Aus dem Deutschen Krebsforschungszentrum Heidelberg
(Geschäftsführender Direktor: Prof. Dr. Michael Baumann)
Abteilung für Molekulare Epidemiologie
(Abteilungsleiterin: Prof. Dr. Barbara Burwinkel)

Identification of Plasma Metabolites Associated with Breast and Ovarian Cancer and Breast Cancer Prognosis

Inauguraldissertation
zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum.)
an der
Medizinischen Fakultät Heidelberg
der
Ruprecht - Karls - Universität

vorgelegt von
Baowen Yuan
aus
Henan, China
2019

Dekan: Prof. Dr. med. Hans-Georg Kräusslich

Doktormutter: Prof. Dr. Barbara Burwinkel

DEDICATION

This thesis is dedicated to my family

Especially to my grandmother: Mrs. Fuqin Liu

And to the memory of my grandfather: Mr. Ronghua Zhang

Contents

I Abbreviations

AA	amino acid
AC	acylcarnitine
AJCC	American Joint Committee on Cancer
ASCO	American Society of Clinical Oncology
AUC	area under the curve
BA	biogenic amine
BCAA	branched-chain amino acids
BMI	body mass index
CA	cancer antigen / carbohydrate antigen / carcinoma antigen
CE	capillary electrophoresis
CEA	carcinoembryonic antigen
Cer	ceramide
CI	confidence interval
CoA	coenzyme A
CSC	cancer stem cells
CTCs	circulating tumor cells
CV	coefficient of variation
DDFS	distant disease-free survival
DFS	disease-free survival
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
EMT	epithelial-mesenchymal transition
EpCAM	epithelial cell adhesion molecule
ER	estrogen receptor

FC	fold change
FDA	US Food and Drug Administration
FDR	false discovery rate
FIA-MS/MS	flow injection analysis coupled with tandem mass spectrometry
FIGO	International Federation of Gynecology and Obstetrics
GC	gas chromatography
H1	the sum of hexoses
HER2	human epidermal growth factor
HMDB	Human Metabolome Database
IDC	invasive ductal carcinoma
ILC	invasive lobular carcinoma
IPE	integrated prediction error
LASSO	least absolute shrinkage and selection operator
LC	liquid chromatography
LOD	limit of detection
(lyso)PC	(Lyso)phosphatidylcholine
MBC	metastatic breast cancer
MS	mass spectrometry
NMR	nuclear magnetic resonance
OS	overall survival
OVCA	ovarian cancer
PBC	primary breast cancer
PCA	principal component analysis
PFS	progression-free survival
PR	progesterone receptor
QC	quality control

RNA	ribonucleic acid
ROC	receiver operating characteristics
ROS	reactive oxygen species
SEER	Surveillance, Epidemiology, and End Results
SM	sphingomyelin
TGF-β	transforming growth factor β
TIC	tumor-initiating cells
TME	tumor microenvironment
UPLC	ultra-high-performance liquid chromatography
UPLC-MS/MS	ultra-high-performance liquid chromatography-tandem mass spectrometry
uPA	urokinase plasminogen activator
WHO	World Health Organization

II List of Figures

III List of Tables

IV Summary

As two leading female cancers, breast cancer, especially metastatic breast cancer, and ovarian cancer, have brought an increasing health and economic burden globally. Biomarkers could improve patient outcomes and quality of life because they play vital roles in cancer screening, diagnosis, prognosis, and prediction. Metabolites are promising cancer biomarkers, as they represent the ultimate phenotypic alteration of the organism and are closely related to cancer. Plasma metabolites can be accessed with minimally invasive procedures. Using plasma metabolites as biomarkers for cancer and other diseases has been widely explored because of the possibility of repeated sampling and periodic monitoring of blood samples. However, metabolic studies are still in their infancy, and only a few studies with large sample sizes are available so far. In this thesis project, we explored the potential of metabolites as putative diagnostic and prognostic markers in breast and ovarian cancer.

