



U SCIENCE TECH
FACULTAT DE CIÈNCIES
I TECNOLOGIA
UVIC-UCC



Norwegian University of
Science and Technology

FINAL THESIS

Heterogeneity characterization of breast cancer biopsies by metabolic profiling using HR MAS MRS analysis

Marina Perea Badia

Biotechnology Degree

UVIC Advisor: Àngels Sierra

NTNU Advisors: Guro F. Giskeødegård and Tone Frost Bathen

Vic, June 2017

Acknowledgements

The work presented in this Project is based on work carried out in the MR Cancer group, Department of Circulation and Medical Imaging, Norwegian University of Science and Technology (NTNU), in the period from September 2016 to December 2016.

There are many persons that have contributed and supported me in achieving this thesis. First of all, I want to give my thanks to Prof. Tone Frost Bathen, who welcomed me and trusted me, giving me the opportunity to carry out my thesis in her group.

A special thanks to my advisor Dr Guro F.Giskeodegard, for her infinite patience and for all the time you put into teaching and helping me. You have been a tremendous mentor for me and I have learned a lot from you, not only in the scientific area but also on a personal level. Thank you so much!

My gratitude goes out to Prof Àngels Sierra, who always taking time to give me a positive feedback and for many helpful advises. You taught me to value the importance of cancer and believe me to put me in contact with MR Cancer Group and make possible to perform this research.

I would like to express my gratitude to all workers in the MR Cancer group. They showed me the meaning and the importance of teamwork. Thanks you to open the doors to me and make me feel like home in Norway. I have learned and enjoyed working with all of you.

Finally, thanks to my friends and my family, especially to you mom, who always supported me and encouraged all my decisions.

Summary

Biotechnology

Heterogeneity characterization of breast cancer biopsies by metabolic profiling using HR MAS MRS analysis

KEYWORDS: Metabolomics, HR MAS MRS, breast cancer heterogeneity, multivariate analysis, PCA

Marina Perea Badia

Angels Sierra, Guro F. Giskeødegård, Tone Frost Bathen

Breast cancer is a heterogeneous disease with a varying prognosis. Today's clinical factors provide some information about prognosis of a breast cancer patient; however there is a need for additional information to stratify patients for improved and more individualized treatment. Breast tumors are highly heterogeneous due to subpopulations of cancer cells that differ in genetic and phenotypic characteristics. Tumor heterogeneity has been associated with treatment resistance and relapse. Additionally, it can be questioned how representative one biopsy is for the whole tumor. Tumor phenotypic heterogeneity cannot be solely attributed to genetic differences, as epigenetics and interaction with the tumor microenvironment also contribute. The aim of this study was to examine intra-tumor and inter-tumor heterogeneity through metabolic profiling in breast cancer tissue.

Fresh frozen tissue slices from the middle of surgically removed breast tumors were used. Five cores throughout the slices were drilled out from 10 tumors, of which six IDC grade 2-3 and four fibroadenomas. Histological examination of each core was performed by HE staining, and metabolic profiling of intact tissue samples were performed by high resolution magic angle spinning (HR MAS) magnetic resonance spectroscopy (MRS). The relative concentrations of 23 metabolites were quantified. Metabolic heterogeneity was measured by coefficient of variation (CoV) and multivariate PCA was used for explain metabolic variance.

The study showed that patients diagnosed with high-grades of IDC had higher metabolic differences compared to patients diagnosed with fibroadenomas. No evidences were found for metabolic intratumor heterogeneity associated with sample location within the tumor. Breast tissue samples examinations by pathologist provide important information in order to characterize intratumor metabolic variations, highlighting the strong role of histology evaluation. Overall, MRS metabolomics contributed to a valuable tool for further characterization of tumor heterogeneity and for identifying opportunities for improving stratification of patients into clinically useful diagnosed groups.

Symbols and abbreviations

^{13}C	Carbon-13 nucleus
^1H	Hydrogen-1 nucleus
I	Spin
μ	Magnetic moment
FID	Free induction decay
B_0	External magnetic field
RF	Radio frequency
Acetyl CoA	Acetyl-coenzyme A
ATP	Adenosine triphosphate
NADH	Nicotinamide adenine dinucleotide
PtdCho	Phosphatidylcholine
Cho	Choline
PCho	Phosphocholine
GPC	glycerophosphocholine
tCho	Total Cholines
TCA	Tricarboxylic acid cycle
LDH	Lactate dehydrogenase
CoV	Coefficient of variation
ER	Estrogen receptor
FN	False negative
FP	False positive
GLS	Glutaminase
GPC	Glycerophosphocholine
GSH	Glutathione
HCA	Hierarchical cluster analysis
HER2	Human epidermal growth factor receptor 2
PgR	Progesterone receptor
HR	High resolution
MAS	Magic angle spinning
MRS	Magnetic resonance spectroscopy
MS	Mass spectrometry
PC	Principal component
PCA	Principal component analysis
ppm	Parts per million
FISH	Fluorescence In Situ Hybridization
TNBC	Triple negative breast cancer
B-glc1	B-glucose peak 1

asc	Ascorbate
Lac	Lactate
mI1, mI2, mI3	Myoinositol peak 1, 2 and 3
tyr	Tyrosine
cr1,cr2	Creatine peak 1 and 2
glu1, glu2	Glutamate peak 1 and 2
y1	Unknown peak
gly	Glycine
tau1, tau2	Taurine peak 1 and 2
succ	Succinate
gluth1, gluth2	Glutathione peak 1 and 2
glmine1, glmine2	Glutamine peak 1 and 2
ala	Alanine

Contents

1. Introduction	8
1.1. Breast cancer.....	9
1.1.1. Anatomy of the breast.....	9
1.1.2. Diagnosis and staging of breast cancer.....	10
1.1.3. Predictive and prognostic factors	11
1.1.4. Breast cancer heterogeneity	11
1.1.5. Breast cancer treatment strategies.....	12
1.2. Tumor metabolism.....	13
1.2.1. Glucose metabolism.....	13
1.2.2. Lipid and choline metabolism.....	14
1.2.3. Amino acid metabolism	15
1.3. Metabolomics	15
1.3.1. MR spectroscopy.....	16
1.3.2. HR MAS MRS.....	16
1.3.3. MRS acquisition.....	17
1.4. Preprocessing of MRS spectra.....	18
1.4.1. Baseline correction.....	18
1.4.2. Peak alignment.....	19
1.4.3. Scaling and normalization.....	19
1.5. Multivariate analysis.....	20
1.5.1. Principal component analysis.....	20
2. Hypothesis	22
3. Main objectives	22
3.1. Overall aim	22
3.2. Specific objectives	22
4. Materials & methods.....	23
4.1. Patients.....	23
4.1.1. Tissue collection	23
4.1.2. Sample drilling.....	25
4.1.3. Histopathology and clinical diagnosis	26
4.2. HR-MAS MRS experiments.....	26
4.2.1. Acquisition protocol.....	27
4.3. Spectral preprocessing and data analysis.....	27
4.3.1. Analysis of individual metabolites.....	27
4.3.2. Statistical analysis	28
4.3.3. Multivariate analysis	28

5. Results & Discussion	29
5.1. Histopathology and clinical results.....	29
5.2. Analysis of individual metabolites results.....	30
5.3. Statistical analysis results	32
5.4. Multivariate analysis results	35
6. Conclusions and Future perspectives	44
References.....	45

Appendix

1. Introduction

The human body is made up by more than fifty billion of cells. Healthy cells are the basic building blocks of all tissue and organs in the body. Hence, cells are commonly referred to as the basic units of life. The division of a cell is governed by a series of tightly regulated events known as the cell cycle.¹ To ensure its proper development, when the cell cycle is impaired, a series of safety mechanisms are responsible to stop its progression or, if the damage is irreparable, to program cell death. Cancer cells exhibit uncontrolled growth and proliferation losing regulation of important cellular processes caused by mutations, which make them capable of avoiding these control mechanisms and reproduce without following pre-wired plans. Additional alterations enable cancer cells to invade surrounding tissue and invade distant locations from the cancer's primary site, also known as metastasis. The accumulation of several errors and genetic alterations in the regulatory pathways over the years increases the risk of cancer. In some cases, genetic alterations are inherited giving a predisposition for cancer. As there are more than 100 distinct types of cancer, the term refers to a collection of diseases, typically classified and named by the organs or tissues where the cancer forms².

Although there is huge complexity and variety in characteristics among different cancer types, cancer cells exhibit common traits or capabilities known as hallmarks of cancer proposed by Hanahan et al. Cancer cells can become 1) self-sufficient of growth signals, be 2) insensitive to anti-growth signals, and 3) avoid programmed cell death (apoptosis). In addition, cancer cells are able to 4) invade adjacent tissues or metastasize, have 5) limitless replicative potential and 6) induce the formation of new blood vessels sustaining angiogenesis (Figure 1.1a). Based on progression in cancer science the aforementioned hallmarks were extended with two additional capabilities: 7) the reprogramming of energy metabolism and 8) evading immune destruction owing to enabling characteristics named as genomic instability and mutation and tumor-promoting inflammation (Figure 1.1b)³.

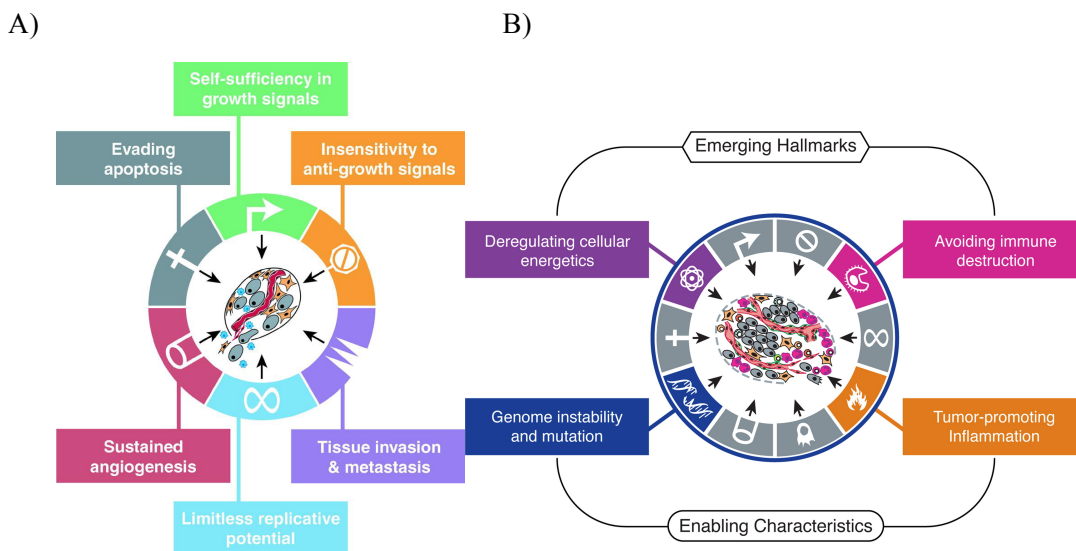


Figure 1.1: Hallmarks and enabling characteristics of cancer. a) The six biological characteristics of cancers acquired during development of human tumors. b) The two emerging hallmarks and enabling characteristics of cancer. Reprinted from Hallmarks of cancer, Vol. 100, 57-70, January 7, 2000, Copyright © 2000 by Cell Press and Hallmarks of Cancer: The Next Generation, Vol. 144, Issue 5, 4 March 2011, Copyright © 2011 with permission from Elsevier Inc. All rights reserved.

1.1 Breast Cancer

Breast cancer is the most frequently diagnosed cancer among women worldwide and in Europe, and it has been estimated that 8% of women will develop this disease by the age of 75. Although the survival in breast cancer patients is high, it is difficult to predict each patient's outcome. Patients with the same diagnosis of breast cancer may have different response to treatment, supporting the need to further characterize breast cancer heterogeneity⁴.

The age is the principal risk factor in breast cancer, with an increased incidence by the age of 35 and stabilization by the age of 55, coinciding with menopause. An important risk factor is the family history of breast cancer, which increases the risk by about 50% compared to women who haven't had blood relatives diagnosed with breast cancer⁵⁶. About 10% of breast cancers are thought to be hereditary, caused by gene mutations. Most inherited cases are associated with two abnormal genes: BRCA1 (Breast Cancer gene one) and BRCA2 (Breast Cancer gene two). Women who have an abnormal *BRCA1* or *BRCA2* gene (or both) can have up to an 80% risk of being diagnosed with breast cancer during their lifetimes. Breast cancers associated with an abnormal *BRCA1* or *BRCA2* gene tend to develop in younger women and occur more often in both breasts than cancers in women without these abnormal genes.

1.1.1 Anatomy of the breast

The female breast consists of fatty tissue, connective tissue, lobes, ducts and lymph nodes (Figure 1.2)⁷. A healthy female breast is made up of 12-20 sections called lobes. Each of these lobes is made up by several small lobules, the gland that produces milk in nursing women. These lobes are connected to ducts that transport the milk from the lobule to the nipple. Lymph nodes and lymph vessels containing immune system cells surround the breast and contribute to removing waste products.

The type of breast cancer is generally determined by the origin of the growth of cancer cell, which is almost always in the lobes, lobules, or ducts. Normally cancers originate from epithelial cells that cover or line the breast and are thus called breast carcinomas. However, in some rare cases (less than 1%) the cancer arises from stromal components (connective tissue) within the breast (i.e. sarcomas). The premalignant changes where the epithelial cells have not broken through the basement membrane are classified into hyperplasia or carcinoma in situ.

Ductal carcinoma in situ (DCIS) is the most common type of non-invasive breast cancer, which is not life-threatening, but having DCIS can increase the risk of developing an invasive breast cancer later on⁸. If cancer cells have broken through the basement membrane and invaded surrounding tissue, it is classified as invasive or infiltrating carcinoma⁸. Invasive carcinoma is by far the most frequent type of breast cancer, where between 72-80% are invasive ductal carcinomas (IDC) and 5-15% are invasive lobular carcinomas (ILC)⁹.

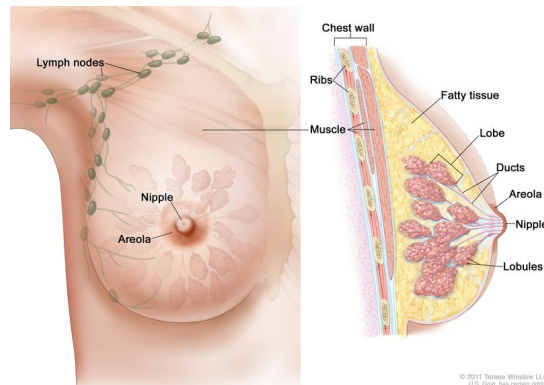


Figure 1.2: Anatomy of the female breast. Adapted with permission from National Cancer Institute © 2011 Therese Winslow, U.S. Govt.

Fibroadenomas are benign breast tumors that are commonly diagnosed in young women and are associated with a slight increase in the risk of breast cancer¹⁰. These benign breast masses are made up of fibrous and granular connective tissue which have grown more rapidly than usual and generally are presented as 2 to 3 cm in size. As they are affected by hormones and tend to fluctuate (or increase) in size during menstrual cycle, pregnancy and breast-feeding, they may increase to > 10 cm. Fibroadenomas are a long-term risk factor for breast cancer as the risk of invasive breast cancer was assessed as 2.17 times higher among women with fibroadenoma than among women without fibroadenoma¹¹. And the risk increases even more if women have complex fibroadenomas, proliferative disease or a family history cancer.

1.1.2 Diagnosis and staging of breast cancer

During the diagnostic process, breast cancer patients are examined by three main strategies: clinical examination, imaging tests (mammogram, ultrasound or magnetic resonance imaging (MRI)), and histological examination of needle biopsies. This approach has been found to be highly precise and accurate when all three modalities are in agreement¹². It has been shown that early diagnosis of breast cancer improves prognosis, thanks to the public health campaigns which offers mammography screening to women, specially in ages from 50 to 74, when the incidence is higher⁶.

Following the diagnosis, breast cancer stage is determined. The stage describes the extend of cancer, particularly the size of a tumor and its dissemination using the TNM- system, which is the most widely used. This classification stages cancer based on primary tumor size (T), the degree of spread to lymph nodes (N), and the existance of distant metastasis (M). Together this determines the patient's stage of breast cancer. Increasing stage indicates a more aggressive disease, with Stage 0 referring to non-invasive breast cancer, Stages I through III describing invasive breast cancer and Stages IV indicating dissemination to other organs. In general, a lower grade number indicates a slower-growing cancer that is less likely to spread, while higher number indicates a faster-growing cancer that is more likely to spread. For instance, cells in IDC grade II breast tumor are moderately differentiated compared to grade I which are well differentiated. Moreover, IDC grade III cells are abnormal and they tend to grow and spread more aggressively than lower grades.

Based on the TNM classification, the tumor is defined as primary operable or inoperable¹³.

1.1.3 Predictive and prognostic factors

Prognostic factors are those capable of predicting clinical outcome of patients irrespective of treatment while predictive factors intend to assess the outcome of patients receiving a certain systemic therapy and thus are intimately associated with sensitivity or resistance to the administered therapy.

The most powerful prognostic factor for early stage breast cancer is the axillary lymph node status, followed by the tumor size, which is the most significant prognostic factor in node-negative patients. These factors are represented in above-mentioned TNM staging system. In addition to finding anatomical features of the tumor, histopathological analysis also provides prognostic value as it gives information of the tumor cells degree of differentiation and include assessment of the tumor's expression of estrogen receptor (ER), progesterone receptor (PgR), human epidermal growth factor (HER2) and, in some cases, proliferation (by the Ki67 marker). The hormone receptors ER and PgR are transcription factors depending on binding of their ligand (the hormones estrogen and progesterone respectively) for activation of important proliferation processes and production of growth factors. An estimated 70% of human breast cancers are ER positive¹⁴. As ER activation also regulates the PgR-gene, less than 1% of PgR-positive (PgR+) cases are ER-negative (ER-). Hormone (estrogen and progesterone) receptor status is the main indicator of the response to endocrine therapy as the presence of ER and/or PgR receptors indicates that tumor progression is driven by these and other related reproductive hormones.

HER2 receptors are proteins that help manage how a breast cell grows, divides, and repairs itself. However, in about a quarter of all breast cancer patients, the HER2 gene is not functioning properly. It makes an excess number of copies of itself in a process known as "HER2 gene amplification." Then these extra genes instruct the cells to make too many HER2 receptors, which is called "HER2 protein overexpression." The ultimate result is that breast cells grow and divide in an uncontrolled fashion¹⁵. Novel targeted therapies that inhibit HER2 have been developed and approved for clinical use, e.g. trastuzumab an antibody which binds to the extracellular domain of the HER receptor resulting in inhibition of cell growth.

HER status is thus a predictive marker for treatment and in addition, has been suggested to play a role for predicting response to chemotherapy¹⁶ and endocrine therapy¹⁷.

Triple negative breast cancer (TNBC) is a heterogeneous subgroup of breast cancer characterized by the absence of expression of estrogen receptor (ER), progesterone receptor (PgR) and human epidermal growth factor receptor-2 (HER-2). TNBC represents approximately 15-20% of all breast cancer cases and is generally considered as the most severe subgroup of breast cancer¹⁸.

1.1.4 Breast cancer heterogeneity

Due to the fact that breast cancer is a heterogeneous and a complex disease, there is a high degree of diversity among patients with similar diagnosis (interheterogeneity). Within a tumor, subclonal population of cancer cells diversity differ in genetic and phenotypic characteristics (intratumor heterogeneity)¹⁹. Tumor heterogeneity has been associated with treatment resistance and relapse, possibly due to subgroups of cancer cells managing the evade treatment²⁰. Previous studies have examined the intra-tumor genetic heterogeneity on breast cancers, however tumor phenotypic heterogeneity cannot be solely attributed to genetic differences, as epigenetics and interactions with the tumor microenvironment

also contribute²¹. Several molecular factors as genetic, transcriptomic, metabolic, proteomic etc further influence tumor heterogeneity and together determine the risk of disease progression and therapeutic resistance.

At the genetic level, germline mutations in BRCA1 and BRCA2 genes account for most hereditary breast cancer cases. At transcriptomic level, five intrinsic subtypes of breast cancer based on gene expression profiles have been classified: luminal A, luminal B, basal-like, HER2 enriched and normal-like. Each of these subtypes has different risk factors for incidence, response to treatment, risk of disease progression and preferential organ sites to metastasize. Luminal tumors have the best prognosis as they are ER+ PgR+ and the majority respond well to hormonal treatments. On the other hand, basal-like tumors in general are deficient in hormone receptors and in HER2. Thus, the majority of these tumors are also called triple negative breast cancer (TNBC) and unfortunately only approximately 20% of these tumors respond well to standard chemotherapy, thus basal-like tumors have the worst prognosis and is considered the most aggressive subtype²².

1.1.5 Breast cancer treatment strategies

After a breast cancer diagnosis, a treatment plan specific to patient situation is developed based on tumor size, histological characterization, grading, receptor status, axillary lymph node status, age of the patients etc. Surgery and radiation therapy are local treatments, which are directed to the tumor. Systemic breast cancer therapies include other tumor directed treatments as endocrine therapy and novel targeted therapy (inhibits angiogenesis growth factors, cell division, promoting cell death), as well as not tumor specific chemotherapies. Table 1.1 provides a simplified overview of breast cancer treatment in Norway²³.

Despite the fact that chemotherapy is more toxic than endocrine or targeted therapy, it is often a mandatory inclusion in the treatment plan when tumors are unresponsive to these treatments, especially in TNBC patients. Moreover, doctors can administrate chemotherapy before surgery, neoadjuvant chemotherapy, due to the size of tumor, since the drugs may shrink the tumor and give patient more surgical options. In contrast, adjuvant (meaning “in addition to”) chemotherapy is administered after surgery to prevent recurrence of the disease, particularly distant recurrence, when cancer is invasive and has spread to nearby lymph nodes or patient has unfavourable prognostic factors^{24 25}.

Table 1.1. Overview of breast cancer treatment strategies in Norway.

