-hydroxyheptadecanoyl)-15-methylhexadecasphing-4-enine	Palmitoyl sphingomyelin
N-(2-Hydroxyhexadecyl)diethanolamine	Palmitoylcarnitine
N-(2-hydroxytetradecanoyl)sphinganine	Palmitylethanolamide
	tetradecaphytosphingosine
	Triethanolamine

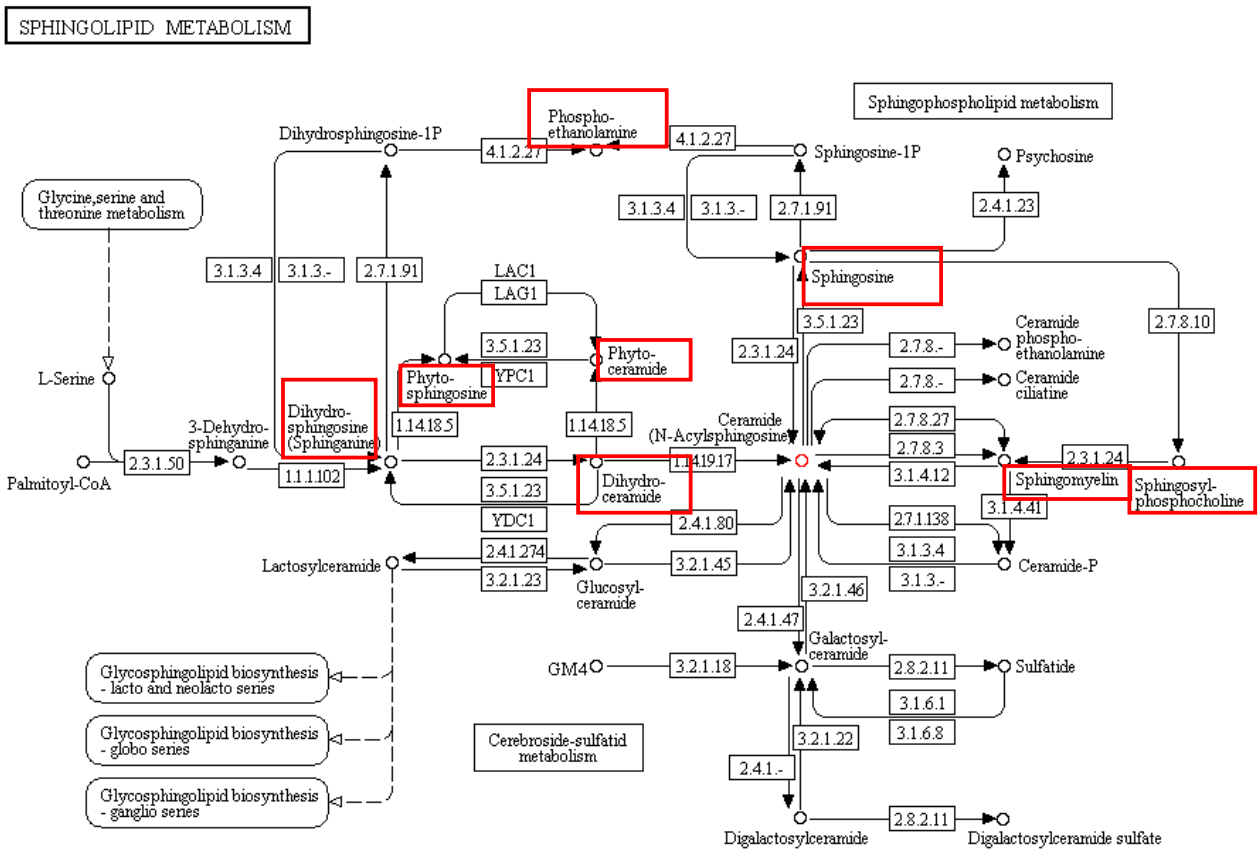

Figure 39- sphingolipid metabolism.
Kegg pathway showing the metabolites involved in the sphingolipid metabolism with marked metabolites identified in breast cancer samples