Plasma metabolite profiling and subsequent validation in **primary breast cancer patients** and healthy controls identified 18 metabolites that were significantly differentially represented (FDR < 0.05). Multivariate logistic regression analysis selected a panel of seven metabolites to discriminate primary breast cancer patients from healthy controls with an AUC of 0.80. If this panel of metabolites identified here could be verified in large prospective study cohorts, this panel, including Glu, Orn, Thr, Trp, Met-SO, C2, and C3, might add value to multi-molecular diagnostic marker sets for breast cancer early detection.

The association of plasma metabolites with **metastatic breast cancer** was investigated as well. Metastatic breast cancer patients with high numbers of circulating tumor cells (termed CTC-positive) and those with low numbers or without CTCs (termed CTC-negative) were analyzed and compared to healthy controls as well as primary breast cancer patients. Lists of 19 and 12 metabolites were identified to significantly distinguish CTC-positive and CTC-negative samples from healthy controls, respectively. A panel comprising His, C4:0, C18:1, lysoPC a C18:2, PC aa C40:6, and PC ae C42:3 for CTC-positive patients with AUC = 0.92, and a combination of Asn, Glu, His, Thr, Trp, C16:0, C18:0, C18:1, C18:2, lysoPC a C18:2, and PC aa C40:6 for CTC-negative patients with AUC = 0.89 were selected to distinguish from healthy controls. Significantly different metabolites between CTC-

positive/CTC-negative and primary breast cancer patients exhibited significant overlaps with those between CTC-positive/CTC-negative patients and healthy controls.

We also investigated the **prognostic value of metabolites in metastatic breast cancer patients**. After successive analysis of the discovery and validation cohorts, four metabolites were found to be significantly negatively correlated with progression-free survival, while 12 metabolites were negatively correlated with overall survival. Amongst these metabolites associated with survival, LASSO Cox regression analysis selected a combination of PC ae C36:1 and PC ae C38:3 to predict progression-free survival, and a combination of lysoPC a C20:3, lysoPC a C20:4, PC aa C38:5, PC ae C38:3, and SM (OH) C22:2 to predict overall survival. Even though the proposed metabolic signatures showed a lower prognostic power than the CTC status, an FDA-approved prognostic marker, the combination of the Cox selected metabolites with the CTC status displayed a lower integrated prediction error than CTC status alone. Therefore, the identified metabolic markers might add prognostic value in combination with other biomarkers such as CTC status determination. The majority of the here identified metabolites have previously shown functional roles in cancer and metastasis development, thus laying a supposed mechanistic basis for their differential levels observed in plasma.

Lastly, comparative profiling of plasma metabolites in **ovarian cancer patients** and healthy controls were applied to identify metabolites associated with ovarian cancer. Remarkably, 71 significantly differentially expressed metabolites were identified (FDR < 0.05). Most of them were down-regulated in ovarian cancer patients. A combination of seven metabolites, including His, Trp, C18:1, lysoPC a C18:2, PC aa C32:2, PC aa C34:4, PC ae C34:3, were identified to differentiate ovarian cancer cases from healthy controls with an AUC of 0.95. Furthermore, this panel could distinguish ovarian cancer from primary breast cancer patients with an AUC of 0.93.

In conclusion, we identified specific signatures of plasma metabolites associated with primary breast cancer, metastatic breast cancer, and ovarian cancer. Further, we identified sets of metabolites correlated with the prognosis of metastatic breast cancer patients. If these

identified metabolic marker signatures can be verified in large, multi-centric, prospective studies, they might add value to the development of blood-based diagnostic tests.

