ZHAO, XUEQING, M.S. A Characteristic Metabolic Signature of Breast Cancer. (2012) Directed by Dr. Wei Jia. 80 pp.

Among women in the U.S., breast cancer is the most commonly diagnosed cancer and the second leading cause of cancer death (after lung cancer). Metabolomics, an approach to the study of small molecules, provides insight of characteristic biochemical phenotypes in disease, and facilitates the development of novel diagnostic tools. This thesis project was to investigate the metabolic signature and to identify potential biomarkers for breast cancer using metabolomics methods. GC-TOFMS and LC-TOFMS spectra were acquired in the plasma collected from 138 breast cancer patients and 76 healthy women. Multivariate and univariate statistics methods were applied to analyze the metabolic alterations in breast cancer. Of the 41 identified differential metabolites, aspartate was the most significantly reduced in breast cancer plasma samples and obtained good predictive power for distinguishing breast cancer patients from healthy controls. An established combination of 7 markers (asparagine, hypotaurine, 5-oxoproline, cysteine, aspartate, glutamate and glutamine) was found to provide even better predictive power than aspartate alone. The altered expression of aspartate was confirmed in an independent set of serum samples and 20 pairs of breast tumor tissue and its adjacent normal tissue. It was also found that the metabolic profiles of stage I and stage IV patients can be separated in the constructed OPLS-DA model. In conclusion, breast cancer exhibits profound metabolic dysregulations and potential biomarkers in breast cancer can be identified using metabolomics approach.

A CHARACTERISTIC METABOLIC SIGNATURE OF BREAST CANCER

by

Xueqing Zhao

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2012

Approved by

_____

Committee Chair

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Among women in the U.S., breast cancer is the most commonly diagnosed cancer

and the second leading cause of cancer death (after lung cancer). Metabolomics, an

approach to the study of small molecules, provides insight of characteristic biochemical

phenotypes in disease, and facilitates the development of novel diagnostic tools. The

purpose of this thesis project is to investigate the metabolic signature and to identify

potential biomarkers for breast cancer using metabolomics methods. The specific aims

for this project, described in Figure 1, include:

*Aim 1. To investigate the metabolic profiles and identify potential biomarkers in breast

cancer.* Working hypothesis of this aim is that breast cancer patients have a characteristic

signature of plasma metabolic profiles, and identifying metabolic shifts helps discover

potential biomarkers. Approach of this aim includes collecting plasma samples from 138

breast cancer patients and 76 healthy controls, dividing them into an age-matched

training set and a test set, constructing the OPLS-DA model using the training set to

identify differential metabolites, verifying the model in the test set, and evaluating the

diagnostic abilities of the potential biomarkers using ROC analysis.

*Aim 2. To validate the alterations of potential biomarkers in an independent sample set

of serum samples.* Working hypothesis of this aim is that the potential biomarkers

identified in the Aim 1 offer predictive power to distinguish breast cancer patients from

healthy controls in another independent sample set. <u>Approach</u> of this aim includes collecting serum samples from 80 breast cancer patients and 70 healthy controls, quantifying the serum levels of the potential biomarkers identified in the Aim 1 and evaluating their diagnostic abilities.

Aim 3. *To validate the alterations of potential biomarkers in breast tumor tissues.* <u>Working hypothesis</u> of this aim is that the potential biomarkers discovered in the blood are also expressed differently in tumor tissues. <u>Approach</u> of this aim includes measuring the tissue levels of the potential biomarkers in twenty pairs of breast tumor tissue and its adjacent normal tissue.

| **Discovery of differential metabolite in plasma** | 138 breast cancer patients and 76 healthy volunteers from California |
| | **Training set** (used to construct the statistical model): 35 patients and 35 age-matched healthy controls \| **Test set** (used to verify the statistical model): 103 patients and 41 healthy controls |

Validation in serum — 80 breast cancer patients and 70 healthy volunteers from Shanghai

Validation in tissue — 20 pairs of breast tumor tissue and its adjacent normal tissue

Figure 1. Study design.

CHAPTER II

REVIEW OF THE LITERATURE

Breast cancer and its diagnosis

Breast cancer is a global health concern, as it is one of the top ten worldwide causes of mortality in women according to statistics by the World Health Organization (World Health Organization, 2012). Breast cancer originates mostly from the lobules which are the glands for milk production, or from the ducts connecting the lobules to the nipple. Breast cancer confined within the lobules or the ducts is called in situ cancer, while breast cancer invading the surrounding tissue of the breast is called invasive breast cancer.

According to global cancer statistics in 2011, breast cancer accounted for 23% (1.38 million) of total new cancer cases and 14% (458,400) of cancer deaths in women in that year, making it the most commonly diagnosed cancer and the leading cause of cancer death among females in the world (Jemal et al., 2011). Among women in the U.S., breast cancer is also the most frequently diagnosed cancer and the second leading cause of cancer death (after lung cancer) (U.S. Cancer Statistics Working Group, 2012). Although breast cancer rates decrease every year because of improvement in early detection and treatment, the American Cancer Society still estimates that in the year 2012, 226,870 new cases of invasive breast cancer and 63,300 new cases of in situ breast cancer will be diagnosed among women, and that around 39,920 U.S. women will die from breast

cancer (American Cancer Society, 2012).

One of the most commonly used classifications of breast cancer is the TNM system, which is based on tumor size (T), lymph node involvement (N) and whether the cancer has spread, or metastasized, to distant organs (M). The stage of a particular case of breast cancer is determined based on the TNM information. Stage 0 describes non-invasive in situ cancer, Stage I describes early stage invasive cancer, and Stage IV describes the most advanced or metastatic cancer.

The stage at diagnosis has been reported to influence breast cancer survival (American Cancer Society, 2011). Early detection of breast cancer improves the chance of successful treatment. Therefore, it has been recommended by experts that women receive annually breast cancer screening beginning at age 40.

Three breast cancer detection methods listed in the American Cancer Society guidelines are mammography, clinical breast examination (CBE) and magnetic resonance imaging (MRI) (American Cancer Society, 2011). Among these, mammography is the most popular screening methods. However, it exposes patients to radiation and is uncomfortable for many women. A relatively low sensitivity of 54% to 77% depending on the type of mammography (Skaane et al., 2009) often necessitates follow-up examinations, such as biopsies.

Breast cancer is a heterogeneous disease, consisting of distinct clinical and histological forms. Because of this heterogeneity, so far no universal biomarker has been found which can diagnose all types of breast cancer. Currently, well-studied biomarkers of breast cancer are categorized into seven groups. They are (i) steroid (hormone)

receptors, (ii) the epidermal growth factor receptor family, (iii) the proliferation marker

Ki67, (iv) cell-cycle regulation and apoptotic markers, (v) angiogenesis-related proteins,

(vi) extracellular matrix-related proteins and (vii) cyclooxygenase-2 (COX-2). Steroid

(hormone) receptors (ER and PR) and the epidermal growth factor receptor family

(HER2/ ErbB2) are the most extensively investigated biomarkers. Despite the fact that

these common biomarkers provide abundant information regarding molecular features

and classifications of breast cancer, none of these biomarkers is abnormally expressed in

all breast cancer patients and thus none of them has satisfactory diagnostic ability.

Therefore, there is a compelling need to identify safer and more sensitive biomarkers that

can distinguish breast cancer patients from normal population.

Metabolomics

Metabolomics is a new approach to the qualitative or quantitative analysis of the

small-molecule metabolites of a biological organism. Metabolic variations, regarded as

the downstream end products of alterations in gene and protein expressions, reflect not

only genetic phenotype but also environmental influences. Because small changes in

enzyme activities can be amplified and detected on the metabolite level, metabolomics

may offer hope for the discovery of potential biomarkers for diseases. It has been widely

used to identify differential metabolites that can distinguish between healthy people and

patients with several types of cancer, such as prostate cancer (Fan et al., 2011), ovarian

cancer (Chen et al., 2011; Garcia et al., 2011), breast cancer (Chen et al., 2009; Nam et

al., 2009; Frickenschmidt et al., 2008), pancreatic cancer (Nishiumi et al., 2010), kidney

cancer (Kim et al., 2011), colorectal cancer (Qiu et al., 2010), gastric cancer (Cai et al., 2010), oral cancer (Wei et al., 2011), lung cancer (An et al., 2010), and bladder cancer (Pasikanti et al., 2010).

Metabolic profiles have been analyzed in various types of biological specimens, including blood, plasma, serum, urine, tissue, cells, saliva, cerebrospinal fluid, and tears. Some biofluid samples, such as blood, urine, and saliva, are non-invasive and easy to acquire, making them appropriate for biomarker discovery research and other clinical studies, while tissue and cell samples may be used for preclinical research.

Two of most commonly used analytical tools in metabolomics are nuclear magnetic resonance (NMR), and mass spectrometry (MS). NMR detects compounds by measuring the magnetic influence on nuclei of compounds by an external magnetic field. The sample preparation process for NMR is easy, but it has low sensitivity. MS detects molecules by creating electrically charged ions in the ion source, separating introduced ions in an analyzer, and then using a detector to detect ions and record information of mass-to-charge ratios. Common ionization methods include electron ionization (EI), chemical ionization (CI), electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI). Common analyzers of MS include quadrupole, triple quadrupoles, time-of-flight (TOF), and ion traps. MS is always coupled with either liquid chromatography (LC) or gas chromatography (GC) to separate different compounds based on the differential partition between the mobile and stationary phases. The time it takes compounds to be eluted from the column (also called retention time) is another parameter to identify the compounds. GC-MS is used to analyze the volatile compounds,

with the advantage of high resolution. However, during the sample preparation process, derivatization may be needed to reduce the polarity of the compounds and render them volatile. LC-MS has an easier sample preparation process and is able to detect a wider range of compounds than GC-MS. The raw data acquired from the instruments must be processed by software programs, with the processing steps like baseline fitting, deconvolution, and peak picking.

Metabolomics is a high-throughput technology and usually produces complex multivariate datasets, and thus chemometric, bioinformatic and statistical methods are required to interpret and visualize the data. Several web-based metabolomics data processing tools have been developed, including MetaboAnalyst and metaP-Server. These tools offer a variety of functions, including univariate statistical methods (such as fold change analysis and t-test), unsupervised and supervised multivariate statistical methods (such as PCA, PLS-DA and hierarchical clustering), metabolite identification and pathway mapping.

Two major metabolomics methods are targeted and non-targeted metabolic profiling. Non-targeted metabolomics studies the profile of the whole metabolome, so that metabolites with unknown chemical structures may still be detected. However, non-targeted metabolomics may produce biased results due to the limitations of current technology to detect all metabolites. Many factors may influence the results, including the sample collection process, sample stability, extraction methods, ion suppression for LC-MS, and derivatization for GC-MS. In contrast, the targeted strategy, which was

developed to quantitatively analyze particular sets of metabolites (e.g. lipids or amino acids), may provide more robust and reliable results (Christians et al., 2011).

Metabolomics in the field of breast cancer research

In the field of breast cancer research, both targeted and non-targeted metabolomics methods have been used for identifying biomarkers for diagnosis and prognosis, predicting treatment activity and toxicity, classifying types of breast cancer, and studying the mechanisms of tumor growth and progression.

Discovery of potential biomarkers for diagnosis

Metabolic profiles of breast cancer between patients and healthy volunteers have been compared in several studies. Urine samples are the most commonly used biofluids in the analysis because they offer an easily acquired and non-biopsy screening approach. Some altered metabolic pathways and dysregulated metabolites have been detected in breast cancer patients by univariate and multivariate statistical model. Receiver operating characteristic (ROC) analysis has been used to select optimal classifiers from these differential metabolites, which may be helpful for the detection of disease.

A study conducted by Nam et al. (2009) analyzed metabolic profiles of the urine samples collected from 50 breast cancer patients and 50 healthy individuals. Four differentially expressed metabolites (homovanillate, 4-hydroxyphenylacetate, 5-hydroxyindoleacetate and urea) were identified as potential biomarkers. This combination of markers yielded a good classification performance in the ROC analysis, with area under the ROC curve (AUC) values of 0.75, 0.79, and 0.79 using three classification

algorithms of a linear discriminate analysis (LDA), a random forest classifier (RF), and the support vector machine (SVM), respectively.

Another study conducted by Chen et al. (2009) combined a metabolomics approach and a metabolic correlation network analysis to investigate the metabolite profiles of 20 breast cancer patients and 18 healthy volunteers. Twelve urine metabolites including dimethylarginine, tyrosine, phenylalanine, pantothenic acid, succinyladenosine, dimethylguanosine, apronal, threonylcarbamoyl adenosine, tryptophan, kynurenic acid, nicotinuric acid and indolelactic acid were identified as potential biomarkers in this study.

Cancer cells show a high rate of RNA turnover to satisfy the metabolic requirements of dysregulated growth. The excretion of modified nucleosides, which are the intermediates of RNA metabolism, has been found increased in the urine in cancer. Therefore, several breast cancer biomarker studies focused on the metabolic phenotype of urinary modified nucleosides. An examination of nucleosides in the urine samples of 113 breast cancer patients and 99 healthy volunteers was performed by Frickenschmidt et al. (2008). This study revealed that a model containing 31 nucleosides could be used to differentiate between breast cancer patients and healthy volunteers with a sensitivity of 87.67% and a specificity of 89.90%. In another study (Bullinger et al., 2008), eleven nucleosides were quantitatively measured in the urine samples collected from 51 breast cancer patients and 65 healthy controls, and the results of this study showed that this panel of eleven nucleosides achieved a sensitivity of 94% and a specificity of 86% in a SVM model.

In addition to urine, saliva has also been used in the biomarker discovery research, because it is easy and cost-effective to collect. The metabolic profiles of saliva from 30 breast cancer and 87 healthy controls were analyzed (Sugimoto et al., 2009). A total of 28 metabolites were identified as biomarker candidates, and they could provide a high AUC value of 0.973 in the ROC analysis to discriminate breast cancer patients from healthy controls.

Although biofluids have shown superiority over other type of samples in the biomarker discovery research, identifying potential markers in the tissue samples may be complementary to the microscopic analysis of biopsies. The biopsy levels of choline related metabolites were analyzed by NMR in 85 breast tumor tissues and 18 adjacent non-tumor tissues (Sitter et al., 2006). Tumor tissues could be distinguished from non-tumor tissues using the relative intensities of glycerophosphocholine, phosphocholine and choline, with a sensitivity of 82% and a specificity of 100%.

A large cohort of 271 breast cancer and 98 normal tissue samples were investigated by GC/TOFMS (Budczies et al., 2012). Using cut-off criteria of sensitivity and specificity greater than 80%, 13 metabolites increased in breast cancer tissues (including cytidine-5-monophosphate, adenosine-5-monophophate, phosphoethanolamine, taurine, pyrazine2,5-dihydroxy, creatinine, N-acetylaspartate, hypoxanthine, glycerol-alpha-phosphate, aminomalonate, glutamic acid, malate and oxoproline) and 7 metabolites increased in normal tissues (including heptadecanoic acid, lignoceric acid, hexadecanol, pentadecanoic acid, glycolic acid, benzoic acid and hydroxylamine) were identified as potential biomarkers. To discriminate breast tumors from normal tissues,

classifiers were constructed from dividing one tumor marker by one normal tissue marker. Among these classifiers, the ratio of cytidine-5-monophosphate/ pentadecanoic acid had the highest discriminant ability with a sensitivity of 94.8% and a specificity of 93.9%.

Discovery of potential biomarkers for prognosis

There are five commonly used treatment options for patients with breast cancer, which are surgery, radiation therapy, chemotherapy, hormone therapy and targeted therapy (American Cancer Society, 2011). However, all the breast cancer patients face risks of recurrence even after successful treatments. A report revealed that the 5-year and 10-year residual risks of recurrence for breast cancer patients were 11% and 20%. Tumor grade, stage, and status of hormone receptors were related with late recurrence (Brewster et al., 2008). Improved prognostic tools are needed for detecting the cancer recurrence early and estimating the survival time. Metabolomics may also help indentify biomarkers for recurrence, besides the occurrence of breast cancer.

Asiago et al. (2010) performed a metabolic profiling analysis of 257 serum samples collected from 20 recurrent breast cancer patients and 36 patients with no clinical evidence of disease. Data were acquired from NMR and GC/GC-MS. A prediction model containing 11 metabolites was established using the logistic regression method. These 11 metabolites included formate, histidine, proline, choline, tyrosine, 3-hydroxybutyrate, lactate, glutamic acid, N-acetyl-glycine, 3-hydroxy-2-methyl-butanoic acid and nonanedioic acid. This panel of 11 markers provided a sensitivity of 86% and a specificity of 84% from ROC analysis.

Sitter ea al. (2010) quantitatively analyzed and compared differential metabolites in the tissue samples of 13 patients with good prognosis and 16 patients with poor prognosis. Prognosis status was defined by tumor size, estrogen and progesterone hormone receptor status, and whether the cancer cells had been spread to axillary lymph nodes. They found that the ratios of taurine/glycine, total cholines/glycine and glycerophosphocholine/glycine were significantly different between the good and poor prognosis groups. Even in the poor prognosis group, patients who had experienced recurrence had significantly lower ratios of taurine/glycine and glycerophosphocholine/glycine, compared to the patients who were healthy five years after surgery.

Alteration of lipid metabolism is a hallmark of breast cancer. A high level of glycerol-3-phosphate acyltransferase (GPAM), a key enzyme catalyzing the biosynthesis of triacylglycerols and phospholipids, was found significantly correlated with a better overall patient survival. Phosphatidylcholines were increased significantly in the high GPAM expression tumors. The other differential metabolites between the groups of high and low expression of GPAM included oxalic acid, glucose, flavin adenine dinucleotide, 2-aminoadipic acid, ribose-5-phosphate, myo-inositol, idonic acid, serine, capric acid, tyrosine, cholesterol, dehydroasorbate, 2-hydroxyglutaric acid, β-alanine and maltotriose (Brockmoller et al., 2012).

Oakman et al. (2011) found that the early stage breast cancer patients have characteristic metabolic profiles in the serum. The established multivariate statistical model showed the early stage cancer patients were clustered together, and could be

separated from the metastatic breast cancer patients. Therefore, they developed a prognostic method called metabolic risk for early stage patients, which was calculated as the inverse distance of each early stage cancer patient from the cluster barycenter of the metastatic patients. However, this method requires further development and validation.