	Before surgery	Surgery	After surgery
Treatment	Neoadjuvant therapy	Mastectomy or breast conserving surgery (tumor excision)	Radiation Therapy Chemotherapy Endocrine Treatment Others
Purpose	Tumor reduction and down staging	Remove primary tumor and lymph nodes	Reduce recurrence probability

1.2 Tumor metabolism

Downstream genomics, transcriptomics and proteomics is metabolomics, a relatively new field that studies small molecular compounds called metabolites. These compounds are end products or intermediates of chemical processes needed for cell viability, e.g energy production and cell signalling. The metabolic profile of a cell, tissue or organism depends on the 'omics' level as well as environmental factors like diet and drugs²⁶, being an amplified output of on-going cellular activity of a biological system. Small alterations in gene expression levels or in the activity of enzymes could have large impact on the concentration of metabolites.

Due to the accumulated alterations within the cancer cells that contribute to their characteristic of uncontrollable growth, they exhibit important metabolic differences compared to normal cells.

As Hanahan et al. suggested, a crucial event of tumor development is deregulation of cellular energetics, characterizing an altered metabolic activity as a malignancy feature²⁷. Metabolic deregulation provides cancer cells with three basic needs: 1) quickly generation of adenosine triphosphate (ATP) as a source of energy, 2) increase synthesis of the four types of macromolecules: lipids, carbohydrates, proteins, and nucleic acids, and 3) suitable redox stability.

Metabolic results can be used to classify breast cancer on the basis of tumor biology, and identify new prognostic and predictive markers to discover new targets for future therapeutic interventions.

1.2.1 Glucose metabolism

Glucose is the major source of energy in living cells. During glycolysis, a small amount of adenosine triphosphate (ATP), the chemical energy transporter essential for cellular processes, is released when glucose is converted into pyruvate together with the reduced form of nicotinamide adenine dinucleotide (NADH). Pyruvate then follows one of two pathways depending on the presence or absence of oxygen. If oxygen is present, pyruvate enters to the mitochondria where it is converted to Acetyl-coenzyme A (Acetyl CoA) and is oxidized in the tricarboxylic acid (TCA) cycle followed by the oxidative phosphorylation, also known as electron transport chain, to produce ATP. Glycolysis, the TCA cycle, and oxidative phosphorylation yield thirty-six molecules of ATP in total.

In hypoxic conditions, i.e. low oxygen concentrations, pyruvate is quickly reduced in the cytoplasm to lactate via the enzyme lactate dehydrogenase (LDH). Through this reaction, the oxidized form of NAD feed glycolysis creating a positive feedback loop. Although the production of ATP via anaerobic glycolysis is 100 times faster than oxidative phosphorylation, it is less efficient as it yields only two ATP molecules per glucose from glycolysis compared to thirty-six in aerobic conditions (Figure 1.3).

In most cancer cells and in proliferative normal cells, the majority of pyruvate produced during glycolysis is converted to lactate even if oxygen is present or not. This switch, discovered in the 1959's, is referred to as the Warburg effect²⁸. To compensate for the inefficient ATP production, which is less relevant than producing new nutrients in cancer cells, most tumors have an increased rate of glucose uptake and an increased glycolysis. It is suggested that the production of lactate stimulate tumor cells, making them more resistant to the immune system and also by generating an acidic microenvironment that is hostile to surrounding normal tissue and promotes metastasis²⁹.

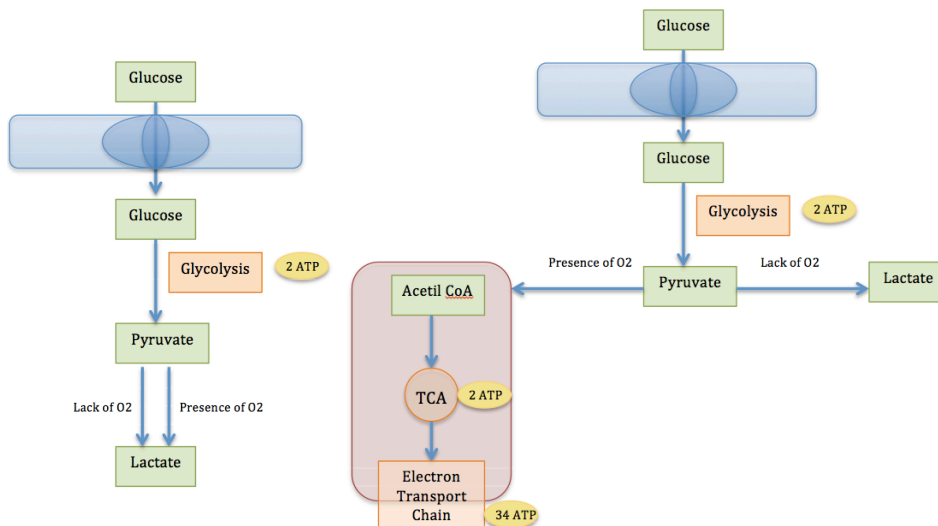


Figure 1.3. Glucose metabolism in proliferative normal cells (left) and in cancer cells (right). *ATP* adenosine triphosphate, *O₂* Oxygen, *TCA* tricarboxylic acid cycle.

1.2.2 Lipid and choline metabolism

In addition to glucose metabolism, several studies have shown that altered choline metabolism as well as altered lipid metabolism is associated with cancer³⁰. Choline is an essential organic compound, which works as a precursor for phosphatidylcholine (PtdCho), the most abundant phospholipid in eukaryotic cellular membranes³¹. Free choline (Cho) is transported into the cell and converted to phosphocholine (PCho) by the enzyme choline kinase in the first step of the Kennedy pathway, described by Kennedy and Weiss in 1956. PCho is then added a cytidyldiphosphate (CDP) group forming cytidine diphosphocholine (CDP-Cho), which combined with 1,2-diacylglycerol synthesize PtdCho. The breakdown products of PtdCho catabolism are glycerophosphocholine (GPC) and 1-acylglycerophosphocholine (Figure 1.4).

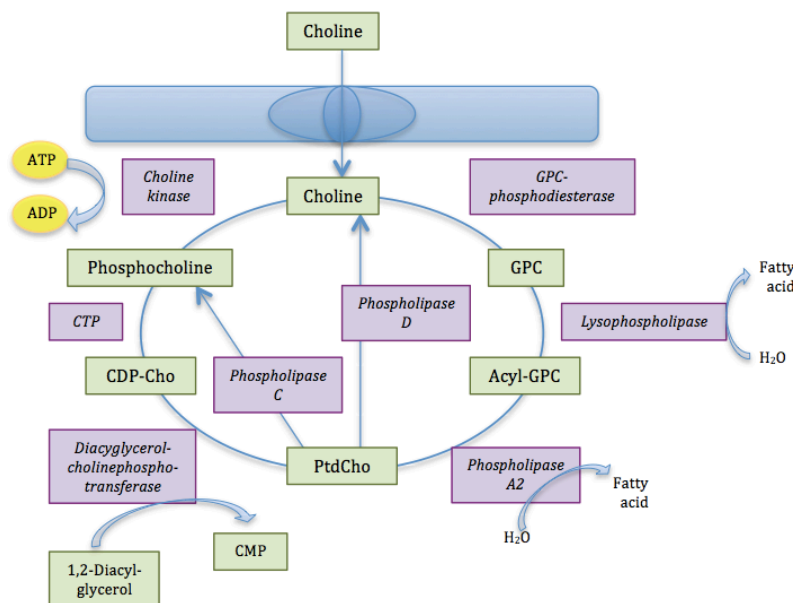


Figure 1.4. Choline metabolism. *GPC* glycerophosphocholine, *PtdCho* phosphatidylcholine, *CMP* cytidylmonophosphate, *CDP-Cho* cytidine diphosphocholine, *CTP* cytidyltriphosphate, *ATP* adenosine triphosphate, *ADP* adenosine diphosphate.

Choline compounds and lipids are considered essential to sustain cell proliferation since cancer cells need substrates for membrane multiplication for new cells. Increased levels of one or more of the metabolites constituting total cholines (tCho: the sum of Cho, GPC and phosphocholine) and an increased lipid biosynthesis have been detected in tumors and are associated with cancer aggressiveness, being an emerging metabolic hallmark of oncogenesis and tumor progression. A reduction in the concentration of tCho has therefore been suggested as a marker of tumor response to cancer treatment.

1.2.3 Amino acid metabolism

Amino acids are structural units to build up proteins in the human body. However, amino acids also have an important role as regulators or intermediate metabolites for several metabolic pathways necessary for the cellular maintenance and growth. Changed levels of different amino acids have been detected in cancers compared to normal cells, and may be a reflection of the high proliferation rate of cancers.

Glutamine is the second principal substrate for cell growth in both normal and cancer cells. It is an essential amino acid, which not only contribute with carbon, but also with reduced nitrogen for *de novo* biosynthesis of a number of diverse nitrogen-containing compounds and nucleotides. Glutamine can be converted, by glutaminase, to glutamate, which further can be used for production of other amino acids or function as a precursor for the important antioxidant glutathione.

Increased levels of the aminoacid taurine have been found in cancerous compared to normal tissue in studies of cervix³², prostate³³, and colon³⁴ tissues. For breast cancer, increased levels of taurine have been detected in cancer tissues³⁵ while the levels were decreased in serum samples of cancer patients compared to healthy volunteers³⁶. Taurine is the most abundant free amino acid in humans and has numerous potential health benefits as it regulates intracellular calcium, homeostasis and maintain the cell membrane stability and protect cells. A study showed that Taurine is an inhibitor of tumor growth and a promoter of apoptosis in human breast cancer cells³⁷. However, limited studies have assessed its effect on tumors and the antitumor mechanism remains unknown.

1.3 Metabolomics

Metabolomics is a newly emerging field of “omics” research concerned with the characterization of the small molecule metabolites and their intermediates present in a biological system. In metabolomics, metabolite levels are measured to obtain the final downstream information of ongoing processes at a specific time (Figure 1.5). Compared to the other “omics” measures, which can be affected by post-modifications and other regulatory mechanism, metabolomics is a powerful approach to directly reflect the final product of gene expression. Thus, metabolomics best represents the molecular phenotype.

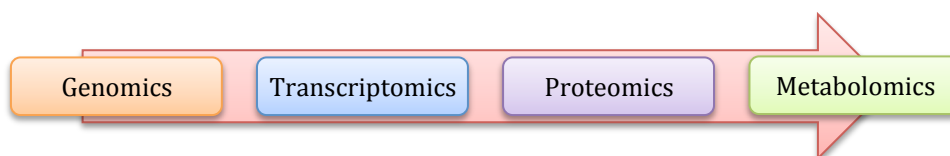


Figure 1.5. The omics cascade. From genomics, the study of DNA and genetic information within a cell, transcriptionist; the study of RNA and differences in mRNA expression and proteomics; the study of proteins, to metabolomics; the study the metabolism substrates and products, which are influenced by both genetic and environmental factors.

Many reactions take place continuously within cells, so concentrations of metabolites are considered to be dynamic, and may change rapidly from one time point to the next. Current analytical techniques used to investigate metabolomics, provides a snapshot in a specific time under a defined condition. The two main techniques are magnetic resonance spectroscopy (MRS) and mass spectrometry (MS). MS is a very sensitive technique, and can detect metabolites present in much lower concentrations than MRS. MRS, however, is highly quantitative and reproducible, suitable for samples in a broad range of conditions independent of acidity and hydrophobicity. In addition, tissue MRS can be performed in a non-destructive technique requiring a minimum of sample preparation³⁸.

1.3.1 MR spectroscopy

MRS is a common tool for examining the metabolic state of a biological system. MR spectroscopy obtains information using the magnetic properties of certain atomic nuclei. These nuclei may possess a quantum mechanical property called spin (I), which can be compared to the rotation of a planet. This property depends on the number of neutrons and protons, and its different combinations giving distinct spin configurations. When a nucleus has the same number of protons and neutrons, they cancel each other and thus they experience no spin ($I=0$). When the number of protons or neutrons is different, nuclei possess non-zero spin ($I \neq 0$) and generate their own magnetic field, called the magnetic moment (μ), and thus giving a signal. ^1H , ^{13}C , ^{14}N , ^{19}F and ^{31}P are used in MRS, of which ^1H is the most commonly used due to his high sensitivity and high natural abundance in the body.

When an external magnetic field (B_0) is applied, the nuclei is oriented in $2I+1$ different energy levels by equilibrium processes, and they precess with a frequency dependent on the type of nuclei and the strength of the magnetic field. ^1H has spin $I = \frac{1}{2}$, and will thus be present in two energy levels; oriented either parallel or anti-parallel to B_0 . If a radio frequency (RF) pulse is applied, the nuclei in a lower energy level will excite to a higher energy level and thereby disrupt the equilibrium. When the RF pulse is switched off, the excited nuclei will return to equilibrium and a signal called the free induction decay (FID) can be detected. The FID can be Fourier transformed into a frequency dependent spectrum where the frequencies are determined by B_0 and the gyromagnetic ratio of the nucleus.

Nuclei in different magnetic environments will experience slightly different magnetic fields, and will therefore appear as peaks at different positions, or chemical shifts, of the spectrum. These chemical shifts are then converted into parts per million (ppm). The nuclei of a molecule will also be influenced by the spins of nuclei located closely, resulting in a division of the peak into multiplets.

1.3.2 HR MAS MRS

The molecules of solids and semi-solid materials (e.g. tissue) have a restricted mobility, which leads to static anisotropic (the property of being directionally dependent) interactions between nuclei. This characteristic results to broad peaks in MR spectra that may conceal relevant spectral information and lead to peak overlapping. After its discovery, the use of MRS experiments was limited to dissolved or extracted solid samples or liquid samples. Andrew and Lowe were the first to describe a solution to this problem in 1958: rapid spinning of the sample (typically 5 kHz) around an axis inclined of 54.7 grades (the magic angle) to the direction of the static magnetic field (B_0) will impose motion on the nuclei and thereby reduce line broadening.

This method, called high resolution (HR) magic angle spinning (MAS) mimics a liquid solution state in which the anisotropy of the interactions is averaged to zero resulting in MR spectra of high resolution with narrow peaks³⁹ (Figure 1.6).

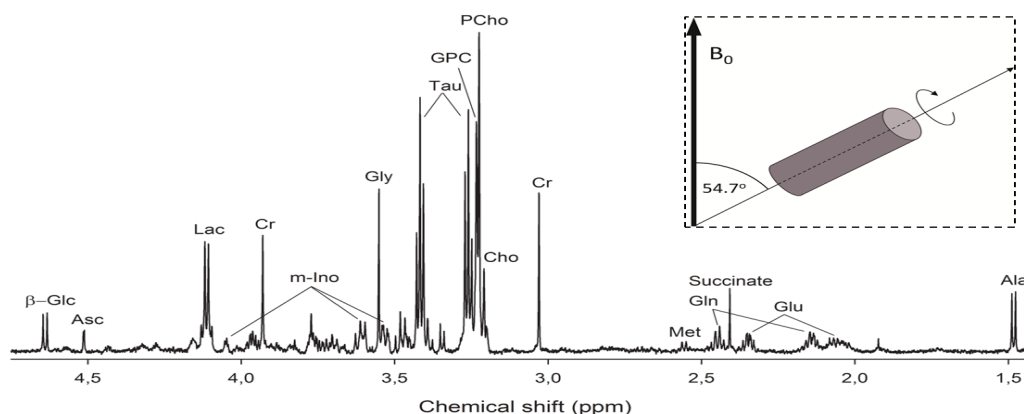


Figure 1.6. A representative HR MAS MR spectrum of breast cancer tissue with assigned metabolites. In frame: schematic representation of a sample in a MAS rotor, inclined in the magic angle (54.7°) to the direction of the static magnetic field B_0 . *Glc* glucose, *Asc* ascorbate, *Lac* lactate, *mI* myo-Inositole, *Cr* creatine, *Gly* glycine, *Tau* taurine, *GPC* glycerophosphocholine, *PCho* phosphocholine, *Cho* choline, *GSH* glutathione, *Gln* glutamine, *Succ* succinate, *Glu* glutamate, *Ala* alanine. The figure is copied with permission from Giskeødegård et al. High-Resolution Magic-Angle spinning NMR Spectroscopy of Intact Tissue⁴⁰.

HR MAS MRS is a non-destructive technique commonly used to obtain the metabolite profile of a tissue sample. The analysis of the intact tissue gives qualitative and quantitative metabolite measures with minimal sample preparation⁴⁰. As the tissue is intact after analysis; it can be used for further analysis as tissue evaluation by histopathology, gene expression profiling or other methods; enabling direct comparisons between spectral and morphological characteristics. Although only a small volume of tissue sample is required for the assessment, it is important to handle the tissue with care and keep it frozen during storage and sample preparation to minimize degradation.

1.3.3 MRS acquisition

Biological tissues contain large amounts of water, and the water signals in an MR spectrum will be several orders of magnitude larger than metabolite signals. Sequences that suppress the water signal are therefore used to give focus to metabolites peaks in MRS acquisition. There are different sequences to suppress water, but the two most common in metabolomics experiments are Nuclear Overhauser Effect Spectrometry (NOESY) and Carr-Purcell-Meiboom-Gill (CPMG). These methods use a pre-saturation of water molecules by exposing the sample to low power continuous wave irradiation before the signal acquisition. Metabolite signals may also be affected by lipids and large molecules, which give broad peaks in the spectrum. CPMG sequences are additionally designed to decrease these signals and obtain all the important metabolites that could be overlapped.

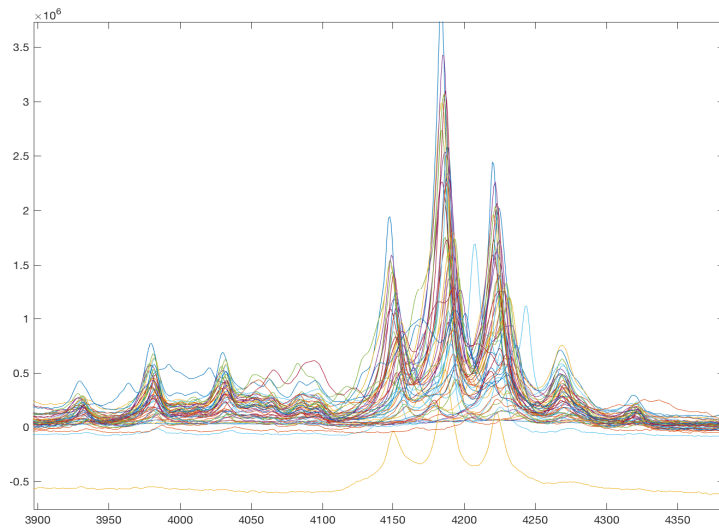
1.4 Preprocessing of MR spectra

HR MAS MR spectra must be preprocessed prior to statistical analysis to remove unwanted technical or biological variation. This procedure involves different computational steps and is carried out to convert the acquired data into a format that is usable to extract the useful information and obtain spectra with high quality, removing irrelevant and unwanted variation such as instrumental or experimental artifacts. The main steps necessary in the preprocessing of HR MAS MRS data are described below.

1.4.1 Baseline correction

Baseline corrections can be applied to correct baseline distortions of the spectra. The baseline consists of broad regions between peaks that do not contain signal of interest. Without correction, baseline will cause errors when performing statistical tests and during quantification and metabolite concentrations. A simple way to correct it is to set the lowest value of each spectrum to zero by subtracting the minimum point (Figure 1.7). This is a safe way of baseline correcting, as it does not change the shape of the spectra⁴¹.

A)



B)

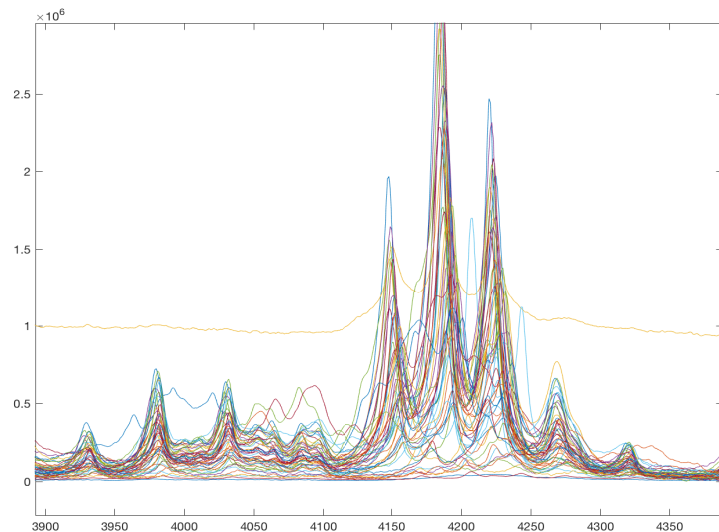


Figure 1.7. Baseline correction based on breast metabolic data. A) Figure shows HR MAS MR region spectra before baseline correction. B) Figure shows the same region spectra after baseline correction. The baseline is clearly less distorted. Yellow spectrum was removed before following procedures.

1.4.2 Peak alignment

The peaks of the MR spectrum can be shifted by differences in pH, temperature, ion-concentrations, molecular interactions or instrumental factors. Peak alignment has the intention to correct these chemical shift differences. Different approaches can be used: either align the entire spectra (global alignment) or separate segments (local alignment). Icoshift is one of the alignments recommended for HR MAS MRS data⁴². The user choose the regions of optional sizes to align, which are shifted to the same segment of a selected reference spectra using Fast Fourier Transformation. This reference spectrum can be a spectrum from the original data set (e.g. the one with the highest correlation to the remaining spectra of the data set) or can be generated by the user (e.g. mean or median spectra of the data set).

1.4.3 Scaling and normalization

After peak alignment, normalization ensures the comparison between spectra, correcting the differences in amount of sample, dilutions (for fluids) or sample weights (for tissue). Several normalization algorithms exist for MR spectra. One frequently used is area normalization, where each variable of the samples is divided by the average of all its variables. This normalization can be considered as a standard normalization approach for MRS metabolic data.

Preprocessing of the variables, or scaling, is performed to bring all the variables into the same range, and regulate the relative importance of each variable to make them finally more comparable. Scaling methods are thus based on variables, not sample-based as normalization.

Mean-centering is a method often applied before other scaling methods. It divides each variable by its own mean resulting in values that fluctuate around zero. Further scaling approaches can be performed, depending on the nature of the data; for instance, autoscaling is a method that divides each variable on its standard deviation. Although autoscaling is commonly used, it may not be optimal for MR spectra because all variables, including noise, are given the same potential to influence the model.

Another approach that focuses to the stable variables of the dataset is variable stability (VAST) scaling method, which divides each variable on its standard deviation and coefficient of variation (CoV)⁴³.