V Zusammenfassung

Als führende Krebserkrankungen bei Frauen haben Brustkrebs und Eierstockkrebs weltweit zu einer zunehmenden gesundheitlichen und wirtschaftlichen Belastung geführt. Ovarialkrebs und metastasierender Brustkrebs weisen zudem eine sehr schlechte Prognose auf. Biomarker könnten den Behandlungserfolg und die Lebensqualität der Patientinnen verbessern, indem sie eine wichtige Rolle bei der Krebsvorsorge, Diagnose, Prognose und Vorhersage spielen. Metaboliten sind vielversprechende Krebs-Biomarker, da sie eine endgültige phänotypische Veränderung des Organismus darstellen und in enger Beziehung zur Krebsentstehung und Progression stehen. Der Zugang zu Plasmametaboliten ist minimal invasiv und aufgrund der Möglichkeit einer wiederholten Probenentnahme und der regelmäßigen Überwachung von Blutproben vielversprechend. Metaboliten-basierte Biomarkerstudien für Krebsdiagnostik und Prognose befinden sich jedoch noch in einem frühen Stadium, da bislang nur wenige Studien mit umfassenden Probenzahlen und /oder optimaler Blutplasmaproben-Prozessierung vorliegen. In diesem Dissertationsprojekt untersuchten wir das Potenzial von Metaboliten als mutmaßliche diagnostische und prognostische Marker für Brust- und Eierstockkrebs.

Das Plasmametaboliten-Profiling und die anschließende Validierung bei **primären Brustkrebspatientinnen** und gesunden Kontrollpersonen identifizierte 18 Metaboliten, die signifikant unterschiedlich in den beiden Gruppen vertreten waren (FDR <0,05). Durch eine multivariate logistische Regressionsanalyse konnte eine Gruppe von sieben Metaboliten (Glu, Orn, Thr, Trp, Met-SO, C2 und C3) ausgewählt werden, die Patientinnen mit primärem Brustkrebs von gesunden Kontrollen mit einer AUC von 0,80 unterscheiden kann. Sofern die hier identifizierten Metaboliten in großen prospektiven Studienkohorten verifiziert werden können, könnten diese zur Entwicklung von multimolekularen diagnostischen Markersätzen für die Brustkrebs Früherkennung beitragen.

Die Assoziation von Plasmametaboliten mit **metastasierendem Brustkrebs** wurde ebenfalls untersucht. Patientinnen mit metastasierendem Brustkrebs, die eine erhöhte Anzahl zirkulierender Tumorzellen (als CTC-positiv bezeichnet) aufwiesen und solche mit einer geringen Anzahl oder ohne CTCs (als CTC-negativ bezeichnet) wurden analysiert und mit gesunden Kontrollpersonen sowie Patientinnen mit primärem Brustkrebs verglichen. Es

wurden 19 bzw. 12 Metabolite identifiziert, um CTC-positive bzw. CTC-negative MBC Proben signifikant von gesunden Kontrollen zu unterscheiden. Zur Unterscheidung von CTC-positiven MBC Patienten von gesunden Kontrollen wurde ein Panel bestehend aus His, C4:0, C18: 1, lysoPCa C18:2, PC aa C40:6 und PC ae C42:3 identifiziert (AUC = 0,92) und für CTC-negative Patienten eine Kombination aus Asn, Glu, His, Thr, Trp, C16:0, C18:0, C18:1, C18:2, lysoPC a C18:2 und PC aa C40:6 (AUC = 0,89) ausgewählt. Signifikant unterschiedliche Metaboliten zwischen CTC-positiven/CTC-negativen metastasierenden Brustkrebspatientinnen und primären Brustkrebspatientinnen zeigten signifikante Überschneidungen mit denen o.g. zwischen CTC-positiven/CTC-negativen metastasierenden Patientinnen und gesunden Kontrollpersonen.

Wir untersuchten auch den **prognostischen Wert von Metaboliten bei Patientinnen mit metastasierendem Brustkrebs**. Nach sukzessiver Analyse der Entdeckungs- und Validierungskohorten wurde festgestellt, dass vier Metaboliten signifikant negativ mit der progressionsfreien Überlebensrate korrelierten, während 12 Metaboliten negativ mit dem Gesamtüberleben korrelierten. Via LASSO Cox-Regressionsanalyse wurde eine Kombination aus PC ae C36:1 und PC ae C38:3 identifiziert, um eine progressionsfreie Überlebensrate vorherzusagen, und eine Kombination aus lysoPC a C20:3, lysoPC a C20:4, PC aa C38:5, PC ae C38:3 und SM (OH) C22:2 um eine Vorhersage zum Gesamtüberleben zu treffen. Obwohl die vorgeschlagenen metabolischen Signaturen eine niedrigere Prognoseleistung als der CTC-Status, ein von der FDA zugelassener Prognosemarker, aufwiesen, konnte durch die Kombination der durch die Cox Analyse ausgewählten Metaboliten mit dem CTC-Status ein geringerer integrierter Prognosefehler als mit dem CTC-Status allein erzielt werden. Daher könnten die identifizierten Metabolitenmarker in Kombination mit anderen Biomarkern wie der CTC-Statusbestimmung die prognostische Aussagekraft erhöhen.