Detection and prediction of treatment efficacy

Breast cancer is a heterogeneous disease, and therefore, patients at the same stage of cancer or with similar symptoms may exhibit different responses to the same treatment. It is important to develop approaches to evaluate the efficacy of treatments for breast cancer. Researchers have utilized metabolomics methods to detect chemotherapy efficiency. By analyzing serum metabolic profiles of 34 HER-2-positive patients who were treated with both paclitaxel (an inhibitor used in breast cancer chemotherapy) and lapatinib (a tyrosine kinase inhibitor targeting HER-2), it was found that patients with longest time to progression (TTP) had significantly distinctive metabolomes, compared to the patients with shortest TTP. The serum levels of glucose were higher in the patients with longest TTP, while the serum levels of glutamate and phenylalanine were higher in the patients with shortest TTP. The constructed model in this study provided a high accuracy of 89.6% to predict the effect of treatment on TTP. Although these results are required to be further validated in the sample sets with larger sample sizes, this study confirms the role of metabolomics in detection and prediction of treatment efficacy (Tenori et al., 2012).

It has been shown that weight gain during the breast cancer treatment is associated with a poor prognosis. Keun et al. (2009) published a study which utilized the

metabolomics methods to investigate the potential mechanism underlying the weight gain induced by the breast cancer chemotherapy. The serum level of lactate was found increased by 63.5% in the weight gain patients, compared to the weight same or loss patients.

Classification of breast cancer

In addition to stages, breast cancer can also be classified into four types based on gene expression profiles: luminal A, luminal B, basal-like, and HER-2 enriched. Patients with luminal A, with a higher expression of estrogen receptor (ER) and low expressions of progesterone receptor (PR) and HER-2, have the best prognosis. Patients with luminal B have a high expression of PR but low expressions of ER and HER-2. Patients with HER-2 enriched expression and with basal-like type (ER PR and HER-2 negative) have the worst prognosis. Although patients with luminal A have a relatively good prognosis, there are still some patients who do not respond to anti-estrogen treatment. Therefore, further study on classification of breast cancer is required. Using metabolomics methods, luminal A breast cancer could be further classified into three subtypes based on hierarchical clustering of the HR MAS MR spectra (Borgan et al., 2010). One of these three subtypes had a lower level of glucose and a higher level of alanine than other subtypes.

Uncovering the altered metabolic pathways

Metabolic reprogramming is recently perceived as another hallmark of cancer, in addition to sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion

and metastasis (Hanahan and Weinburg, 2011; Ward and Thompson, 2012). Tumor cells reorganize their metabolic phenotypes to support rapid growth and proliferation, and thus exhibit a characteristic metabolic signature. Therefore, unraveling the alterative metabolism of cancer is another research focus in metabolomics studies. Understanding the fundamental metabolic pathways altered by cancer may facilitate the discovery of new therapeutic targets.

The most noticeable metabolic feature of tumor cells is the Warburg effect. The Warburg effect, which was initially found by Nobel laureate Otto Warburg in 1924, describes that cancer cells tend to take up more glucose, and more glucose is converted into lactate through pyruvate instead of entering into the TCA cycle to produce energy. Normal cells produce a large amount of lactate only when oxygen is absent, but cancer cells maintain this metabolic process even though oxygen is present, so this process is called aerobic glycolysis. 2-[^{18}F]Fluoro-2-deoxyglucose positron emission tomography (FDG-PET) scanning, which assesses the Warburg effect by measuring glucose uptake in the local tissue, has been widely applied to diagnosis and staging of cancer. FDG-PET scanning also plays an important role in diagnosis, staging, and prognosis in breast cancer (Kumar and Alavi, 2004).

Aerobic glycolysis produces excess lactate and generates much less ATP than oxidative phosphorylation, which is an inefficient way to consume glucose. One possible reason explaining why cancer cells remain this inefficient metabolism is because ATP is not a limiting resource for the cancer cells and it is necessary for cancer cells to balance energy generation and macromolecule production in the cancer cells (Heiden et al., 2009).

It has been shown that aerobic glycolysis produces a great amount of precursors for biosynthesis of biomass, such as lipids, nucleotides and amino acids, and fulfills the metabolic requirements of cell growth and proliferation.

In addition to the Warburg effect, various other metabolic alterations that tumors cells undergo have been found in breast cancer using metabolomics methods.

Nucleoside metabolism

Increased biosynthesis of nucleotides and nucleosides was detected in the breast cancer tissues when analyzing metabolic profiles of 271 breast cancer tissues and 98 normal tissues using GC/TOFMS. The levels of CMP and AMP were significantly increased in the breast cancer tissues, with fold changes of 10.3 and 7.8 respectively compared to the normal tissues (Budczies et al., 2012).

Cancer cells have also shown a high rate of RNA turnover to satisfy the metabolic requirements of dysregulated growth. The excretion of modified nucleosides, the intermediates of RNA metabolism, is increased in the urine in cancer, because of the lack of specific phosphorylases to recycle them. Woo et al. (2009) analyzed the urine levels of nucleosides among 10 breast cancer patients and 22 normal controls, and found that two nucleosides, 8-hydroxy-2-deoxyguanosine and 5-hydroxymethyl-2-deoxyuriding, were significantly increased in the urines of breast cancer patients. The profiles of modified nucleosides of MCF-7 breast cancer cells and MCF10A normal breast cells were measured and analyzed using LC-MS (Bullinger et al., 2007). 5-methyluridine, $N^2,N^2,7$-trimethylguanosine, N^6-methyl-N^6-threonylcarbamoyladenosine and 3-(3-aminocarboxypropyl)-uridine were significantly elevated in the medium of MCF-7 cells.

Lipid metabolism

The change of lipid metabolism is another characteristic feature observed in breast cancer. Fatty acids have essential functions as energy source and compositions of cell membrane lipids. Fatty acids in human body are obtained by diet or *de novo* synthesis. Breast cancer cell growth was impeded by silencing the gene ELOVL1 (elongation of very long chain fatty acid-like1), and it suggests that *de novo* biosynthesis of fatty acid plays an important role in breast cancer cells (Hilvo et al., 2011).

The free fatty acid profiles of serum have been compared in breast cancer patients and healthy controls by GC-MS (Lv and Yang, 2012). This work found that breast cancer patients had significantly increased levels of C16:0, C18:0 and C18:2, and had significantly decreased levels of C18:3, C20:5 and C22:5 compared to the healthy controls.

The elevated biosynthesis of cell membrane lipids has been observed in breast cancer as well. A lipidomic profiling analysis (Hilvo et al., 2011), conducted in 257 breast cancer tissue samples and 10 adjacent breast tissue samples, showed that phosphatidylcholines, phosphatidylethanolamines, phosphatidylinositols, sphingomyelins and ceramides were significantly increased in tumor tissues. The mRNA expression of glycerol-3-phosphate acyltransferase (GPAM), a key enzyme catalyzing the biosynthesis of triacylglycerols and phospholipids, was also found higher in breast cancer and gynecological cancer than that in other types of cancer by analyzing the GeneSapiens in silico transcriptiomics database. In addition, it was reported that breast cancer tissues had

a higher level of sn-glycerol-3-phosphate which is a substrate of GPAM compared to normal tissues.

Amino acid metabolism

Amino acid metabolism is also altered in cancer. The plasma free amino acid profiles of five types of cancer (lung, gastric, colorectal, breast and prostate cancer) were measured by HPLC-MS (Miyagi et al., 2011). It was found that different types of cancer shared several similar changes of plasma free amino acid profiles, such as decreased concentrations of tryptophan and histidine in all of the cancers except prostate cancer. Each cancer exhibited a unique characteristics of metabolic profiles as well, for instance, breast cancer had an increased concentration of threonine, which was decreased in gastric and colorectal cancer. But differences between cancer patients and controls were larger than those among different types of cancer.

Glutamine is a nonessential amino acid. However, cancer cells show a distinct characteristic that they cannot survive in the absence of exogenous glutamine. This metabolic signature of cancer is regarded as glutamine addiction. Glucose tends to be converted into lactate instead of producing energy through TCA cycle and oxidative phosphorylation in cancer cells as described in the Warburg effect. Glutamine has been considered as another essential source of energy and nitrogen for cancer cells. It has been shown that glutamine can be converted into α-ketoglutarate to replenish TCA cycle intermediates and provide energy for the cancer cells (Wise and Tompson, 2010).

Unraveling the mechanisms of progression

As a tumor rapidly grows, the center of tumor may undergo hypoxia condition because of deprived blood supply. Weljie et al. (2011) studied metabolic alterations induced by hypoxia condition in breast cancer. They analyzed the metabolite patterns of MDA-MB-231 breast cancer cells in both hypoxic and normoxic conditions *in vitro*, and they also *in vivo* compared the serum metabolic profiles of the mice introduced with MDA-MB-231 cells and control animals using NMR. When comparing the *in vitro* and *in vivo* results, it was found that the concentrations of threonine, leucine, lysine, phenylalanine, 1-methylhistidine and 2-hydroxybutrate were increased, while 2-oxoglutarate was decreased, in both the medium of cells under hypoxic condition and the serum of mice with tumors.

Richardson et al. (2008) conducted a stable isotope flux analysis of carbon metabolism in the MCF10 model of breast cancer which included normal breast cells MCF10A and three MCF10A derived breast cancer cell lines in different stages of tumor progression (transformation, tumorigenicity and metastasis). A total of 22 key metabolites and 15 metabolic pathways were quantitatively measured using 2D HSQC-NMR and GC-MS. It was found that in the breast cancer cell lines, more glucose was metabolized and converted to pyruvate via pentose phosphate pathway, rather than via glycolysis, and the production of ribose through pentose phosphate pathway was also increased in the breast cancer cells compared to normal cells. Carbon flux through TCA cycle was increased and the level of succinate, a key metabolic intermediate in the TCA cycle, was higher in the breast cancer cells. Compared to the normal cells, *de novo*

synthesis of palmitate (C16:0), stearate (C18:0) and oleate (C18:1) were elevated in the tumor cells. Transformed and tumorigenic cells had higher pool size of glutamate, while metastatic cells had smaller pool size of glutamate but higher proline.

Conclusion

Breast cancer has been the most frequently diagnosed cancer and the second leading cause of cancer death among American women. There is a compelling need to identify novel biomarkers that can distinguish breast cancer patients from normal population.

Metabolomics, an approach to study of small molecules, provides high-resolution insight of characteristic biochemical phenotypes in disease, and may offer opportunity for the discovery of new diagnostic tools for breast cancer. Numerous targeted or non-targeted metabolomics studies have revealed that breast cancer undergoes profound metabolic alterations in glycolysis, TCA cycles, lipid metabolism, amino acid metabolism and nucleic acid metabolism. Metabolomics has also been used for identification of biomarkers for diagnosis and prognosis in breast cancer, prediction of treatment activity and toxicity, and classification of breast cancer.

Blood and urine are two biofluids commonly used in the metabolomics research. However, most of potential biomarkers for breast cancer diagnosis were identified in the urine. Analysis of metabolic patterns and identification of potential biomarkers of breast cancer have never been performed in a large cohort of plasma samples. The metabolites and nutrients are transported by blood and excreted into urine only when their

concentrations exceed the relevant kidney threshold. As the main function of urine is to remove unwanted compounds from the body, the levels of nonnutrient compounds found in urine will naturally be higher than those found in blood. Therefore, blood and urine of breast cancer patients may obtain distinct metabolic profiles, and further studies are required to investigate whether there are any differentially expressed metabolites in the circulating blood that can achieve remarkable diagnostic power and serve as biomarkers for breast cancer.

CHAPTER III

ANALYSIS OF METABOLIC PROFILES AND IDENTIFICATION OF POTENTIAL BIOMARKERS IN BREAST CANCER

Abstract

Metabolomics provides insight of characteristic biochemical phenotypes in disease, and facilitates the development of novel diagnostic tools. In order to identify potential biomarkers for breast cancer, GC-TOFMS and LC-TOFMS spectra were acquired in the plasma collected from 138 breast cancer patients and 76 healthy women. Multivariate and univariate statistics methods were applied to analyze the metabolic alterations between groups. Of the 41 identified differential metabolites, aspartate was significantly reduced in breast cancer and obtained good predictive power for distinguishing breast cancer patients from healthy controls. An established combination of 7 markers (asparagine, hypotaurine, 5-oxoproline, cysteine, aspartate, glutamate and glutamine) was found to provide even better predictive power than aspartate alone. The altered expression of aspartate was confirmed in an independent set of serum samples and 20 pairs of breast tumor tissue and its adjacent normal tissue. It was also found that the metabolic profiles of stage I and stage IV patients can be separated in the constructed OPLS-DA model. In conclusion, metabolomics offers a powerful opportunity to detect profound metabolic dysregulations and to identify potential biomarkers in breast cancer.

Introduction

Among women in the U.S., breast cancer has been the most commonly diagnosed cancer and the second leading cause of cancer death (after lung cancer) for most of the past decade (U.S. Cancer Statistics Working Group, 2010). One of the most popular screening methods for breast cancer is mammography. However, mammography exposes patients to radiation and is uncomfortable for many women. Furthermore, it has a sensitivity of only 54% to 77% depending on the type of mammography (Skaane, 2009). A safer and more sensitive diagnostic method is thus required. Biomarker in the urine or blood is an indicator of health status, and it has advantage in easy acquisition. However, breast cancer is heterogeneous disease, consisting of distinct clinical and histological forms. This heterogeneity makes it lacking of the universal diagnosis biomarker. Therefore, there is a compelling need to identify novel biomarkers that can distinguish breast cancer patients from healthy people.

Metabolomics is a new approach to the study of small molecules, and may facilitate to identify the metabolic alterations of cancer and offer opportunity to discover biomarkers for disease diagnosis. Three common analytical tools used in metabolomics are nuclear magnetic resonance (NMR), liquid chromatography-mass spectrometry (LC-MS) and gas chromatography-mass spectrometry (GC-MS). Metabolomics method has been widely used to identify differential metabolites between disease patients and healthy people in the field of cancer research, such as prostate cancer (Fan et al., 2011), ovarian cancer (Chen et al., 2011; Garcia et al., 2011), breast cancer (Chen et al., 2009; Nam et al., 2009; Frickenschmidt et al., 2008), pancreatic cancer (Nishiumi et al., 2010), kidney

cancer (Kim et al., 2011), colorectal cancer (Qiu et al., 2010), gastric cancer (Cai et al., 2010), oral cancer (Wei et al., 2011), lung cancer (An et al., 2010), and bladder cancer (Pasikanti et al., 2010).

Metabolic reprogramming is recently perceived as a hallmark of cancer. The most noticeable metabolic feature of tumor cells is the Warburg effect, which describes that cancer cells tend to take up more glucose and convert it into lactate via aerobic glycolysis. 2-[^{18}F]Fluoro-2-deoxyglucose positron emission tomography (FDG-PET) scanning, which assesses the Warburg effect by measuring glucose uptake in the local tissue, plays an important role in diagnosis, staging, and prognosis in breast cancer (Kumar et al., 2004). In addition to the Warburg effect, breast cancer patients have also been reported to show the perturbed metabolic patterns of lipids (Hilvo et al., 2011; Lv and Yang, 2012), nucleosides (Budczies et al., 2012; Woo et al., 2009; Bullinger et al., 2007) and amino acids (Miyagi et al., 2011). Previous metabolomics studies have identified a couple of potential biomarkers for breast cancer diagnosis in the urine and saliva samples. However, the metabolic pattern of breast cancer has never been analyzed in a large cohort of plasma samples.

The aim of the present study was to investigate metabolic signature of breast cancer and to identify potential biomarkers by analyzing plasma samples from 138 breast cancer patients and 76 healthy women using LC-TOFMS and GC-TOFMS. The metabolic profiles of patients in different stages were also compared. A target analysis was performed in the serum samples to validate the potential biomarker. In order to investigate whether the altered expression of the potential biomarker in the plasma can be

also observed in tumor, 20 pairs of breast tumor tissue and its adjacent normal tissue were analyzed as well.

Materials and Methods

Plasma and serum sample collection

We obtained a batch of plasma samples to discover differential metabolites for breast cancer in this study. These plasma samples were collected from 138 breast cancer patients aged 30-70 years and 76 healthy women aged 20-40 years at the City of Hope Cancer Center in California. Their TNM stages were recorded from pathological result with: stage I, 19 patients; stage II, 50 patients; stage III, 49 patients; stage IV, 20 patients. The subjects were divided into a training set, which was used to construct the statistical model to discriminate the patients from the healthy controls, as well as a test set to verify the statistical model. The training set included 35 patients and 35 age-matched healthy controls, and the remaining patients and healthy controls were analyzed as the test set.

Another batch of serum samples was used to validate the discovered differential metabolites. They were collected from 80 newly diagnosed breast cancer patients, including 10 stage I patients, 45 stage II patients and 25 stage III patients, and 70 healthy volunteers at Ruijin Hospital in Shanghai, China. The patients ranged in age from 36 to 78, with an average age of 49.5, and the healthy volunteers ranged in age from 35 to 76, with an average age of 58.3.

All the patients were newly diagnosed as breast cancer and were not on any medication prior to sample collection. All the samples were collected in the morning

before breakfast and were stored in the clean tubes at -80°C until analysis. The characteristics of patients were summarized in Table 1. This study was approved by the local institutional review boards and all patients gave informed written consent before they were involved in the study.

Breast cancer tissue samples

Twenty pairs of frozen tissue were purchased from Biochain (CA, USA). Each pair of tissue consisted of breast tumor tissue and its adjacent normal tissue from the same patient. These samples were excised from 20 patients in stage I (n=1), stage II (n=10), stage III (n=8), stage IV (n=1) .The average age of the patients was 62.5 (from 46 to 75 years). The characteristics of patients were summarized in Table 1.

Table 1. Summary of the characteristics of patients used in this study.