Additional preprocessing operations such as variable selection might also be done. Since decisions on what variables or samples to include will affect the result of multivariate analysis, each step should be carefully evaluated and optimized for the specific data.

1.5 Multivariate analysis

MR spectra consist of a vast amount of variables, making the data sets quite complex to analyse. Additionally, many of these variables are collinear, i.e. an approximate linear relationship exists between some of the variables in X-matrix. Moreover, many peaks can represent one single metabolite and in addition, a high number of variables often exceed the number of samples. Multivariate analysis methods can handle several variables simultaneously and are commonly used for the analysis of MR spectra. Two

approaches are defined based on the purposes they are used for. Unsupervised methods, or clustering, are exploratory and useful tools to identify the data set in terms of possible groupings, patterns and outliers, without taking a response variable into account. Examples of common unsupervised methods are principal component analysis (PCA) and hierarchical cluster analysis (HCA). Supervised multivariate methods are used to classify the data based on a specific property (class membership or a measured variable) and predict this property for new data. These methods model the relationship between independent variables (e.g. spectral data) and a response variable (e.g. clinical characteristics) by identifying patterns in the input data that can be following applied to new data⁴⁴.

1.5.1 Principal component analysis (PCA)

Principal component analysis explains the variance structure of a data set making linear combinations of the variables. Thus, PCA simplifies the data calculating new variables, or latent variables, called principal components (PCs), and they form a new coordinate system of reduced dimensionality. PCs are orthogonal, independent from each other, which is graphically translated as being perpendicular between them. Each new PC explains less variation than the previous one, with the first PC describing the direction of the maximum variation. Thus, the first few PCs usually describe the relevant data variation and the interesting aspects of the data, revealing possible relationships previously hidden in the data.

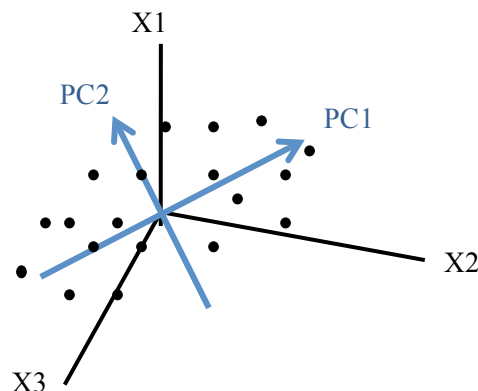


Figure 1.8. Principal components representation Three variables represented: x1, x2 and x3. PC1 explain the maximum variance and PC2 is orthogonal to PC1.

Mathematically, a PCA can be described by equation 1.1:

$$X = TP^T + E \quad (1.1)$$

Where X is the original data set, T is the score matrix, P is the loadings matrix and E is the matrix of residuals. Scores define the position of the samples in the new PC coordinate system and loadings describe how important each of the original variables have been in construction of the specific PC (Figure 1.8). The variation not explained by the PCA model, or matrix of residuals (E), decreases, as more PCs are included.

The combined use of scores and loadings plots, which can be plotted together as a biplot, allows the association or dissociation of samples with variables, observing the positions in the new coordinate system to detect natural clusters and outliers (Figure 1.9).

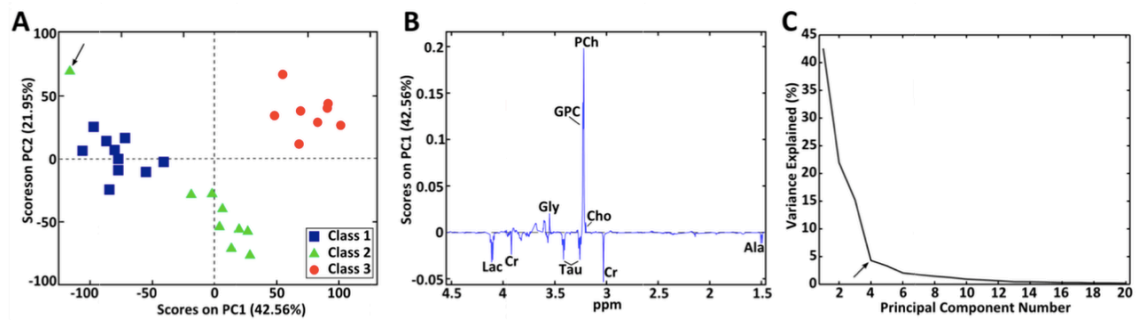


Figure 1.9. Hypothetical example of PCA A) The scores plot shows the three sample classes clearly separated, with scores on PC1 increasing with increasing class number. The arrow indicates a potential outlier. B) The loadings plot shows lactate (Lac), creatine (Cr), taurine (Tau), and alanine (Ala) having lower values for PC1 (i.e. higher in Class 1 samples) while glycine (Gly), glycerophosphocholine (GPC), phosphocholine (PCh), and choline (Cho), have higher values for PC1 (i.e. higher in Class 3 samples). C) Scree plots with arrow indicating the bend in the curve (optimal dimensionality) at four principal components. The figure is copied from Leslie W. Euceda's PhD thesis with permission.

2. Hypothesis

Are breast tumors metabolically heterogeneous, with cancer tumors having larger intra-heterogeneity than non-malignant fibroadenomas?

Have high-grade tumors larger intra-tumor heterogeneity than lower grade tumors?

3. Main objectives

3.1 Overall aim

The main aim of the research presented in this project is to characterize intra- and inter-heterogeneity of breast cancer biopsies by metabolic profiling through HR MAS MRS analysis.

3.2 Specific Objectives

- Optimize a sample handling protocol for metabolic studies of breast tumor tissue to HR MAS MR analysis
- Investigate metabolic differences between different located tumor samples from the same patient and determine if a needle biopsy examination is representative for the whole breast tumor
- Examine metabolic differences between patients with IDC grade 2 and IDC grade 3 with regard to benign disease (fibroadenoma)

4. Materials & Methods

4.1 Patients

4.1.1 Tissue collection

Table 4.1: Patients included in the study

Patient	Operation Date	Age	Diagnose	Histological diagnose	Grade	Surgery
1	2013.06.27	56	Malignant breast tumor	IDC	3	Breast conserving Sentinel node + Axillary clearance
2	2013.04.04	71	Malignant breast tumor	IDC	2	Axillary clearance Mastectomy
3	2012.11.13	40	Fibroadenoma			
4	2013.07.23	48	Malignant breast tumor	IDC	2	Axillary clearance Mastectomy
5a	2014.01.23	26	Fibroadenoma (left)			Tumor excision
5b	2014.01.23	26	Fibroadenoma (right)			Tumor excision
6	2012.11.01	25	Fibroadenoma			Tumor excision
7	2012.06.28	56	Malignant breast tumor	IDC	3	Axillary clearance Mastectomy
8	2012.06.12	43	Malignant breast tumor	IDC	2	Axillary clearance Mastectomy
9	2012.05.15	51	Malignant breast tumor	IDC	3	Axillary clearance Breast conserving Sentinel node

Breast tissue samples were collected from the period 2012-2014 during surgeries at St Olav's Hospital (Trondheim, Norway) (Table 4.1). A tissue slices from each patient was obtained from the middle of the tumor: three fibroadenoma, three ductal carcinoma grade 2 and three ductal carcinoma grade 3 diagnosed patients, in total nine patients aged between 25 and 71 (mean = 44,2). Tumors were between 13.00 and 40.00 mm of size and were examined to conclude free tumor cells resection margins and negative metastasis diagnose. One breast cancer slice was taken from each patient, except for two slices which were taken from the same patient (left breast and right breast). Therefore, ten breast tissue slices were initially collected for the study. Subsequently, five tissue samples were randomly drilled from each sample slice, concluding with fifty frozen tissue sample to analyse by HR MAS MRS.

Information on the expression status of estrogen receptor (ER) and progesterone receptor (PgR) was obtained from previous sample examinations, as well as HER2. Samples from patients 1 and 7 had zero per cent of ER receptor status while samples from patients 2, 4 and 8 had high percentage values between 95.0 and 100.0 %. Human epithelial grow factor receptor was analysed by Fluorescence In Situ Hybridization (FISH) test. FISH find out if there are increased copies of HER2 gene in cancer cells. HER2 results were concluded as negative in all samples, i.e. no HER2 gene amplification was in any IDC diagnosed samples (Table 4.2).

Table 4.2: ER, PgR percentage and HER2 examination provided by clinical data

Patient	ER status %	PgR status %	HER2	HER2 method
1	0.0	0.0	Negative	FISH
2	95.0	75.0	Negative	FISH
4	95.0	95.0	Negative	FISH
7	0.0	0.0	Negative	FISH
8	90.0	0.0	Negative	FISH
9	100.0	100.0	Negative	FISH

Diagnose details and additional clinical information of IDC diagnose patients are showed in table 4.3. Tumor sizes and types of treatment the patient was given are also showed in table 4.2. Patient’s number 1, 2, 3, 8 and 10 were treated with adjuvant chemotherapy, while patient’s number 4 and 9 were treated with neoadjuvant chemotherapy as well.

Table 4.3: Additional clinical information

Patient	Metastasis Diagnose	Resection Margins	Tumor Size (mm)	Treatment
1	Negative	Free	23.0	Adjuvant chemotherapy, Adjuvant radiation therapy
2	Negative	Free	40.0	Adjuvant Endocrine Therapy Adjuvant Chemotherapy Adjuvant Radiation Treatment
4	Negative	Free	20.0	Adjuvant Endocrine Therapy Adjuvant Radiation Treatment Neoadjuvant Chemotherapy
7	Negative	Free	23.0	Adjuvant chemotherapy Adjuvant radiation therapy
8	Negative	Free	13.0	Adjuvant Endocrine Therapy Adjuvant Radiation Treatment Neoadjuvant Chemotherapy
9	Negative	Free	20.0	Adjuvant Endocrine Therapy Adjuvant Chemotherapy Adjuvant Radiation Treatment

4.1.2 Sample storage

Breast samples slices were frozen in -80 degrees immediately after dissection during surgery and stored until further sample handling. To perform an HR MAS MRS analysis and provide a snapshot of tissue metabolism at the time of sampling it is important to preserve the sample in correct temperature conditions as robustly as possible during sampling and storage. As degradation starts immediately once the blood circulation to the tissue is cut, keeping the sample cold reduced both chemical and enzymatic degradation (Figure 4.1).

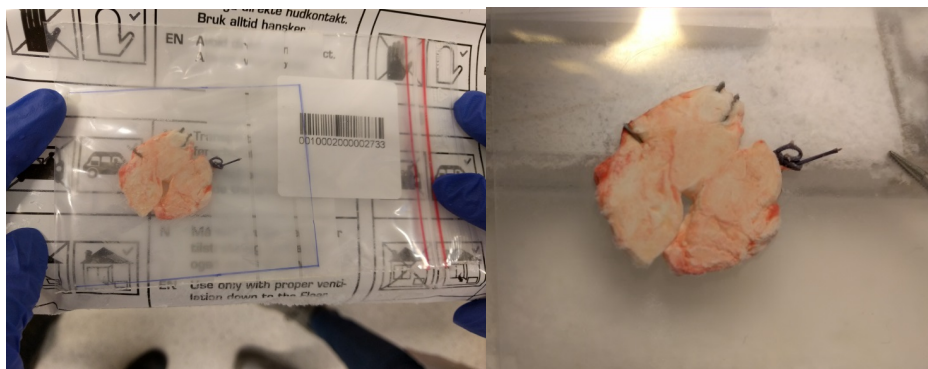


Figure 4.1. Breast tumor slices. Left: Breast tumor slice sample 3 placed on dry ice to preserve cold temperature. Right: breast tumor slice sample 3 over the workstation prepared to drill.

4.1.3 Sample drilling

With the purpose of describing metabolic heterogeneity within a tumor sample, metabolic profiles of different slice positions had to be done. To minimize tissue degradation, the samples should preferably not be thawed during sample preparation. In order to achieve it, frozen samples were prepared on a cooling workstation filled with liquid nitrogen (Fig 4.2). The preparation time did not exceed five minutes, and consisted in drilling randomly five different tissue samples spots of each tissue slice (see figure 4.3) to examine if location may play an important role in metabolic profile.

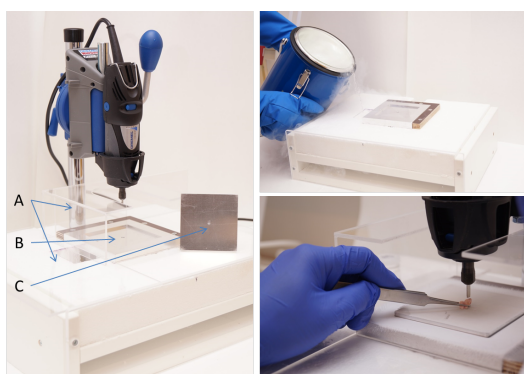


Figure 4.2. Workstation and drill (left) The workstation was filled with nitrogen (*top right*), ensuring the samples to be kept frozen throughout sample preparation. The tissue was cut to fit into an insert using a drill with a hollow core drill bit (*bottom right*). (a) Plexiglas for minimizing nitrogen evaporation, (b) cooling block, (c) metal plate. The figure is copied with permission from Guro et al. High-Resolution Magic-Angle_spinning NMR Spectroscopy of Intact Tissue.

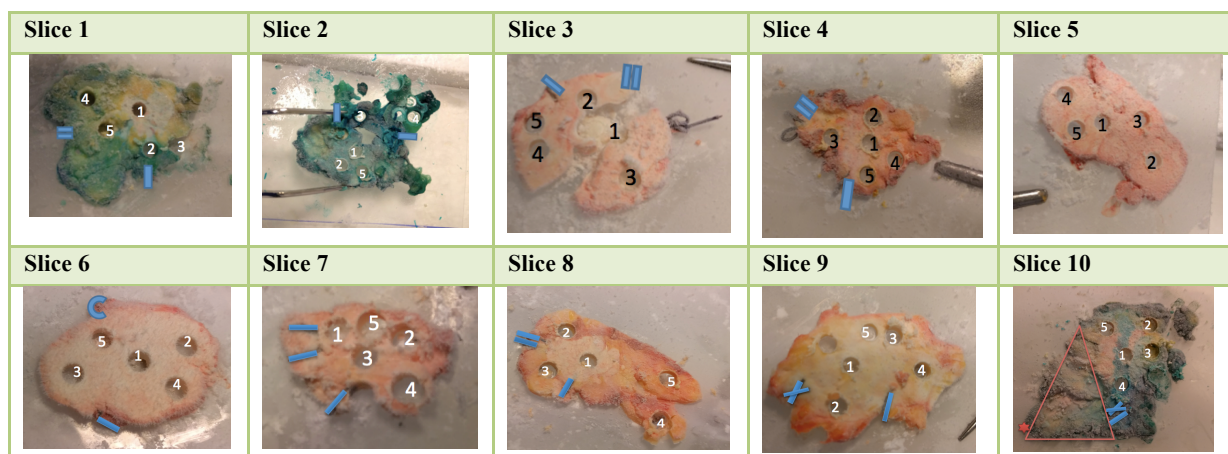


Figure 4.3. Breast tumor slices samples drilled. Numbers indicates the number of slices spots. Slice 3 was broken during drilling. Area marked by a triangle in slice 10 was too thin to drill. Slices 1, 2 and 10 were stained with cytokeratin for histologic examination of sentinel lymph nodes (SLN).

4.1.4 Histopathology and clinical diagnosis

After drilling, the resulting tissue samples were cryosectioned for histological analysis by the CMIC core facility, NTNU. A cryosection from each HR MAS MRS sample was HES stained for measuring relative amounts of tumor tissue, fat, connective tissue etc in each sample. To ensure the content of tumor cells within the spot, histological slides were scored by an experienced pathologist. Twenty-one spots were excluded after being diagnosed as not containing tumor cells.

4.2 HR MAS MRS experiments

After cryosectioning, the tissue samples were cut to fit leak-proof disposable 30 μL inserts (Bruker, Biospin Corp, USA) containing 3 μL cold buffer solution of sodium formate (HCOONa) in D_2O at a concentration of 24.29 mM. The insert was weighed before and after the sample was put into it. This step was carried out very carefully as air bubbles might have caused problems with line shape in MR analysis and made the samples difficult to shim in the magnet. The insert was transferred in a 4-mm diameter zirconium MAS rotor and kept at -20°C until spectral acquisition (Figure 4.4). The insert should have not be able to rotate inside the rotor. To avoid it, the screw cap was adjusted by tightening or loosening up it before to put on the rotor cap. All procedures were done under RNase free conditions using RNase decontamination solution (*RNaseZap*) for future RNA experiments.

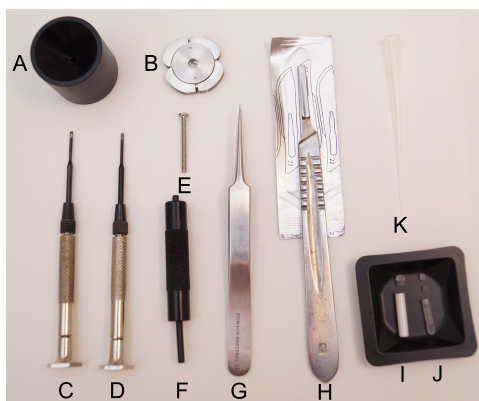


Figure 4.4 HR-MAS equipment (a) Filling funnel, (b) rotor cap remover, (c) blunt-ended screwdriver, (d) forked screwdriver, (e) extraction screw, (f) rotor packer, (g) tweezers, (h) scalpel, (i) rotor with spinning cap, (j) insert with taper and screw cap, (k) a narrow pipette tip for placing D₂O with formate into insert. The figure is copied with permission from Guro et al. High-Resolution Magic-Angle-spinning NMR Spectroscopy of Intact Tissue.

4.2.1 Acquisition protocol

Samples were analysed on a Bruker Avance DRX600 spectrometer (Bruker, Biospin, GmbH, Germany) equip with a ¹H/¹³C MAS probe with gradient. Before acquisition, samples were spun for 5 minutes to allow for temperature acclimation. ¹H spectra were acquired using a water and lipid suppressing spin-echo CPMG (Bruker) sequence through standard acquisition parameters. Two additional experiments, namely NOESY and two-dimensional J-resolved pulse sequences were acquired, but were not used for the purposes of this thesis. Magic angle adjustment and temperature calibration were performed regularly, especially after maintenance, e.g. after changing the NMR probe. HR MAS protocol (*Version October 2016, Generated by Tonje Haukaas and Maria Dung Cao, NTNU, ISB, MR Cancer Group*) used in the study is attached in the appendix.

4.3 Spectral preprocessing and data analysis

The acquired spectral data was Fourier transformed into 64k real points after modification by an exponential line-broadening factor of 0.30 Hz. Each spectrum was both automatically and manually phase corrected using TopSpin 3.5 pl6 (Bruker Biospin). Subsequent preprocessing was performed in Matlab R2017a (The Mathworks, Inc., USA). Chemical shifts were referenced to the creatine peak at 3.03 ppm. The spectral region between 1.11-4.67 ppm, containing the majority of low-molecular weight metabolites, was selected as the region of interest. Additional baseline correction was achieved by subtracting each spectrum with the lowest value. Peak alignment was performed using icoshift and normalization was carried out by mean normalization.

4.3.1 Analysis of individual metabolites

Metabolite peaks were assigned based on previous identification in HR MAS MR spectra of human breast tumor tissue and by identifying each metabolite in Chenomix software. As breast tissue is composed of high lipid concentrations due to normal adipose tissue, HR-MAS spectra show strong lipid contributions

that usually overlap metabolite regions and minimize metabolite signals caused by its high peaks in the spectra, which hamper metabolite estimation. In order to take it into account, relative quantification was calculated in data before and after removing lipid peaks, by integrating fixed spectral regions corresponding to the metabolite of interest in Matlab R2017a. Concluding with two different dataset: first one with lipids and second one without lipids. In order to identify an unknown peak in spectra, STOCSY script was run and subsequently a Heteronuclear Single Quantum Coherence experiment was performed. HSQC is a highly sensitive 2D-NMR experiment which resulting spectrum is two-dimensional (2D) with one axis for proton (^1H) and the other for a heteronucleus (usually ^{13}C or ^{15}N) so further information was provided.

Using in vivo MRS, signals from GPC, PCho and Cho are gathered as a single peak termed total choline (tCho). Due to overlapping lipid peaks in most spectra in the first creatine peak (3.928-3.914 ppm) and in the second glutamine peak (2.177-2.114 ppm), the regions to quantify were creatine 2 peak (3.045-3.024) and glutamine 1 peak (2.485-2.430 ppm). And since there was more than one peak for some metabolites (Glucose, Myo-inositol, Taurine and Glutathione), means of integrated peaks were calculated.

4.3.2 Statistical analysis

Interclass correlation for each metabolite was calculated between samples originating from the same tumor slice in R studio (2009-2015 RStudio, Inc.) and adjusted from multiple testing by the Benjamini Hochberg methods for future analysis. Coefficient of variation (CoV) was calculated in Matlab R2017a to compare metabolites quantification integration between patients and cancer diagnoses. CoV represents the ratio of the standard deviation to the expected value shown in the equation 1.1, in this case, the mean of each metabolite; thus, the amount of variability relative to the mean is described.

$$\text{Coefficient of variation} = \frac{\text{standard deviation}}{\text{mean}} \quad (1.1)$$

4.3.3 Multivariate analysis

Multivariate analysis on spectra was performed in Matlab R2017a and PCA was employed using PLS Toolbox 8.2.1 (Eigenvector Research Inc., U.S.A) to explore naturally occurring differences in data. Spectra were mean-centered prior to multivariate model building. The number of PCs for PCA was selected based on visual inspection of residual explained variance plots. Principal component analysis was performed in both data, with lipids and without lipid data, comparing sample's location and clinical diagnosis.

5. Results & Discussion

5.1 Histopathology and clinical results

All breast cancer cores were pathologically diagnosed and tumor percentage scored from cryosections acquired before HR MAS analysis. Figures with histology detailed results and additional pathologist's comments of following percentages: tumor tissue (divided into tumor epithelial tissue and tumor stromal tissue), connective tissue, normal breast tissue, fat, necrosis and an overall assignment of the sample classified as either cancer, normal tissue or fibroadenoma, are included in the appendix. After histology results, statistical analysis were repeated after removing samples without tumor cells marked in Table 5.1 to investigate how including samples without tumor cells were affected the results.