Zuletzt wurden Plasmametabolite identifiziert, die mit **Ovarialkrebs** assoziiert sind. Bemerkenswerterweise wurden zwischen Proben von Ovarialkrebspatientinnen und gesunden Kontrollen 71 signifikant unterschiedlich exprimierte Metaboliten identifiziert (FDR <0,05). Die meisten von ihnen waren bei Patientinnen mit Ovarialkarzinom herunterreguliert. Eine Kombination von sieben Metaboliten (His, Trp, C18:1, lysoPC a C18:2, PC aa C32:2, PC aa C34:4, PC ae C34:3) wurde identifiziert, um Ovarialkrebsfälle von gesunden Kontrollen zu

unterscheiden (AUC = 0,95). Darüber hinaus konnte dieses Panel Eierstockkrebs von primären Brustkrebspatientinnen mit einer AUC von 0,93 unterscheiden.

Zusammenfassend konnten wir spezifische Signaturen von Plasmametaboliten identifizieren, die mit primärem Brustkrebs, metastasiertem Brustkrebs und Eierstockkrebs assoziiert sind. Darüber hinaus identifizierten wir Metabolitensätze, die mit der Prognose metastasierender Brustkrebspatientinnen korrelieren. Sofern diese identifizierten metabolischen Markersignaturen in großen, multizentrischen, prospektiven Studien verifiziert werden können, könnten sie zur Entwicklung von Blut-basierten diagnostischen Markersätzen beitragen.

1. Introduction

1.1 Cancer

Cancer is a collection of more than 100 related diseases that can start almost anywhere in the body and characterize by the feature of the abnormal and uncontrollable proliferation of cells. Carcinogenesis involves a series of complex and dynamic processes consisting of three stages: initiation, progression, and metastasis (Wang *et al.*, 2017). Biological capabilities acquired during the complex, multistep development of human tumors are defined as the hallmarks of cancer, which was proposed by Hanahan and Weinberg in 2000 (Hanahan and Weinberg, 2000). Later in 2011, another four emerging hallmarks were supplemented (Hanahan and Weinberg, 2011), and the ten hallmarks are shown in Figure 1.1. Cancer is a genetic disease. Genetic changes, which regulate cell cycle and cell homeostasis, result in an imbalance between cell proliferation and cell death, leading to tumorigenesis (Weinberg, 2014). Cancer-causing mutations primarily include point mutations, as well as chromosomal structural alterations and loss of heterozygosity (Loeb and Loeb, 2000). Cells that gain these cancer-causing mutations and stimulate tumor formation are defined as tumor-initiating cells (TIC) (Qureshi-Baig *et al.*, 2017). Tumors can be either benign or malignant. The ability to invade nearby tissues and spread to distant sites is an important criterion to distinguish malignant tumors from benign tumors. Metastases spawned by malignant tumors are responsible for almost all deaths from cancer (Sopik and Narod, 2018). In addition to the tumor cells, tumor niche within the tumor microenvironment (TME) is closely connected to the process of tumorigenesis, from pre-malignant to a malignant state and facilitates tumor growth (Wang *et al.*, 2017). Based on the tissue of origin, cancer in humans can be classified into five principal types: carcinoma (epithelial tissues), sarcoma (connective tissue), leukemia (blood cells of bone marrow), lymphoma and multiple myeloma (immune system), and neuroectodermal tumors (brain and spinal cord) (Weinberg, 2014).