		Plasma samples from City of Hope Cancer Center, California, USA Training set		Test set		Serum samples from Ruijin Hospital, Shanghai, China		Tissue samples puchased from Biochain
		breast cancer patients	healthy volunteers	breast cancer patients	healty volunteers	breast cancer patients	healty volunteers	patients
N		35	35	103	41	80	70	20
Age (Average ± SD)		39.6 ± 4.5	38.1 ± 1.4	57.7 ± 7.4	28.7 ± 4.5	49.5 ± 8.5	58.3 ± 8.6	62.6 ± 9.4
TNM Stage	**I**	1		18		10		1
	II	16		34		45		10
	III	16		33		25		8
	IV	2		18				1
Race	**Asian**	4		14		80	70	
	White	23	17	77	20			20
	Black	3	8	6	5			
	Latino		10		16			
	Native			1				
	Others	5		5				

Plasma and serum sample preparation and data acquisition by GC-TOFMS

Plasma and serum metabolites extraction and derivatization were performed following our lab's previously published procedure with minor modifications (Bao et al.,2009; Qiu et al., 2009). Briefly, a 50 µL aliquot of plasma or serum sample was spiked

with two internal standard solutions (10 µL p-chlorophenylalanine in water, 0.1 mg/mL; 10 µL heptadecanoic acid in methanol, 1 mg/mL). The mixed solution was extracted with 175 µL of methanol: chloroform (3:1) and vortexed for 30 seconds. After storing for 10 minutes at -20°C, the samples were centrifuged at 13,200 rpm for 10 minutes at 4°C. An aliquot of 200 µL supernatant was transferred to a glass sampling vial to vacuum dry at room temperature. The dried extracts were derivatized using a two-step procedure. First, 50 µL methoxyamine (15 mg/mL in pyridine) was added to the vial and derivatized at 30°C for 90 minutes. After adding 10 µL C10-C40 (all even alkanes, 12.5µg/mL) as retention index, the samples were added 50 µL BSTFA (1%TMCS) and derivatized at 70°C for 60 minutes.

Each 1 µL aliquot of the derivatized solution was injected in splitless mode into an Agilent 7890A gas chromatograph with the Agilent 7683 autosampler coupled to a Pegasus HT time-of-flight mass spectrometer (Leco Corporation, St Joseph, USA). To minimize systematic error, each healthy control sample was run for every one or two breast cancer samples, and one quality control sample was run for every ten analyzed samples. Breast cancer samples from different stages were also run evenly in the whole experiment. Separation was achieved on an Rxi-5 ms capillary column (Crossbond® 5% diphenyl/ 95% dimethyl polysiloxane; 30m × 0.25mm i.d. × 0.25µm; Restek, PA, USA), with helium as the carrier gas at a constant flow rate of 1.0 mL/min. The temperatures of injection, transfer interface, and ion source were set to 260°C, 260°C, and 210°C, respectively. The GC temperature programming was set to 2 min isothermal heating at 80°C, followed by 10°C/min oven temperature ramps to 220°C, 5°C/min to 240°C, and

25°C/min to 290°C, and a final 8 min maintenance at 290°C. Electron impact ionization (70 eV) at full scan mode (m/z 40-600) was used, with an acquisition rate of 20 spectra/second in the TOFMS setting.

The raw data generated in GC-TOFMS instrument were processed by ChromaTOF software (v4.43, Leco Co., CA, USA), including baseline computation, peak detection and peak area calculation. Compound identification was performed by comparing the mass fragments (or retention time and retention index if available) with our in-house standard library and NIST 11 mass spectral library. Those metabolites identified from our in-house library were viewed as identities confirmed by standards. Peak areas of unique mass were normalized to the internal standard. Internal standards and any known artificial peaks, such as peaks caused by noise, column bleed and derivatization procedure, were removed from the data set.

Plasma and serum sample preparation and data acquisition by LC-TOFMS

The plasma or serum samples were thawed and centrifuged at 13,200 rpm for 5 min. A volume of 100 μL supernatant was mixed with 10 μL of p-chlorophenylalanine (10 μg/mL, internal standard) and 400 μL of a mixture of methanol and acetonitrile (5:3). The mixture was vortexed for 2 min, allowed to stand for 10 min, then centrifuged at 13,200 rpm for 20 min, and the supernatant was used for liquid chromatography-time of flight mass spectrometry (LC-TOFMS) analysis.

An Agilent HPLC 1200 system equipped with a binary solvent delivery manager and a sample manager (Agilent Corporation, Santa Clara, CA, USA) was used with chromatographic separations performed on a 4.6 × 150 mm 5 μm Agilent ZORBAX

Eclipse XDB-C18 chromatography column. The LC elution conditions were optimized as follows: isocratic at 1% B (0–0.5 min), linear gradient from 1% to 20% B (0.5-9.0 min), 20-75% B (9.0-15.0 min), 75-100% B (15.0-18.0 min), isocratic at 100% B (18–19.5 min); linear gradient from 100% to 1% B (19.5-20.0 min) and isocratic at 1% B (20.0–25.0 min). Here, A = water with 0.1% formic acid and B = acetonitrile with 0.1% formic acid. The column was maintained at 30 °C. A 10 μL aliquot sample was injected onto the column. Mass spectrometry was performed using an Agilent model 6220 MSD TOF mass spectrometer equipped with a dual sprayer electrospray ionization source (Agilent Corporation, Santa Clara, CA, USA). The system was tuned for optimum sensitivity and resolution using an Agilent ESI-L low concentration tuning mix in both positive (ES+) and negative (ES-) electrospray ionization modes. Agilent API-TOF reference mass solution kit was used to obtain accurate mass time-of-flight data in both positive and negative mode operation. The TOF mass spectrometer was operated with the following optimized conditions: (1) ES+ mode, capillary voltage 3500 V, nebulizer 45 psig, drying gas temperature 325 °C, drying gas flow 11 L/min, and (2) ES- mode, similar conditions as ES+ mode except the capillary voltage was adjusted to 3000 V. The TOF mass spectrometer was calibrated routinely in ES+ and ES- modes using the Agilent ESI-L low concentration tuning mix. During metabolite profiling experiments, both plot and centroid data were acquired for each sample from 50 to 1,000 Da over a 25 min analysis time.

The resulting .d files were then centroided, deisotoped, and converted to mzData xml files using the MassHunter Qualitative Analysis Program (vB.05.00, Agilent).

Following conversion, xml files were analyzed using the open source XCMS package (v1.24.1), which runs in the statistical package R (v.2.15.1) to pick, align, and quantify features (chromatographic events corresponding to specific m/z values and retention times). The software was used with default settings as described except for xset (bw = 5) and rector (plottype = "m", family = "s"). The created .tsv file was opened using Excel software and saved as .xls file. The resulting data sheet normalized to the internal standard was used for the further analysis. Metabolites annotation was performed by comparing the accurate mass (m/z) and retention time (RT) of reference standards in our in-house library and the accurate mass of compounds obtained from the web-based resources such as the Human Metabolome Database.

Profiling analysis of tissue samples by GC-TOFMS and LC-TOFMS

Approximately 40 mg of tissue samples were prepared using two-step extraction. The tissue sample was first added with 50 μL of the first-step extraction solvent (chloroform : methanol : water = 1:2.5:1) and homogenized for 6 min in a bullet blender (Next Advance, Inc., BIOBOX^{TM}). Then the sample was extracted with another 250 μL of first-step extraction solvent, and centrifuged at 13,200 rpm for 20 minutes at 4°C. An aliquot of the 100 μL supernatant was transferred to a GC sampling vial and another 100 μL supernatant was transferred to a LC sampling vial. The rest of supernatant was used as quality control (QC) samples. At the second step, the deposit of tissue was extracted with 300 μL methanol. After centrifugation at 13,200 rpm for 20 minutes at 4°C, two aliquots of 100 μL supernatant were transferred to the GC and LC vials, and the rest was pooled into QC samples.

For the GC-TOFMS analysis, the samples spiked with two internal standard solutions (10 µL p-chlorophenylalanine in water, 0.1 mg/mL; 10 µL heptadecanoic acid in methanol, 1 mg/mL) were vacuum-dried at room temperature. The dried extracts were derivatized using a two-step procedure. First, 50 µL methoxyamine (15 mg/mL in pyridine) was added to the vial and derivatized at 30°C for 90 minutes. Then, the samples were added 50 µL BSTFA (1%TMCS) and derivatized at 70°C for 60 minutes. The samples were separated through an Rxi-5 ms capillary column (Crossbond® 5% diphenyl/ 95% dimethyl polysiloxane; 30m × 0.25mm i.d. × 0.25µm; Restek, PA, USA), and analyzed by an Agilent 7890A gas chromatograph coupled to a Pegasus HT time-of-flight mass spectrometer (Leco Corporation, St Joseph, USA). Helium was used as the carrier gas at a constant flow rate of 1.0 mL/min. Each 1 µL aliquot of the derivatized solution was injected in splitless mode. The GC temperature programming was set to 2 min isothermal heating at 80°C, followed by 10°C/min oven temperature ramps to 220°C, 5°C/min to 240°C, and 25°C/min to 290°C, and a final 8 min maintenance at 290°C. To minimize systematic error, each normal control sample was run for every one breast cancer samples, and one quality control sample was run for every ten analyzed samples. GC-MS data, processed by ChromaTOF software (v4.43, Leco Co., CA, USA), was normalized to QC samples and then to total peak area of each subject. Metabolites were identified by our in-house standard library or annotated by the NIST library.

For the LC-TOFMS, the samples were mixed with 10 µL of p-chlorophenylalanine (10 µg/mL) as internal standard, and each 10 µL aliquot sample was injected onto a Agilent ZORBAX Eclipse XDB-C18 chromatography column (150 × 4.6

mm, 5 μm). The samples were analyzed by an Agilent HPLC 1200 system coupled with

an Agilent model 6220 MSD TOF mass spectrometer (Agilent Corporation, Santa Clara,

CA, USA). The HPLC was performed at a flow rate of 0.4 mL/min with mobile phase A

(water with 0.1% formic acid) and mobile phase B (acetonitrile with 0.1% formic acid).

The gradient in both positive and negative mode is isocratic at 1% B (0–0.5 min), linear

gradient from 1% to 20% B (0.5-9.0 min), 20-75% B (9.0-15.0 min), 75-100% B (15.0-

18.0 min), isocratic at 100% B (18–19.5 min); linear gradient from 100% to 1% B (19.5-

20.0 min) and isocratic at 1% B (20.0–25.0 min). The TOF mass spectrometer was

operated with the following optimized conditions: (1) ES+ mode, capillary voltage 3500

V, nebulizer 45 psig, drying gas temperature 325 °C, drying gas flow 11 L/min, and (2)

ES- mode, similar conditions as ES+ mode except the capillary voltage is adjusted to

3000 V. During metabolite profiling experiments, both plot and centroid data were

acquired for each sample from 50 to 1,000 Da over a 25 min analysis time. LC-MS data,

processed by the MassHunter Qualitative Analysis Program (vB.05.00, Agilent) and

XCMS package (v1.24.1), was normalized to QC samples and then to total peak area of

each subject. Metabolites were identified by our in-house standard library or annotated by

the HMDB library.

Data analysis

All annotated variables from GC-TOFMS and LC-TOFMS were combined into

one excel file. The combined data sheet were imported into SIMCA-P software 12.0

(Umetrics, Umeå, Sweden) for multivariate statistical analysis. Principal component

analysis (PCA), projection to latent structures-discriminant analysis (PLS-DA) and

orthogonal projection to latent structures-discriminant analysis (OPLS-DA) were carried out to discriminate between breast cancer patients and healthy controls. Based on a variable importance in the projection (VIP) threshold of 1 from the 7-fold cross-validated OPLS-DA model, differential metabolites were identified. In parallel, the metabolites identified by the OPLS-DA model were validated at a univariate level using the nonparametric Mann-Whitney test with the critical p-value set to 0.05 for the plasma samples and using the paired t test with the critical p-value set to 0.05 for the paired tissue samples. The corresponding fold change shows how these selected differential metabolites varied between the cancer and healthy control groups. The quantification of metabolites in serum was performed based on six-point calibration curves. The concentrations of metabolites were expressed in μg/mL serum.

Using the differential metabolites identified from GC-TOFMS and LC-TOFMS analysis, Receiver Operating Characteristic (ROC) curve analysis was conducted by SPSS 18.0 (SPSS Inc.). The optimal cut point was determined for each biomarker by searching for those that yielded the maximum corresponding sensitivity and specificity. ROC curves were then plotted on the basis of the set of optimal sensitivity and specificity values. Area under the curve (AUC) was computed via numerical integration of the ROC curves. The biomarker with the largest AUC value was identified as having the strongest predictive power for detecting breast cancer.

Results

Plasma metabolic profiling of breast cancer patients

In the plasma samples, 812 features were detected using GC-MS and 1662 features were detected using LC-MS, including 1023 features from ES+ and 639 features from ES-. A total of 227 metabolites were identified in either GC-MS (128 metabolites) or LC-MS (99 metabolites), including amino acids, amines, organic acids, carbohydrates, fatty acids and nucleic acids (Appendix A). A total of 102 out of 227 metabolites (44.9%, 69 metabolites from GC-MS and 33 metabolites from LC-MS) were confirmed by standard compounds, while others were annotated by the available libraries (either NIST or HMDB).

In the training set, separation tendency could be observed between breast cancer patients and healthy controls in the PCA score plot using 6 components (R^2X_{cum}=0.407, Q^2_{cum}=0.115) (Figure 2). The 7-fold cross-validated PLS-DA model was obtained with 2 components showing the difference between breast cancer samples and healthy controls (R^2X_{cum}=0.14, R^2Y_{cum}= 0.861, Q^2_{cum}=0.697) (Figure 3). Validation model using the 999 random permutation tests demonstrated the robustness of the PLS-DA model, with the Q^2Y-intercept of -0.269 (Figure 4). Clear separation could be observed in the 7-fold cross-validated OPLS-DA model using one predictive component and one orthogonal components (R^2X_{cum}=0.14, R^2Y_{cum}= 0.861, Q^2_{cum}=0.717) (Figure 5).

In order to validate the predictive ability, this OPLS-DA model was applied to an external test set consisting of plasma samples for 103 breast cancer patients and 41 healthy controls, which were not matched in age and excluded in the model building