Table 5.1: Overview of histology diagnose of patients. Samples without tumor content are marked in blue. Sclerosing adenosis (SA) is a benign breast lesion, which demonstrates increased numbers of distorted lobules accompanied by stromal fibrosis.

Patient 1 IDC3	1	X	Patient 5a Fibroad	1	Fibroad
	2	Cancer		2	Sclerosing adenosis
	3	Fat		3	Sclerosing adenosis
	4	Not diagnostic		4	Fibroadenoma
	5	Cancer		5	Sclerosing adenosis
Patient 2 IDC2	1	Cancer	Patient 6 Fibroad	1	Normal
	2	Cancer		2	Fibroad
	3	Cancer		3	Normal
	4	Fat		4	Normal
	5	Normal		5	Fibroad
Patient 3 Fibroad	1	Fibroad	Patient 7 IDC3	1	Cancer
	2	Fibroad		2	Normal
	3	Fibroad		3	Cancer
	4	Fibroad		4	X
	5	Fibroad		5	X
Patient 4 IDC2	1	Cancer	Patient 8 IDC2	1	Normal
	2	Tumor tissue		2	Fat
	3	Tumor cells infiltration in connective tissue		3	Normal
	4	Normal		4	Normal
	5	Normal		5	Normal
Patient 5a Fibroad	1	Fibroad	Patient 9 IDC3	1	Cancer
	2	Fibroad		2	Fat+connective tissue
	3	Normal		3	Fat+connective tissue
	4	Fibroad		4	Cancer
	5	Fibroad		5	Cancer

As metabolite quantification and multivariate analysis were performed before and after histology results, four final datasets were obtained (Table 5.2): Dataset 1 (with all samples and lipids), dataset 2 (with all samples but with lipid removal), dataset 3 (with only twenty-nine included samples and lipids) and dataset 4 (with only final included samples and with lipid removal).

Table 5.2: Datasets classification

Dataset 1	With lipids	All initial samples included
Dataset 2	Without lipids	
Dataset 3	With lipids	Only final 29 samples included
Dataset 4	Without lipids	

5.2 Analysis of individual metabolites results

Metabolite quantifications were performed integrating 26 metabolites peaks and 3 lipid peaks: *B-glc1*, *asc*, *lac*, *m11*, *tyr*, *cr1*, *glu1*, *glu2*, *y1*, *m12*, *gly*, *m13*, *tau1*, *s1*, *tau2*, *gpc*, *pcho*, *cr2*, *gluth1*, *lipid 12*, *gluth2*, *gmline1*, *succ*, *gmate*, *acetate*, *lipid 9*, *ala*, *lipid 7* and *lactate*. Quantification data was used to calculate CoV.

In addition to HR MAS MRS experiments of all samples, additional running experiment was made only with RNase zap free content mixed with buffer (Figure 5.1), in order to obtain a single spectrum and compare it to the spectra ensuring that any contamination occurring from RNase kit were not present in the spectra.

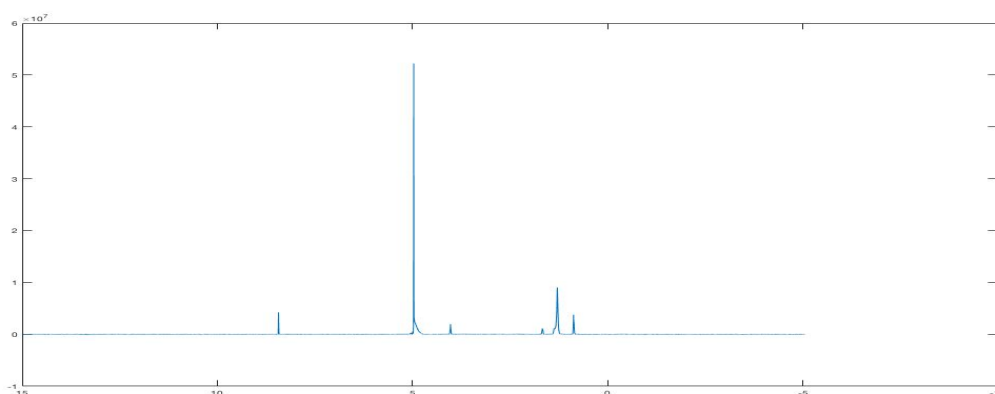


Figure 5.1 RNase zap free spectrum mixed with buffer.

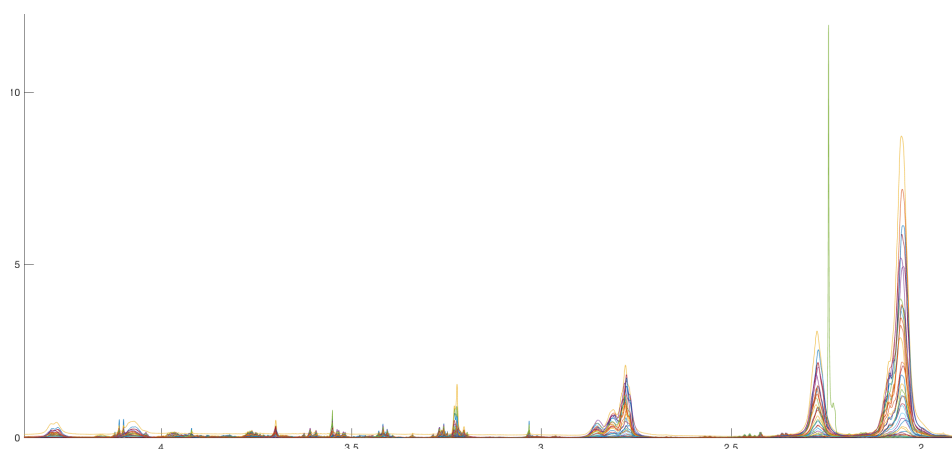


Figure 5.2 General spectra of all samples.

Highest green peak at 2,22 ppm (Figure 5.2) corresponds to a contamination peak from acetone and was removed in pre-processing procedures before statistical and multivariate analysis; it was caused by acetone contamination after cleaning the rotors, and was only present in one sample. Moreover, an additional unknown peak was observed at 3,7 ppm (Figure 5.3). It was significantly high in one sample but it appeared in most of the samples, with greater or less intensity. It could not be identified by Chenomix or searching it in bibliography, so correlation between this peak and all metabolite spectra peaks was performed by STOCSY script, which confirm that unknown peak did not belong to other metabolite and it was independent from other metabolites, there was not correlation (Figure 5.4).

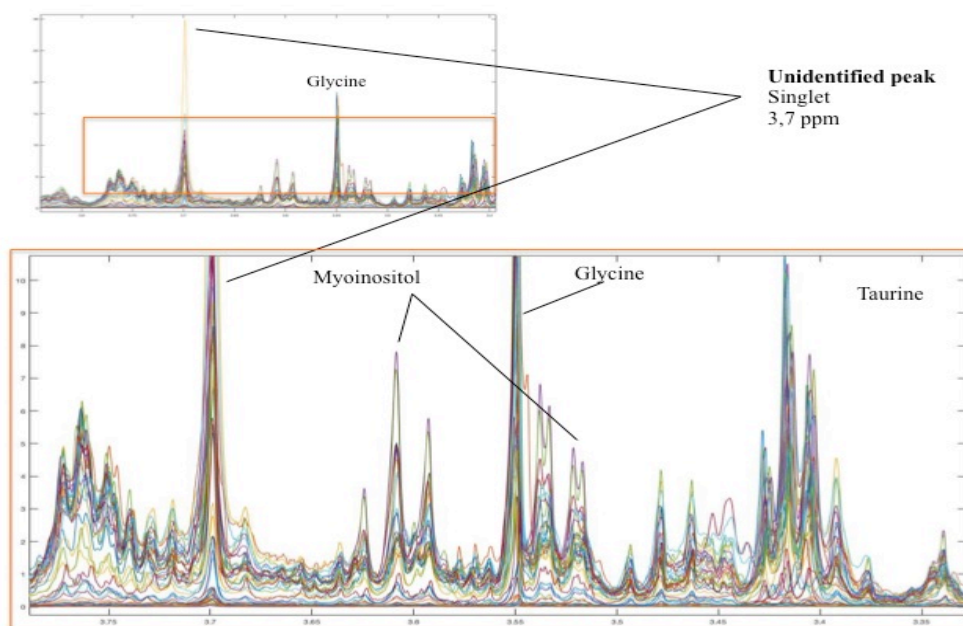


Figure 5.3 Unidentified contamination peak at 3,7 ppm.

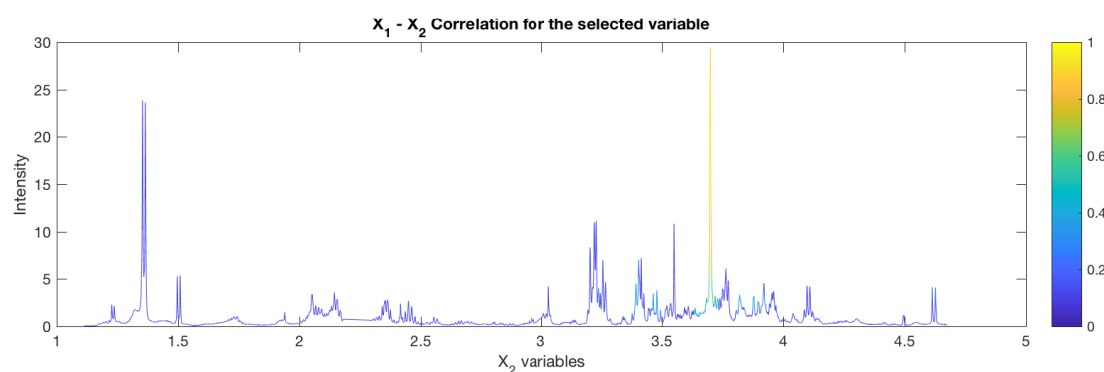


Figure 5.4 STOCSY graph indicating no correlation was between unknown peak and metabolites.

Subsequently, HSQC was performed to identify it. Although this experiment provided more information, the unknown peak could not be identified, and it was preprocessed and integrated as other metabolites, called y1.

5.3 Statistical analysis results

Metabolite quantifications and coefficient of variation calculations (CoV) were first performed with dataset 1 corresponding to dataset containing lipids and dataset 2 corresponding to removed lipid peaks dataset. After histology results, CoV calculations were done with dataset 3 and dataset 4 (without lipid peaks removal and with lipid peaks removal respectively). To remove the effect of lipid signals after removing the lipid peaks, mean normalization was repeated for datasets 2 and 4.

➤ Dataset 1

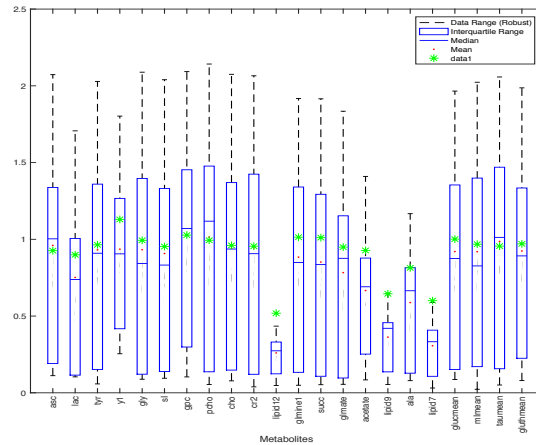


Figure 5.5 Dataset 1. CoV of all metabolites from normalized spectra, including the lipid peaks., presented in boxplots. The central red mark corresponds to the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme datapoints the algorithm considers to be not outliers, and the outliers are plotted individually.

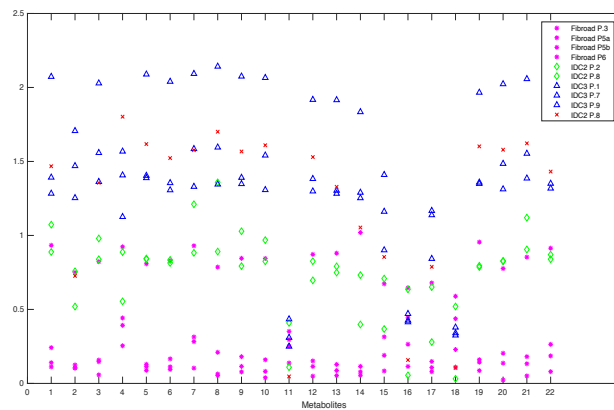


Figure 5.6 Dataset 1. CoV of all metabolites from normalized spectra, including the lipid peaks, colored by cancer diagnose. Blue triangles correspond to IDC 3 patients, green diamonds to IDC2 patients and magenta stars to fibroadenoma patients. Red crosses correspond to patient 8 with IDC2. X axis numbers correspond to: 1 *asc*, 2 *lac*, 3 *tyr*, 4 *yl*, 5 *gly*, 6 *sI*, 7 *gpc*, 8 *pcho*, 9 *cho*, 10 *cr2*, 11 *lipid12*, 12 *glnme1*, 13 *succ*, 14 *glnme*, 15 *acetate*, 16 *lipid9*, 17 *ala*, 18 *lipid7*, 19 *glucmean*, 20 *mImean*, 21 *taumea*, 22 *gluthmean*.

Dataset 1 with lipids shows high CoV in the majority of metabolites, except for lipid 12, lipid 9 and lipid 7 which coefficient variation was significantly lower. Several metabolites as ascorbate, glycine, phosphocholine, GPC, and taurine had the high CoV values compared to the remaining metabolites (Figure 5.5). Figure 5.6 show CoV values colored depending on clinic diagnose (magenta stars: fibroadenoma, green diamonds: IDC2, blue triangles: IDC3 and red crosses: IDC2 Patient 8).

Three main groups of different CoV were exhibited corresponding to three clinical diagnoses. It can be appreciated that IDC3 patients had higher CoV than IDC2, and these higher CoV values than fibroadenomas. Patient 8 diagnosed with IDC2 had higher CoV values than others IDC2 patients, similar to patients diagnosed with IDC grade 3. Patient 8 received neoadjuvant therapy before surgery and all samples were excluded after pathologist examination owing to non-tumoral cells content. These results did not correspond to expected results, as metabolic heterogeneity should be lower after neoadjuvant treatment.

After lipid removal, dataset 2 showed less metabolic variation compared to dataset 1, thus, less metabolic heterogeneity. Prior to lipid elimination, metabolites from dataset 1 had an interval of CoV between 0.1 and 2, and in dataset 2 the CoV were between 0.1 and 1.2, which indicate more metabolic differences if lipids are included than if they are removed from the spectra. But still three main groups were described depending on which clinical diagnose the patients belonged to, supporting that malignant tumors (IDC2-3) had higher CoV than benign tumors (fibroadenoma). Figures of CoV for dataset 2 are attached in appendix.

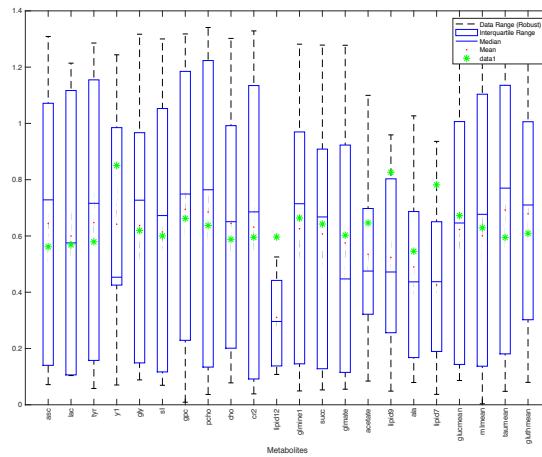


Figure 5.7 Dataset 3. CoV of all metabolites from normalized spectra, including the lipid peaks, presented as boxplots. Samples without tumor content have been removed from the data. The central red mark corresponds to the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme datapoints the algorithm considers to be not outliers, and the outliers are plotted individually.

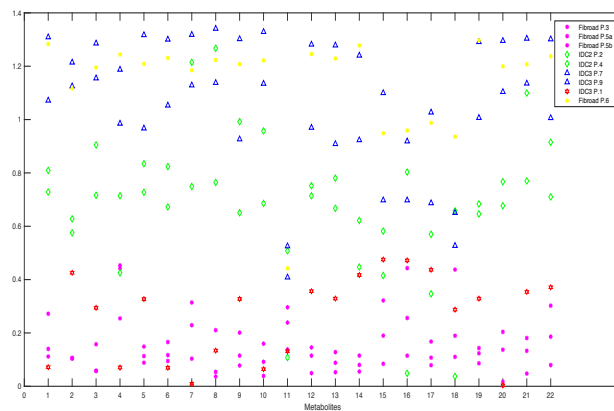


Figure 5.8 Dataset 3. CoV of all metabolites from normalized spectra, including the lipid peaks, colored by cancer diagnose. Samples without tumor content have been removed from the data. Blue triangles correspond to IDC 3 patients, green diamonds to IDC2 patients, magenta stars to fibroadenoma patients, red hexagram to IDC3 patient 1 and yellow stars to Fibroadenoma patient 6. X axis numbers correspond to: 1 *asc*, 2 *lac*, 3 *tyr*, 4 *y1*, 5 *gly*, 6 *sl*, 7 *gpc*,

8 *pcho*, 9 *cho*, 10 *cr2*, 11 *gmine1*, 12 *succ*, 13 *gmate*, 14 *acetate*, 15 *ala*, 16 *glucmean*, 20 *mImean*, 21 *taumean*, 22 *gluthmean*.

After removing non-tumor samples, CoV values of all metabolites in dataset 3 went down due to less heterogeneity between samples from the same patient (Figure 5.7). This initial high CoV range was caused by inpatient metabolic differences that were present before sample removing, as one patient could have several spots with tumor cells and other without tumor content.

Despite this reduction in metabolites CoV, still three distinct groups could be distinguished depending on clinical diagnose: IDC3 patients with the highest CoV values, then IDC2 patients and finally fibroadenoma patients with the lowest CoV values (Figure 5.8).

Interestingly, patient 6 diagnosed with fibroadenoma (colored with yellow stars) had high values of CoV and Patient 1 with IDC3 (red hexagrams) had lower CoV values similar to fibroadenoma patients. Samples 1,3 and 4 were excluded from both patients for not containing tumor cells. Samples 2 and 5 from patient 1 were examined with Tumor content: 80%, Epithelial tissue: 80%, Stromal: 20% and Tumor content: 60%, Epithelial tissue: 80% and Stroma:20% respectively. As histology results were similar, this could be the cause of not high resulting CoV. On the other hand, sample 2 from patient 6 was analysed with Tumor content: 25%, Epithelial: 5, Stromal: 95% and Fat: 20% and sample 5 from patient 6 with T: 75%, Epithelial: 25% and Stromal: 75%. Tumor percentage had 40% of difference between samples from the same patient, and thus could explain the high coefficient of variation showed in results.

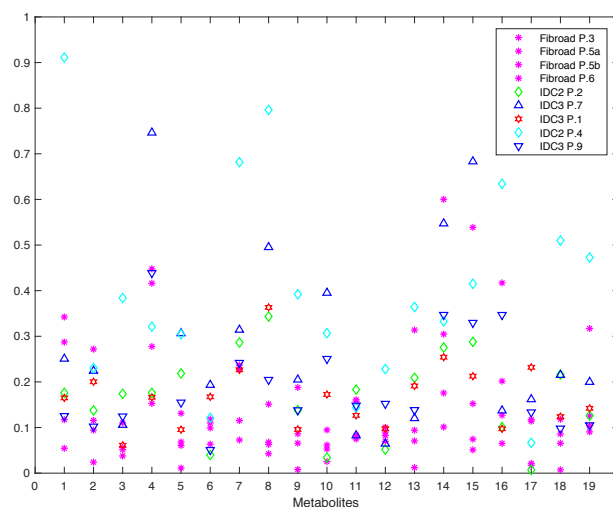


Figure 5.8. Dataset 4. CoV of all metabolites from normalized spectra, with lipid peaks removed, colored according to clinical diagnose. Magenta stars correspond to fibroadenoma patients, green diamonds to IDC2 Patient 2, cyan diamonds to IDC2 Patient 4, red hexagram to IDC3 Patient 1, blue triangle to IDC3 Patient 7 and inverted blue triangle to IDC3 Patient 9. X axis numbers correspond to: 1 *asc*, 2 *lac*, 3 *tyr*, 4 *yl*, 5 *gly*, 6 *sl*, 7 *gpc*, 8 *pcho*, 9 *cho*, 10 *cr2*, 11 *gmine1*, 12 *succ*, 13 *gmate*, 14 *acetate*, 15 *ala*, 16 *glucmean*, 17 *mImean*, 18 *taumean*, 19 *gluthmean*.

Comparing the CoV values among datasets, dataset 4 had the lowest CoV value range (Figure 5.8). The three main CoV groups were not as clear as identified in previous datasets, but still patients diagnosed with fibroadenomas had the lowest metabolic CoV.

Patient 4 diagnosed with IDC2 was the patient that showed highest CoV values compared to other patients. Samples 4 and 5 from patient 4 were excluded after histology results while samples 1, 2 and 3 were included. Sample 1 had Tumor content of 70%, Epithelial: 40%, Stromal: 60% and Fat: 30% but sample 3 had only 10% of Tumor content, 10% of epithelial, 5% of fat and none of stromal. This

histology discrepancy answers why patient 4 showed high CoV values after lipid peaks removal and excluding its non-tumor samples.

Generally, coefficient of variations values were expressed depending on which clinical diagnose the patient had, showing IDC patients a higher CoV compared to fibroadenoma patients, but to an extent of this difference in CoV between clinical diagnoses could be explained by the histologic composition of the samples analysed.

5.4 Multivariate analysis results

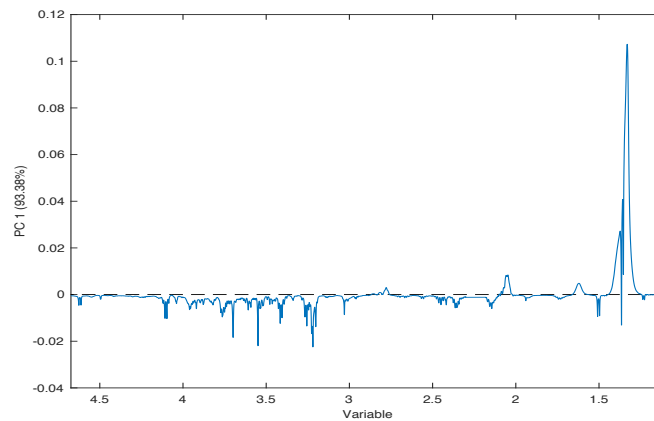


Figure 5.9. Dataset 1 loadings including lipid peaks. Principal component 1 explains 93.38% of variation.