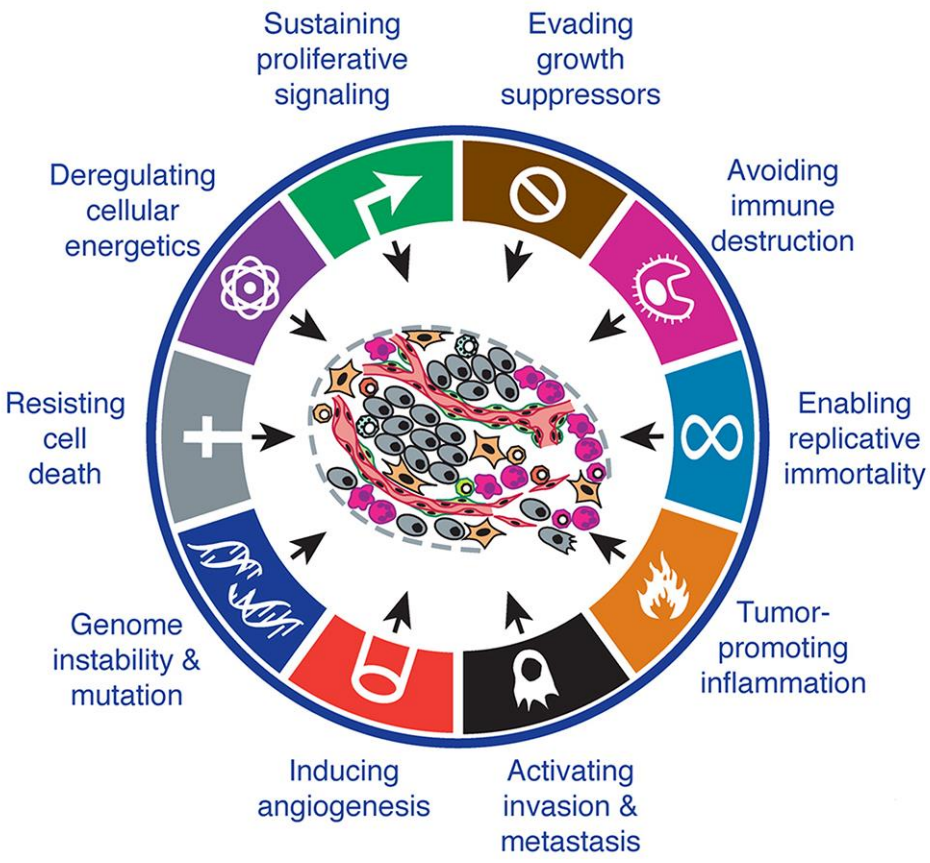


Figure 1.1: The Hallmarks of Cancer. Ten functional properties of tumor cells required for complete malignant transformation are essential. (Adapted from Hanahan and Weinberg, Cell, 2011, Hallmarks of cancer: the next generation).

Cancer is a major public health problem worldwide and is the second leading cause of death in the United States (Siegel *et al.*, 2018). Cancer was responsible for an estimated 8.1 million new cancer cases, and 9.6 million deaths in 2018 alone, about 1 in 6 deaths is due to cancer globally (Bray *et al.*, 2018). The commonly diagnosed cancer cases were lung cancer (11.6%), female breast cancer (11.6%), prostate cancer (7.1%), and colorectal cancer (6.1%) (Bray *et al.*, 2018). Prevention, screening, diagnosis, prognosis, and tumor development or spread targeted therapies are strategies for cancer management and treatment (Gralow *et al.*, 2008).

1.2 Breast Cancer

1.2.1 Epidemiology and Detection

Breast cancer is cancer that arises from the breast tissue, and most of the breast cancer cases are adenocarcinomas. Breast cancer continues to be the most prevalent malignancy and a major cause of human mortality in women worldwide. It accounts for 24.2% of all newly diagnosed cancer cases and 15% of all cancer-related deaths among females in 2018 (Figure 1.2) (Bray *et al.*, 2018). Due to the availability of early detection and treatment, the incidence rates in high-income countries are higher than those in low- and middle-income countries, whereas the mortality rates are lower in high-income countries than those in low- and middle-income countries (Torre *et al.*, 2017). In Germany, breast cancer causes the highest cancer-related morbidity and mortality rates in females, with percentages of 25.9% and 17.6%, respectively (Figure 1.2) (Bray *et al.*, 2018). These high numbers have indicated the urgency and importance of breast cancer study in the field of cancer research.