process. In the training set, the samples with t-score > 0 were assigned to the breast cancer group and the samples with t-score < 0 were assigned to the healthy control group. T-predicted scatter plot shows the predicted t-score (tPS) for the test set (Figure 6). With the cutoff of 0, a total of 100 out of 103 patients and 36 out of 41 healthy controls were correctly predicted, with a sensitivity of 97.1% and specificity of 87.8%. This suggests that the constructed OPLS-DA model provides great predictive ability between breast cancer and healthy control groups.
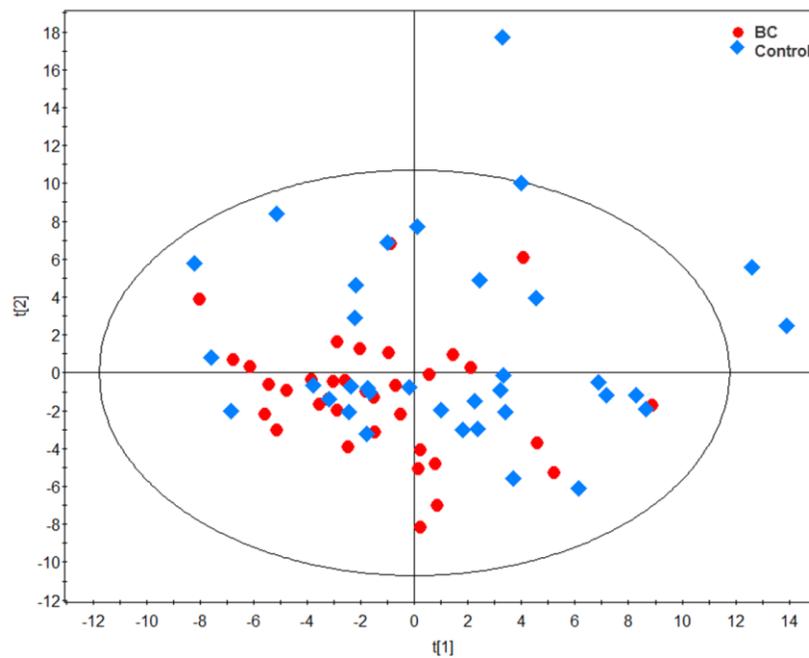
Figure 2. PCA scores plot for the training set. In the training set, the PCA scores plot using 6 components (R^2X_{cum}=0.407, Q^2_{cum}=0.115) was constructed between breast cancer (BC) patients (n=35) and healthy controls (n=35). Separation tendency could be observed between breast cancer patients and healthy controls.
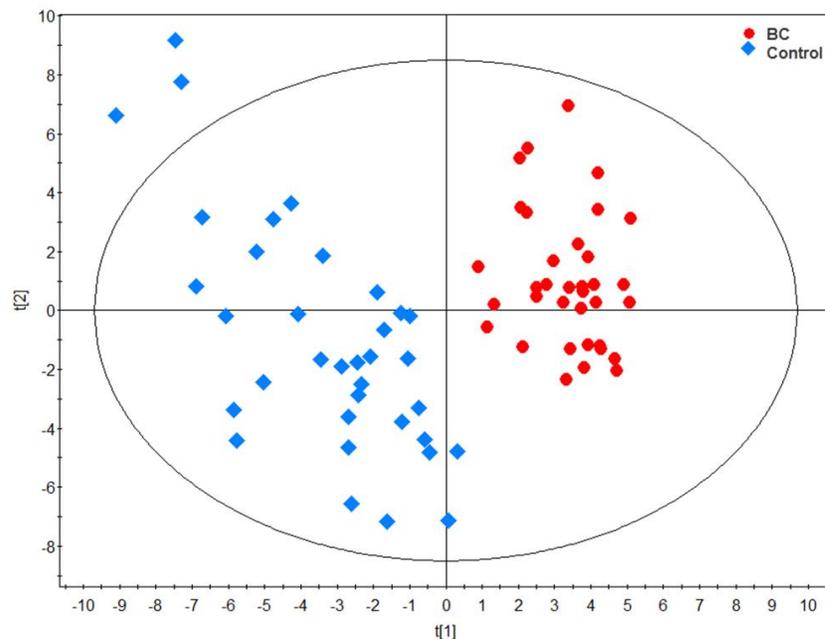
Figure 3. PLS scores plot for the training set. In the training set, The 7-fold cross-validated PLS-DA model (R^2X_{cum}=0.14, R^2Y_{cum}= 0.861, Q^2_{cum}=0.697) was obtained with 2 components showing the difference between breast cancer (BC) patients (n=35) and healthy controls (n=35).
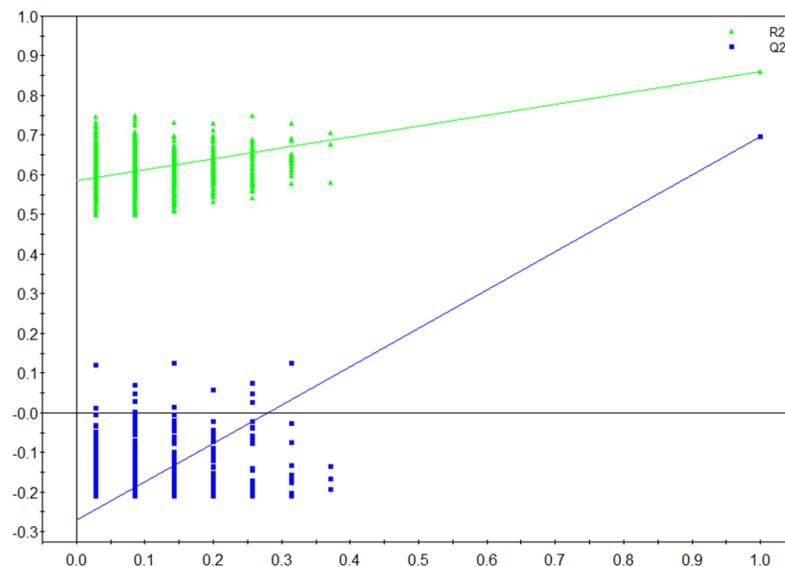
Figure 4. Validation model. The validation model was constructed using the 999 random permutation tests with 2 components, with R^2Y-intercept of 0.583 and Q^2Y-intercept of -0.269.
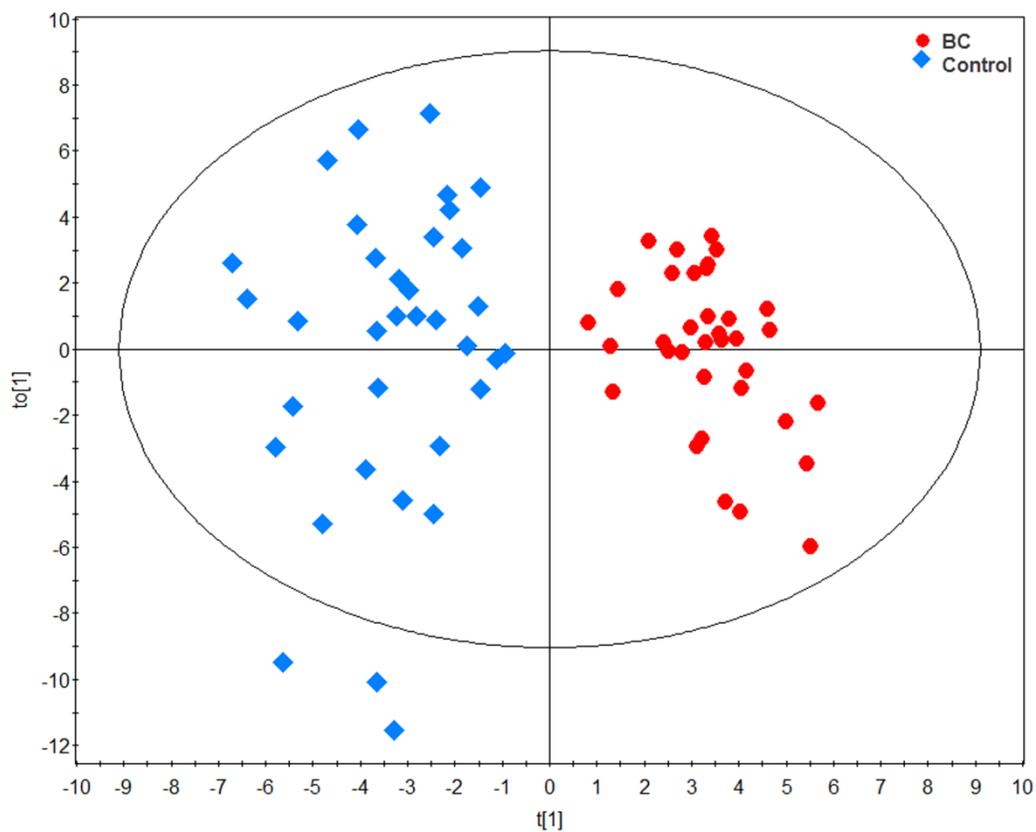
Figure 5. OPLS-DA scores plot for the training set. In the training set, the 7-fold cross-validated OPLS-DA model using one predictive component and one orthogonal components (R^2X_{cum}=0.14, R^2Y_{cum}= 0.861, Q^2_{cum}=0.717) was constructed and clear separation could be observed between breast cancer (BC) patients (n=35) and healthy controls (n=35).
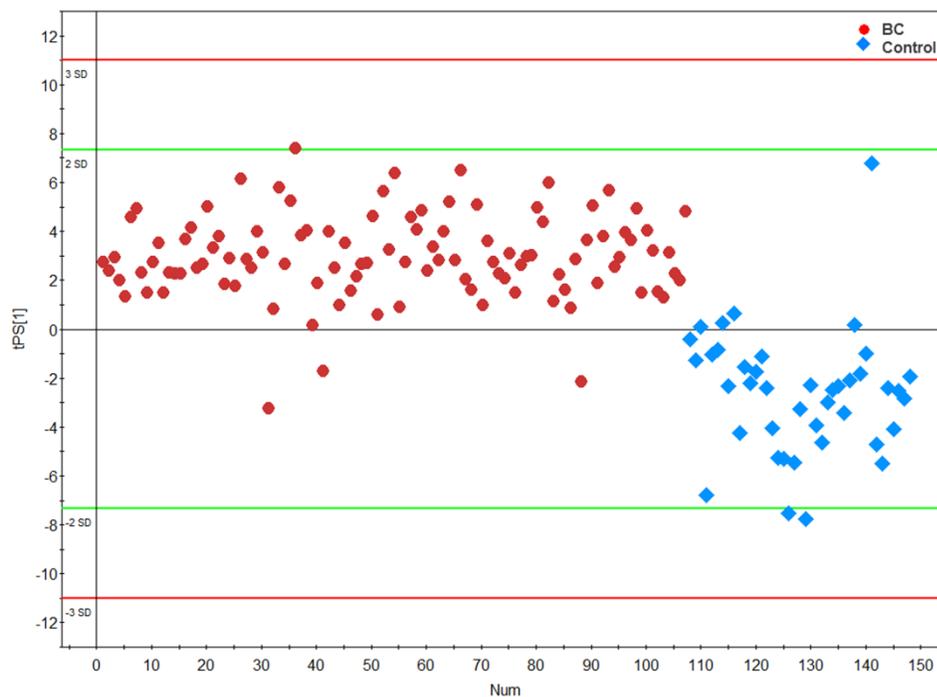
Figure 6. T-predicted scatter plot for the test set. The test set containing 103 breast cancer (BC) patients and 41 healthy controls was predicted using the OPLS-DA model constructed with the training set. In the training set, breast cancer patients and healthy controls were assigned to t-score > 0 and < 0, respectively. T-predicted scatter plot shows the predicted t-score (tPS) for the test set. A total of 100 out of 103 patients and 36 out of 41 healthy controls were correctly predicted, with a sensitivity of 97.1% and specificity of 87.8%.

Potential biomarker identification for breast cancer

Variable importance in the projection (VIP) value, which was obtained from OPLS-DA model, describes the contribution of a variable to the model. Using the VIP values (VIP > 1) and p-values from Mann-Whitney test (p-value < 0.05), a total of 41 metabolites were selected as differential variables between breast cancer patients and controls (Table 2). To visualize the variations of those differential metabolites in all the cancer samples, the heatmap was carried out with the ratio of the individual value of each metabolite to the mean value of healthy controls (Figure 7). To further interpret the

38

significant differences in plasma metabolites between breast cancer patients and healthy

controls, the metabolites were linked to metabolic pathways using the Kyoto

Encyclopedia of Genes and Genomes (KEGG) database. Several pathways were found

dysregulated in breast cancer patients, including the TCA cycle, γ-glutamyl cycle, amino

acid metabolism, lipid metabolism and nucleotide metabolism. Of these, the amino acid

metabolism pathway had the most differential metabolites and showed a profound change.

The receiver operating characteristics (ROC) analysis was used to select potential

biomarkers for breast cancer based on the predictive performance of the 41 differential

metabolites. When analyzing each individual metabolite, aspartate, whose plasma levels

were significantly decreased in breast cancer patients (FC = 0.34, p-value = 6.27E-13),

obtained the best predictive performance with an area under the ROC curves (AUC) of

1.000, a sensitivity of 100% and a specificity of 100% in the training set. In the test set,

aspartate did not show a perfect predictive performance, only obtaining an AUC of 0.935

(95% confidence interval: 0.884-0.987), a sensitivity of 85.4% and a specificity of 95.1%.

However, in the test set, the plasma levels of aspartate were still significantly lower in

breast cancer patients than those in the normal controls (FC = 0.47, p-value = 3.99E-16).

Table 2. A list of 41 differential metabolites between breast cancer patients and healthy controls. The 41 metabolites selected as differential variables between breast cancer patients and healthy controls. Variable importance in the projection (VIP) was obtained from OPLS-DA with a threshold of 1.0; p-value was calculated from the Mann-Whitney test; Fold change (FC) with a value more than 1 indicates a relatively higher concentration in the breast cancer samples, while a value less than 1 means a relatively lower concentration compared to healthy controls.

Pathway	Name	Training Set			Test Set	
		VIP	FC	p	FC	p
Amino Acid Metabolism	Asparagine	3.66	2.78	6.10E-12	2.92	2.79E-14
	Hypotaurine	3.41	2.51	3.17E-10	2.71	4.48E-13
	5-Oxoproline	3.33	0.36	6.62E-12	0.49	4.50E-14
	Cysteine	3.14	2.42	1.40E-09	2.57	3.48E-12
	Aspartate	3.13	0.34	6.27E-13	0.47	3.99E-16
	Glutamate	2.93	0.35	2.25E-12	0.47	1.32E-15
	Glutamine	2.57	2.07	3.77E-07	2.48	1.28E-11
	Lysine	2.22	1.80	3.13E-05	2.37	7.10E-11
	Cystine	2.17	1.76	5.47E-05	2.94	2.05E-14
	Isoleucine	1.90	0.59	1.92E-04	0.71	5.78E-04
	N-acetyl-glutamine	1.48	0.62	5.91E-04	0.78	1.20E-02
	5-Hydroxy-tryptophan	1.30	1.36	2.76E-02	1.09	4.24E-01
	Aminoacetone	1.26	0.72	1.91E-02	0.84	8.96E-02
	3-Amino-2-piperidone	1.25	0.73	2.52E-02	0.73	1.28E-03
	Homocysteic acid	1.23	1.42	1.18E-02	1.08	4.77E-01
	Delta-hydroxylysine	1.18	0.72	1.97E-02	0.81	4.38E-02
Lipid Metabolism	Glycerolphosphate	3.12	0.34	9.63E-13	0.44	1.39E-18
	Glycerophosphocholine	3.06	0.38	5.81E-11	0.49	1.79E-14
	Arachidonic acid	2.33	0.43	4.43E-09	0.51	7.79E-13
	Nicotinuric acid	1.88	0.69	7.02E-03	0.74	3.31E-03
	Propionylcarnitine	1.61	1.44	8.66E-03	1.58	1.21E-04
	Butyrylcarnitine	1.25	1.38	1.94E-02	1.71	9.20E-06
	Choline	1.15	0.73	2.30E-02	0.81	4.19E-02
	Myristoleic acid	1.11	0.76	4.40E-02	0.75	3.36E-03
	Octadecanoic acid	1.09	0.68	5.28E-03	0.70	2.76E-04
Carbohydrate Metabolism	Glyoxylic acid	1.93	2.03	7.38E-07	2.07	1.11E-08
	Lactate	1.92	0.69	7.02E-03	0.66	2.43E-05
	α-ketoglutarate	1.79	0.50	1.34E-06	0.80	2.41E-02
	Pyruvate	1.65	1.62	5.66E-04	2.16	2.44E-09
	Malate	1.59	0.71	1.18E-02	0.71	6.91E-04
	Oxaloacetate	1.58	0.67	3.38E-03	0.81	4.47E-02
	6-Phosphogluconic acid	1.46	0.67	4.23E-03	0.93	5.17E-01
	4-Hydroxy-2-oxoglutaric acid	1.26	0.67	3.93E-03	0.80	2.64E-02
	Hydroxyacetic acid	1.19	1.40	1.58E-02	1.56	1.79E-04
	Succinate	1.15	0.63	1.07E-03	0.65	8.19E-06
Nucleotide Metabolism	Pentosidine	1.89	0.45	2.58E-08	0.48	4.99E-15
	Pseudo uridine	1.38	0.74	3.21E-02	0.94	5.66E-01
	Uracil	1.29	1.47	5.28E-03	1.42	2.55E-03
other	Quinic acid	2.22	0.51	1.91E-06	0.92	4.14E-01
	N-Acetylneuraminic acid	1.41	1.83	1.86E-05	1.50	4.98E-04
	Epinephrine glucuronide	1.01	0.75	3.90E-02	0.88	2.08E-01
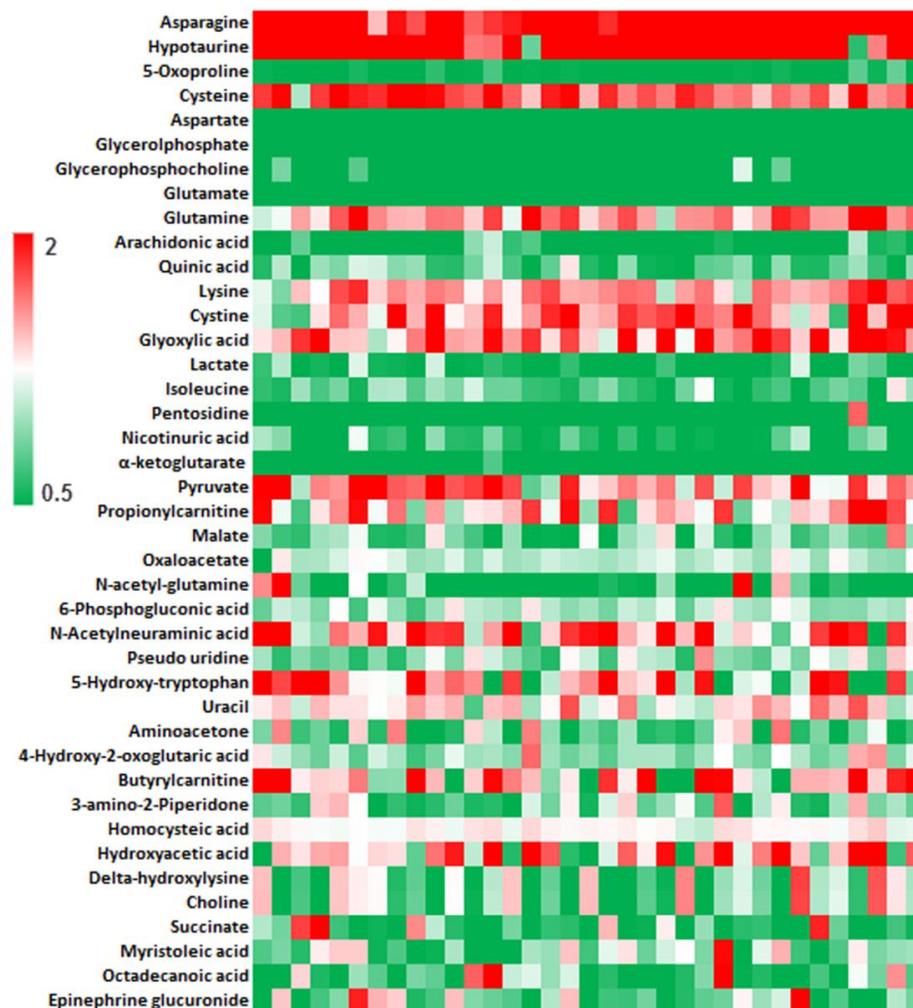
Figure 7. Heatmap for 41 differential metabolites.