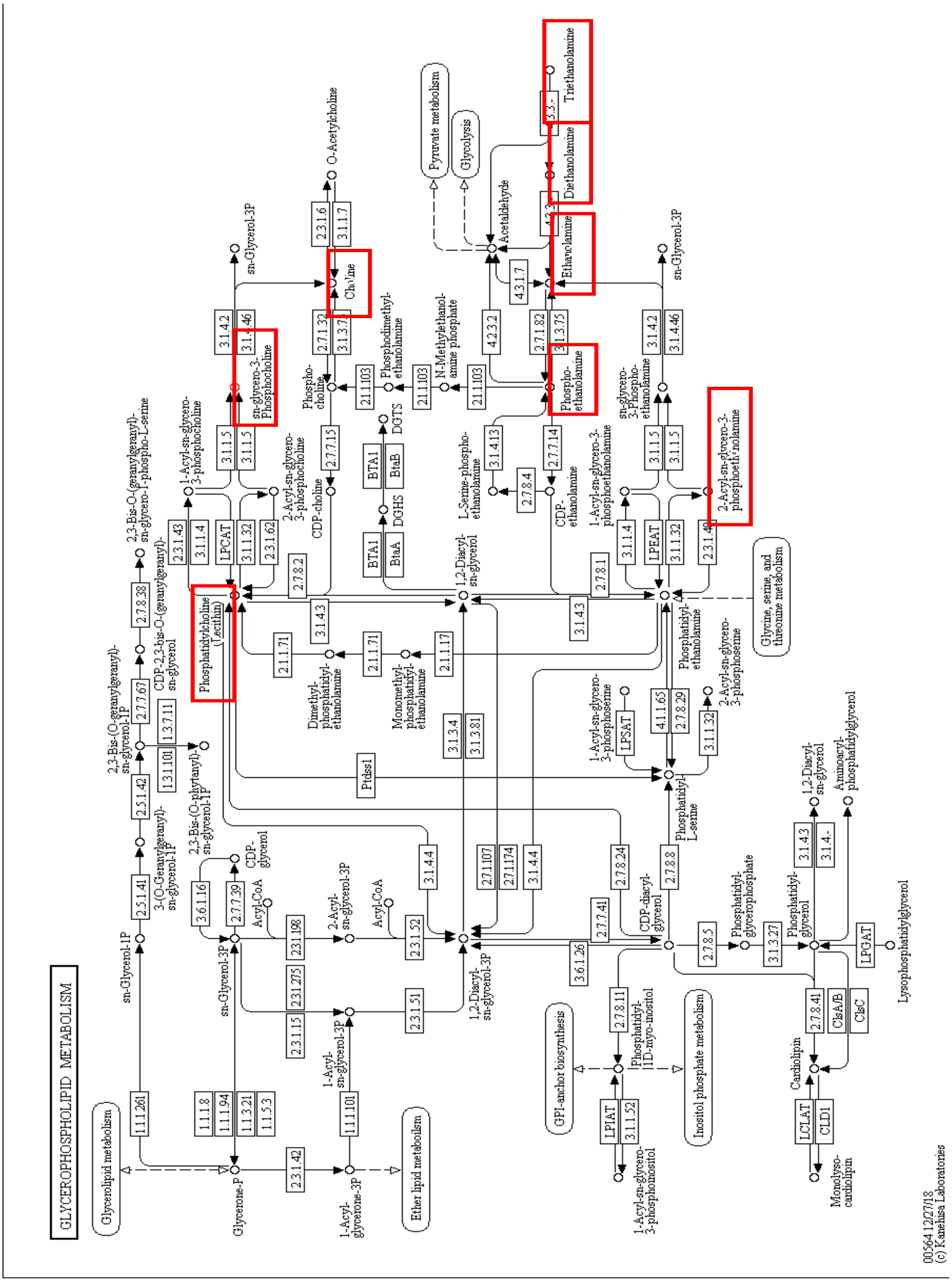

Figure 40- glycerophospholipid metabolism

Kegg pathway showing the metabolites involved in the glycerophospholipid metabolism with marked metabolites identified in breast cancer samples

The list of metabolites (table 10) were analysed in a PCA (figure 41) with Perseus to investigate if the sphingolipid and glycerophospholipid metabolism is responsible for causing a separation between the breast cancer cell lines or subtypes.

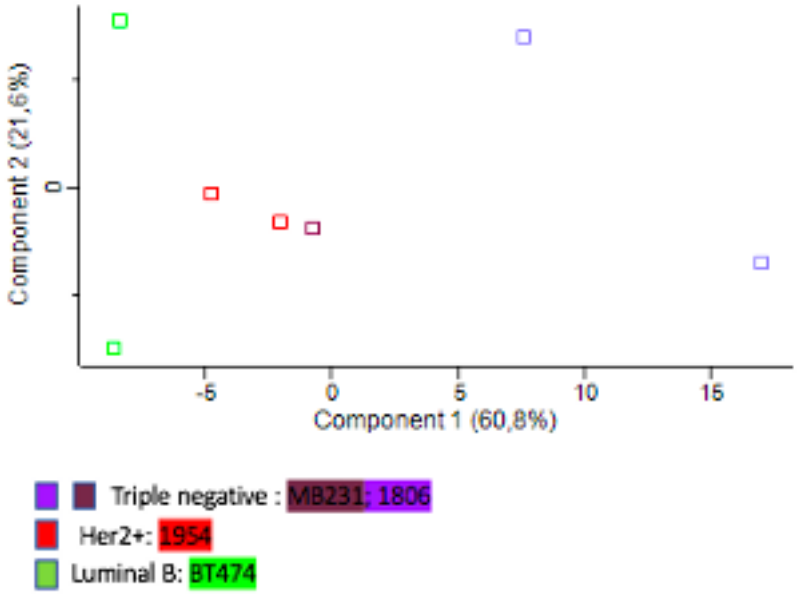

Figure 41- PCA of lipid compounds
The two biological replicates of cell lines BT474,1954,1806 cluster together in PC1 and a separation between the breast cancer subtypes can be observed in PC1.

The PCA shows a clear clustering dependent on the breast cancer cell line in PC1. It may suggest a malignant shift in lipid profile characteristic for the cell line or breast cancer subtype. The following table 9, shows the upper percentiles meaning all lower than 2,5% and higher than 97,5% of the PC loadings of the metabolites involved in sphingolipid and glycerophospholipid metabolism. The biological replicates of the cell lines are marked with different colours, and the one's symbolizing the same BC subtype are grouped together. The numbers in the columns represent the mean value of fold 2 changes from day 0 to day 4, of the three technical replicates of each sample, with negative values showing a relative increase in concentration and positive numbers showing a relative decrease over time stressed with a colour change from green (positive) to red (negative). Cell line HCC 1954 Her 2 positive, circled in red, shows a stronger decrease in lipids compared to TNB and Luminal B cell lines. Previous studies showed that alterations in phospholipid level like , phosphoethanolamine ,glycerophosphoplipids ,sphingolipids indicate membrane remodelling across breast cancer subtypes, with greater differences in Her2 positive tumours due to the stronger tumour growth [58].

Table 11- heatmap of identified lipids

Upper percentiles of the PC loadings of the metabolites involved in sphingolipid and glycerophospholipid metabolism. The biological replicates of the cell lines symbolizing the same BC subtype are grouped together and the relative increase or decrease is expressed by the mean value of fold 2 changes from day 0 to day 4, of the three technical replicates of each sample ,with negative values (red) showing a relative increase in concentration and positive numbers showing a relative decrease (green).