Currently, mammography is the standard technique for breast cancer screening and detection, and it was reported to have high detection sensitivity (93%) for breast cancer patients with symptoms (Jiang *et al.*, 2016). However, mammography is far less effective for early-stage breast cancer detection and is less sensitive among women with high breast density and young women (Buist *et al.*, 2004; Tabar *et al.*, 1995). Moreover, mammography has only a minor effect on the reduction of breast cancer mortality (Bleyer and Welch, 2012; Broeders *et al.*, 2012). Besides, the x-ray radiation from mammography adds up over time does harm to the breast (Miglioretti *et al.*, 2016). Other screening methods, including self or clinical examination, breast ultrasound, and magnetic resonance imaging, have their pros and cons (Elmore *et al.*, 2005).

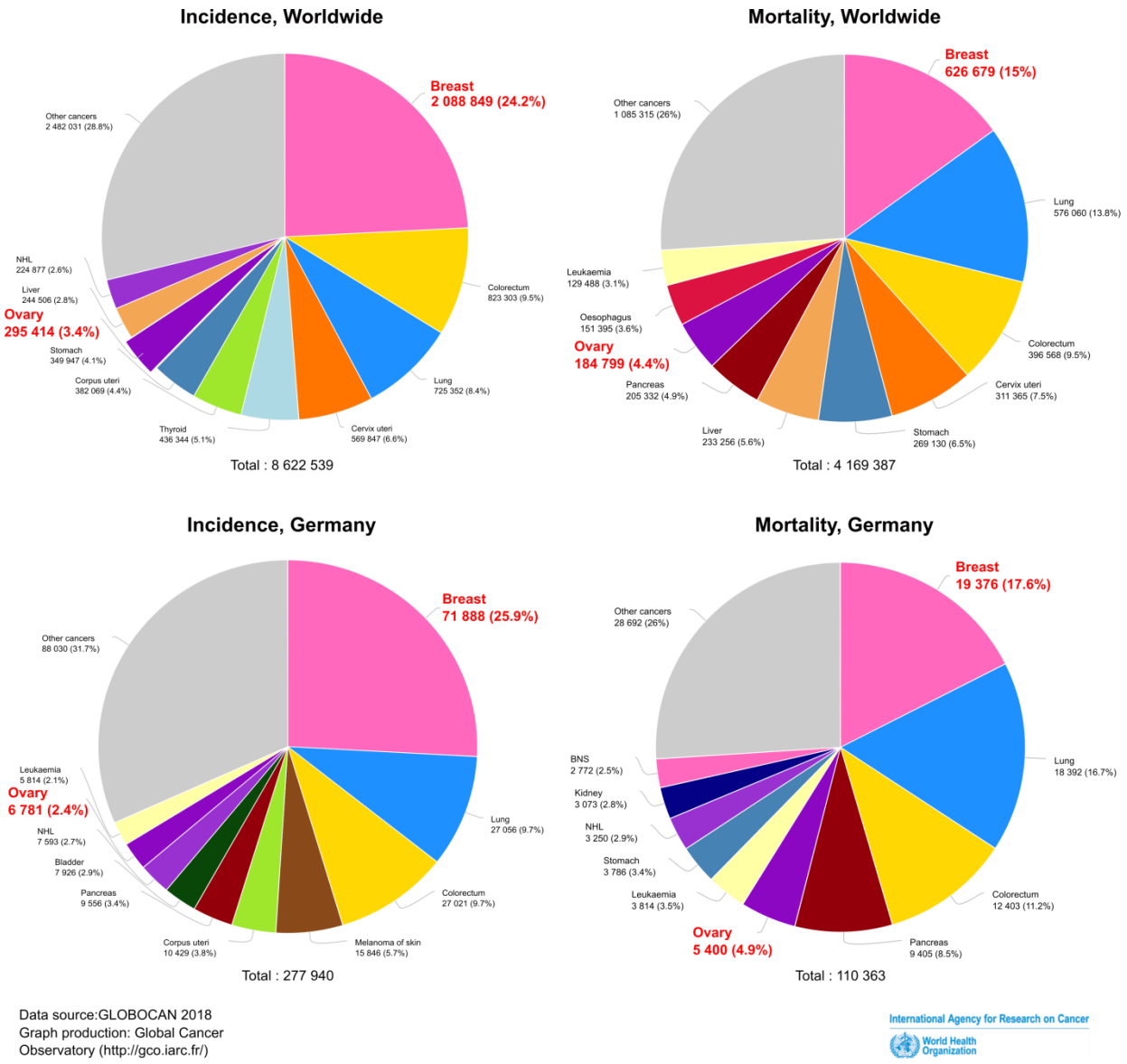


Figure 1.2: Incidence and mortality numbers and rates of female cancer worldwide and in Germany in 2018. Data source: GLOBOCAN 2018 (http://gco.iarc.fr/).

1.2.2 Classification

Breast cancer is a highly heterogeneous disease that has long been noted in histology and clinical outcomes, which serve as the basis for disease classification (Polyak, 2011). Since the subtypes of breast cancer differ a lot in outcomes and responses to treatment (Taherian-Fard

et al., 2015; Viale, 2012), breast cancer classification and characterization are substantially crucial for a better understanding of the disease and the treatment of individuals with precision medicine. There are four major routinely used breast cancer classification methods in the clinic, as shown in Figure 1.3.

Based on the tumor cells of origin, breast cancer can be classified into different histological subtypes. As reported by the latest World Health Organization (WHO) guidelines, 20 major and 18 minor tumor types, including benign, carcinoma in situ, malignant invasive carcinomas, and sarcomas, are recognizable (Tavassoéli, 2003). Among these histological subtypes, invasive ductal carcinoma (IDC) is the most common, representing 75% of all cancer cases. The second common is invasive lobular carcinoma (ILC), accounting for around 10% of all cancer patients (Li *et al.*, 2005).