Since the combination of several biomarkers may provide more information than the individual one, we tried to construct a model containing multiple markers. Of ten most differential metabolites with the highest VIP values, seven were related to amino acid metabolism. Therefore, these seven plasma metabolites (asparagine, hypotaurine, 5-oxoproline, cysteine, aspartate, glutamate and glutamine) were selected to construct the combination model. This combination model obtained an AUC of 1.000 with a sensitivity

of 100% and a specificity of 100% in the training set (Figure 8A) and an AUC of 0.956

(95% confidence interval: 0.902-1.000) with a sensitivity of 98.1% and a specificity of

90.2% in the test set (Figure 8B). The boxplots of these seven metabolites in the training

set is shown in Figure 8C- I.
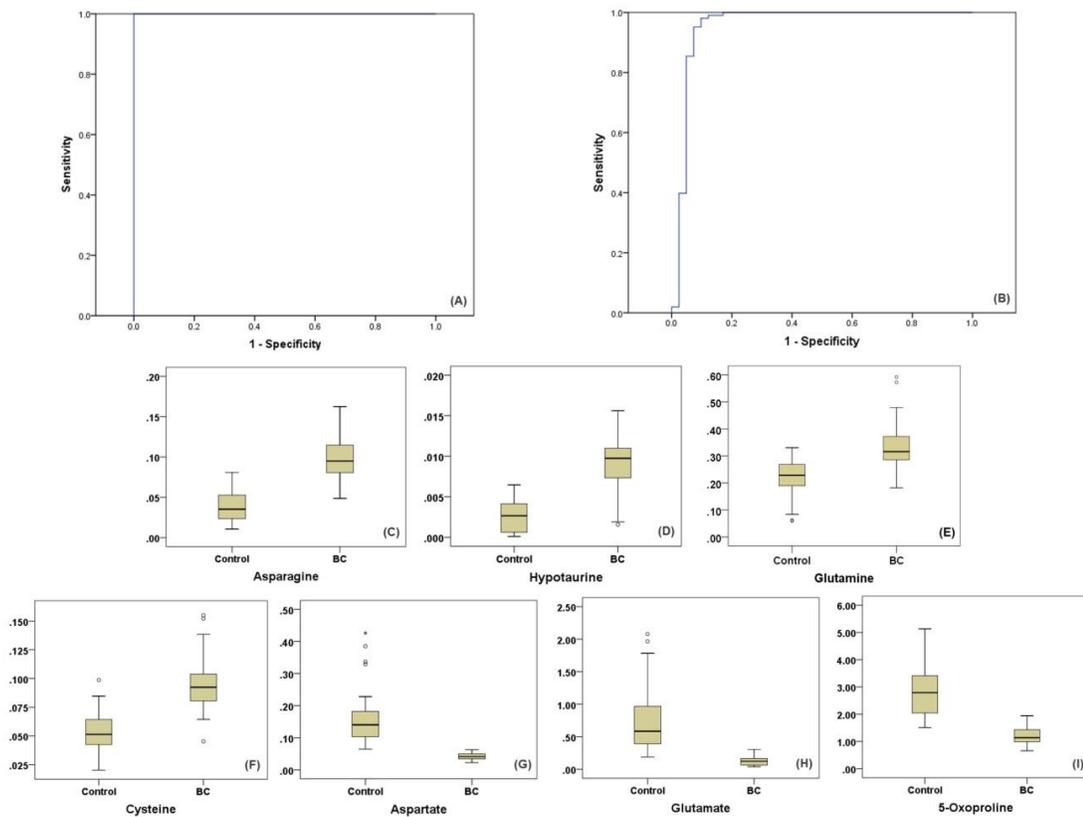
Figure 8. Analysis of seven combined potential biomarkers. A ROC curve analysis for predictive power of seven combined potential breast cancer biomarkers. (A) The calculated AUC in the training set was 1.000, with a sensitivity of 100% and a specificity of 100%. (B) The AUC in the test set was 0.956 (95% confidence interval: 0.903-1.000), with a sensitivity of 98.1% and a specificity of 90.2% in the test set. (C-I) Box plots of seven potential biomarkers distinguishing breast cancer (BC) from healthy controls in the training set.

Comparison of metabolic profiles among different stages of breast cancer

The plasma metabolic profiles of breast cancer patients in different stages, to my knowledge, have not been investigated. Therefore, multivariate statistics based on 227 annotated metabolites was performed in breast cancer patients in various stages, including 19 patients in stage I, 50 patients in stage II, 49 patients in stage III, and 20 patients in stage IV. An OPLS-DA model was established with different stages of cancer (R^2X_{cum}=0.0796, R^2Y_{cum}= 0.577, Q^2_{cum}=0.138). It showed that the metabolic profiles of stage II and stage III were merged together, and they couldn't be separated from either stage I or stage IV (Figure 9). However, separation tendency could be observed clearly between the metabolic profiles of stage I and stage IV in the constructed OPLS-DA model (R^2X_{cum}=0.193, R^2Y_{cum}= 0.778, Q^2_{cum}=0.309; Figure 10).

The significantly differential metabolites, selected using the VIP values (VIP > 1) and *p*-values from Mann-Whitney test (*p*-value < 0.05), have been listed in the Table 3. Interestingly, aspartate, one of the most differential metabolites in breast cancer patients, also had significantly different levels between stage I and stage IV patients. The plasma levels of aspartate in stage IV patients were 1.86 times higher than that in stage I patients (*p*-value = 0.001), 1.57 times higher than that in stage II patients (*p*-value = 0.001), and 1.58 times higher than that in stage III patients (*p*-value = 0.001).
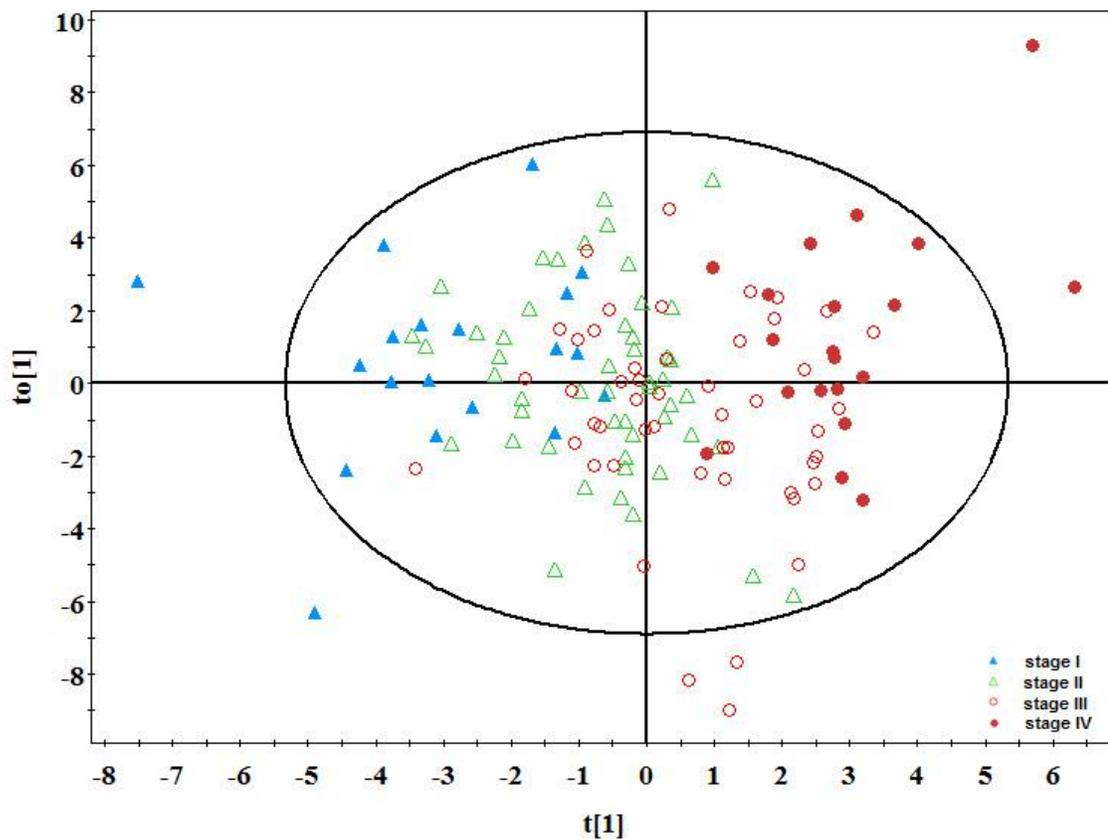
Figure 9. Analysis of metabolic profiles among four stages of breast cancer. An OPLS-DA model (R^2X_{cum}=0.0796, R^2Y_{cum}= 0.577, Q^2_{cum}=0.138) was established with breast cancer patients in stage I (n=19, blue triangle), stage II (n=50, green open triangle), stage III (n=49, red circle) and stage IV (n=20, red dot). It showed that the metabolic profiles of stage II and stage III were merged together, and they could not be separated from either stage I or stage IV.
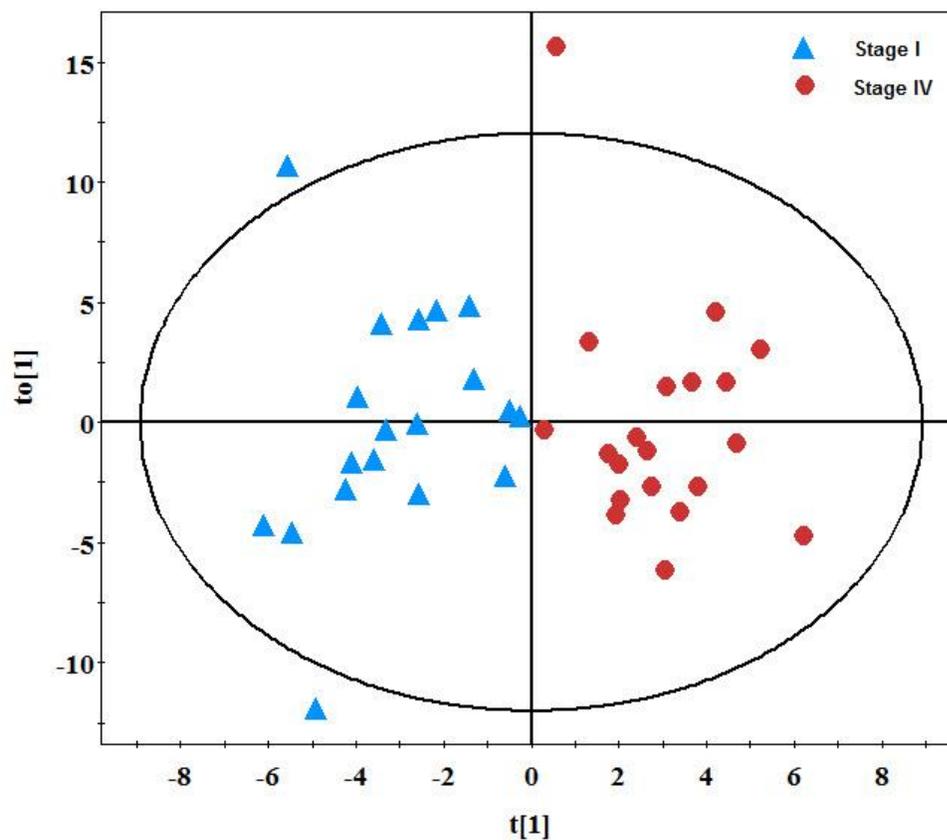
Figure 10. OPLS-DA scores plot established between the breast cancer patients in stage I and stage IV. Separation tendency could be observed clearly between the metabolic profiles of stage I and stage IV in the constructed OPLS-DA model (R^2X_{cum}=0.193, R^2Y_{cum}= 0.778, Q^2_{cum}=0.309).

Table 3. A list of differential metabolites between stage I and stage IV patients.
Variable importance in the projection (VIP) was obtained from OPLS-DA with a threshold of 1.0; Fold change (FC) with a value more than 1 indicates a relatively higher concentration in stage IV patients, while a value less than 1 means a relatively lower concentration compared to stage I patients; p-value was calculated from the Mann-Whitney test (* means p-value < 0.05, ** means p-value < 0.01, and *** means p-value < 0.001).

	Compound name	FC	VIP
Increased in the stage IV	Aspartate**	1.86	1.93
	3-hydroxyoxyisovaleric acid**	1.85	1.25
	Glutamate**	1.72	2.12
	L-Homoserine**	1.62	1.85
	Pyruvate*	1.48	1.56
	α-ketoglutarate*	1.46	1.62
	Lactate*	1.45	1.81
	Nicotinuric acid*	1.45	1.78
Decreased in the stage IV	Myristoleic acid***	0.49	2.92
	3-amino-2-Piperidone***	0.52	2.61
	Cholesterol**	0.55	2.38
	á-Amino isobutyric acid**	0.58	2.32
	Phenylglyoxylic acid**	0.6	2.22
	Hydroxycarbamic acid**	0.6	1.87
	3,6-Dioxa-2,7-disilaoctane, 2,2,4,7,7-pentamethyl-**	0.61	1.61
	3, 4 dehydroproline**	0.61	2.17
	6-Phosphogluconic acid**	0.62	1.47
	Anthranilic acid**	0.62	2.15
	4-Deoxypyridoxine**	0.62	2.16
	Guanidineacetic acid*	0.63	1.62
	4-Aminohippuric acid*	0.63	1.63
	6-Dehydrotestosterone glucuronide*	0.64	1.33
	4-Hydroxy-2-oxoglutaric acid*	0.64	1.33
	3-Phosphoglyceric acid, TMS*	0.65	1.56
	4,8 dimethylnonanoyl carnitine*	0.65	1.12
	Oxalic acid*	0.67	1.91
	3-Succinoylpyridine*	0.67	1.62
	d-Galactose*	0.68	1.98
	Decanoylcarnitine*	0.68	1.16
	N-acetyl-glycine*	0.68	1.27
	L-Tryptophan*	0.69	1.24
	Hydroxyacetic acid*	0.7	1.77

Validation in serum samples

Because aspartate was one of the most dysregulated metabolites and offered the best predictive performance in the plasma samples, quantification of aspartate was performed by GC-TOFMS and the diagnosis ability was evaluated in the serum samples collected from 80 breast cancer patients (aged from 36 to 78, with an average age of 49.5) and 70 healthy volunteers (aged from 35 to 76, with an average age of 58.3) at Ruijin Hospital in Shanghai, China.

Using a six-point calibration curve, the concentrations of aspartate were determined. Breast cancer patients had an average concentration of 1.06 μg/mL, whereas healthy volunteers had an average concentration of 2.79 μg/mL. The serum level of aspartate in breast cancer patients was 0.38 times lower than the level in healthy volunteers (p value = 7.25E-24, Figure 11A). ROC analysis showed that it obtained an AUC of 0.978 (95% confidence interval: 0.949-1.000) with a sensitivity of 97.1% and a specificity of 97.5% (Figure 11B). These results confirmed the altered expression of aspartate in breast cancer.

(A)   **(B)**
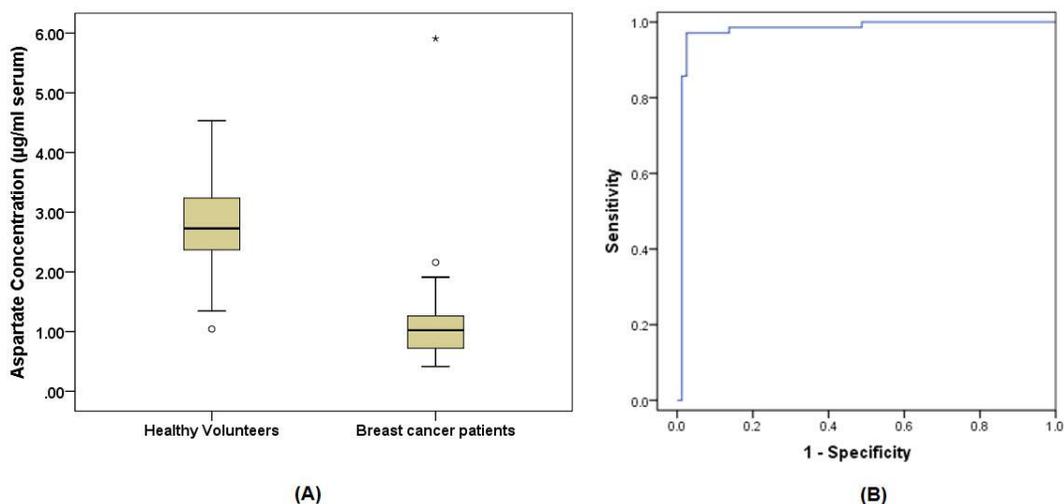
Figure 11. Quantitative measure and diagnosis ability analysis of aspartate in the serum samples. (A) Box plot of aspartate concentrations in the serum samples of breast cancer patients and healthy volunteers. (B) A ROC curve analysis for predictive power of aspartate. The AUC was 0.978 (95% confidence interval: 0.949-1.000), with a sensitivity of 97.1% and a specificity of 97.5%.

Profiling analysis and differential metabolite identification in tissue samples

In this current study, 20 pairs of breast tumor tissue and its adjacent normal tissue were also analyzed by GC-TOFMS and LC-TOFMS. A total of 206 metabolites were identified in either GC-TOFMS (143 metabolites) or LC-TOFMS (63 metabolites). A total of 155 out of 206 metabolites (122 from GC-TOFMS, and 33 from LC-TOFMS) were confirmed by reference standards, while others were annotated by the available libraries (either NIST or HMDB). Multivariate statistical models were established using these 206 annotated metabolites. The PCA score plot using 6 components (R^2X_{cum}=0.665, Q^2_{cum}=0.406) showed that tumor tissue could be separated from normal controls (Figure 12A), and a clear separation tendency between these two groups was obtained in the OPLS-DA model controls (R^2X_{cum}=0.476, R^2Y_{cum}= 0.914, Q^2_{cum}=0.725; Figure 12B).
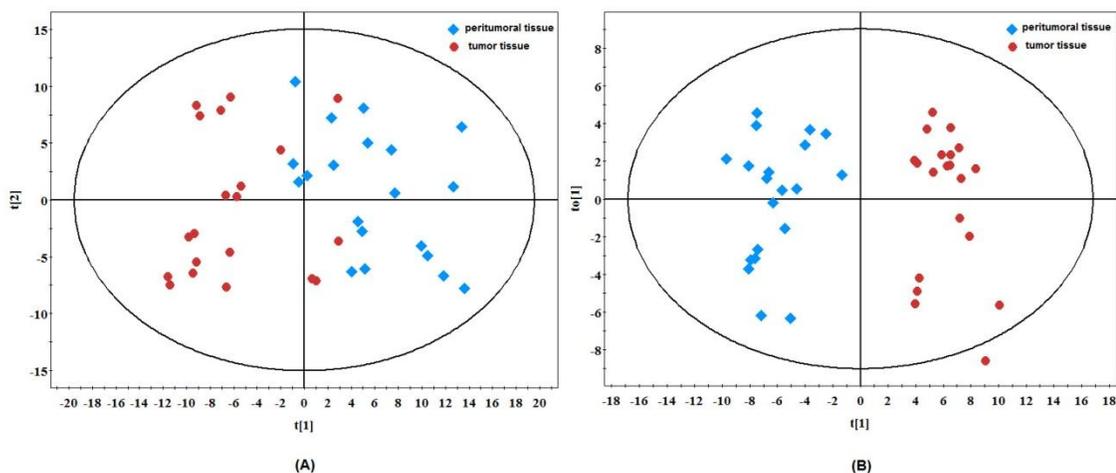
Figure 12. Multivariate statistical analysis of tissue samples. (A) PCA scores plot (R^2X_{cum}=0.665, Q^2_{cum}=0.406) was constructed between the breast tumors (n=20) and the peritumoral tissues (n=20). (B) A clear separation tendency between breast tumors and the peritumoral tissues was obtained in the OPLS-DA model (R^2X_{cum}=0.476, R^2Y_{cum}= 0.914, Q^2_{cum}=0.725).