A	B	C	BZ	CA	CB	CC	CD	CE	CF
			TNB			Her 2 pos		Luminal B	
Component 1	componen	T: Name	MB-231	1806-1	1806-2	1954-1	1954-2	BT474-1	BT474-2
-0,0908234	PC1 neg	N-(icosanoyl)ethanolamine	0,14	-0,08	-1,38	-6,57	-6,74	7,07	0,37
-0,0877978	PC1 neg	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine	0,29	1,11	1,36	0,77	-0,6	0,93	1,06
-0,0861071	PC1 neg	1-hexadecanoyl-2-(13Z-docosenoyl)-sn-glycero-3-phosphoethanolamine	0,92	1,2	1,4	0,84	-0,34	0,83	1,02
-0,0861004	PC1 neg	N-(9Z-hexadecenoyl)sphingosine-1-phosphocholine	0,79	1,09	1,51	0,86	-0,5	0,7	1,33
0,0979956	PC1 pos	N-(2-hydroxytetradecanoyl)sphinganine	0,35	-0,57	-1,58	8,27	8,15	-2,28	0,2
0,0980842	PC1 pos	N-Octylethanolamine	0,23	-0,07	-0,69	8	9,27	-6,48	0,15
0,0984067	PC1 pos	HYDROXYPALMITOYL SPHINGANINE	0,42	-0,39	-1,63	8,19	8,31	-2,77	0,36
0,0989803	PC1 pos	Caprylic diethanolamide	-0,16	-0,08	0,19	4,72	4,66	-1,71	-0,26
0,0990382	PC1 pos	Dihydroceramide C2	0,04	-0,26	-0,32	8,33	8,62	-2,11	0,26
-0,127835	PC2 neg	Caprylic diethanolamide	0,93	1,66	1,72	1,34	4,2	2,4	-3,77
-0,127163	PC2 neg	N,N-Diethylethanolamine	-1	-0,36	-0,68	-0,44	2,88	1	-3,93
-0,119818	PC2 neg	Dihydroceramide C2	1,92	2,66	3,16	1,84	6,26	2,45	-4,68
-0,118025	PC2 neg	N-lauroylethanolamine	0,17	0,19	0,25	0,11	4,39	0,88	-4,09
0,100409	PC2 pos	1-stearoyl-sn-glycero-3-phosphoethanolamine	2,78	3,72	5,66	4,58	2,01	4,01	5,51
0,101416	PC2 pos	tetradecaphytosphingosine	0,27	-0,03	-0,94	4,93	4,78	-2,96	1,87
0,102691	PC2 pos	O-(4,8-dimethylnonanoyl)carnitine	0,43	-0,78	-0,52	3,98	3,59	-2,91	1,08
0,116842	PC2 pos	Palmitoylcarnitine	0,38	-0,1	-0,52	4,53	4,17	-1,93	2,89
0,134647	PC2 pos	N-methylethanolamine phosphate	-1,8	2,53	0	2,28	-4,16	-0,76	6,27

Triple negative : MB231; 1806

Her2+: 1954

Luminal B: BT474

A heatmap with all lipids identified and involved in the sphingolipid and glycerophospholipid metabolism was generated in figure 42. The biological replicates cluster well together demonstrating the similarity between the samples of the same cell line. Furthermore, there is a clustering of the cell lines belonging to the same BC subtype as HCC1806 and MDA-MB231 are both representatives of TNB. Especially the lower part shows a strong differentiation between Her2+ cell line 1954 and the other ones.