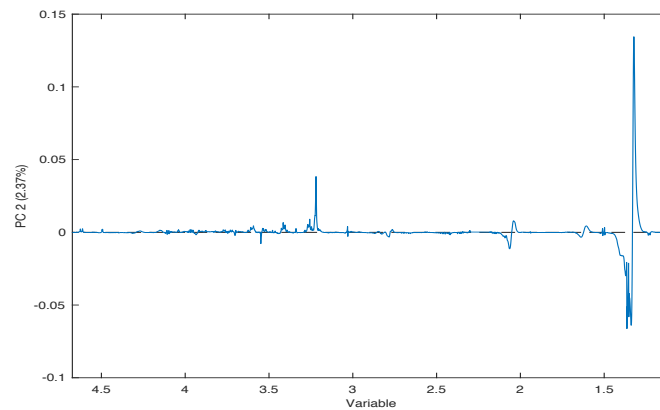


Figure 5.10. Dataset 1 loadings including lipid peaks. Principal component 2 explains 2,37% of variation.

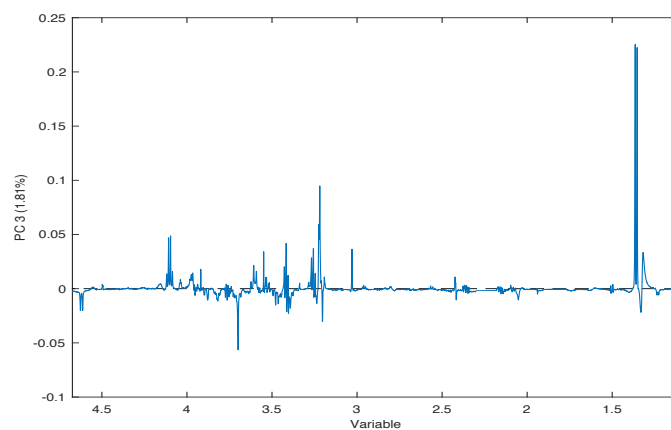


Figure 5.11. Dataset 1 loadings including lipid peaks. Principal component 3 explains 1,81% of variation.

Variance captured was plotted against number of PCs in order to decide the number of principal components for the study. PC1, PC2 and PC3 from dataset 1 explained 93.38%, 2.37% and 1.81% the data variance respectively, and three PCs were considered sufficient to represent the data. Principal component 1 was positive for lactate and lipid peaks (Figure 5.9). PC2 was positive for cholines and lipids and negative for lactate (Figure 5.10). And PC3 was positive for lactate, taurine, cholines and lipids, and negative for glucose (Figure 5.11). Loadings of dataset 2 were positive for the same PC1, PC2 and PC3 metabolites as dataset 1. Loadings plots of dataset 2 are included in the appendix.

Scores plot of all datasets were colored according to first, cancer type (blue: fibroadenoma, green: IDC2 and yellow: IDC3) and secondly according to tumor location (green: core of the tumor, yellow: tumor margin and blue: intermediate region).

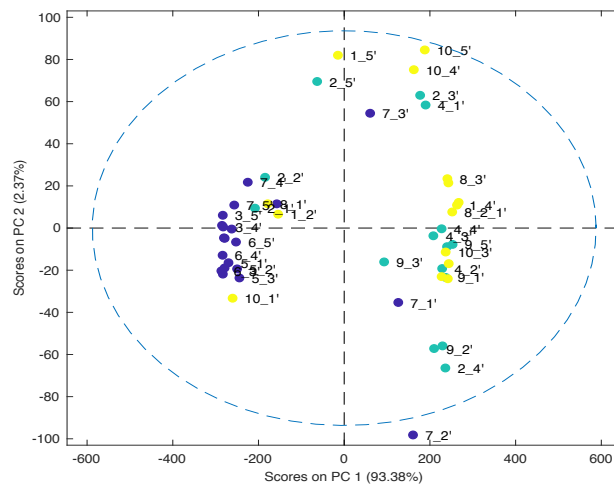


Figure 5.12. Dataset 1. PC1 vs. PC2 scores plot. Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3).

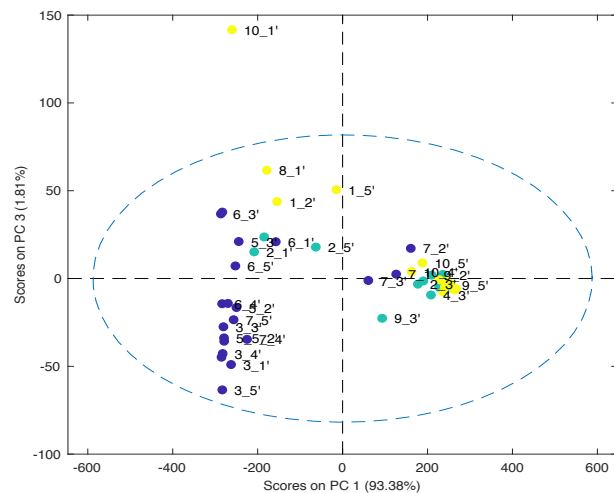


Figure 5.13. Dataset 1. PC1 vs. PC3 scores plot. Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3).

When PC1 was analysed versus PC2, two distinctly groups were shown. The majority of fibroadenomas (17 samples of 20) were had lower levels of lipid compared to the remaining samples (a negative score on PC1), and the majority of IDC (24 of 30) were positive on PC1 (Figure 5.12). These results give support

to the tumor metabolic characteristic of containing high lipids in order to sustain cell proliferation and tumor progression. Conversely, there were not shown clearly groups depending on PC2 metabolic variance.

Fibroadenomas were mostly negative (13 of 20) for PC1 as well as PC3, i.e. had lower levels of lactate, taurine, cholines and creatine (Figure 5.13) compared to the remaining samples. The majority of IDC had positive values for PC1 but neither extremely positive nor extremely negative. Sample 10_1 corresponding to sample 1 from patient 9 was extraordinarily positive of PC3 metabolic variance. The slice from that sample analysed by pathologist was concluded with 100% tumor cells content. As increased cholines and lactate have been identified as malignant cancer traits, these metabolic differences confirm previous studies results.

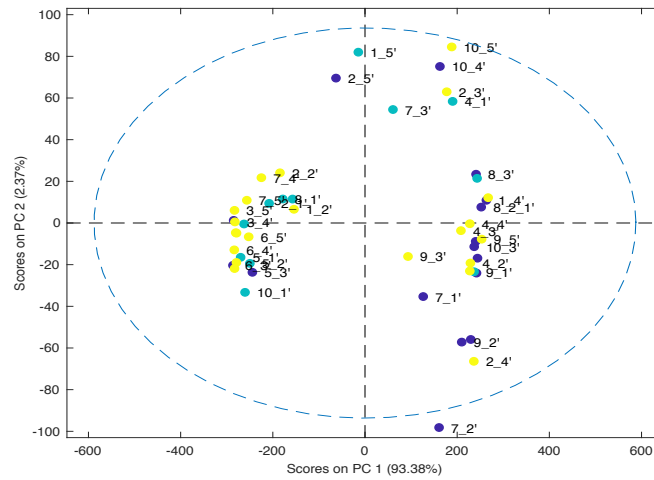


Figure 5.14. Dataset 1. PC1 vs. PC2 scores plot. Colored by location (green: core, yellow: margin and blue: intermediate region).

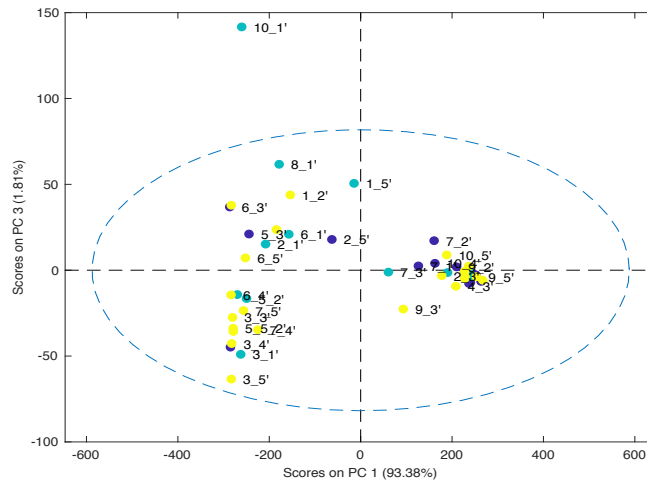


Figure 5.15. Dataset 1. PC1 vs. PC3 scores plot. Colored by location (green: core, yellow: margin and blue: intermediate region).

Results of PC1 versus PC2 (Figure 5.14) and PC1 versus PC3 (figure 5.15) scores plots did not reveal metabolic variance differences between samples taken from the core and samples taken from the margins. For example, samples taken from the margins (yellow) were as positive as negative for PC1, PC2 and PC3. The same happened with samples taken from cores and samples taken from intermediate locations.

Metabolic differences from all patients were showed between samples taken from the same location e.g. core, confirming an existence metabolic interpatient heterogeneity between samples from the same location.

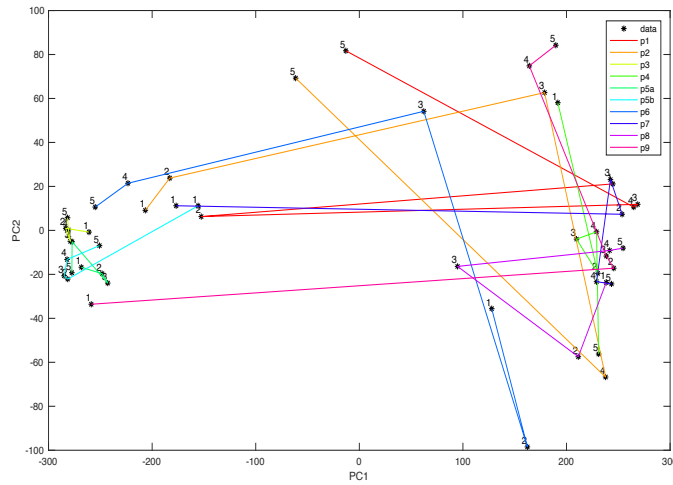


Figure 5.15. Dataset 1. PC1 vs.. PC2 trajectory plot. Samples from the same patient are joined by straight lines. Each colour represents one patient.

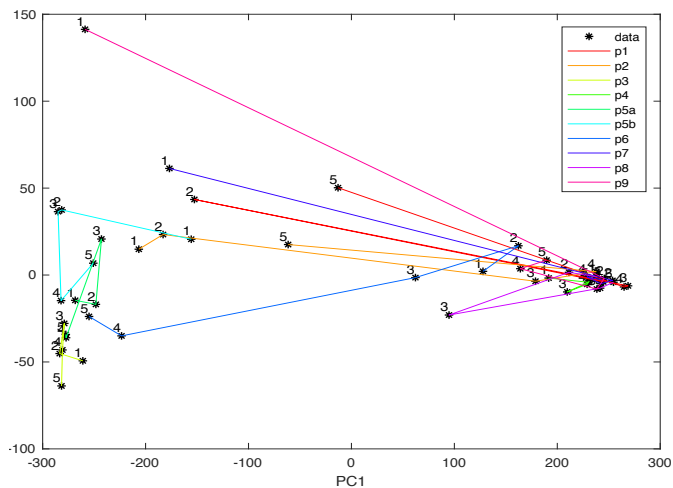


Figure 5.16. Dataset 1. PC1 vs.. PC3 trajectory plot. Samples from the same patient are joined by straight lines. Each colour represents one patient.

In order to facilitate metabolic comparisons between patient’s samples, trajectory plots were performed. Figure 5.15 comparing PC1 and PC2 metabolic differences from dataset 1 show how some samples from the same patient are well grouped while other patient’s samples are metabolically distant.

Samples from patients 3, 5a and 5b (fibroadenoma) and patient 4 (IDC2) were grouped while samples from patients 1, 2, 6 and 9 were notoriously separated. Samples 4 and 5 from patient 6, who was diagnosed with fibroadenoma, were negative for PC1 and quite positive for PC2. Even so, these samples were totally metabolic different from samples 1, 2 and 3 taken from the same patient. After histology results (skive 7 in appendix), samples 1, 3 and 4, were concluded as not containing tumor cells. In spite of samples 2 and 5 (from patient 6) were concluded as fibroadenoma, the percentage of tumor cells were 25% and 75% respectively, which can be attributed to these distant metabolic variation in trajectory plot.

A clear example of metabolic inpatient heterogeneity is clearly showed in Figure 5.16. All samples from patient 9 (diagnosed with IDC3) were positive for PC1 except sample 1, which was negative for PC1 and extremely positive for PC3. Pathologist classified samples 2 and 3 as not tumor content and samples 1, 3 and 5 as positive for tumor content. Even though the samples 1 and 5 had the same clinical diagnose (malignant cancer), the metabolic variation content was completely different. This metabolic difference could be partly explained by the fact that sample 1 had 100% of tumor content in the entire sample and sample 5 had 70% of tumor content in a half of the sample.

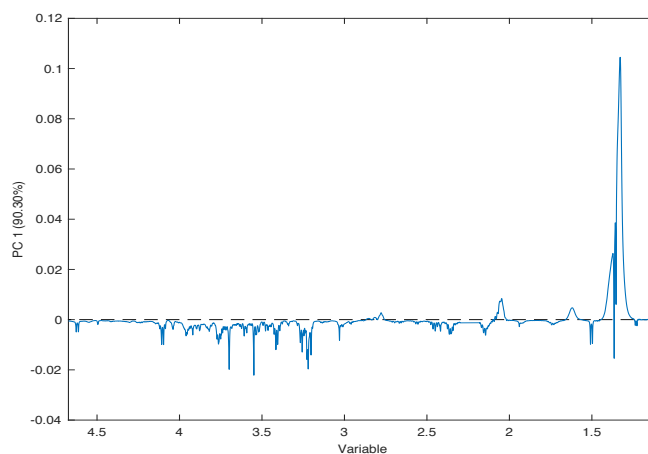


Figure 5.17. Dataset 3. Loadings including lipid peaks .Principal component 1 explains 90,30% of variation.

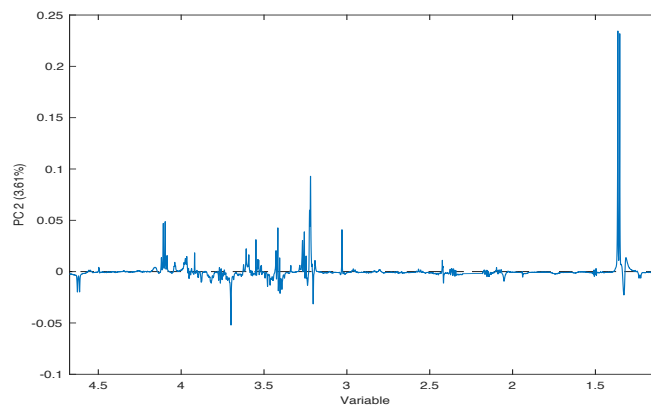


Figure 5.18. Dataset 3. Loadings including lipid peaks. Principal component 2 explains 3,61% of variation.

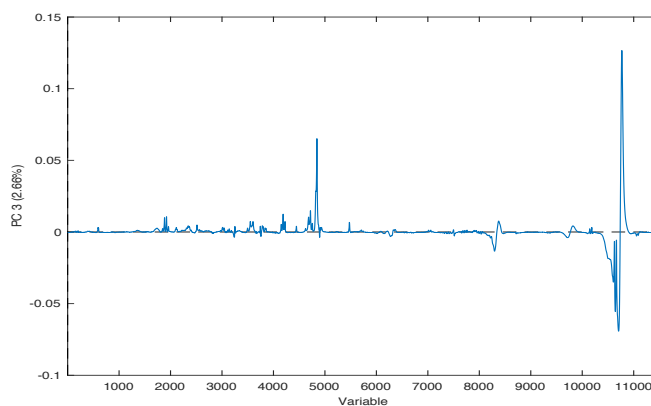


Figure 5.19. Dataset 3. Loadings including lipid peaks. Principal component 3 explains 2,66% of variation.

PC1, PC2 and PC3 loadings from dataset 3 (lipid removal and non-tumor samples removed) were different from dataset 1 and 2. PC1, PC2 and PC3 from dataset 3 explained 90,30%, 3,61% and 2,66% of the data variance respectively. Principal component 1 was positive for lactate and lipid peaks (Figure 5.17). PC2 was positive for lactate, taurine, cholines and lipids, and negative for glucose (Figure 5.18). And PC3 was positive for cholines and lipids (Figure 5.19). The loadings were similar for dataset 4 (attached in appendix).

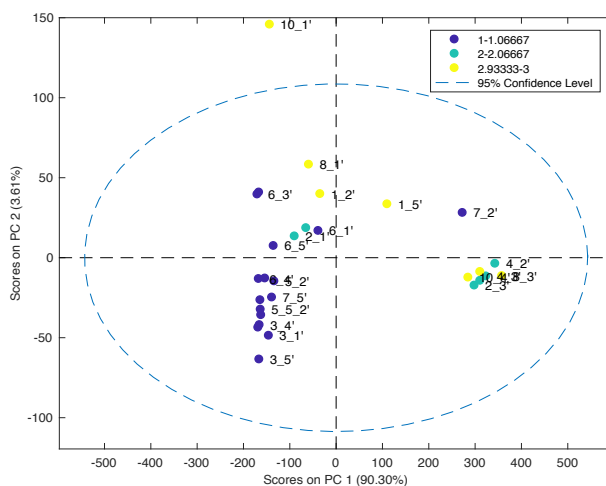


Figure 5.20. Dataset 3. PC1 vs. PC2 scores plot including lipid peaks. Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3).

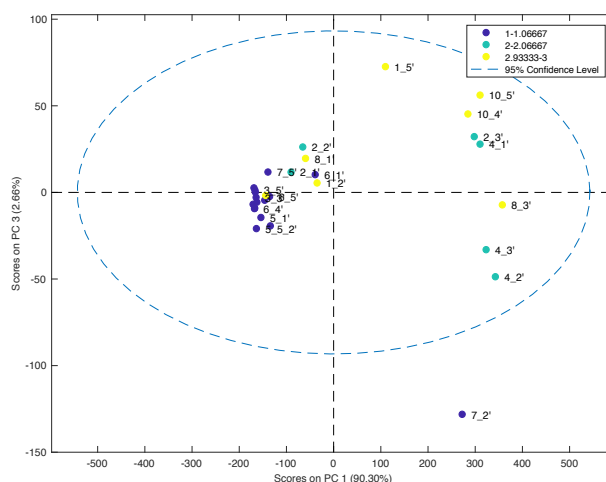


Figure 5.21. Dataset 3. PC1 vs. PC3 scores plot including lipid peaks. Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3).

The same PCA analysis results were obtained in both non-tumor samples removed datasets: dataset 3 and dataset 4. However lipid peaks removal was done for dataset 2 and 4, some lipid peaks after alanine peak (1.5 ppm) were not eliminated. All figures from dataset 3 and dataset 4 are attached in appendix.

As seen from the above discussion of dataset 1, two distinctly main groups were shown in PC1 versus PC2 scores plot (Figure 5.20). Fibroadenomas had a negative score on for PC1 (15 patients from 16) as well as negative for PC2 (12 patients from 16), while IDC patients were divided between positive PC1 and positive PC2.

These two groups negative and positive for PC1 were flagrant showed in Figure 2.21. Nevertheless, a group of samples positive and a group of samples negative for PC3 were not distinguished such as

positive and negative samples for PC1. Sample named as 7_2 corresponded to sample 2 from patient 6, who was diagnosed with fibroadenoma. Sample 2 and 5 from patient 6 were the ones included in the second analysis after histology results. Sample 2, which was extremely negative for PC3 (negative for lactate, taurine, cholines and lipids), had a tumor percentage of 25 % and sample 5, which was positive for PC3, had a tumor percentage of 75%. This contrast metabolic variation between samples 2 and 5 would be explained by histological tumor content cells difference.

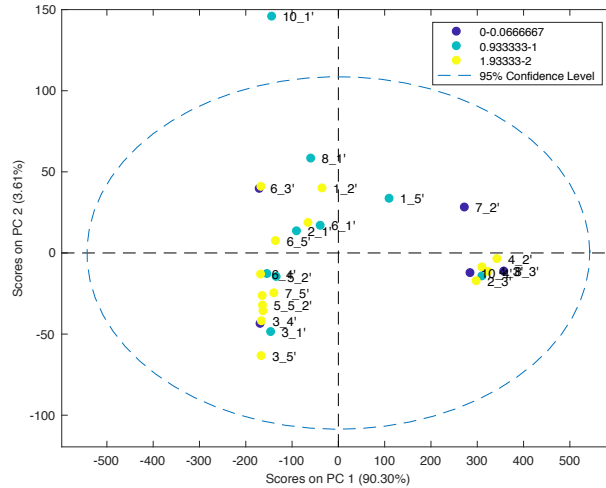


Figure 5.22. Dataset 3. PC1 vs. PC2 scores plot including lipid peaks. Colored by location (green: core, yellow: margin and blue: intermediate region).

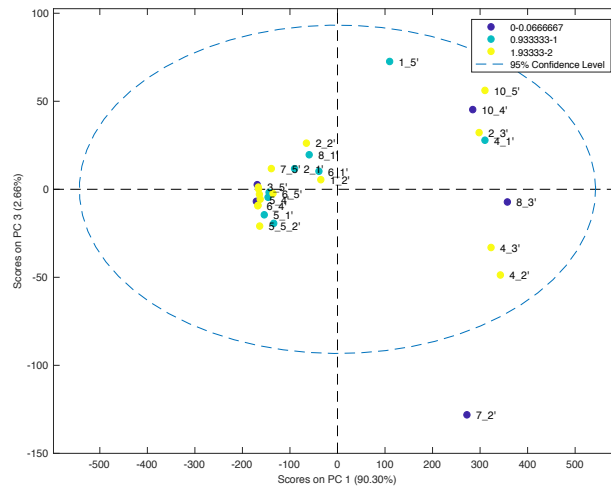


Figure 5.23. Dataset 3. PC1 vs. PC3 scores plot including lipid peaks. Colored by location (green: core, yellow: margin and blue: intermediate region).