Using the criteria of VIP values (VIP > 1) and *p*-values from paired t test (*p*-value < 0.05), 83 metabolites were found significantly altered and they provided great contribution to discriminate tumor samples from normal controls in the OPLS-DA model. These 83 differential metabolites, which have been listed in the Appendix B, mainly contained amino acids, free fatty acids, carbohydrates and nucleosides.

Among these 83 differential metabolites, we observed that aspartate and a couple of metabolites involved aspartate metabolism were significantly altered in breast tumor tissue. Aspartate and asparagine were elevated by 1.92 and 2.18 fold, respectively, in the tumor tissues compared to the adjacent normal tissues (Figure 13). Seven nucleosides or nucleobases, including uridine, uracil, guanosine, orotidine, dihydrouracil, hypoxanthine, and 8-hydroxy-deoxyguanosine were increased in breast tumor as well, while the level of cytosine was decreased in the tumor tissues.
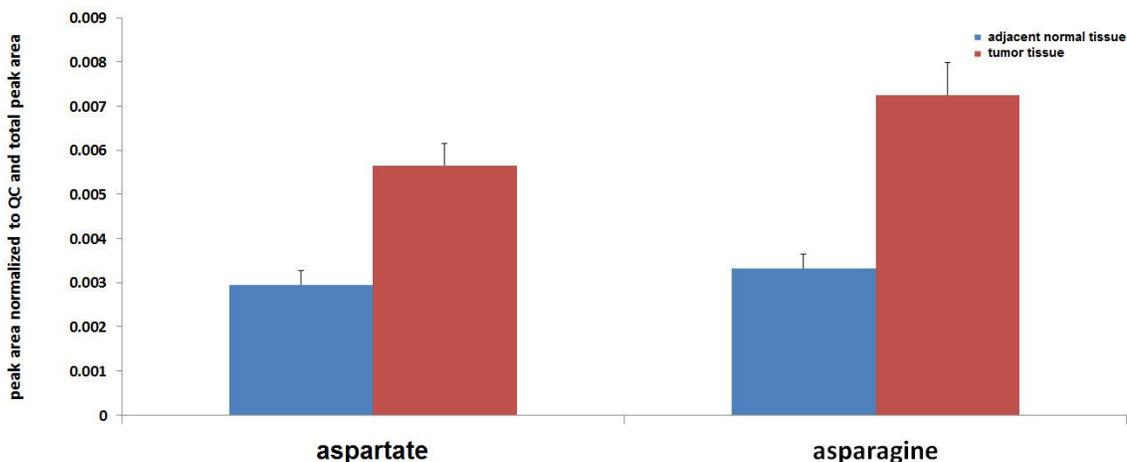
Figure 13. Tumor levels of aspartate and asparagine. Aspartate and asparagine were elevated by 1.92 and 2.18 fold, respectively, in the tumor tissues compared to the adjacent normal tissues. Data are presented as mean ± SEM of twenty samples.

Discussion

Potential biomarkers for breast cancer

Although a number of potential biomarkers have been found in the urine and tissue samples of breast cancer patients, metabolic patterns of breast cancer have never been analyzed in a large cohort of plasma samples. In the present study, we acquired GC-TOFMS and LC-TOFMS spectra of plasma metabolites from 138 breast cancer patients and 76 healthy controls, and we discovered a novel potential biomarker, aspartate, whose expression has never been reported altered in the urine samples in the previous studies. Aspartate obtained the best predictive power among the 41 differential metabolites to distinguish breast cancer from normal controls, with both sensitivity and specificity of 100% in the age-matched training set.

In the non-age-matched test set, aspartate was significantly reduced in the breast cancer patients as well, and offered a lower but still good predictive power, with a

sensitivity of 85.4% and a specificity of 95.1%. The difference in predictive power of the training and test set might be due to a larger sample size of the test set with a greater intra-group variation. Although the test set is non-age-matched, no difference of aspartate level was found between young and old subjects in either breast cancer patients or healthy volunteers (data not shown). But the difference might be caused by the test set containing more stage IV patients, whose aspartate level was significantly different from the levels in the patients at other stages.

The diagnostic ability of aspartate was also validated in an independent set of serum samples and it obtained a sensitivity of 97.1% and a specificity of 97.5%. The plasma samples were collected from American women, while the serum samples were collected from Chinese women. The metabolic profiles among various races of people may be distinct due to different genetic phenotypes, diets and lifestyles. In order to minimize the effect of diet, we collected the plasma and serum samples in the morning before breakfast. Aspartate exhibited good diagnostic abilities in both American and Chinese patients, and this may reflect that the altered expression of aspartate is not affected by race.

The current study showed that the metabolism of amino acids underwent great alterations in breast cancer. Of ten most differential metabolites with the highest VIP values, seven were related to amino acid metabolism. These seven metabolites (asparagine, hypotaurine, 5-oxoproline, cysteine, aspartate, glutamate and glutamine) were thus selected to construct a combination model to distinguish breast cancer patients from normal controls. This model offered good classifying performance with an AUC of

1 in the training set, and better performance in the test set than aspartate alone, providing a sensitivity of 98.1% and a specificity of 90.2%.

Aspartate metabolism

In the training set, aspartate was significantly reduced by 0.34 times in the plasma of breast cancer patients with the smallest p-value of 6.27E-13, and in the test set aspartate was decreased by 0.47 times in breast cancer with a p-value of 3.99E-16. In the independent set of serum samples, the level of aspartate in breast cancer patients was also found 0.38 times lower than the level in healthy volunteers. In order to investigate whether the decrease of aspartate in the blood was due to the metabolic alterations of breast cancer tissues, we analyzed 20 pairs of breast tumor and its adjacent tissue, and we found that aspartate levels in breast cancer tissue were 1.92 times higher than the levels in the adjacent normal tissues.

Aspartate is a non-essential amino acid, synthesized from oxaloacetate which is an immediate of TCA cycle. Oxaloacetate was found decreased in the plasma of breast cancer patients in the current study. In addition to protein biosynthesis, aspartate is involved in many other metabolic pathways as well, including generating asparagine, functioning as a substrate of *de novo* biosynthesis of pyrimidine, participating in urea cycle to facilitate the removal of ammonia, playing a role in translocating NADH (produced from glycolysis) into mitochondria across inner mitochondria membrane for oxidative phosphorylation by aspartate-malate shuttle, and undergoing gluconeogenesis by being converted to alanine.

There is a high demand for nucleotides in dividing cells. In this current study, we observed increased levels of seven nucleotides or nucleobases in breast tumor. The *de novo* pyrimidine biosynthesis activity was found elevated by 4.4 folds in the MCF-7 cells compared to that in the MCF-10A cells (Sigoillot et al., 2004). They also found that *de novo* pyrimidine biosynthesis pathway in the MCF-10A cells was tightly regulated in each growth phase, while this pathway remained elevated in all growth phases in the MCF-7 cells. A large multifunctional protein CAD, which has carbamoyl phosphate synthetase, aspartate transcarbamoylase and dihydroorotase activities, is the rate-limiting enzyme in the *de novo* pyrimidine biosynthesis pathway. The activity of CAD has been found increased in the tumor cells (Calva et al., 1959; Aoki et al., 1982), also revealing an enhanced *de novo* pyrimidine biosynthesis pathway in the cancer cells. It was found that H-Ras oncogene, which is activated in many cancers, increased the conversion of aspartate from ^{13}C-glucose via the transamination of oxaloacetate and the produced aspartate entered pyrimidine nucleotide biosynthesis in the H-Ras transformed cells (Telang et al., 2007). Biosynthesis may not satisfy the high demand for aspartate, so cancer cells may absorb aspartate from the bloodstream, leading to low level of aspartate in the plasma and high level in tumor, to satisfy its high rate of *de novo* pyrimidine biosynthesis.

In addition, asparagine synthetase, which catalyzes the biosynthesis of asparagine from aspartate, was found overexpressed under glucose-deprived condition in the pancreatic cancer cells, and had a protective capability against apoptosis for the cancer cells (Cui et al., 2007). The overexpression of insulin-like growth factor (IGF) 1 and 2

(an essential regulator of breast cancer development) in the MCF-7 breast cancer cells led to an increased expression of asparagine synthetase (Pacher et al., 2007). In the current study, asparagine was elevated in the plasma of breast cancer patients compared to normal controls, and was increased in breast tumor relative to adjacent normal tissue. It is possible as well that the elevated demand for aspartate in breast cancer cells is due to the increased biosynthesis of asparagine.

Other amino acid metabolism

Glutamate performs an essential role in the metabolism of amino acids. It provides nitrogen for amino acid synthesis and collects nitrogen to form urea during the degradation of other amino acids. A decreased plasma level of glutamate may suggest an increased turnover of amino acids. Glutamine is produced from glutamate by glutamine synthetase. Besides glycolysis, glutaminolysis has been suggested as another important energy source in cancer cells. Glutaminolysis is a process that converts glutamine to glutamate and then to α-ketoglutarate, which is a substrate for the TCA cycle. The increased level of plasma glutamine may indicate a higher requirement of glutamine for cancer development and an increased synthesis of glutamine in non-cancerous tissues.

Glutathione plays a role in protecting cells from oxidative damage due to its abilities to form oxidized glutathione by interacting directly with free radicals and to act as a cofactor for antioxidant enzymes such as GSH peroxidases. Glutathione also participates in the phase II detoxification process in which glutathione S-transferases catalyze the conjugation of glutathione with electrophilic metabolites, including many carcinogens (Balendiran et al., 2004). 5-oxoproline is an intermediate of the γ-glutamyl

cycle, which is involved in glutathione biosynthesis and metabolism. In this cycle, γ-glutamyl moiety is transfered from glutathione to other amino acids, which then are converted to 5-oxoproline by the action of γ-glutamylcyclotransferase. 5-oxoproline is converted back to glutamate, and glutathione is synthesized again from glutamate, glycine and cysteine. In the tumor cells, increased fluxes from 3-phosphoglycerate to glycine, and then to glutathione have been found (Richardson et al., 2008). The expression of γ-glutamylcyclotransferase was reported significantly increased in breast cancer tissue in the previous study (Gromov et al., 2010). These findings showed that breast cancer may have an increased glutathione metabolism. Therefore, the abnormal low level of 5-oxoproline in the plasma of breast cancer patients may be due to the higher uptake 5-oxoproline of tumor cells drived by this increased glutathione metabolism.

Although L-cysteine is a precursor of glutathione (which acts as an antioxidant and detoxifying agent as mentioned above), it itself may produce an excess amount of free radicals and hydrogen peroxide, and therefore may increase the oxidative stress of the cells, damage their DNA and lead to cancer development. It was shown that elevated plasma L-cysteine was associated with an increased risk of developing breast cancer in a prospective case-control study (Lin et al., 2010). The current study showed a similar result in that the plasma levels of L-cysteine of breast cancer patients were higher than those of the control group. L-cystine, which is formed by the oxidation of cysteine, was also found increased.

Glycolysis and TCA cycle

The plasma level of four TCA cycle intermediates, 2-Ketoglutarate, succinate, oxaloacetate and malate, were decreased, while the pyruvate level was increased in the breast cancer patients. These findings were consistent with the Warburg effect, which describes the higher conversion of glucose to pyruvate and lactate via the aerobic glycolysis pathway in the cancer cells. Surprisingly, the plasma level of lactate was decreased in our study. It has been shown that certain cancer cell lines can take up lactate by monocarboxylate transporters and utilize lactate. Lactate levels have been demonstrated to vary between tumor samples and within individual tumors. This variation may be caused by different levels of monocarboxylate transporter activities and varied abilities of lactate reutilization (Kennedy et al., 2010).

Lipid Metabolism

Three free fatty acids (arachidonic acid C20:4, octadecanoic acid C18:0 and 9-tetradecenoic acid C14:1) were observed at decreased levels in the breast cancer patients, suggesting an alternative metabolism of fatty acids. C18:0 was reported increased in the serum of breast cancer patients (Lv and Yang, 2012), while we observed an opposite trend in our study. This may be caused by the variations of different sample sets.

Carnitine and its esters have a crucial role in transporting long chain fatty acids into the mitochondria for β-oxidation, and facilitating the excretion of accumulated acyl-CoA out of mitochondria in a reverse mode. In the breast cancer patients, the plasma level of two carnitine esters butyrylcarnitine (C4) and propionylcarnitine (C3) were significantly increased, which may demonstrate the alteration of fatty acid oxidation.

Choline is converted into phosphatidylcholine, an essential phospholipid in the cell membrane bilayer, by the Kennedy pathway. Phosphocholine is an intermediate of this pathway, whose production is catalyzed by choline kinase α. It has been shown that choline kinase α is overexpressed in the breast cancer cells and the production of phosphocholine is increased as well (de Molina, 2004). It has also been reported that choline and choline containing metabolites positively correlated with the rate of breast cancer cell proliferation (Miyake and Parsons, 2011). The decreased plasma level of choline in the breast cancer patients in the present study may suggest the high requirement of choline for tumor growth. Glycerophosphocholine, another major derivative of choline, has been shown reduced in the breast cancer tumor (Smith et al., 1991), which is consistent with what we found in our study.

Conclusion

In the current study, we used the GC-TOFMS and LC-TOFMS analytical platforms to characterize the metabolic profiles in the plasma of breast cancer patients. The constructed OPLS-DA model showed a clear separation of breast cancer patients and healthy controls, and this model was validated by a test set as well. We also found stage I and stage IV patients can be separated using their metabolic profiles.

Among the 41 identified differential metabolites between breast cancer and normal controls, 7 metaboites (asparagine, hypotaurine, 5-oxoproline, cysteine, aspartate, glutamate and glutamine) were selected to establish a combination of biomarkers to

discriminate breast cancer patients from healthy controls. However, these biomarkers need to be validated in a larger size of samples.

We also noticed that aspartate had the most significantly altered expression, and it alone provided a good diagnostic power in the training set of plasma samples. The alteration of aspartate was validated in an independent set of serum samples. The current study also revealed that breast tumor tissues had an elevated level of aspartate, and several metabolites involved in the aspartate metabolism, such as asparagine and nucleosides, were differently expressed in breast tumor as well. Further studies are needed to find out which pathway drives the altered expressions of aspartate in the blood and breast tumor.

CHAPTER IV

EPILOGUE

Metabolic transformation is recently regarded as a hallmark of cancer. Tumor cells reorganize their metabolic phenotypes to support rapid growth and proliferation, and thus exhibit a characteristic metabolic signature. Identification of these metabolic shifts may provide opportunity to discover novel biomarkers for cancer diagnosis.

In this thesis project, global metabolic profiling analysis of breast cancer was performed in the plasma samples collected from 138 breast cancer patients and 76 healthy women. 41 differential metabolites were identified using multivariate and univariate statistical methods. A predictive model containing seven markers was constructed to distinguish breast cancer patients from healthy controls with good diagnostic performance. Aspartate, which was significantly depleted in the plasma of breast cancer patients, provided a good diagnostic power as well. The decreased expression of aspartate and its diagnostic power were validated in an independent set of serum samples. We also observed an increased level of aspartate in breast tumor tissue by metabolic analysis of 20 pairs of breast cancer tissue and its adjacent normal tissue. However, we still do not know whether aspartate depletion in the blood and its accumulation in the tumor are due to the higher uptake of aspartate in cancer cells, and if it is the case, how aspartate is utilized and which metabolic pathway drives cancer cells to absorb more aspartate need to be studied. Therefore, it would be worthwhile to expand

this study in the future by investigating the mechanisms behind these metabolic alterations at the cell culture level.

Metabolomics, especially untargeted profiling analysis, is a high throughput approach and often generates a high volume of data. In order to identify the underlying features and complex relationships of the data, data mining and machine methods are critical in the metabolomics study. In this project, by applying unsupervised (such as PCA) and supervised (such as OPLS-DA) learning methods to interpret our data, we observed a distinct metabolic pattern of breast cancer in both plasma and tissue samples, and discovered numerous differential metabolites contributing much to this characteristic pattern. How to interpret these differential metabolites is very important as well. We categorized these metabolites based on their related metabolic processes, but we only focused on several ones with great diagnostic power. However, there is still abundant information hidden in these data requiring further interpretations using methods such as pathway mapping and metabolic network modeling.

Considerable previous studies have dedicated to the discovery of diagnostic biomarkers in the urine samples using metabolomics methods, and lots of differential metabolites or combinations of differential metabolites have been found with good diagnostic abilities. However, many confounding factors such as age, diet, and race may increase the variations among the sample sets from different studies, resulting that the potential biomarkers with great diagnostic abilities discovered in one study might not obtain consistent performances in another study. In our current study, an age-matched training set was used to minimize the influence of age, and we collected the plasma and

serum samples from the subjects in the morning before breakfast to reduce the variations introduced by diet. We identified aspartate as a potential biomarker for breast cancer in American women and validated its diagnostic power in Chinese women, which may reflect the alteration of aspartate in breast cancer is universal in different races. However, in order to transform pre-clinical studies into therapeutic application, the potential biomarkers found in this study still require to be validated in more independent sample sets with large sample sizes.

Different types of cancer may share some similar metabolic patterns. For instance, the Warburg effect is a universal metabolic signature for cancer. Also, the analysis of free amino acid profiles in the plasma of patients with five types of cancer (lung, gastric, colorectal, breast and prostate cancer) showed decreased ratios of tryptophan and histidine in all of the cancers except prostate cancer (Miyagi et al. 2011). Therefore, the potential biomarkers, which have good predictive power to discriminate patients with one particular type of cancer from healthy controls, may not be able to distinguish them from patients with other types of cancer. Aspartate, which was identified as one of potential biomarkers, has also been found increased in multiple types of tumor, including prostate cancer (Taylor et al., 2007), colorectal cancer (Piotto et al., 2009) and lung cancer (Chen et al., 2011). And thus the alteration of aspartate may be a universal phenomenon in cancer, and may not be a unique biomarker specifically for breast cancer.

However, some previous studies have showed that one particular type of cancer may exhibit unique metabolic shifts as well. Slupsky et al. (2010) performed metabolomics analysis using NMR in the urine samples collected from 48 breast cancer

patients and 50 ovarian cancer patients. They observed a clear separation of metabolic

profiles between breast cancer and ovarian cancer based on the 67 annotated metabolites.