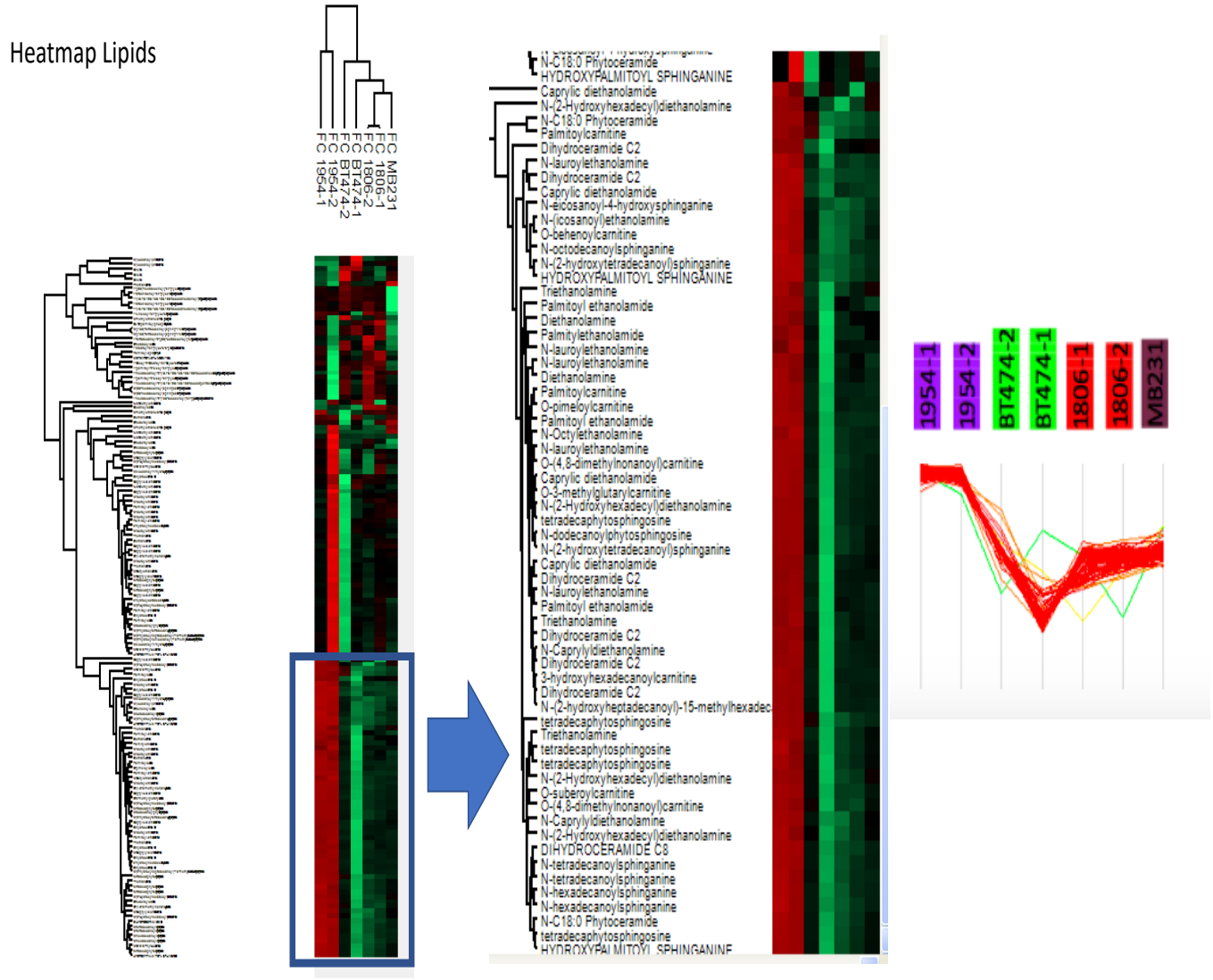

Figure 42- heatmap lipids
All lipids identified and involved in the sphingolipid and glycerophospholipid metabolism identified in the samples BT474, 1954,1806, MB231 were considered. A clustering between the biological replicates and a clustering of the cell lines belonging to the same BC subtype, as HCC1806 and MB231 are both representatives of TNB, can be observed. Especially the lower part, marked with a rectangle, shows a strong differentiation between Her2+ cell line 1954 and the other ones.

The lipids appearing in the most differentiating part of the heat map, marked with a blue rectangle, include metabolites of the sphingolipid metabolism, including phosphoethanolamine and phosphocholine derivates, phytoceramide, sphinganine derivates. A difference between cell line HCC1954 and the other cell lines can be observed, which is further underlined in the scatterplot next to the heatmap.

Since breast cancer belongs to the most spread diseases among women, there is a high interest in decreasing mortality rate especially trough early detection and effective treatment. To improve treatment outcomes metabolomics tries to target individual and cancer type specific pathways that are altered throughout malignancy. It has long been discovered that cancer cells rewire their metabolic survival strategies in order to

tolerate the hypoxic and unhostile environment due to rapid cell growth. Apart from the well-known shift from aerobic to anaerobic glycolysis described as the Warburg effect, changes in glutaminolysis and nucleotide synthesis, another phenotypic alteration in fatty acid and lipid metabolism has been observed and investigated as therapeutic target [98,99]. From the untargeted metabolomics approach of different breast cancer cell lines, a hint of altered lipid metabolism including sphingolipids and glycerophospholipids could be detected. Lipids are known to have various important functions in cell metabolism thus indicating pathological changes. In previous studies alteration of lipid metabolism in particular phospholipids has been reported in breast cancer patients compared to healthy controls [30,81,98–108]. Fatty acids are overproduced to generate membrane phospholipids, which are strongly required in proliferating cells, and furthermore function as signalling mediators [98] As phospholipids are the main builiding block for membranes sphingolipids are important for cell membrane and, glycosphingolipids pattern the surface of the cell , enabling cell-cell interaction and intracellular signalling. Other sphingolipids, such as ceramide, ceramide 1-phosphate (Cer1P), sphingosine, and sphingosine 1-phosphate (S1P) are capable of influencing cancer cell fate as signalling mediators [109].

Phospholipids are a class of lipids built from two fatty acyl molecules esterified with glycerol and phosphate group and linked to the hydrophilic head group such as choline, ethanolamine, serine, glycerol or inositol, which define the type of phospholipid (figure 43). There is a structural difference when glycerol is exchanged with ceramide e.g. in sphingosines or sphingomyelins composed of phosphocholines bound to ceramide.

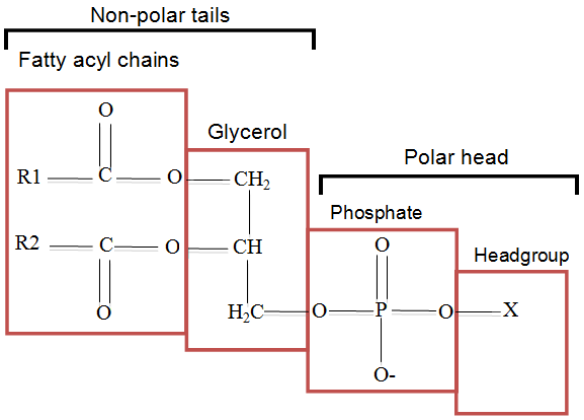

Figure 43- structure of fatty acids [99]
The main structure consists of an apolar tail build of glycerol esterified with fatty acyl molecules with a polar head consisting of choline, ethanolamine, serine, glycerol or inositol, which define the type of phospholipid.

De novo synthesis of these phospholipids occurs via the so-called Kennedy pathway. Among this compound class phosphatidylcholine (PC), phosphatidylethanolamine (PE), and sphingomyelin are the most common phospholipids of biological membranes. Especially PC and PE or their precursor molecules (PCho and phosphoethanolamine) are significantly altered in breast cancer patients. The alteration suggests a change in lipid profile progressing carcinogenesis. [81,99]

Complex lipids are vital not only as energy source or membrane compounds but also important drivers of major oncogenic pathways. Accumulated evidence indicates that sphingolipids are important mediators in several cancers like breast, colon, lung, gastric and prostate cancer and play important roles in tumour formation [104,105]. Deriving from the phospholipid pathway ceramide and sphingosine are apoptotic molecules, while sphingosine-1-phosphat promotes cell survival and glycosphingolipids are associated with drug resistance[104]. Ceramide and substrates to generate ceramide, like sphingomyelin or dihydroceramide, has been identified as differentiating metabolite between breast cancer tissue and healthy tissue and associated less aggressiveness breast cancer subtypes [104]. Previous studies have found that cancer cells displaying specific changes in choline and lipid metabolism were linked to more aggressive carcinomas. High levels of choline-containing compounds including choline, phosphocholine and glycerophosphocholine have been measured in cancer cells as well as tumour tissues in correlation with cancer transformation [58,98,103,110–114].