Also in dataset 3, metabolic variations results were not associated directly with sample location as previously discussed for in dataset 1. Two main groups were clearly showed in Figure 2.22 checking metabolic variation on PC1 and PC2, but these differences were not associated on wherein samples were collected. The same results were obtained in Figure 5.23, which PC1 versus PC2 scores were plotted.

Although location of the samples were not an important role in the metabolic heterogeneity, it can be observed that after eliminating samples that did not contain tumor content, the inpatient heterogeneity was smaller than inpatient heterogeneity of all samples included, since samples were more grouped in PC1 vs. PC2 and PC1 vs. PC3 dataset 3 scores plots. This could understood by the histologically

differences between samples in dataset 1 and 2, that the same patient could contain samples with tumor cells and samples without tumor content. Although the inpatient heterogeneity was smaller in datasets 3 and 4, there was still a metabolic difference, thus, metabolic heterogeneity between samples from the same patient.

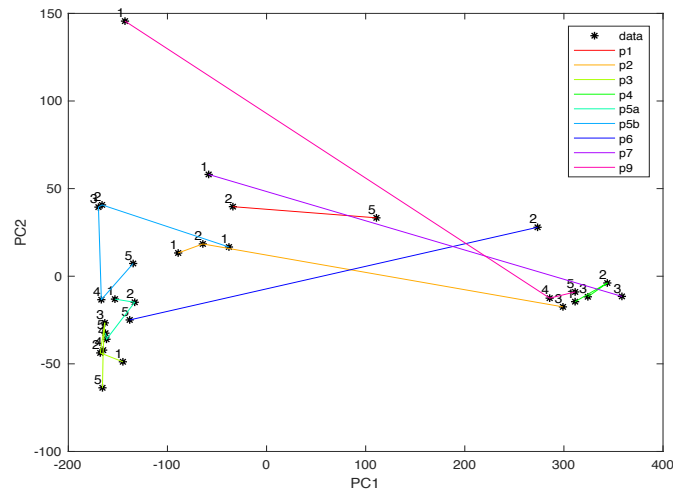


Figure 5.24. Dataset 3. PC1 vs. PC2 trajectory plot including lipid peaks. Samples from the same patient are joined by straight lines. Each colour represents one patient.

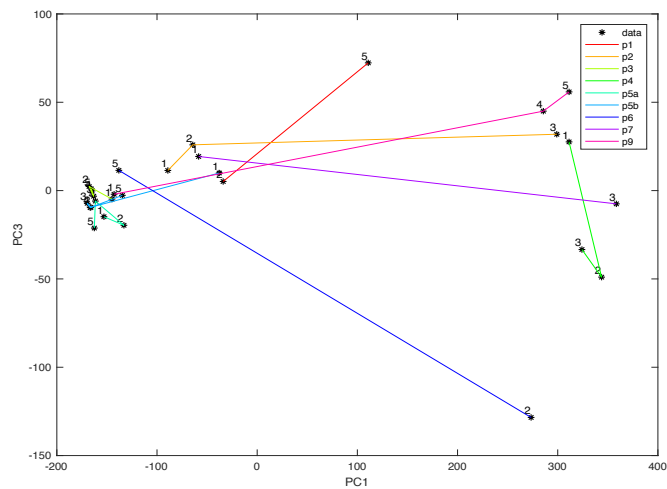


Figure 5.25. Dataset 3. PC1 vs. PC3 trajectory plot including lipid peaks. Samples from the same patient are joined by straight lines. Each colour represents one patient.

Trajectory plots of PC1 versus PC2 from dataset 3 show high metabolic variation between samples from patients: 2, 6, 7, 9 and low metabolic variation between samples from patients 1, 3, 4, 5a and 5b (Figure 5.24).

For instance, two samples (1 and 3) those were included after histology results from patient 7. Sample 1 was negative for PC1 and positive for PC2, contrary to sample 3, which was negative for PC1 and negative for PC2. Sample 1 was concluded by pathologist as cancer tissue with 80% of tumor content, but sample 3 was not certainly concluded due to cells were in poor conditions and it could might be tumor epithelium or tumor associated lymphocytes. Thus, this metabolic differences could be explained by histological tissue content differences.

PC1 vs. PC3 trajectory plot from dataset 3 show the same samples patients with a higher metabolic variation as PC1 vs. PC2 plot (Figure 5.25).

For example, samples 2 and 3 from patient 4 were truly closed compared to sample 1 from the same patient. The samples 2 and 3 were negative for PC3 but sample 1 was positive for PC3. Sample 1 was examined by the pathologist and was concluded with 70% of tumor content and samples 2 and 3 with a 3 and a 10 per cent of tumor content.

Thus, it can be said again that, this metabolic variations could be explained by results on histological tissue content examinations.

6. Conclusions and future perspectives

In this thesis, MR-based metabolomics was used to examine metabolic features contributing to breast cancer heterogeneity and to assess the existence of intra-tumor and inter-tumor metabolic differences depending on sample tumor location. After evaluate the statistical and multivariate results, the data from this project suggest that:

- Sample handling protocol of breast cancer tissue by HR MAS MRS analysis was correctly optimized for further studies, as we achieved high-quality MR spectra.
- MR spectrometry is a promising tool for stratification of patients into clinically useful treatment groups.
- Higher intra-tumor heterogeneity has been characterized in high-grade IDC tumors compared to non-malignant tumors (fibroadenomas).
- Metabolic content variation within a tumor was associated with histology tissue results.
- Tumor location (if samples were targeted to the tumor core, margin or in between) matters less than histology examination for characterization of tumor heterogeneity.
- Breast tumors are metabolically heterogeneous; however a needle biopsy may be representative for a patient's tumor as long as tumor histology is taken into account.

Further studies will try to identify if larger intratumor heterogeneity is linked with resistance to treatment. To achieve it, tumor samples heterogeneity from patients treated with chemotherapy will be analysed and larger patient cohort will be included in study.

Metabolic results will be combined with other platforms (e.g. transcriptomics and proteomics) to further provide targets for investigation of new treatment strategies at different molecular levels.

The results performed in this final Bachelor Thesis will be included in a future publication.

References

1. Cooper, G. M. & Hausman, R. E. *The Cell: A Molecular Approach 2nd Edition*. Sinauer Associates (2007).
2. Understanding cancer - National Cancer Institute. Available at: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>. (Accessed: 25th May 2017)
3. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
4. Jaime Galceran, Alberto Ameijide, Marià Carulla, Antonio Mateos, J. R. Q. & et al. Estimaciones de la incidencia y la supervivencia del cáncer en España y su situación en Europa. *Red Española Regist. Cáncer* 1–58 (2014).
5. Breast Cancer Organization. Breast Cancer Risk Factors: Genetics. Available at: <http://www.breastcancer.org/risk/factors/genetics>. (Accessed: 3rd June 2017)
6. American Cancer Society. Breast Cancer Prevention and Early Detection. *Cancer.org* 1–23 (2016).
7. Breast Anatomy - National Breast Cancer Foundation. Available at: <http://www.nationalbreastcancer.org/breast-anatomy>. (Accessed: 25th May 2017)
8. Types of Breast Cancer | Breastcancer.org. Available at: <http://www.breastcancer.org/symptoms/types>. (Accessed: 25th May 2017)
9. Breast Cancer Organization. IDC — Invasive Ductal Carcinoma. Available at: <http://www.breastcancer.org/symptoms/types/idc>. (Accessed: 25th May 2017)
10. Cerrato, F. & Labow, B. I. Diagnosis and management of fibroadenomas in the adolescent breast. *Semin. Plast. Surg.* **27**, 23–5 (2013).
11. Dupont, W. D. *et al.* Long-Term Risk of Breast Cancer in Women with Fibroadenoma. *N. Engl. J. Med.* **331**, 10–15 (1994).
12. Kachewar, S. S. & Dongre, S. D. Role of triple test score in the evaluation of palpable breast lump. *Indian J. Med. Paediatr. Oncol.* **36**, 123–7 (2015).
13. Tavassoéli, F.A., Devilee, P. *Pathology and Genetics Tumours of the Breast and Female Genital Organs - WHO - OMS - No4* (2003).
14. Tokunaga, E. *et al.* Molecular mechanisms regulating the hormone sensitivity of breast cancer. *Cancer Sci.* **105**, 1377–1383 (2014).
15. Growth of Cancer - National Breast Cancer Foundation. Available at: <http://www.nationalbreastcancer.org/growth-of-breast-cancer>. (Accessed: 26th May 2017)
16. Tsutsui, S., Ohno, S., Murakami, S., Hachitanda, Y. & Oda, S. Prognostic value of c-erbB2 expression in breast cancer. *J. Surg. Oncol.* **79**, 216–223 (2002).
17. Paik, S. *et al.* HER2 and choice of adjuvant chemotherapy for invasive breast cancer: National Surgical Adjuvant Breast and Bowel Project Protocol B-15. *J. Natl. Cancer Inst.* **92**, 1991–8 (2000).
18. Cao, M. D. *et al.* Metabolic characterization of triple negative breast cancer. *BMC Cancer* **14**, 941 (2014).
19. Martelotto, L. G., Ng, C. K. Y., Piscuoglio, S., Weigelt, B. & Reis-Filho, J. S. Breast cancer intra-tumor heterogeneity. *Breast Cancer Res.* **16**, 210 (2014).
20. Turner, N. C. & Reis-Filho, J. S. Genetic heterogeneity and cancer drug resistance. *Lancet Oncol.* **13**, e178–e185 (2012).

21. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345 (2013).
22. Polyak, K. Heterogeneity in breast cancer. *J. Clin. Invest.* **121**, 3 (2011).
23. Norwegian Ministry of Health Care Services National Cancer. Together – against cancer Norwegian cancer strategy 2013-2017. (2017).
24. American Cancer Society. Targeted Therapy for Breast Cancer. Available at: <https://www.cancer.org/cancer/breast-cancer/treatment/targeted-therapy-for-breast-cancer.html>. (Accessed: 26th May 2017)
25. Breast Cancer Organization. Treatment & Side Effects. Available at: <http://www.breastcancer.org/treatment>. (Accessed: 26th May 2017)
26. Bjerrum, J. T. in *Methods in molecular biology (Clifton, N.J.)* **1277**, 1–14 (2015).
27. N.Pavlova, Natalia and B.Thompson, C. The emerging hallmarks of cancer metabolism. *Cell Metab.* **23**, 27–47 (2017).
28. Mendiola, A. V. & Cruz, I. S. Energetic metabolism and cancer. *Vertientes Rev. Espec. en Ciencias la Salud* **17**, 108–113 (2014).
29. Hirschhaeuser, F., Sattler, U. G. A. & Mueller-Klieser, W. Lactate: A metabolic key player in cancer. *Cancer Res.* **71**, 6921–6925 (2011).
30. Kristine Glunde, Zaver M. Bhujwalla, and S. M. R. Choline metabolism in malignant transformation. *Nat. Rev. Cancer* **23**, 1–7 (2008).
31. Vance, J. E. & Vance, D. E. Phospholipid biosynthesis in mammalian cells. *Biochem. Cell Biol.* **82**, 113–128 (2004).
32. Sitter, B. *et al.* Cervical cancer tissue characterized by high-resolution magic angle spinning MR spectroscopy. *MAGMA Magn. Reson. Mater. Physics, Biol. Med.* **16**, 174–181 (2004).
33. Swanson, M. G. *et al.* Proton HR-MAS spectroscopy and quantitative pathologic analysis of MRI/3D-MRSI-targeted postsurgical prostate tissues. *Magn. Reson. Med.* **50**, 944–954 (2003).
34. Tessem, M.-B. *et al.* Discrimination of Patients with Microsatellite Instability Colon Cancer using ¹H HR MAS MR Spectroscopy and Chemometric Analysis. *J. Proteome Res.* **9**, 3664–3670 (2010).
35. Beckonert, O., Monnerjahn, J., Bonk, U. & Leibfritz, D. Visualizing metabolic changes in breast-cancer tissue using ¹H-NMR spectroscopy and self-organizing maps. *NMR Biomed.* **16**, 1–11 (2003).
36. El Agouza, I. M., Eissa, S. S., El Houseini, M. M., El-Nashar, D. E. & Abd El Hameed, O. M. Taurine: a novel tumor marker for enhanced detection of breast cancer among female patients. *Angiogenesis* **14**, 321–330 (2011).
37. Zhang, X. *et al.* Taurine induces the apoptosis of breast cancer cells by regulating apoptosis-related proteins of mitochondria. *Int. J. Mol. Med.* **35**, 218–226 (2015).
38. Gowda, G. A. N. *et al.* Metabolomics-based methods for early disease diagnostics. *Expert Rev. Mol. Diagn.* **8**, 617–33 (2008).
39. S., M., B., S., T.F., B., M.-B., T. & I.S., G. HR MAS MR spectroscopy in metabolic characterization of cancer. *Curr. Top. Med. Chem.* **11**, 2–26 (2011).
40. Giskeødegård, G. F., Cao, M. D. & Bathen, T. F. in *Methods in molecular biology (Clifton, N.J.)* **1277**, 37–50 (2015).
41. Euceda, L. R., Giskeødegård, G. F. & Bathen, T. F. Preprocessing of NMR metabolomics data. *Scand. J. Clin. Lab. Invest.* **75**, 193–203 (2015).

42. Savorani, F., Tomasi, G. & Engelsen, S. B. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *J. Magn. Reson.* **202**, 190–202 (2010).
43. Giskeødegård, G. F. Multivariate data analyses tools applied in clinical MRS. 1–13
44. Euceda, L. R., Haukaas, T. H., Bathen, T. F. & Giskeødegård, G. F. Prediction of clinical endpoints in breast cancer using NMR metabolic profiles. 1–24

Appendix

1. HR MAS protocol

Version October 2016

Generated by Tonje Haukaas and Maria DC

NTNU, ISB, MR Cancer Group

1. Insert sample

First: make sure there is no sample inside the magnet:

<sx ej ↵

>sx (position of the sample)↵

2. Turn on the spin in MAS display

Check the spin rate (5000 Hz)

Click GO

3. Set the temperature in Temperature control suite

Check temperature and wait until it is green

Click ON

4. Tune & match:

Open a **zgpr** experiment. Wait for the spinning speed and temperature to be stabilized before acquiring one spectra

>zg

Adjust tune and match:

>woob ↵

Adjust with yellow screws under the magnet until woob-signal is sharp and centered.

>stop ↵

5. Lock:

>lock ↵

Choose: D2O_TISSUE_new

6. Shim:

I. Acquire a zpg- spectrum and perform automatic phase:

>zgpf ↵

>apk ↵

>absn ↵

II. Evaluate the shape and half width of a selected metabolite peak e.g. formate or lactate

To check the half width using build-in commands:

Zoom in on the selected peak

>pps ↵

>hwcal ↵

To check the half width manually:

Zoom in on the selected peak.

Click on the left side of the peak at half height and drag the pointer to the right side of the peak to measure the distance.

III. Perform shim in GS mode if the half width is larger than 1.5 Hz or the shape of the peak is not optimal:

>gs ↵

ADJUST E.G Y, YZ, YZ² and Z⁴ using BSMS Control Suite

>stop ↵

>zgfp ↵

>apk ↵

IV. Control the shape and half width. Repeat step III if necessary.

7. Determine P1 (90° pulse):

I. Open a zg experiment and enter P1 from previous experiment (should be $\approx 10 \mu\text{sec}$):

>P1 ↵

II. Acquire and phase the spectrum to make sure that the signals are positive at 90° pulse:

>zgfp ↵

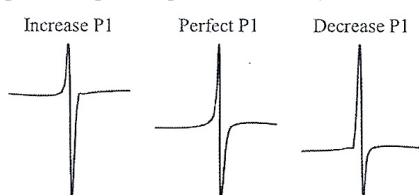
>apk ↵

III. To find the optimal P1, change the value to P1 x 4 instead of P1:

>P1 (enter the calculated value)

>zgfp ↵

IV. Try out different P1 x 4 values to find the optimal P1 where the signal is equal to zero at 360° pulse. Repeat step III if necessary.



V. Note down P1 (90° pulse), P2 = P1x2 (180° pulse), and P4 = P1x4 (360° pulse).

8. Determine O1:

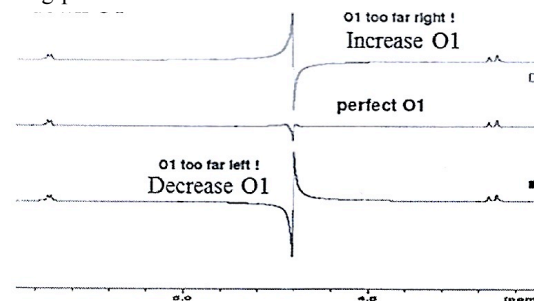
I. Open a zpg experiment,

II. Set in the P1 value from the previous step and acquire a spectrum:

III. Find the optimal O1 value for good suppression of the water signal:

>O1 (try different values for O1)

>zgfp ↵



9. Acquire standard spectrum

I. Chose and experiment folder you want to copy and make sure you have the correct values for NS, DS etc.

II. Open each experiment you want to copy from the folder and change the name:

>edc ↵

Type new name, OK

III. Open the first experiment in your new folder and define P1 and O1:

>O1 (enter O1 value from step 8)

>getprosol 1H P1 -6.60 (enter P1 value from step 7)

Open the next experiment and type:

>O1 (enter O1 value from step 8)

>getprosol 1H P1 -6.60 (enter P1 value from step 7)

Repeat this step for all experiments in the folder

IV. To run all experiments in the same folder, open the first experiment and write:

>multizg ↵ (check if you have the correct total acquisition time)

Fill the number of experiments (3)

V. To stop the experiment during acquisition:

> kill ↵

Click on your experiment and KILL

Perform kill again if the acquisition is still running

10. Eject sample

Stop spin in MAS display by clicking HALT

Wait until the spin is below 100 Hz

> sx ej ↵

11. Post processing

I. Post processing commands for 1D experiment:

II. Post processing commands for 2D experiment (e.g. JRES):

III. If you want to post process several spectra simultaneously:

>qmulti ↵

Select experiments for processing in the window.

Insert the commands you want to perform separated by semi colon

e.g. epf;apk;abs;xfb

12. Clean the workstation and all equipment after use

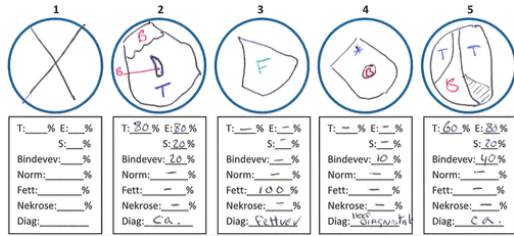
I. Tweezers, scapel: Ethanol baths

II. Rotors, caps: In acetone baths

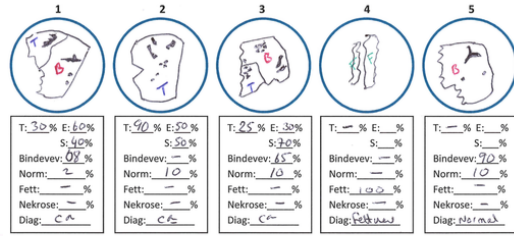
2. Histology results in Norwegian

T Tumor tissue, which is either E epithelial or S stromal tissue. Bidev means connective tissue, Norm Normal cells, Fett Fat, Nekrose Necrosis and Diag Diagnosis.