Woo et al. (2008) analyzed urine levels of hormone metabolites and nucleosides among

10 breast cancer patients, 9 ovarian cancer patients, 12 cervical cancer patients and 22

normal controls. They also observed a clear separation of metabolic patterns between

breast cancer and other types of cancer on the PLS-DA model. Therefore, there is still

hope for identifying the unique metabolic biomarker for one particular type of cancer, but

great effort will be needed in the future.

REFERENCES

American Cancer Society. Breast Cancer Facts & Figures 2011-2012. Atlanta: American Cancer Society, Inc; 2011.

American Cancer Society. Cancer Facts & Figures 2012. Atlanta: American Cancer Society; 2012.

An ZL, Chen YH, Zhang RP, et al. Integrated Ionization Approach for RRLC-MS/MS-based Metabonomics: Finding Potential Biomarkers for Lung Cancer. Journal of Proteome Research. 2010;9(8):4071-81.

Aoki T, Morris HP, Weber G. REGULATORY PROPERTIES AND BEHAVIOR OF ACTIVITY OF CARBAMOYL PHOSPHATE SYNTHETASE-II (GLUTAMINE-HYDROLYZING) IN NORMAL AND PROLIFERATING TISSUES. Journal of Biological Chemistry. 1982;257(1):432-8

Asiago VM, Alvarado LZ, Shanaiah N, et al. Early Detection of Recurrent Breast Cancer Using Metabolite Profiling. Cancer Research. 2010;70(21):8309-18.

Balendiran GK, Dabur R, Fraser D. The role of glutathione in cancer. Cell Biochemistry and Function. 2004;22(6):343-52.

Bao YQ, Zhao T, Wang XY, et al. Metabonomic Variations in the Drug-Treated Type 2 Diabetes Mellitus Patients and Healthy Volunteers. Journal of Proteome Research. 2009;8(4):1623-30.

Borgan E, Sitter B, Lingjaerde OC, et al. Merging transcriptomics and metabolomics - advances in breast cancer profiling. Bmc Cancer. 2010;10:628.

Brewster AM, Hortobagyi GN, Broglio KR, et al. Residual risk of breast cancer recurrence 5 years after adjuvant therapy. Journal of the National Cancer Institute. 2008;100(16):1179-83.

Brockmoller SF, Bucher E, Muller BM, et al. Integration of Metabolomics and Expression of Glycerol-3-phosphate Acyltransferase (GPAM) in Breast Cancer-Link to Patient Survival, Hormone Receptor Status, and Metabolic Profiling. Journal of Proteome Research. 2012;11(2):850-60.

Budczies J, Denkert C, Muller BM, et al. Remodeling of central metabolism in invasive breast cancer compared to normal breast tissue - a GC-TOFMS based metabolomics study. Bmc Genomics. 2012;13:334.

Bullinger D, Froehlich H, Klaus F, et al. Bioinformatical evaluation of modified nucleosides as biomedical markers in diagnosis of breast cancer. Analytica Chimica Acta. 2008;618(1): 29-34.

Bullinger D, Neubauer H, Fehm T, Laufer S, Gleiter CH, Kammerer B. Metabolic signature of breast cancer cell line MCF-7: profiling of modified nucleosides via LC-IT MS coupling. BMC biochemistry. 2007;8:25.

Cai Z, Zhao JS, Li JJ, et al. A Combined Proteomics and Metabolomics Profiling of Gastric Cardia Cancer Reveals Characteristic Dysregulations in Glucose Metabolism. Molecular & Cellular Proteomics. 2010;9(12):2617-28.

Calva E, Lowenstein JM, Cohen PP. CARBAMYL PHOSPHATE-ASPARTATE TRANSCARBAMYLASE ACTIVITY IN TUMORS. Cancer Research. 1959;19(1):101-3.

Chen J, Zhang XY, Cao R, et al. Serum 27-nor-5 beta-Cholestane-3,7,12,24,25 Pentol Glucuronide Discovered by Metabolomics as Potential Diagnostic Biomarker for Epithelium Ovarian Cancer. Journal of Proteome Research. 2011;10(5):2625-32.

Chen W, Zu Y, Huang Q, et al. Study on Metabonomic Characteristics of Human Lung Cancer Using High Resolution Magic-angle Spinning H-1 NMR Spectroscopy and Multivariate Data Analysis. Magnetic Resonance in Medicine. 2011;66(6):1531-40.

Chen Y, Zhang R, Song Y, et al. RRLC-MS/MS-based metabonomics combined with in-depth analysis of metabolic correlation network: finding potential biomarkers for breast cancer. Analyst. 2009;134(10):2003-11.

Christians U, Klawitter J, Hornberger A, Klawitter J. How Unbiased is Non-Targeted Metabolomics and is Targeted Pathway Screening the Solution? Current Pharmaceutical Biotechnology. 2011;12(7):1053-66.

Cui HY, Darmanin S, Natsuisaka M, et al. Enhanced expression of asparagine synthetase under glucose-deprived conditions protects pancreatic cancer cells from apoptosis induced by glucose deprivation and cisplatin. Cancer Research. 2007;67(7):3345-55.

de Molina AR, Banez-Coronel M, Gutierrez R, et al. Choline kinase activation is a

    critical requirement for the proliferation of primary human mammary epithelial

    cells and breast tumor progression. Cancer Research. 2004;64(18):6732-9.

Fan Y, Murphy TB, Byrne JC, Brennan L, Fitzpatrick JM, Watson RWG. Applying

    Random Forests To Identify Biomarker Panels in Serum 2D-DIGE Data for the

    Detection and Staging of Prostate Cancer. Journal of Proteome Research.

    2011;10(3):1361-73.

Frickenschmidt A, Froehlich H, Bullinger D, et al. Metabonomics in cancer diagnosis:

    mass spectrometry-based profiling of urinary nucleosides from breast cancer

    patients. Biomarkers. 2008;13(4):435-49.

Garcia E, Andrews C, Hua J, et al. Diagnosis of Early Stage Ovarian Cancer by H-1

    NMR Metabonomics of Serum Explored by Use of a Microflow NMR Probe.

    Journal of Proteome Research. 2011;10(4):1765-71.

Gromov P, Gromova I, Friis E, et al. Proteomic Profiling of Mammary Carcinomas

    Identifies C7orf24, a gamma-Glutamyl Cyclotransferase, as a Potential Cancer

    Biomarker. Journal of Proteome Research. 2010;9(8):3941-53.

Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell.

    2011;144(5):646-74.

Heiden MGV, Cantley LC, Thompson CB. Understanding the Warburg Effect: The

    Metabolic Requirements of Cell Proliferation. Science. 2009;324(5930):1029-33.

Hilvo M, Denkert C, Lehtinen L, et al. Novel Theranostic Opportunities Offered by

    Characterization of Altered Membrane Lipid Metabolism in Breast Cancer

    Progression. Cancer Research. 2011;71(9):3236-45.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global Cancer Statistics. Ca-

    a Cancer Journal for Clinicians. 2011;61(2):69-90.

Kennedy KM, Dewhirst MW. Tumor metabolism of lactate: the influence and therapeutic

    potential for MCT and CD147 regulation. Future Oncology. 2010;6(1):127-48.

Keun HC, Sidhu J, Pchejetski D, et al. Serum Molecular Signatures of Weight Change

    during Early Breast Cancer Chemotherapy. Clinical Cancer Research.

    2009;15(21):6716-23.

Kim K, Taylor SL, Ganti S, Guo LN, Osier MV, Weiss RH. Urine Metabolomic Analysis

    Identifies Potential Biomarkers and Pathogenic Pathways in Kidney Cancer.

    Omics-a Journal of Integrative Biology. 2011;15(5):293-303.

Kumar R, Alavi A. Fluorodeoxyglucose-PET in the management of breast cancer.

    Radiologic Clinics of North America. 2004;42(6):1113-22, ix.

Lin J, Lee IM, Song YQ, et al. Plasma Homocysteine and Cysteine and Risk of Breast

    Cancer in Women. Cancer Research. 2010;70(6):2397-405.

Lv WW, Yang TS. Identification of possible biomarkers for breast cancer from free fatty

    acid profiles determined by GC-MS and multivariate statistical analysis. Clinical

    Biochemistry. 2012;45(1-2):127-33.

Miyagi Y, Higashiyama M, Gochi A, et al. Plasma Free Amino Acid Profiling of Five

    Types of Cancer Patients and Its Application for Early Detection. Plos One.

    2011;6(9): e24143.

Miyake T, Parsons SJ. Functional interactions between Choline kinase alpha, epidermal

    growth factor receptor and c-Src in breast cancer cell proliferation. Oncogene.

    2012;31(11):1431-41.

Nam H, Chung BC, Kim Y, Lee K, Lee D. Combining tissue transcriptomics and urine

    metabolomics for breast cancer biomarker identification. Bioinformatics.

    2009;25(23): 3151-7.

Nishiumi S, Shinohara M, Ikeda A, et al. Serum metabolomics as a novel diagnostic

    approach for pancreatic cancer. Metabolomics. 2010;6(4):518-28.

Oakman C, Tenori L, Claudino WM, et al. Identification of a serum-detectable

    metabolomic fingerprint potentially correlated with the presence of

    micrometastatic disease in early breast cancer patients at varying risks of disease

    relapse by traditional prognostic methods. Annals of Oncology. 2011;22(6):1295-

    301.

Pacher M, Seewald MJ, Mikula M, et al. Impact of constitutive IGF1/IGF2 stimulation

    on the transcriptional program of human breast cancer cells. Carcinogenesis.

    2007;28(1):49-59.

Pasikanti KK, Esuvaranathan K, Ho PC, et al. Noninvasive Urinary Metabonomic

    Diagnosis of Human Bladder Cancer. Journal of Proteome Research.

    2010;9(6):2988-95.

Piotto M, Moussallieh FM, Dillmann B, et al. Metabolic characterization of primary

   human colorectal cancers using high resolution magic angle spinning H-1

   magnetic resonance spectroscopy. Metabolomics. 2009;5(3):292-301.

Qiu YP, Cai GX, Su MM, et al. Urinary Metabonomic Study on Colorectal Cancer.

   Journal of Proteome Research. 2010;9(3):1627-34.

Qiu YP, Cai GX, Su MM, et al. Serum Metabolite Profiling of Human Colorectal Cancer

   Using GC-TOFMS and UPLC-QTOFMS. Journal of Proteome Research.

   2009;8(10):4844-50.

Richardson AD, Yang C, Osterman A, Smith JW. Central carbon metabolism in the

   progression of mammary carcinoma. Breast Cancer Research and Treatment.

   2008;110(2):297-307.

Sigoillot FD, Sigoillot SM, Guy HI. Breakdown of the regulatory control of pyrimidine

   biosynthesis in human breast cancer cells. International Journal of Cancer.

   2004;109(4):491-8.

Sitter B, Lundgren S, Bathen TF, Halgunset J, Fjosne HE, Gribbestad IS. Comparison of

   HR MAS MR spectroscopic profiles of breast cancer tissue with clinical

   parameters. Nmr in Biomedicine. 2006;19(1):30-40.

Sitter B, Bathen TF, Singstad TE, et al. Quantification of metabolites in breast cancer

   patients with different clinical prognosis using HR MAS MR spectroscopy. Nmr

   in Biomedicine. 2010;23(4):424-31.

Skaane P. Studies Comparing Screen-Film Mammography and Full-Field Digital

    Mammography in Breast Cancer Screening: Updated Review. Acta Radiologica.

    2009;50(1):3-14.

Slupsky CM, Steed H, Wells TH, et al. Urine Metabolite Analysis Offers Potential Early

    Diagnosis of Ovarian and Breast Cancers. Clinical Cancer Research.

    2010;16(23):5835-41.

Smith TAD, Eccles S, Ormerod MG, Tombs AJ, Titley JC, Leach MO. THE

    PHOSPHOCHOLINE AND GLYCEROPHOSPHOCHOLINE CONTENT OF

    AN ESTROGEN-SENSITIVE RAT MAMMARY-TUMOR CORRELATES

    STRONGLY WITH GROWTH-RATE. British Journal of Cancer.

    1991;64(5):821-6.

Sugimoto M, Wong DT, Hirayama A, Soga T, Tomita M. Capillary electrophoresis mass

    spectrometry-based saliva metabolomics identified oral, breast and pancreatic

    cancer-specific profiles. Metabolomics. 2010;6(1):78-95.

Taylor BS, Pal M, Yu J, et al. Humoral response profiling reveals pathways to prostate

    cancer progression. Molecular & Cellular Proteomics. 2008;7(3):600-11.

Telang S, Lane AN, Nelson KK, Arumugam S, Chesney J. The oncoprotein H-Ras(V12)

    increases mitochondrial metabolism. Molecular Cancer. 2007;6:77.

Tenori L, Oakman C, Claudino WM, et al. Exploration of serum metabolomic profiles

    and outcomes in women with metastatic breast cancer: A pilot study. Molecular

    Oncology. 2012;6(4):437-44.

U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2007 Incidence and Mortality Web-based Report. Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute; 2012. Available at: http://www.cdc.gov/uscs.

Ward PS, Thompson CB. Metabolic Reprogramming: A Cancer Hallmark Even Warburg Did Not Anticipate. Cancer Cell. 2012;21(3):297-308.

Wei JE, Xie GX, Zhou ZT, et al. Salivary metabolite signatures of oral cancer and leukoplakia. International Journal of Cancer. 2011;129(9):2207-17.

Weljie AM, Bondareva A, Zang P, Jirik FR. H-1 NMR metabolomics identification of markers of hypoxia-induced metabolic shifts in a breast cancer model system. Journal of Biomolecular Nmr. 2011;49(3-4):185-93.

Wise DR, Thompson CB. Glutamine addiction: a new therapeutic target in cancer. Trends in Biochemical Sciences. 2010;35(8):427-33.

Woo HM, Kim KM, Choi MH, et al. Mass spectrometry based metabolomic approaches in urinary biomarker study of women's cancers. Clinica Chimica Acta. 2009;400(1-2):63-9.

World Health Organization. Ten leading causes of death in females by age and country income group, 2008. WHO; 2012. Available at: http://gamapserver.who.int/gho/interactive_charts/women_and_health/causes_death/chart.html.

LIST OF 227 METABOLITES INDENTIFIED BY EITHER GC-TOFMS

OR LC-TOFMS IN THE PLASMA SAMPLES

Name	Database	Instrument	Mass	RT
1,2-dimethylpropanol	NIST	GC-MS	117	4.44
18-Hydroxycorticosterone	HMDB	LC-MS	362.209	19.71
1H-Indole-3-acetic acid	NIST	GC-MS	202	16.50
1-hydroxy-1-cyclohexen	NIST	GC-MS	155	5.25
1-Pyrroline-2-carboxylic acid	HMDB	LC-MS	113.048	3.75
2,2'-Bipyridine	NIST	GC-MS	156	10.65
2,3-Diaminopropionic acid	Standard	LC-MS	105.065	3.40
2,3-Dihydroxybutanoic acid	NIST	GC-MS	117	9.56
2,3'-Dipyridyl	NIST	GC-MS	156	11.83
2,4-bishydroxybutanoic acid	NIST	GC-MS	103	10.30
2-amino-6-methylaminohexanoic acid	NIST	GC-MS	116	14.29
2-aminobutyric acid	Standard	GC-MS	130	6.96
2-Butenedioic acid	Standard	GC-MS	245	9.34
2-Hydroxycinnamic acid	HMDB	LC-MS	164.047	11.22
2-hydroxypyridine	NIST	GC-MS	152	5.01
2-methyl-butyric acid	NIST	GC-MS	159	4.22
2-oxo-3-methyl-pentanoic acid	Standard	GC-MS	151	7.07
2-oxo-4-methylvaleric acid	Standard	GC-MS	200	7.55
2-Piperidinecarboxylic acid	NIST	GC-MS	156	9.65
3, 4 dehydroproline	NIST	GC-MS	208	15.22
3,4-bishydroxybutanoic acid	NIST	GC-MS	73	10.55
3,6-Dioxa-2,7-disilaoctane, 2,2,4,7,7-pentamethyl-	NIST	GC-MS	117	4.52
3-amino-2-Piperidone	NIST	GC-MS	243	10.85
3-Aminosalicylic acid, TMS	Standard	GC-MS	186	11.64
3-Hydroxybutyric acid	Standard	GC-MS	147	6.77
3-Hydroxydodecanedioic acid	HMDB	LC-MS	246.147	18.92
3-hydroxyoxyisovaleric acid	NIST	GC-MS	131	7.48
3-hydroxypyridine	Standard	GC-MS	152	6.41
3-Indolepropionic acid	HMDB	LC-MS	189.079	20.14
3-methyl-2-oxo-butanoic acid	NIST	GC-MS	202	6.43
3-Phosphoglyceric acid, TMS	NIST	GC-MS	211	19.46
3-Pyridylacetic acid	Standard	LC-MS	138.053	3.86
3-Succinoylpyridine	HMDB	LC-MS	179.058	18.25

4,8 dimethylnonanoyl carnitine	HMDB	LC-MS	329.257	19.66
4-Aminohippuric acid	HMDB	LC-MS	194.069	5.57
4-Deoxypyridoxine	NIST	GC-MS	282	13.67
4-Hydroxy-2-oxoglutaric acid	HMDB	LC-MS	162.016	3.14
4-hydroxy-proline	Standard	LC-MS	132.073	3.88
5-Acetylamino-6-formylamino-3-methyluracil	HMDB	LC-MS	226.07	17.22
5-Hydroxyindoleacetic acid	HMDB	LC-MS	191.058	3.14
5-Hydroxylysine	HMDB	LC-MS	162.1	3.63
5-Hydroxy-tryptophan	NIST	GC-MS	174	21.61
5-Methoxytryptophol	HMDB	LC-MS	191.095	17.93
5-Oxoproline	Standard	GC-MS	156	11.66
5-Phosphoribosylamine	HMDB	LC-MS	229.035	24.78
6-Dehydrotestosterone glucuronide	HMDB	LC-MS	462.225	21.35
6-deoxy-mannose	NIST	GC-MS	204	17.36
6-Phosphogluconic acid	Standard	LC-MS	275.012	3.15
9,12-Octadecadienoic acid (Z,Z)-	Standard	GC-MS	337	19.06
á-Amino isobutyric acid	NIST	GC-MS	211	7.21
à-Aminoadipic acid	NIST	GC-MS	260	12.06
Acetylcarnitine	Standard	LC-MS	204.123	4.14
à-Hydroxyisobutyric acid	Standard	GC-MS	131	6.31
Alanine	Standard	GC-MS	116	5.95
Alloisoleucine	Standard	GC-MS	158	8.70
Aminoacetone	HMDB	LC-MS	73.0528	3.82
Aminomalonic acid	NIST	GC-MS	218	11.05
Androstenedione	HMDB	LC-MS	286.193	18.43
Anthranilic acid	NIST	GC-MS	208	12.67
Arabinofuranose	NIST	GC-MS	217	14.58
Arachidonic acid	Standard	GC-MS	91	20.80
Arginine	Standard	LC-MS	175.116	3.49
Asparagine	Standard	GC-MS	116	13.42
Aspartate	Standard	GC-MS	232	11.66
Benzaldehyde	HMDB	LC-MS	106.042	15.04
Benzocaine	HMDB	LC-MS	165.079	15.00
Benzoic acid	Standard	GC-MS	179	7.98
Beta-alanine	Standard	GC-MS	174	10.46
beta-D-Glucopyranuronic acid	HMDB	LC-MS	314.064	3.79
Bilirubin	HMDB	LC-MS	584.263	19.32
Bisnorcholic acid	HMDB	LC-MS	380.256	22.24
Butyrylcarnitine	HMDB	LC-MS	231.147	16.24
Cadaverine	NIST	GC-MS	174	18.36