As molecular understanding of choline and phospholipid metabolism increases, studying the effect of the enzymes will be necessary to reveal potential treatment targets[113,115]. Supporting evidence of the importance of fatty acid metabolism for carcinogenesis is shown in the decrease of tumour progression when knocking out main metabolic enzymes involved in lipid metabolism, like fatty acid synthase (FASN) or choline kinase (ChoK). It is noted that Choline kinase (ChoK) is overexpressed in approximately 40% of all breast cancers and inhibitors have a therapeutic value in antitumor activity [98]. Its overexpression contributes to the recognized increase of PC and tCho in breast cancer and other tumours [113,114]. FASN converts fatty acids to various lipids. Lipid metabolism may also indicate breast cancer subtypes due to an observed increase in FASN correlated with Her2 expression. Cancer cells overexpressing Her2 activate translation of FASN which enhances Her2 signalling, cell

growth and increase proliferation. FASN overexpression guarantees breast cancer cell survival and correlates to poor prognosis and aggressiveness [116].

Based on these considerations' association of lipids and cancer promotion has become a target of investigation in cancer research [117]. Identification of biomarkers is valuable in cancer diagnosis, treatment targeting and response, cancer reoccurrence and clinical outcome [118]. Nevertheless, consistency of the study outcomes should be questioned, and clinical relevance further examined.

4 Conclusion

In conclusion, this thesis provides a method for analysing cancer cell supernatant by LC-MS. By untargeted profiling of cell culture media, the best conditions for sample preparation and LC-MS analysis for obtaining the highest amount of metabolite identifications have been evaluated. Based on these results, the first untargeted approach, experiment 2, comparing fresh and spent media showed that significant changes in metabolite concentrations can be detected. The second untargeted profiling, experiment 3, used relevant samples from different breast cancer subtypes. Comparing the metabolite concentrations over time clearly stated that differences in metabolite profiles exist. Based on this analysis LC-MS metabolomics can be considered as a method to reveal metabolite changes and observe differences in metabolic signatures occurring due to the pathological state. This enables a deeper understanding of molecular pathways involved in disease progression and propose new treatment targets and diagnosing approaches. Metabolomic biomarker discovery is still at an early stage but many studies spotted changes in glycolysis, glutaminolysis, lipid, glycolipid and amino acid metabolism and correlations to breast cancer cell line and subtypes. Our results show that different breast cancer subtypes employ cancer-associated metabolic pathways such as sphingolipid and phospholipid pathway.

Since breast cancer remains second leading cause of cancer deaths among women more reliable screening tests apart from mammography are required to recognize tumour formation at an early stage. Phospholipids have been identified in disease progression and could be investigated for breast cancer detection [99]. Despite the recognition of LC-MS metabolomics as powerful tool to detect altered metabolites and metabolic pathways in cancer, clinical development of metabolic profiling remains a challenge. In order to evaluate their clinical advantages and applications, adequate sample size, standardized methods to increase statistical power and verification of the identified biomarkers in human studies, are the next inevitable steps. To increase the power of omics in depth analysis by combining different omics approaches is gaining more importance [119].

To conclude, metabolomics is a potential tool to identify discriminative variables and now the challenge is to validate the potential biomarkers.

Bibliography

1. Pavlova, N. N. & Thompson, C. B. The Emerging Hallmarks of Cancer Metabolism. *Cell Metab.* **23**, 27–47 (2016).

2. Peng, B., Li, H. & Peng, X. X. Functional metabolomics: from biomarker discovery to metabolome reprogramming. *Protein Cell* **6**, 628–637 (2015).

3. Patel, S. & Ahmed, S. Emerging field of metabolomics: Big promise for cancer biomarker identification and drug discovery. *J. Pharm. Biomed. Anal.* **107**, 63–74 (2015).

4. Deberardinis, R. J. & Chandel, N. S. Fundamentals of cancer metabolism INTRODUCTION AND OVERARCHING PRINCIPLES. (2016). doi:10.1126/sciadv.1600200

5. Armitage, E. G. & Barbas, C. Metabolomics in cancer biomarker discovery: Current trends and future perspectives. *J. Pharm. Biomed. Anal.* **87**, 1–11 (2014).

6. Lander, E. S. . L. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

7. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).

8. Franklin, S. & Vrondiska, T. M. Genomes, Proteomes and the Central Dogma. **4**, 1–17 (2012).

9. Hart, C. D., Tenori, L., Luchinat, C. & Leo, A. Di. Novel Biomarkers in the Continuum of Breast Cancer. **882**, 217–234 (2016).

10. Monteiro, M. S., Carvalho, M., Bastos, M. L. & Guedes de Pinho, P. Metabolomics analysis for biomarker discovery: advances and challenges. *Curr. Med. Chem.* **20**, 257–71 (2013).

11. Oliver, S. G., Winson, M. K., Kell, D. B. & Baganz, F. Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **16**, 373–378 (1998).

12. Verma, M. & Banerjee, H. Metabolomic Approaches in Cancer Epidemiology. *Diseases* **3**, 167–175 (2015).

13. Aboud, O. A. & Weiss, R. H. New opportunities from the cancer metabolome. *Clin. Chem.* **59**, 138–146 (2013).

14. Giera, M., Spilker, M. E. & Siuzdak, G. Metabolomics : the Superglue of Omics. *Anal. Sci.* 1–7 (2018).

15. Bujak, R., Struck-Lewicka, W., Markuszewski, M. J. & Kaliszan, R. Metabolomics for laboratory diagnostics. *J. Pharm. Biomed. Anal.* **113**, 108–120 (2014).

16. Martín-Martín, N., Carracedo, A. & Torrano, V. Metabolism and Transcription in Cancer: Merging Two Classic Tales. *Front. Cell Dev. Biol.* **5**, 1–8 (2018).

17. Lindahl, A. *Exploring Disease Biomarkers and Mechanisms Using Metabolomics*. (2017).

18. van der Knaap, J. A. & Verrijzer, C. P. Undercover: Gene control by metabolites and metabolic enzymes. *Genes Dev.* **30**, 2345–2369 (2016).

19. Zhou, Z., Ibekwe, E. & Chornenkyy, Y. Metabolic Alterations in Cancer Cells and the Emerging Role of Oncometabolites as Drivers of Neoplastic Change. *Antioxidants* **7**, 16 (2018).

20. Armitage, E. G. & Southam, A. D. Monitoring cancer prognosis, diagnosis and treatment efficacy using metabolomics and lipidomics. *Metabolomics* **12**, 1–15 (2016).