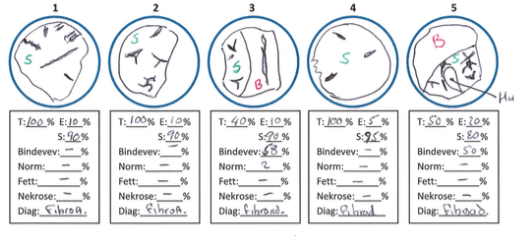
Skive 1



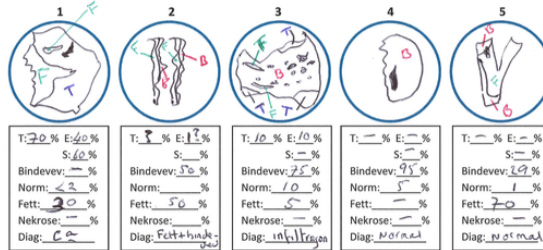
Skive 2



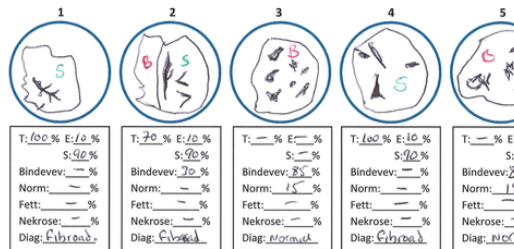
Skive 3



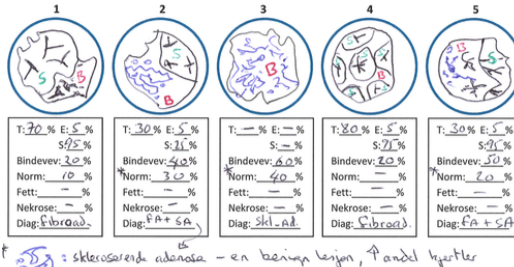
Skive 4



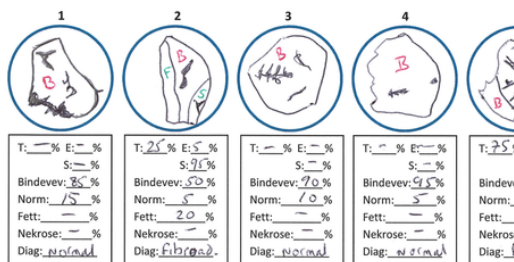
Skive 5



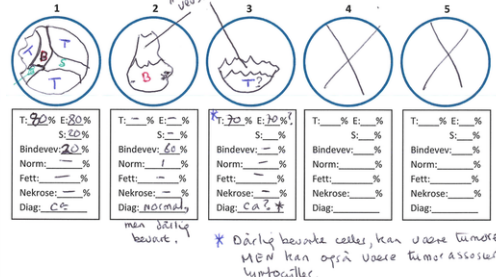
Skive 6




Skive 7






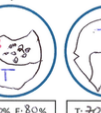

Skive 8



Skive 9

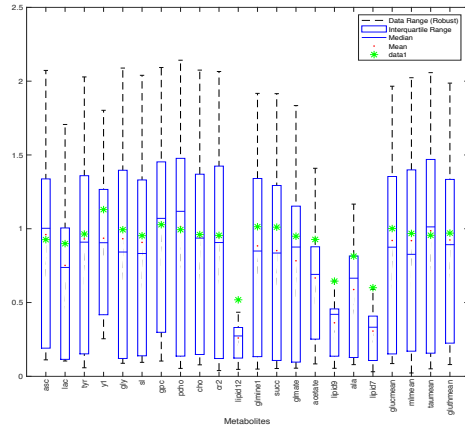
 <p>T: — % E: — % S: — % Bindevev: 75 % Norm: 5 % Fett: — % Nekrose: — % Diag: Normal</p>	 <p>T: — % E: — % S: — % Bindevev: 10 % Norm: — % Fett: 90 % Nekrose: — % Diag: Fettver</p>	 <p>T: — % E: — % S: — % Bindevev: 90 % Norm: 5 % Fett: — % Nekrose: — % Diag: Normal</p>	 <p>T: — % E: — % S: — % Bindevev: 20 % Norm: — % Fett: 20 % Nekrose: — % Diag: Normal</p>	 <p>T: — % E: — % S: — % Bindevev: 100 % Norm: — % Fett: — % Nekrose: — % Diag: Normal</p>
--	--	--	---	---

Skive 10

 <p>T: 100 % E: 50 % S: 20 % Bindevev: — % Norm: — % Fett: — % Nekrose: — % Diag: C 6</p>	 <p>T: — % E: — % S: — % Bindevev: 30 % Norm: — % Fett: 70 % Nekrose: — % Diag: Fett + Bindev</p> <p><i>bindev. beschr. vel.</i></p>	 <p>T: — % E: — % S: — % Bindevev: 25 % Norm: — % Fett: 75 % Nekrose: — % Diag: Fett + Bindev</p>	 <p>T: 60 % E: 80 % S: 20 % Bindevev: 35 % Norm: 5 % Fett: — % Nekrose: — % Diag: C 6</p> <p><i>lobul. (az. r. l.?)</i></p>	 <p>T: 70 % E: 80 % S: 10 % Bindevev: 15 % Norm: — % Fett: 5 % Nekrose: — % Diag: C 6</p>
--	--	--	--	--

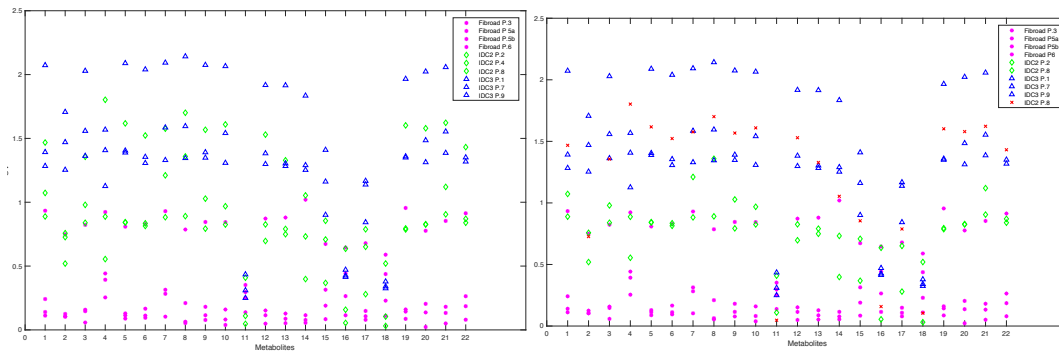
2. Statistical analysis & figures

➤ Dataset 1

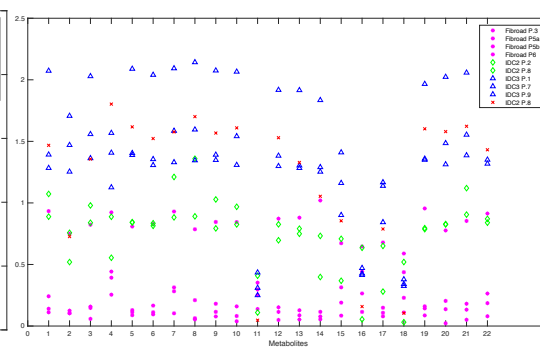


Dataset 1. CoV of all metabolites from normalized spectra, including the lipid peaks, presented in boxplots. The central red mark corresponds to the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme datapoints the algorithm considers to be not outliers, and the outliers are plotted individually.

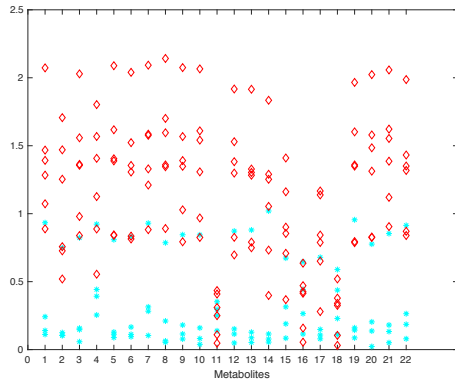
A)



B)

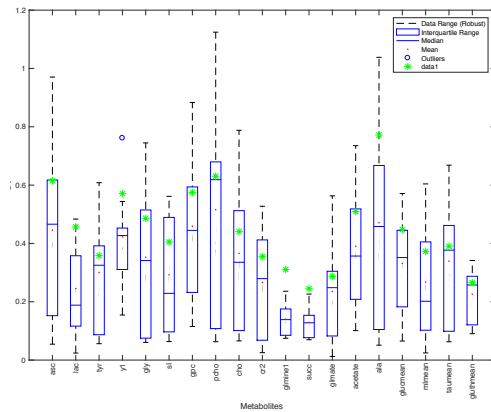


Dataset 1. CoV of all metabolites from normalized spectra, including the lipid peaks, colored according to clinical diagnoses: A) Blue triangles correspond to IDC 3 patients, green diamonds to IDC2 patients and magenta stars to fibroadenoma patients. B) Red crosses correspond to patient 8 with IDC2. X axis numbers correspond to: 1 *asc*, 2 *lac*, 3 *tyr*, 4 *yl*, 5 *gly*, 6 *sl*, 7 *gpc*, 8 *pcho*, 9 *cho*, 10 *cr2*, 11 *lipid12*, 12 *glmine1*, 13 *succ*, 14 *glmate*, 15 *acetate*, 16 *lipid9*, 17 *ala*, 18 *lipid7*, 19 *glucmean*, 20 *mImean*, 21 *taumean*, 22 *gluthmean*.

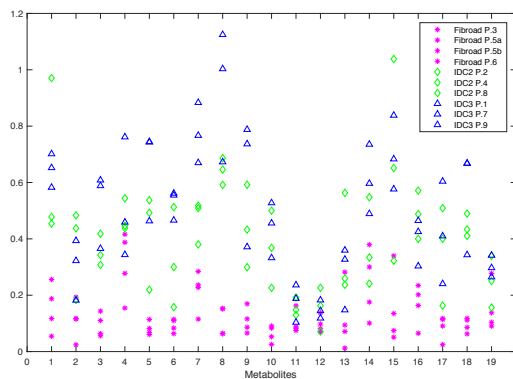


Dataset 1. CoV of all metabolites from normalized spectra, including the lipid peaks, colored according to clinical diagnoses: Red diamonds correspond to IDC patients and cyan stars to fibroadenoma patients. X axis numbers correspond to: 1 *asc*, 2 *lac*, 3 *tyr*, 4 *yl*, 5 *gly*, 6 *sI*, 7 *gpc*, 8 *pcho*, 9 *cho*, 10 *cr2*, 11 *lipid* 12, 12 *glmine1*, 13 *succ*, 14 *glmate*, 15 *acetate*, 16 *lipid9*, 17 *ala*, 18 *lipid7*, 19 *glucmean*, 20 *mImean*, 21 *taumean*, 22 *gluthmean*.

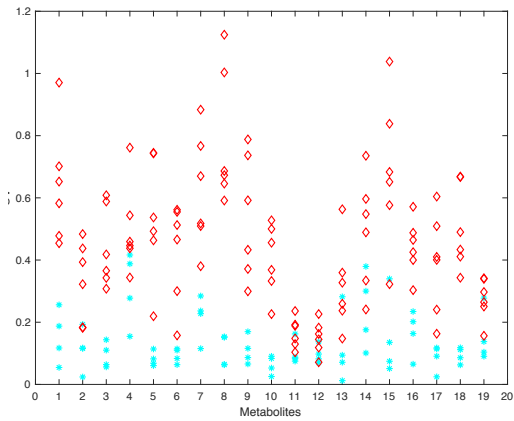
➤ Dataset 2



Dataset 2. CoV of all metabolites from normalized spectra, not including the lipid peaks, presented in boxplots. Colored according to clinical diagnose: The central red mark corresponds to the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme datapoints the algorithm considers to be not outliers, and the outliers are plotted individually.

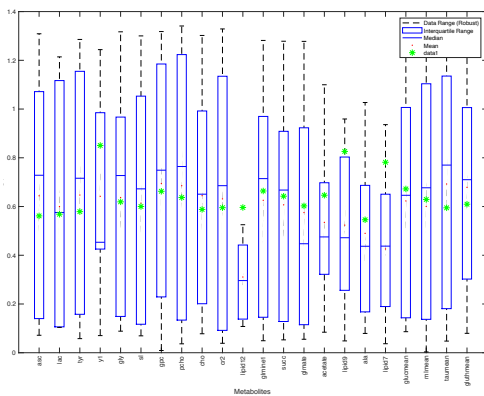


Dataset 2. CoV of all metabolites from normalized spectra, not including the lipid peaks, colored according to clinical diagnose: Blue triangles correspond to IDC 3 patients, green diamonds to IDC2 patients and magenta stars to fibroadenoma patients. X axis numbers correspond to: 1 *asc*, 2 *lac*, 3 *tyr*, 4 *yl*, 5 *gly*, 6 *sI*, 7 *gpc*, 8 *pcho*, 9 *cho*, 10 *cr2*, 11 *lipid* 12, 12 *glmine1*, 13 *succ*, 14 *glmate*, 15 *acetate*, 16 *lipid9*, 17 *ala*, 18 *lipid7*, 19 *glucmean*, 20 *mImean*, 21 *taumean*, 22 *gluthmean*.

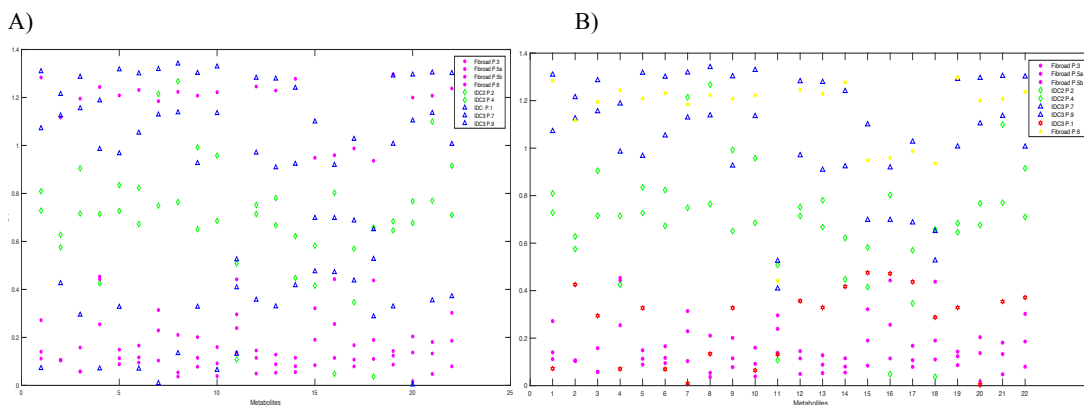


Dataset 2. CoV of all metabolites from normalized spectra, not including the lipid peaks, colored according to clinical diagnoses: Red diamonds correspond to IDC patients and cyan stars to fibroadenoma patients. X axis numbers correspond to: 1 *asc*, 2 *lac*, 3 *tyr*, 4 *yl*, 5 *gly*, 6 *sl*, 7 *gpc*, 8 *pcho*, 9 *cho*, 10 *cr2*, 11 *lipid12*, 12 *gmline1*, 13 *succ*, 14 *glmate*, 15 *acetate*, 16 *lipid9*, 17 *ala*, 18 *lipid7*, 19 *glucmean*, 20 *mImean*, 21 *taumean*, 22 *gluthmean*.

➤ Dataset 3

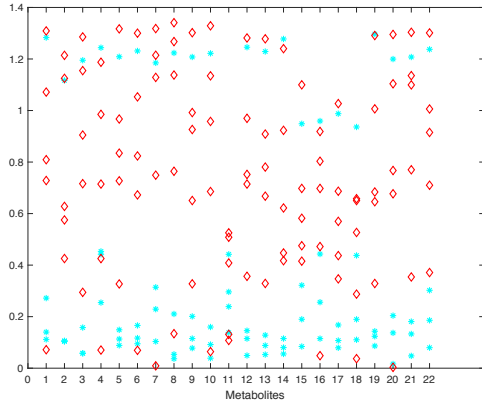


Dataset 3. CoV of all metabolite from normalized spectra, including the lipid peaks, presented in boxplots. Samples without tumor content have been removed from the data. The central red mark corresponds to the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme datapoints the algorithm considers to be not outliers, and the outliers are plotted individually.



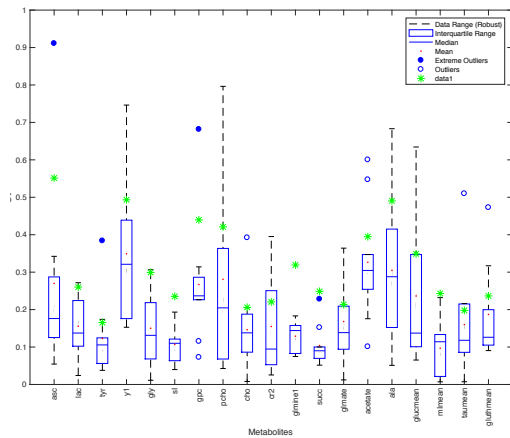
Dataset 3. CoV of all metabolite from normalized spectra, including the lipid peaks, colored according to clinical diagnose. Samples without tumor content have been removed from the data. A) Blue triangles correspond to IDC 3 patients, green diamonds to IDC2 patients and magenta stars to fibroadenoma patients. B) Blue triangles correspond

to IDC 3 patients, green diamonds to IDC2 patients, magenta stars to fibroadenoma patients, red hexagram to IDC3 patient 1 and yellow stars to Fibroadenoma patient 6. X axis numbers correspond to: 1 *asc*, 2 *lac*, 3 *tyr*, 4 *yl*, 5 *gly*, 6 *sI*, 7 *gpc*, 8 *pcho*, 9 *cho*, 10 *cr2*, 11 *lipid* 12, 12 *gmine1*, 13 *succ*, 14 *gmate*, 15 *acetate*, 16 *lipid9*, 17 *ala*, 18 *lipid7*, 19 *glucmean*, 20 *mI*mean, 21 *taumean*, 22 *gluthmean*.

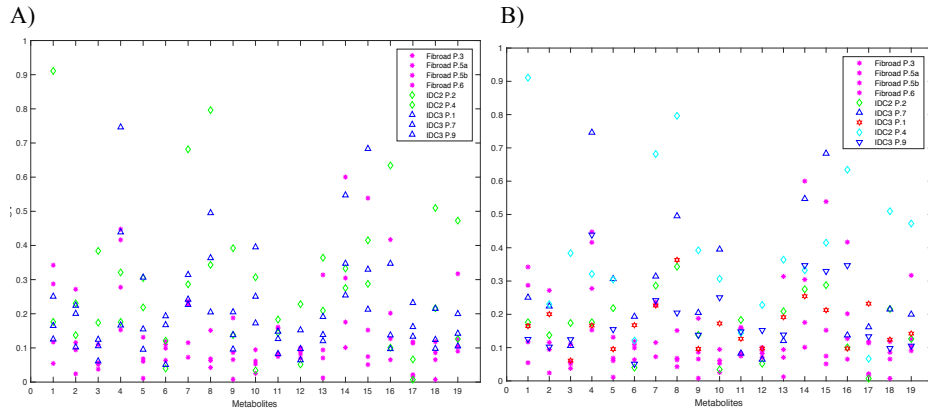


Dataset 3. CoV of all metabolite from normalized spectra, including the lipid peaks, colored according to clinical diagnose. Samples without tumor content have been removed from the data. Red diamonds correspond to IDC3 patients and cyan stars to fibroadenoma patients. X axis numbers correspond to: 1 *asc*, 2 *lac*, 3 *tyr*, 4 *yl*, 5 *gly*, 6 *sI*, 7 *gpc*, 8 *pcho*, 9 *cho*, 10 *cr2*, 11 *lipid* 12, 12 *gmine1*, 13 *succ*, 14 *gmate*, 15 *acetate*, 16 *lipid9*, 17 *ala*, 18 *lipid7*, 19 *glucmean*, 20 *mI*mean, 21 *taumean*, 22 *gluthmean*.

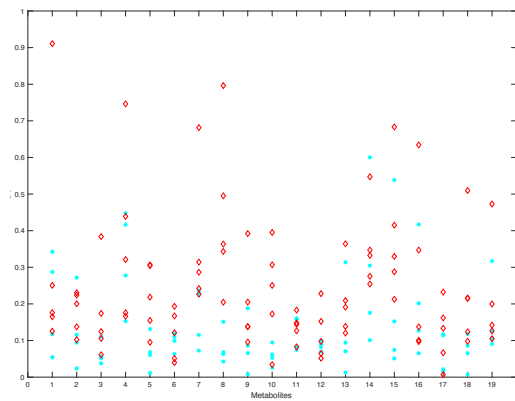
➤ Dataset 4



Dataset 4. CoV of all metabolite from normalized spectra, with lipid peaks removed, presented in boxplots. Samples without tumor content have been removed from the data. The central red mark corresponds to the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme datapoints the algorithm considers to be not outliers, and the outliers are plotted individually.



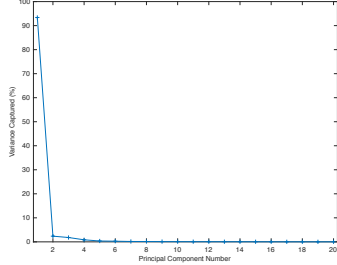
Dataset 4. CoV of all metabolite from normalized spectra, with lipid peaks removed, colored according to clinical diagnose. Samples without tumor content have been removed from the data. Blue triangles correspond to IDC 3 patients, green diamonds to IDC2 patients and magenta stars to fibroadenoma patients. B) Magenta stars correspond to fibroadenoma patients, green diamonds to IDC2 Patient 2, cyan diamonds to IDC2 Patient 4, red hexagram to IDC3 Patient 1, blue triangle to IDC3 Patient 7 and inverted blue triangle to IDC3 Patient 9. X axis numbers correspond to: 1 *asc*, 2 *lac*, 3 *tyr*, 4 *yl*, 5 *gly*, 6 *sI*, 7 *gpc*, 8 *pcho*, 9 *cho*, 10 *cr2*, 11 *lipid 12*, 12 *glmine1*, 13 *succ*, 14 *glmate*, 15 *acetate*, 16 *lipid9*, 17 *ala*, 18 *lipid7*, 19 *glucmean*, 20 *mImean*, 21 *taumeana*, 22 *gluthmean*.



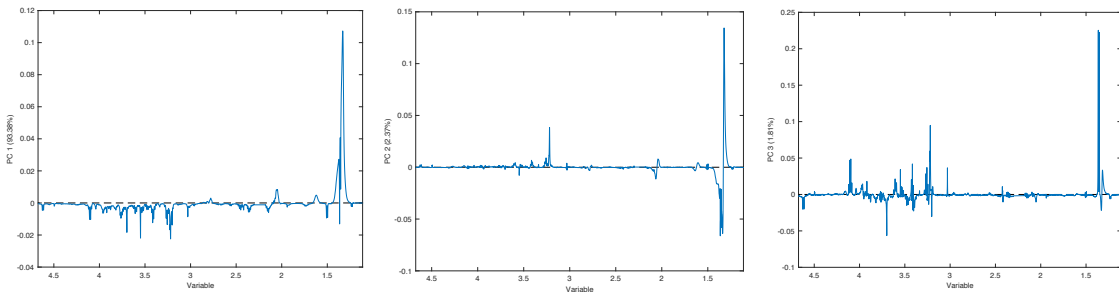
Dataset 4. CoV of all metabolite from normalized spectra, with lipid peaks removed, colored according to clinical diagnose. Samples without tumor content have been removed from the data. Red diamonds correspond to IDC patients and cyan stars to fibroadenoma patients. X axis numbers correspond to: 1 *asc*, 2 *lac*, 3 *tyr*, 4 *yl*, 5 *gly*, 6 *sI*, 7 *gpc*, 8 *pcho*, 9 *cho*, 10 *cr2*, 11 *lipid 12*, 12 *glmine1*, 13 *succ*, 14 *glmate*, 15 *acetate*, 16 *lipid9*, 17 *ala*, 18 *lipid7*, 19 *glucmean*, 20 *mImean*, 21 *taumeana*, 22 *gluthmean*.