Calcidiol	HMDB	LC-MS	400.334	21.84
Caproic acid	NIST	GC-MS	173	5.47
Carnitine	Standard	LC-MS	162.107	3.64
Carnosine	Standard	LC-MS	225.106	3.67
Cholesterol	Standard	GC-MS	129	27.22
Choline	Standard	LC-MS	104.105	3.54
cis-2-Methylaconitate	HMDB	LC-MS	188.032	3.08
cis-3-Hexenyllactate	NIST	GC-MS	83	4.67
Citric acid	Standard	GC-MS	273	15.10
Citrulline	Standard	LC-MS	176.093	3.66
Creatine	Standard	GC-MS	115	12.11
Cyclohexanone	NIST	GC-MS	58	4.61
Cyclohexyloxy	NIST	GC-MS	157	4.64
Cystathionine	Standard	LC-MS	223.072	3.67
Cysteine	Standard	GC-MS	220	12.08
Cystine	Standard	GC-MS	218	20.30
Decanoic acid	Standard	GC-MS	229	10.74
Decanoylcarnitine	HMDB	LC-MS	315.241	19.43
Dehydroascorbic acid	HMDB	LC-MS	174.016	5.48
Delta-hydroxylysine	Standard	LC-MS	163.108	3.63
Deoxycholic acid	Standard	LC-MS	393.291	23.11
D-Fructose	Standard	GC-MS	217	15.75
d-Galactose	Standard	GC-MS	157	15.77
D-Glucuronic acid	NIST	GC-MS	333	16.43
Dihydroxyacetone phosphate	HMDB	LC-MS	169.998	3.15
Dodecanoic acid	Standard	GC-MS	211	13.07
Dodecanoylcarnitine	HMDB	LC-MS	343.272	20.17
D-Ribofuranose	NIST	GC-MS	217	14.73
d-Xylose	NIST	GC-MS	217	13.43
Elaidic acid	Standard	GC-MS	339	19.19
Epinephrine	Standard	LC-MS	184.09	3.60
Epinephrine glucuronide	HMDB	LC-MS	359.122	4.44
Erythrose	NIST	GC-MS	205	14.38
Galactonic acid	Standard	GC-MS	73	17.12
Glucopyranose	NIST	GC-MS	204	16.82
Glucose 6-phosphate	Standard	LC-MS	259.033	3.61
Glutamate	Standard	GC-MS	246	12.82
Glutamine	Standard	LC-MS	147.074	3.57
Gluticol	NIST	GC-MS	73	13.58
Glyceraldehyde	Standard	GC-MS	192	7.53

Glyceraldehyde 3-phosphate	Standard	LC-MS	171.004	3.15
Glyceric acid, TMS	Standard	GC-MS	189	9.23
Glycerol	Standard	GC-MS	218	8.45
Glycerolphosphate	Standard	GC-MS	299	14.50
Glycerophosphocholine	HMDB	LC-MS	257.103	3.58
Glycerylphosphorylethanolamine	HMDB	LC-MS	215.056	18.35
Glycine	Standard	GC-MS	174	8.88
glycol	NIST	GC-MS	147	4.32
Glycolaldehyde	HMDB	LC-MS	60.0211	3.76
Glycylprolylhydroxyproline	HMDB	LC-MS	285.132	22.57
Glyoxylic acid	NIST	GC-MS	160	4.33
Guanidineacetic acid	Standard	LC-MS	118.073	3.63
Guanine	HMDB	LC-MS	151.049	5.00
Heneicosanoic acid	NIST	GC-MS	117	20.52
Hexadecanoic acid	Standard	GC-MS	117	17.19
Hexanoylcarnitine	HMDB	LC-MS	259.178	18.11
Hexanoylglycine	HMDB	LC-MS	173.105	18.45
Homoanserine	HMDB	LC-MS	254.138	23.27
Homocysteic acid	HMDB	LC-MS	183.02	17.74
Hordenine	HMDB	LC-MS	165.115	15.36
Hydroxyacetic acid	Standard	LC-MS	75.0239	4.27
Hydroxycarbamic acid	NIST	GC-MS	221	9.58
Hypotaurine	Standard	GC-MS	188	13.01
Hypoxanthine	Standard	LC-MS	137.046	5.22
Indoleacrylic acid	HMDB	LC-MS	187.063	17.31
Inositol	NIST	GC-MS	318	17.38
Isoleucine	Standard	GC-MS	158	8.65
Isopentenyl pyrophosphate	HMDB	LC-MS	246.006	3.11
Isovalerylcarnitine	HMDB	LC-MS	245.163	17.89
Lactate	Standard	GC-MS	117	5.36
L-Homoserine	NIST	GC-MS	218	13.75
Linoelaidic acid	HMDB	LC-MS	280.24	23.59
Linoleyl carnitine	HMDB	LC-MS	423.335	21.44
L-Leucine	Standard	GC-MS	158	8.39
L-Lysine	Standard	GC-MS	156	16.10
L-methionine	Standard	GC-MS	176	11.62
L-Ornithine	NIST	GC-MS	142	15.04
L-phenylalanyl-L-proline	HMDB	LC-MS	262.132	17.74
L-Threonic acid	Standard	GC-MS	292	12.22
L-Tryptophan	Standard	GC-MS	202	19.39

L-Tyrosine	Standard	GC-MS	218	16.27
Lysine	Standard	LC-MS	145.09	3.56
Malate	Standard	GC-MS	73	11.26
Maleimide	NIST	GC-MS	154	5.77
Mannitol	Standard	GC-MS	319	16.36
Methionine	Standard	LC-MS	150.058	5.00
Methylcysteine	Standard	GC-MS	218	10.40
Methylguanidine	Standard	LC-MS	74.0583	3.82
Myo-Inositol	Standard	GC-MS	217	18.07
Myristic acid	Standard	GC-MS	117	15.19
Myristoleic acid	NIST	GC-MS	256	15.07
N,N-Dimethyl-2-isopropoxyethylamine	NIST	GC-MS	58	4.31
N6-Acetyl-L-lysine	Standard	LC-MS	187.109	4.30
N-Acetyl glucosamine	NIST	GC-MS	274	14.32
N-acetyl-glutamine	Standard	LC-MS	187.082	5.04
N-acetyl-glycine	Standard	GC-MS	144	9.67
N-Acetylneuraminic acid	NIST	GC-MS	362	21.73
N-formyl-glycine	Standard	GC-MS	160	10.05
Nicotine	NIST	GC-MS	163	5.90
Nicotinic acid	NIST	GC-MS	180	8.31
Nicotinuric acid	HMDB	LC-MS	180.053	5.02
Nonanoic acid	Standard	GC-MS	215	9.49
Norleucine	Standard	LC-MS	132.102	10.48
Octadecanoic acid	Standard	GC-MS	117	19.40
Octanoic acid	NIST	GC-MS	201	8.18
Octanoylcarnitine	HMDB	LC-MS	287.21	18.75
oleoylcarnitine	HMDB	LC-MS	425.35	22.00
Olic acid	Standard	GC-MS	339	19.12
Ornithine	Standard	GC-MS	204	15.05
o-Tyrosine	HMDB	LC-MS	181.074	11.22
Oxalic acid	Standard	GC-MS	73	6.52
Oxaloacetate	Standard	LC-MS	130.992	3.15
Oxanilic acid	NIST	GC-MS	147	13.22
Palmitoleic acid	Standard	GC-MS	129	16.99
Palmitoylcarnitine	Standard	LC-MS	400.342	21.84
Parabanic acid	NIST	GC-MS	243	10.15
Pentosidine	HMDB	LC-MS	378.202	17.37
Phenylalanine	Standard	GC-MS	218	12.92
Phenylglyoxylic acid	HMDB	LC-MS	150.032	3.15
Phenyllactic acid	Standard	LC-MS	165.053	18.33

Phosphoserine	Standard	LC-MS	186.019	3.12
Picolinic acid	NIST	GC-MS	180	9.03
Progesterone	HMDB	LC-MS	314.225	19.14
Proline	Standard	GC-MS	142	8.74
Proline betaine	HMDB	LC-MS	143.095	4.04
Propionylcarnitine	HMDB	LC-MS	217.131	13.17
Pseudo uridine	NIST	GC-MS	217	20.79
Pyrrole-2-carboxylic acid	Standard	GC-MS	240	9.52
Pyruvate	Standard	GC-MS	174	5.18
Quinic acid	HMDB	LC-MS	192.063	4.32
Ribitol	Standard	GC-MS	217	14.11
sarcosine	NIST	GC-MS	116	6.44
Serine	Standard	GC-MS	204	9.62
Sorbose	NIST	GC-MS	103	14.73
Sphingosine	HMDB	LC-MS	299.282	24.41
Sphingosine 1-phosphate	HMDB	LC-MS	379.249	22.24
Stearoylcarnitine	HMDB	LC-MS	427.366	22.68
Succinate	Standard	GC-MS	247	8.91
Succinyladenosine	HMDB	LC-MS	383.108	3.61
Tetracosahexaenoic acid	HMDB	LC-MS	356.272	23.51
Tetradecanoylcarnitine	HMDB	LC-MS	371.304	21.00
Thiamine	HMDB	LC-MS	265.112	18.01
Threitol	Standard	GC-MS	217	11.58
Threonine	Standard	GC-MS	219	9.99
Tiglylglycine	HMDB	LC-MS	157.074	15.98
trans-Hexadec-2-enoyl carnitine	HMDB	LC-MS	397.319	21.22
Trimethylamine N-oxide	Standard	LC-MS	76.065	3.67
Undecanoic acid	HMDB	LC-MS	186.162	21.44
Uracil	Standard	LC-MS	113.022	5.08
Urea	Standard	GC-MS	189	7.97
Uric acid	Standard	GC-MS	441	18.10
Valine	Standard	LC-MS	118.092	4.33
Xylitol	NIST	GC-MS	217	13.98
α-ketoglutarate	Standard	GC-MS	198	12.30

APPENDIX B

LIST OF 83 DIFFERENTIAL METABOLITES BETWEEN BREAST TUMORS

AND NORMAL CONTROLS

	Name	Library	Instrument	FC	p	VIP
Amino acid metabolism	4-hydroxyphenylacetic acid	STD	GCMS	0.34	6.2E-09	1.79
	N-formyl-glycine	STD	GCMS	0.39	1.7E-08	1.71
	pyrrole-2-carboxylic acid	STD	GCMS	0.35	1.8E-08	1.70
	picolinic acid	STD	GCMS	0.37	4.8E-08	1.69
	4-hydroxy-proline	STD	GCMS	3.40	6.1E-08	1.86
	glutamate	STD	GCMS	2.01	1.1E-06	1.62
	asparagine	STD	GCMS	2.18	5.9E-06	1.45
	aspartate	STD	GCMS	1.92	7.4E-06	1.39
	tyrosine	STD	GCMS	1.69	1.7E-05	1.41
	gamma-aminobutyric acid	STD	GCMS	4.73	1.8E-05	1.59
	phenylalanine	STD	GCMS	1.69	2.6E-05	1.37
	methionine	STD	GCMS	1.73	2.9E-05	1.30
	isoleucine	STD	GCMS	1.68	3.3E-05	1.37
	5-oxoproline	STD	GCMS	1.58	3.3E-05	1.44
	tryptophan	STD	GCMS	1.68	7.0E-05	1.28
	histidine	STD	GCMS	1.57	1.0E-04	1.20
	proline	STD	GCMS	1.52	1.2E-04	1.30
	methylcysteine	STD	GCMS	1.48	1.5E-04	1.17
	aminomalonic acid	NIST	GCMS	2.42	3.8E-04	1.29
	carnosine	STD	LCMS	0.42	3.9E-04	1.05
	serine	STD	GCMS	1.46	4.0E-04	1.10
	threonine	STD	GCMS	1.45	4.3E-04	1.15
	cysteine	STD	GCMS	2.02	9.9E-04	1.23
	leucine	STD	GCMS	1.39	1.0E-03	1.11
	2-aminobutyric acid	STD	GCMS	1.91	1.4E-03	1.08
	3-methylhistidine	HMDB	LCMS	0.36	2.0E-03	1.10
Carbohydrate metabolism	citric acid	STD	GCMS	0.28	1.6E-05	1.55
	myo-inositol	STD	GCMS	0.47	4.0E-05	1.55
	xylulose	STD	GCMS	1.54	4.3E-05	1.24
	sucrose	STD	GCMS	0.32	7.5E-05	1.32
	xylose	STD	GCMS	1.96	3.6E-04	1.12
	N-acetyl-D-glucosamine	STD	GCMS	2.00	4.4E-04	1.29
	rhamnose	STD	GCMS	1.74	5.1E-03	1.06

	Name	Library	Instrument	FC	*p*	VIP
Lipid metabolism	nonadecanoic acid (C19:0)	NIST	GCMS	0.35	3.3E-09	1.83
	glycerol	STD	GCMS	0.40	5.1E-09	1.79
	capric acid (C10:0)	STD	GCMS	0.45	2.4E-08	1.70
	hexanoic acid (C6:0)	STD	GCMS	0.45	4.7E-07	1.54
	adrenic acid (22:4(n-6))	NIST	GCMS	0.58	4.9E-06	1.57
	nonanoic acid	STD	GCMS	0.45	7.6E-06	1.46
	phosphoethanolamine	STD	GCMS	3.21	1.7E-05	1.45
	palmitoleic acid (C16:1)	STD	GCMS	0.29	1.9E-05	1.54
	myristic acid (C14:0)	STD	GCMS	0.31	4.6E-05	1.48
	squalene	STD	GCMS	0.14	5.0E-05	1.44
	lauric acid (C12:0)	STD	GCMS	0.30	6.7E-05	1.43
	lysoPC(16:0)	HMDB	LCMS	0.35	1.1E-04	1.01
	cholesterol	STD	GCMS	0.48	1.2E-04	1.26
	linolic acid (18:2(n-6))	STD	GCMS	0.24	1.2E-04	1.34
	hexadecanoic acid (C16:0)	STD	GCMS	0.41	1.7E-04	1.38
	octanoic acid	NIST	GCMS	0.60	2.2E-04	1.15
	palmitoylcarnitine	STD	LCMS	4.21	9.5E-04	1.13
	octadecanoic acid (C18:0)	STD	GCMS	0.53	9.6E-04	1.16
	glyceraldehyde 3-phosphate	STD	GCMS	2.50	1.4E-03	1.21
	hexanoylcarnitine	STD	LCMS	3.85	1.7E-03	1.15
	1-stearoylglycerol	STD	GCMS	0.65	1.8E-03	1.06
	3-dehydrosphinganine	HMDB	LCMS	3.72	1.9E-03	1.06
	arachidyl carnitine	HMDB	LCMS	12.09	3.8E-03	1.04
	palmitin	STD	GCMS	0.60	3.8E-03	1.03
Nucleic acid metabolism	cytosine	NIST	GCMS	0.45	5.2E-08	1.68
	ribose	NIST	GCMS	2.54	3.5E-07	1.61
	uracil	STD	GCMS	2.95	2.9E-06	1.69
	guanosine	STD	GCMS	3.81	3.3E-05	1.55
	orotidine	STD	GCMS	4.42	4.8E-05	1.42
	dihydrouracil	STD	GCMS	3.52	8.4E-05	1.26
	hypoxanthine	STD	GCMS	1.54	2.5E-04	1.25
	8-hydroxy-deoxyguanosine	HMDB	LCMS	3.75	4.5E-04	1.20
	ribonic acid	NIST	GCMS	1.62	9.1E-04	1.13
	uridine	STD	GCMS	2.10	2.5E-03	1.01

	Name	Library	Instrument	FC	*p*	VIP
	3-hydroxypyridine	STD	GCMS	0.37	7.9E-09	1.80
	2-hydroxypyridine	STD	GCMS	0.35	1.2E-08	1.80
	phenol	STD	GCMS	0.28	1.4E-08	1.79
	cyclohexanol	NIST	GCMS	0.30	1.6E-08	1.78
	3-octenoic acid	NIST	GCMS	0.34	3.2E-08	1.73
	phosphate	NIST	GCMS	0.51	2.5E-07	1.70
	benzoic acid	STD	GCMS	0.37	3.1E-06	1.48
Other	threonic acid	STD	GCMS	0.36	3.0E-05	1.60
	diphosphoric acid	NIST	GCMS	0.36	1.2E-04	1.43
	methylsuccinic acid	STD	GCMS	0.54	1.3E-04	1.13
	hydroxyacetic acid	STD	GCMS	0.43	1.3E-04	1.32
	1,4-butanediamine	STD	GCMS	3.28	4.2E-04	1.29
	alpha-tocopherol (vitamin E)	STD	GCMS	0.37	1.5E-03	1.12
	pantothenic acid (vitamin B5)	STD	GCMS	1.84	1.8E-03	1.13
	spermidine	STD	LCMS	2.65	2.0E-03	1.02
	5-hydroxyindoleacetic acid	HMDB	LCMS	3.16	2.4E-03	1.03