21. Buescher, J. M. & Driggers, E. M. Integration of omics: More than the sum of its parts. *Cancer Metab.* **4**, 1–8 (2016).

22. Patel, S. & Ahmed, S. Emerging field of metabolomics: Big promise for cancer biomarker identification and drug discovery. *J. Pharm. Biomed. Anal.* **107**, 63–74 (2015).

23. Mordente, A., Meucci, E., Martorana, G. E. & Silvestrini, A. Cancer Biomarkers Discovery and Validation: State of the Art, Problems and Future Perspectives. **867**, 9–26 (2015).

24. Abu Bakar, M. H. *et al.* Metabolomics - the complementary field in systems biology: a review on obesity and type 2 diabetes. *Mol. Biosyst.* **11**, 1742–1774 (2015).

25. Dona, A. C., Coffey, S. & Figtree, G. Translational and emerging clinical applications of metabolomics in cardiovascular disease diagnosis and treatment. *Eur. J. Prev. Cardiol.* **23**, 1578–1589 (2016).

26. Wishart, D. S. Emerging applications of metabolomics in drug discovery and precision medicine. *Nat. Rev. Drug Discov.* **15**, 473–484 (2016).

27. IARC. Latest Global Cancer Data, 2018. *World Heal. Organ.* 13–15 (2018).

28. Institute, N. N. C. Cancer Facts & Figures 2020. *CA. Cancer J. Clin.* 1–76 (2020).

29. Society, A. C. Cancer Facts & Figures 2018. *Am. Cancer Soc.* (2018). doi:10.1182/blood-2015-12-687814

30. Ogrodzinski, M. P., Bernard, J. J. & Lunt, S. Y. Deciphering Metabolic Rewiring in Breast Cancer Subtypes. *Transl. Res.* **189**, 105–122 (2017).

31. R, L., A, V. & Estrada R, P. The Application of Metabolomics in the Development of Novel Diagnostic and Therapeutic Tools for Breast Cancer. *Metabolomics (Los Angels)* **08**, (2018).

32. Günther, U. L. Metabolomics Biomarkers for Breast Cancer. *Pathobiology* **82**, 153–165 (2015).

33. Dai, X., Cheng, H., Bai, Z. & Li, J. Breast cancer cell line classification and Its relevance with breast tumor subtyping. *J. Cancer* **8**, 3131–3141 (2017).

34. Cappelletti, V. *et al.* Metabolic Footprints and Molecular Subtypes in Breast

Cancer. *Dis. Markers* **2017**, (2017).

35. Jasbi, P. *et al.* Breast cancer detection using targeted plasma metabolomics. *J. Chromatogr. B* **1105**, 26–37 (2018).

36. Tan, J. & Le, A. The Heterogeneity of Cancer Metabolism. *Springer Int. Publ. AG, part Springer Nat.* **1063**, (2018).

37. Lulu Wang. Early diagnosis of breast cancer. *Sensors* **17**, 1572 (2017).

38. Duarte, I. F. & Gil, A. M. Metabolic signatures of cancer unveiled by NMR spectroscopy of human biofluids. *Prog. Nucl. Magn. Reson. Spectrosc.* **62**, 51–74 (2012).

39. Seyfried, T. N. & Shelton, L. M. Cancer as a metabolic disease. *Nutr. Metab.* **7**, (2010).

40. Seyfried, T. N. Cancer as a mitochondrial metabolic disease. *Front. Cell Dev. Biol.* **3**, 1–12 (2015).

41. Seyfried, T. N., Flores, R. E., Poff, A. M. & D'Agostino, D. P. Cancer as a metabolic disease: Implications for novel therapeutics. *Carcinogenesis* **35**, 515–527 (2014).

42. Granchi, C., Fancelli, D. & Minutolo, F. An update on therapeutic opportunities offered by cancer glycolytic metabolism. *Bioorganic Med. Chem. Lett.* **24**, 4915–4925 (2014).

43. Sciacovelli, M., Gaude, E., Hilvo, M. & Frezza, C. *The metabolic alterations of cancer cells*. *Methods in Enzymology* **542**, (Elsevier Inc., 2014).

44. Chen, J. Q. & Russo, J. Dysregulation of glucose transport, glycolysis, TCA cycle and glutaminolysis by oncogenes and tumor suppressors in cancer cells. *Biochimica et Biophysica Acta - Reviews on Cancer* **1826**, 370–384 (2012).

45. Nowicki, S. & Gottlieb, E. Oncometabolites: Tailoring our genes. *FEBS J.* **282**, 2796–2805 (2015).

46. Sciacovelli, M. & Frezza, C. Oncometabolites: Unconventional triggers of oncogenic signalling cascades. *Free Radic. Biol. Med.* **100**, 175–181 (2016).

47. Sévin, D. C., Kuehne, A., Zamboni, N. & Sauer, U. Biological insights through nontargeted metabolomics. *Curr. Opin. Biotechnol.* **34**, 1–8 (2015).

48. Kyle Strimbu and Jorge A. Tavel, M. D. What are biomarkers? (2010).

49. Nagana Gowda, G. & Raftery, D. Biomarker Discovery and Translation in Metabolomics. *Current Metabolomics* **1**, 227–240 (2013).

50. Wishart, D. S. *et al.* HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Res.* **41**, 801–807 (2013).

51. Trezzi, J.-P., Vlassis, N. & Hiller, K. The Role of Metabolomics in the Study of Cancer Biomarkers and in the Development of Diagnostic Tools. **867**, 41–57 (2015).

52. Gika, H. G., Theodoridis, G. A., Plumb, R. S. & Wilson, I. D. Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. *J. Pharm. Biomed. Anal.* **87**, 12–25 (2014).

53. Spratlin, J. L., Serkova, N. J. & Eckhardt, S. G. Clinical applications of metabolomics in oncology: A review. *Clin. Cancer Res.* **15**, 431–440 (2009).

54. Sitter, B. *et al.* Comparison of HR MAS MR spectroscopic profiles of breast cancer tissue with clinical parameters. *NMR Biomed.* **19**, 30–40 (2006).