4. Multivariate analysis figure

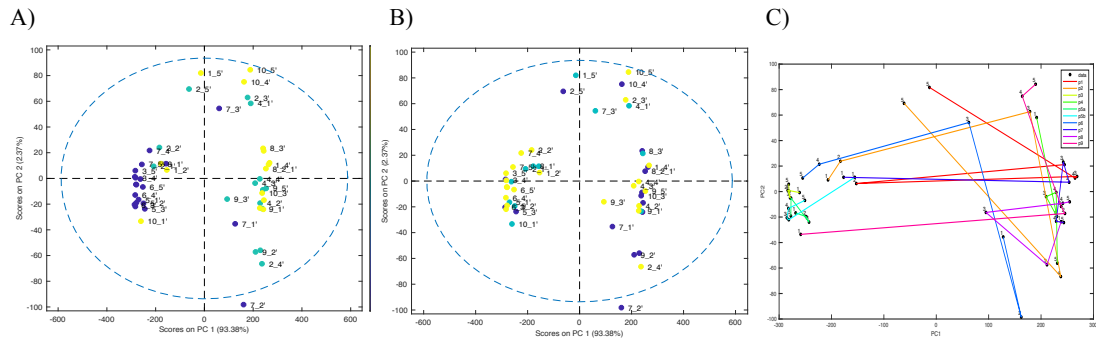
➤ Dataset 1



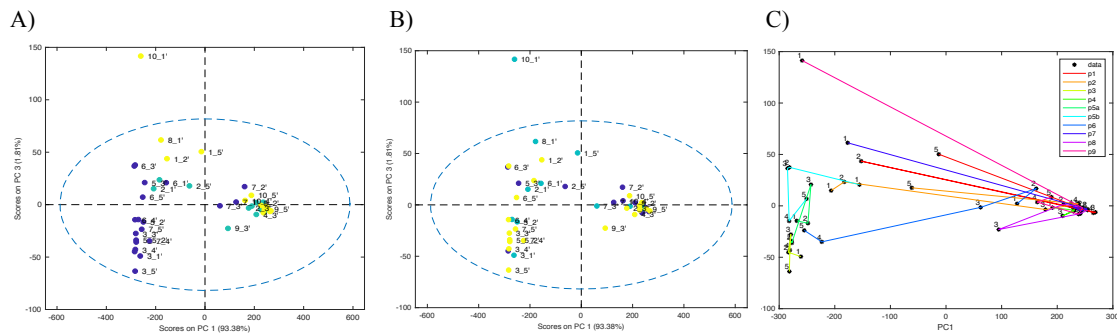
Dataset 1. Variance captured vs Principal component number plot. PC1, PC2 and PC3 were chosen for PCA multivariate analysis.



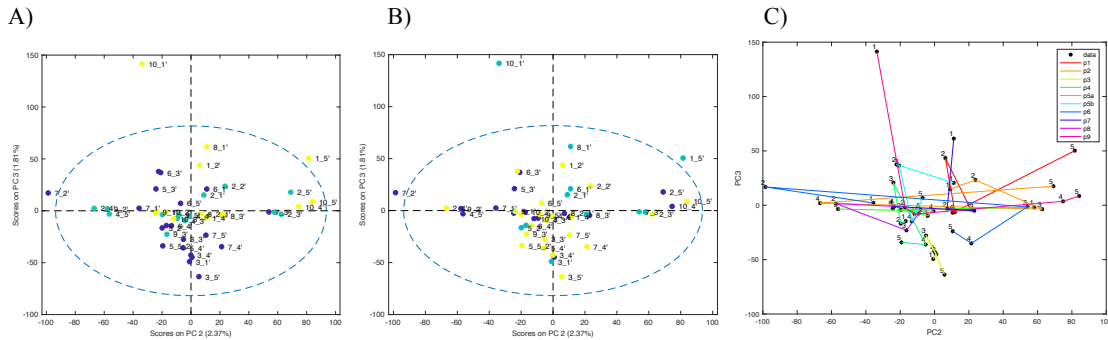
Dataset 1. Loadings including lipid peaks: Principal component 1 explains 90.30% of variation, principal component 2 explains 3.61% of variation and principal component 3 explains 2.66% of variation.



Dataset 1. PC1 vs. PC2 scores plot including lipid peaks. A) Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3). B) Colored by location within the tumor (Green: core, yellow: margins and blue: other regions). C) Trajectory plot showing all of each patient's spots united.

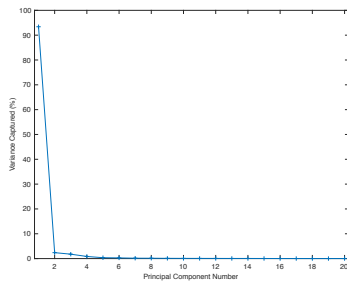


Dataset 1. PC1 vs. PC3 scores plot including lipid peaks. A) Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3). B) Colored by location within the tumor (green: core, yellow: margins and blue: other regions). C) Trajectory plot showing all of each patient's spots united.

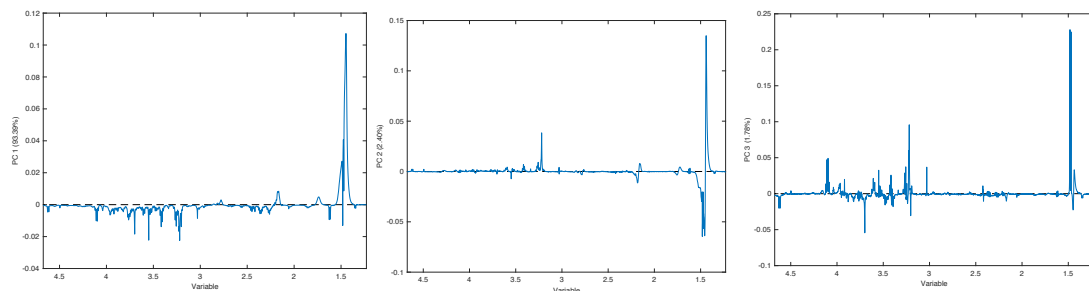


Dataset 1. PC2 vs. PC3 scores plot including lipid peaks. A) Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3). B) Colored by location within the tumor (green: core, yellow: margins and blue: other regions). C) Trajectory plot showing all of each patient's spots united.

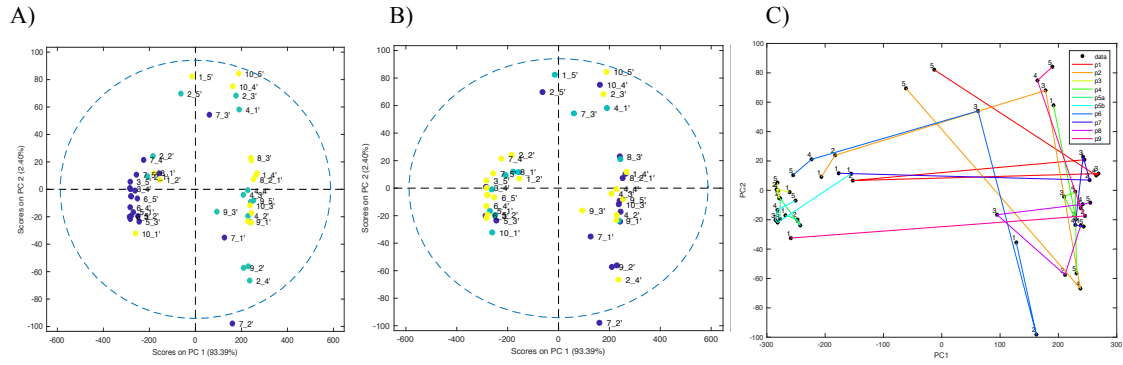
➤ Dataset 2



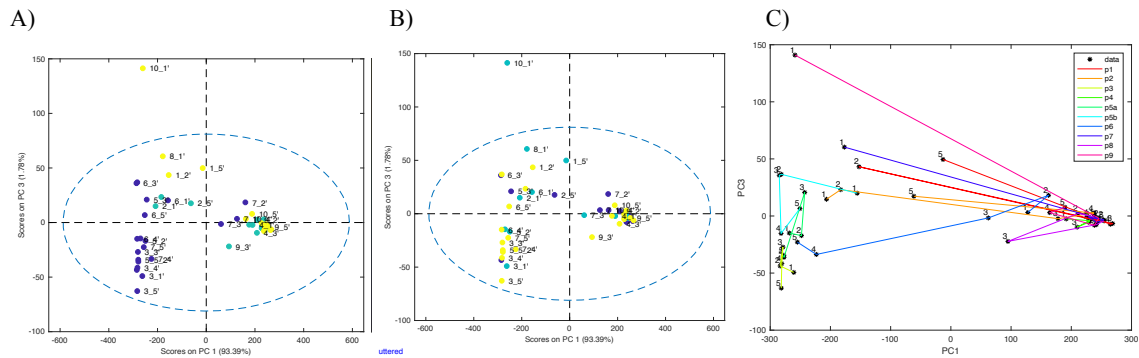
Dataset 2. Variance captured vs Principal component number plot. PC1, PC2 and PC3 were chosen for PCA multivariate analysis.



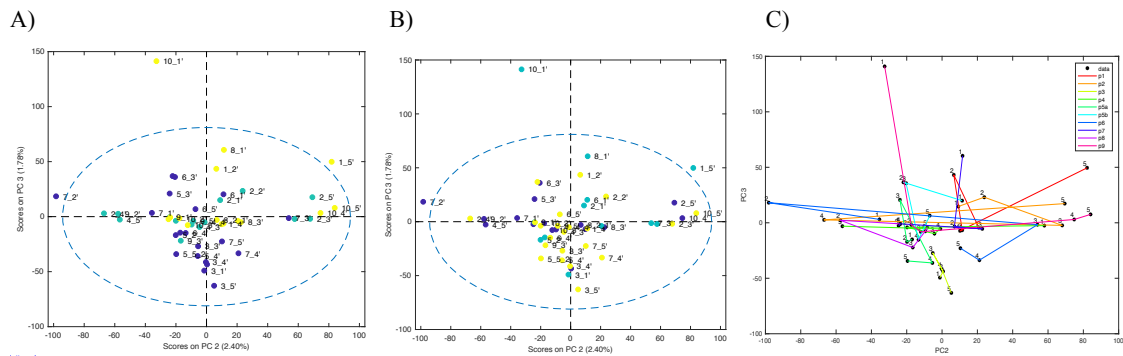
Dataset 2. Loadings with lipid peaks removal: Principal component 1 explains 93.39% of variation, principal component 2 explains 2.40% of variation and principal component 3 explains 1.78% of variation.



Dataset 2. PC1 vs. PC2 scores plot with lipid peaks removal. A) Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3). B) Colored by location within the tumor (green: core, yellow: margins and blue: other regions). C) Trajectory plot showing all of each patient's spots united.

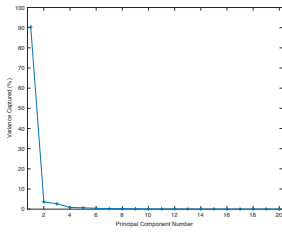


Dataset 2. PC1 vs. PC3 scores plot with lipid peaks removal. A) Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3). B) Colored by location within the tumor (green: core, yellow: margins and blue: other regions). C) Trajectory plot showing all of each patient's spots united.

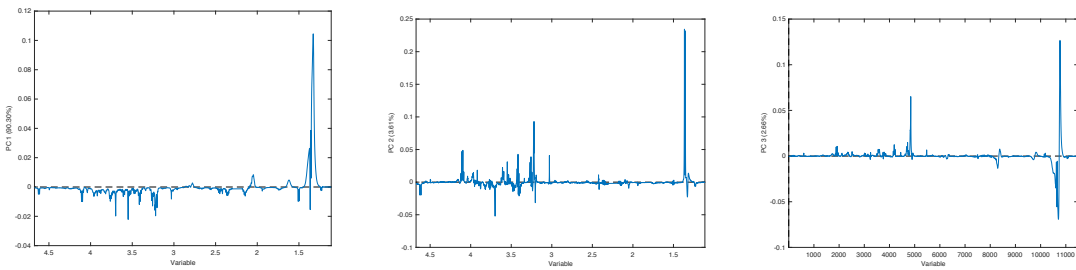


Dataset 2. PC2 vs. PC3 scores plot with lipid peak removal. A) Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3). B) Colored by location within the tumor (green: core, yellow: margins and blue: other regions). C) Trajectory plot showing all of each patient's spots united.

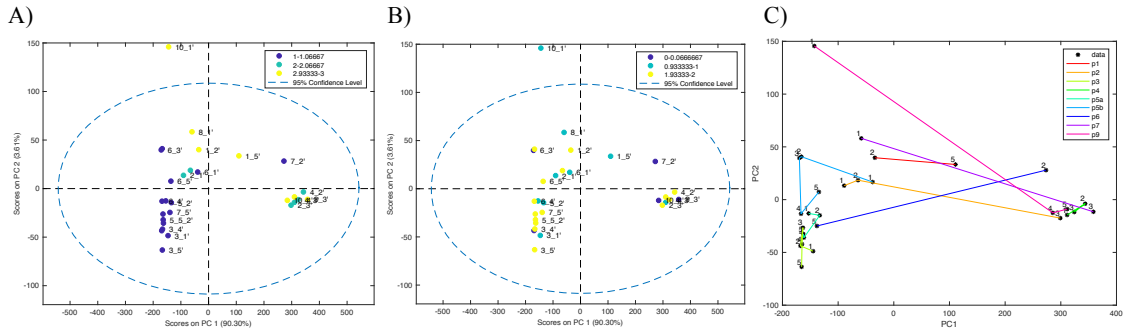
➤ Dataset 3



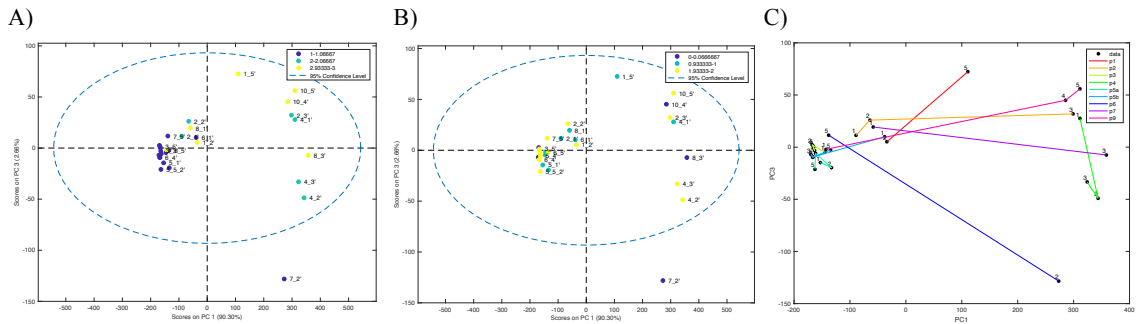
Dataset 3. Variance captured vs Principal component number plot. PC1, PC2 and PC3 were chosen for PCA multivariate analysis.



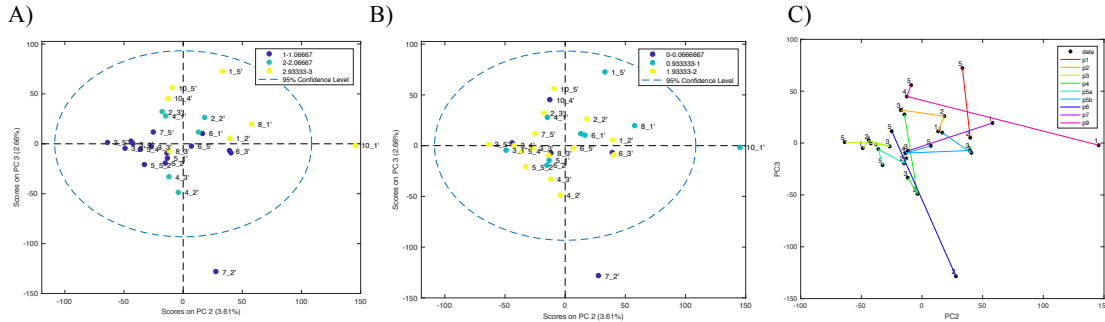
Dataset 3. Loadings including lipid peaks: Principal component 1 explains 90.30% of variation, principal component 2 explains 3.61% of variation and principal component 3 explains 2.66% of variation. Samples without tumor content have been removed from the data.



Dataset 3. PC1 vs. PC2 scores plot including lipid peaks. A) Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3). B) Colored by location within the tumor (green: core, yellow: margins and blue: other regions). C) Trajectory plot showing all of each patient's spots united. Samples without tumor content have been removed from the data.

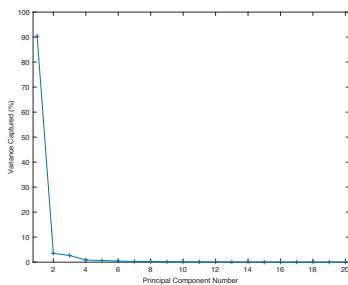


Dataset 3. PC1 vs. PC3 scores plot including lipid peaks. A) Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3). B) Colored by location within the tumor (green: core, yellow: margins and blue: other regions). C) Trajectory plot showing all of each patient's spots united. Samples without tumor content have been removed from the data.

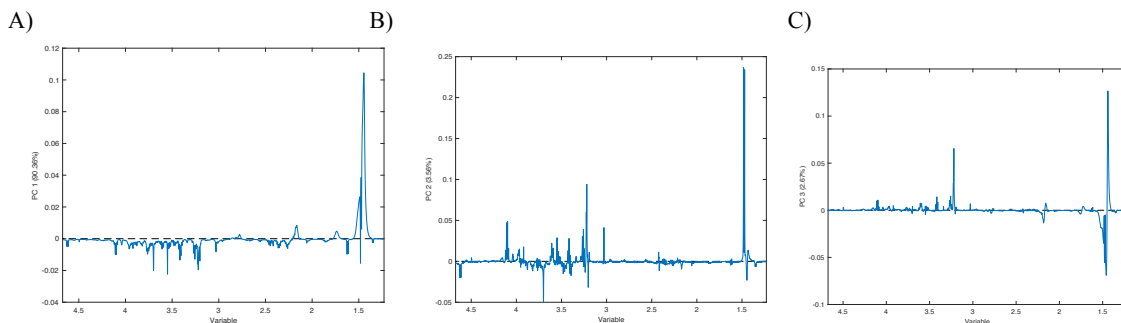


Dataset 3. PC2 vs. PC3 scores plot including lipid peaks. A) Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3). B) Colored by location within the tumor (green: core, yellow: margins and blue: other regions). C) Trajectory plot showing all of each patient's spots united. C) Trajectory plot showing all of each patient's spots united. Samples without tumor content have been removed from the data.

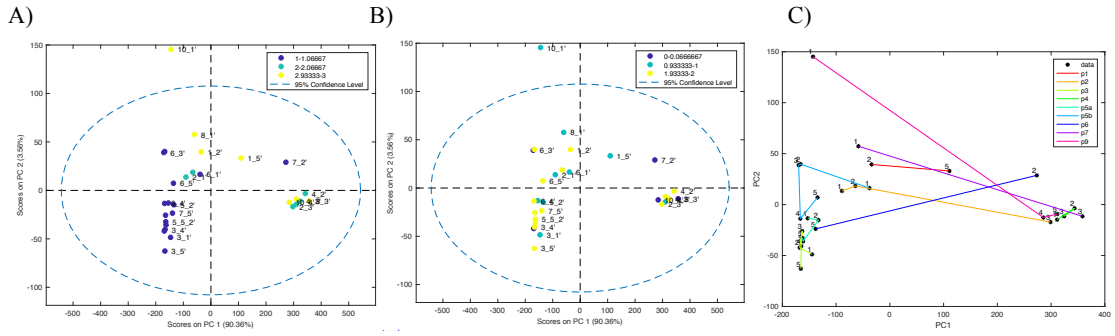
➤ **Dataset 4**



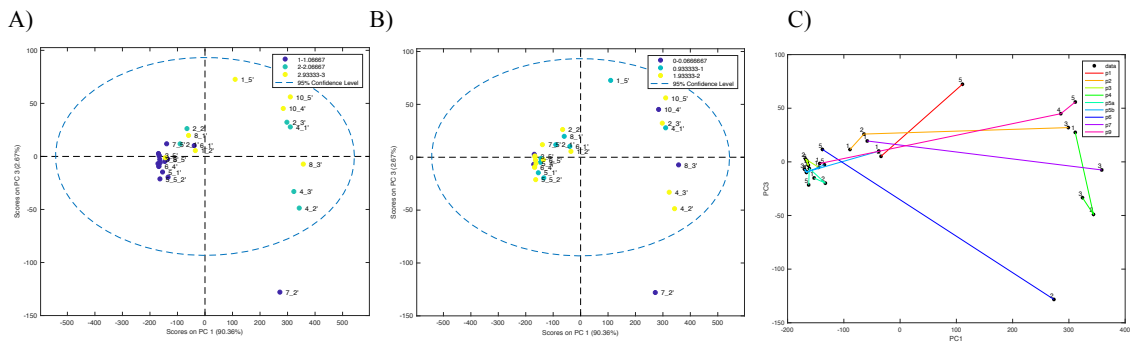
Dataset 4. Variance captured vs Principal component number plot. PC1, PC2 and PC3 were chosen for PCA multivariate analysis.



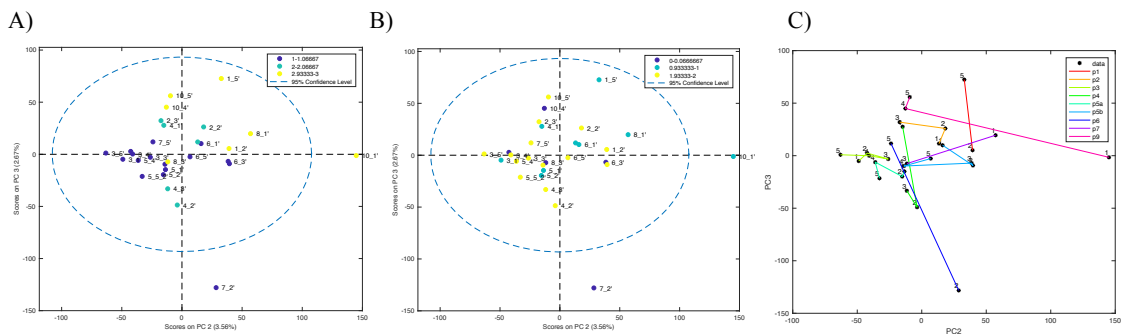
Dataset 4. Loadings with lipid peaks removal: Principal component 1 explains 90.36% of variation, principal component 2 explains 3.56% of variation and principal component 3 explains 2.67% of variation. Samples without tumor content have been removed from the data.



Dataset 4. PC1 vs. PC2 scores plot with lipid peaks removal. A) Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3). B) Colored by location within the tumor (green: core, yellow: margins and blue: other regions). C) Trajectory plot showing all of each patient's spots united. Samples without tumor content have been removed from the data.



Dataset 4. PC1 vs. PC3 scores plot, with lipid peaks removal. A) Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3). B) Colored by location within the tumor (green: core, yellow: margins and blue: other regions). C) Trajectory plot showing all of each patient's spots united. Samples without tumor content have been removed from the data.



Dataset 4. PC2 vs. PC3 scores plot with lipid peaks removal. A) Colored by cancer type (blue: Fibroadenoma, green: IDC2 and yellow: IDC3). B) Colored by location within the tumor (green: core, yellow: margins and blue: other regions). C) Trajectory plot showing all of each patient's spots united. Samples without tumor content have been removed from the data.