55. Glunde, K., Jie, C. & Bhujwalla, Z. M. Molecalar causes of tile aberrant choline phospholipid metabolism in breast cancer. *Cancer Res.* **64**, 4270–4276 (2004).

56. Oakman, C. *et al.* Uncovering the metabolomic fingerprint of breast cancer. *Int. J. Biochem. Cell Biol.* **43**, 1010–1020 (2011).

57. Jagannathan, N. R. & Sharma, U. Breast tissue metabolism by magnetic resonance spectroscopy. *Metabolites* **7**, (2017).

58. Cappelletti, V. *et al.* Metabolic Footprints and Molecular Subtypes in Breast Cancer. *Dis. Markers* **2017**, 1–19 (2017).

59. Ganti, S. *et al.* Urinary acylcarnitines are altered in human kidney cancer. *Int. J. Cancer* **130**, 2791–2800 (2012).

60. Kim, K. *et al.* Urine metabolomic analysis identifies potential biomarkers and pathogenic pathways in kidney cancer. *Omi. A J. Integr. Biol.* **15**, 293–303 (2011).

61. Kline, E. E. *et al.* Citrate Concentrations in Human Seminal Fluid and Expressed Prostatic Fluid Determined via 1H Nuclear Magnetic Resonance Spectroscopy Outperform Prostate Specific Antigen in Prostate Cancer Detection. *J. Urol.* **176**, 2274–2279 (2006).

62. DeFeo, E. M., Wu, C. L., McDougal, W. S. & Cheng, L. L. A decade in prostate cancer: From NMR to metabolomics. *Nat. Rev. Urol.* **8**, 301–311 (2011).

63. Swanson, M. G. *et al.* Quantitative analysis of prostate metabolites using 1H HR-MAS spectroscopy. *Magn. Reson. Med.* **55**, 1257–1264 (2006).

64. Cheng, L. L., Wu, C. lee, Smith, M. R. & Gonzalez, R. G. Non-destructive quantitation of spermine in human prostate tissue samples using HRMAS 1H NMR spectroscopy at 9.4 T. *FEBS Lett.* **494**, 112–116 (2001).

65. Roberts, M. J., Schirra, H. J., Lavin, M. F. & Gardiner, R. A. NMR-based Metabolomics: Global Analysis of Metabolites to Address Problems in Prostate Cancer. *Cervical, Breast and Prostate Cancer* 1–43 (2014). doi:10.1161/01.CIR.0000056107.05724.40

66. Sreekumar, A. *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457**, 910–914 (2009).

67. Denkert, C. *et al.* Mass spectrometry-based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors. *Cancer Res.* **66**, 10795–10804 (2006).

68. Rocha, C. M. *et al.* Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of blood plasma. *J. Proteome Res.* **10**, 4314–4324 (2011).

69. Ritchie, S. A. *et al.* Reduced levels of hydroxylated, polyunsaturated ultra long-chain fatty acids in the serum of colorectal cancer patients: Implications for early screening and detection. *BMC Med.* **8**, (2010).

70. Ritchie, S. A. *et al.* Low-serum GTA-446 anti-inflammatory fatty acid levels as a new risk factor for colon cancer. *Int. J. Cancer* **132**, 355–362 (2013).

71. Qiu, Y. *et al.* Urinary metabonomic study on colorectal cancer. *J. Proteome Res.* **9**, 1627–1634 (2010).

72. Yu, Z. Metabolomics analyses to better understand complex phenotypes Metabolomics analyses to better understand complex phenotypes. (2013). doi:10.1007/s10530-017-1439-8

73. Claudino, W. M. *et al.* Metabolomics: Available results, current research projects in breast cancer, and future applications. *J. Clin. Oncol.* **25**, 2840–2846 (2007).

74. Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–269 (2012).

75. Theodoridis, G. A., Gika, H. G., Want, E. J. & Wilson, I. D. Liquid chromatography-mass spectrometry based global metabolite profiling: A review. *Anal. Chim. Acta* **711**, 7–16 (2012).

76. Sandås, E. M. Liquid chromatography – Orbitrap mass spectrometry is a useful tool in untargeted metabolomics analysis of dried blood spots in clinical chemistry. (2018).

77. Kohler, I., Verhoeven, A., Derks, R. J. & Giera, M. Analytical pitfalls and challenges in clinical metabolomics. *Bioanalysis* **8**, 1509–1532 (2016).

78. chemyx. c. (2017). Available at: https://www.chemyx.com/terms-of-service/. (Accessed: 30th March 2020)

79. Wong, P. S. H. & Cooks, R. G. Ion trap mass spectrometry. 85–92 (1997). doi:10.1002/rcm.1290080114

80. B Zhou, J F Xiao, L Tuli, and H. W. R. LC-MS-based metabolomics. *Natl. Instute Heal.* **8**, 470–481 (2013).

81. Jelonek, K., Ros, M., Pietrowska, M. & Widlak, P. Cancer biomarkers and mass spectrometry based analyses of phospholipids in body fluids. *Clin. Lipidol.* **8**, 137–150 (2013).

82. Shimadzu Corporation. Fundamental Guide to Liquid Chromatography Mass Spectrometry (LCMS). 61 (2019).

83. Watson, D. G. A rough guide to metabolite identification using high resolution liquid chromatography mass spectrometry in metabolomic profiling in metazoans. *Comput. Struct. Biotechnol. J.* **4**, e201301005 (2013).

84. Strupat, K., Scheibner, O. & Bromirski, M. Orbitrap Mass Spectrometry – Definitions , Opportunities , and Advantages. (2013).

85. Makarov, A. & Scigelova, M. Coupling liquid chromatography to Orbitrap mass spectrometry. *J. Chromatogr. A* **1217**, 3938–3945 (2010).

86. Olsen, J. V. *et al.* Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010–2021 (2005).

87. Lee, D. Y., Bowen, B. P. & Northen, T. R. Mass spectrometry–based metabolomics, analysis of metabolite- protein interactions, and imaging. *Biotechniques* **49**, 130–134 (2010).

88. Armitage, E. G. & Barbas, C. Metabolomics in cancer biomarker discovery: Current trends and future perspectives. *J. Pharm. Biomed. Anal.* **87**, 1–11 (2013).

89. Zhang, A., Sun, H., Wang, P., Han, Y. & Wang, X. Recent and potential developments of biofluid analyses in metabolomics. *J. Proteomics* **75**, 1079–1088 (2012).

90. Vuckovic, D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. *Anal. Bioanal. Chem.* **403**, 1523–1548 (2012).

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.* **27**, 1897–1905 (2016).

92. Schrimpe-rutledge, A. C., Codreanu, S. G., Sherrod, S. D. & Mclean, J. A. Untargeted Metabolomics Strategies- Challenges and Emerging Directions. *Journal of american society of mass spectrometry* **27**, 1897–1905 (2017).

93. Scheltema, R. A. *et al.* Increasing the mass accuracy of high-resolution LC-MS

data using background ions - A case study on the LTQ-Orbitrap. *Proteomics* **8**, 4647–4656 (2008).

94. Kapoore, R. V. & Vaidyanathan, S. Towards quantitative mass spectrometry-based metabolomics in microbial and mammalian systems. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **374**, (2016).

95. Broadhurst, D. *et al.* Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **14**, 1–17 (2018).

96. Matsuda, F. Technical Challenges in Mass Spectrometry-Based Metabolomics. *Mass Spectrom.* **5**, S0052–S0052 (2016).

97. Lindahl, A., Sääf, S., Lehtiö, J. & Nordström, A. Tuning Metabolome Coverage in Reversed Phase LC-MS Metabolomics of MeOH Extracted Samples Using the Reconstitution Solvent Composition. *Anal. Chem.* **89**, 7356–7364 (2017).

98. Hilvo, M. & Orešič, M. Regulation of lipid metabolism in breast cancer provides diagnostic and therapeutic opportunities. *Clin. Lipidol.* **7**, 177–188 (2012).

99. Mistry, D. A. H. & French, P. W. Circulating phospholipids as biomarkers of breast cancer: A review. *Breast Cancer Basic Clin. Res.* **10**, 191–196 (2016).

100. Ryland, L. K., Fox, T. E., Liu, X., Loughran, T. P. & Kester, M. Dysregulation of sphingolipid metabolism in cancer. *Cancer Biology and Therapy* **11**, 138–149 (2011).

101. Glunde, K., Ackerstaff, E., Mori, N., Jacobs, M. A. & Bhujwalla, Z. M. Choline phospholipid metabolism in cancer: Consequences for molecular pharmaceutical interventions. *Mol. Pharm.* **3**, 496–506 (2006).

102. Jia, M. *et al.* Estrogen receptor a promotes breast cancer by reprogramming choline metabolism. *Cancer Res.* **76**, 5634–5646 (2016).

103. Mimmi, M. C. *et al.* Absolute quantification of choline-related biomarkers in breast cancer biopsies by liquid chromatography electrospray ionization mass spectrometry. *Anal. Cell. Pathol.* **36**, 71–83 (2013).

104. Moro, K. *et al.* Ceramide species are elevated in human breast cancer and are associated with less aggressiveness. *Oncotarget* **9**, 19874–19890 (2018).

105. Nagahashi, M. *et al.* High levels of sphingolipids in human breast cancer. *J. Surg. Res.* **204**, 435–444 (2016).

106. Cui, M., Wang, Q. & Chen, G. Serum metabolomics analysis reveals changes in signaling lipids in breast cancer patients. *Biomedical Chromatography* **30**, 42–47 (2016).

107. Qiu, Y. *et al.* Mass spectrometry-based quantitative metabolomics revealed a distinct lipid profile in breast cancer patients. *Int. J. Mol. Sci.* **14**, 8047–8061 (2013).

108. Huang, C. & Freter, C. Lipid metabolism, apoptosis and cancer therapy. *Int. J. Mol. Sci.* **16**, 924–949 (2015).

109. Don, A. S., Lim, X. Y. & Couttas, T. A. Re-configuration of sphingolipid metabolism by oncogenic transformation. *Biomolecules* **4**, 315–353 (2014).

110. Mori, N., Wildes, F., Takagi, T., Glunde, K. & Bhujwalla, Z. M. The Tumor Microenvironment Modulates Choline and Lipid Metabolism. *Frontiers in Oncology* **6**, (2016).

111. Aboud, O. A. & Weiss, R. H. New opportunities from the cancer metabolome. *Clin. Chem.* **59**, 138–146 (2013).

112. Moestue, S. A., Giskeodegard, G. F., Cao, M. D., Bathen, T. F. & Gribbestad, I. S. Glycerophosphocholine (GPC) is a poorly understood biomarker in breast cancer. *Proc. Natl. Acad. Sci.* **109**, E2506–E2506 (2012).

113. Cheng, M., Bhujwalla, Z. M. & Glunde, K. Targeting Phospholipid Metabolism in Cancer. *Front. Oncol.* **6**, 1–17 (2017).

114. Eliyahu, G., Kreizman, T. & Degani, H. Phosphocholine as a biomarker of breast cancer: Molecular and biochemical studies. *Int. J. Cancer* **120**, 1721–1730 (2007).

115. Cao, M. D. *et al.* Targeting choline phospholipid metabolism: GDPD5 and GDPD6 silencing decrease breast cancer cell proliferation, migration, and invasion. *NMR Biomed.* **29**, 1098–1107 (2016).

116. Jin, Q. *et al.* Fatty acid synthase phosphorylation: A novel therapeutic target in HER2-overexpressing breast cancer cells. *Breast Cancer Res.* **12**, R96 (2010).

117. Corominas-Faja, B. *et al.* Clinical and therapeutic relevance of the metabolic oncogene fatty acid synthase in HER2+ breast cancer. *Histology and Histopathology* **32**, 687–698 (2017).

118. McCartney, A. *et al.* Metabolomics in breast cancer: A decade in review. *Cancer Treat. Rev.* **67**, 88–96 (2018).

119. Hwang, V. J. & Weiss, R. H. Metabolomic profiling for early cancer detection: current status and future prospects. *Expert Opinion on Drug Metabolism and Toxicology* **12**, 1263–1265 (2016).

Table of Figures

Table of Tables