The grading system, which indicates the differentiation extent of tumor cells, is based on the microscopic appearance of the tumor section. The Nottingham grading system is currently the most widely recommended method for breast cancer grading (Bloom and Richardson, 1957). According to tubule formation, nuclear pleomorphism, and mitotic count, tumors are classified into well-differentiated (low grade, Grade 1), moderately differentiated (intermediate grade, Grade 2), and poorly differentiated (high grade, Grade 3) (Bloom and Richardson, 1957).

The staging system was proposed by the American Joint Committee on Cancer (AJCC) in 1959 for the first time, and from then on, it has been updated continuously as understandings of disease refreshed (Edge S *et al.*, 2010). The AJCC system classified cancers with three criteria, tumor size (Tx, Tis, T0, T1-T4), lymph node number (Nx, N0-N3), and distant metastasis (Mx, M0, M1), which is known as the TNM staging system (Edge S *et al.*, 2010). Breast tumors include stages 0, I, II, III, and IV according to these three TNM pathological features, and the prognosis becomes worse as the stages increase. The five-year survival rate is 100% for stage 0 and stage I patients, 93% for stage 2, 72% for stage 3, and decrease to 22% for patients with stage IV breast cancer (Howlader N *et al.*, 2019).

The immunohistochemical recognition of hormone receptors, including estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2

(HER2/neu), classifies breast cancer into four molecular subtypes (Schnitt, 2010), i.e. Luminal A (ER+ and/or PR+, HER2-, low grade), Luminal B (ER+ and/or PR+, HER2-, high grade), HER2 amplified, and basal or triple-negative (ER-/PR-/HER2-). The molecular subtype of breast cancer determines its progression and treatment response. From Luminal A, Luminal B, HER2 amplified to triple-negative, cancer growth rate increase, as well as cancer aggressiveness, while cancer prognosis becomes continuously worse (Schnitt, 2010). Besides these four molecular subtypes, some other subtypes, such as Claudin-low and normal-like breast cancer, have been proposed (Malhotra *et al.*, 2010). Determination of the subtype of breast cancer is vital for making treatment decisions, specifically for those that target hormone receptors, and improving patient management.

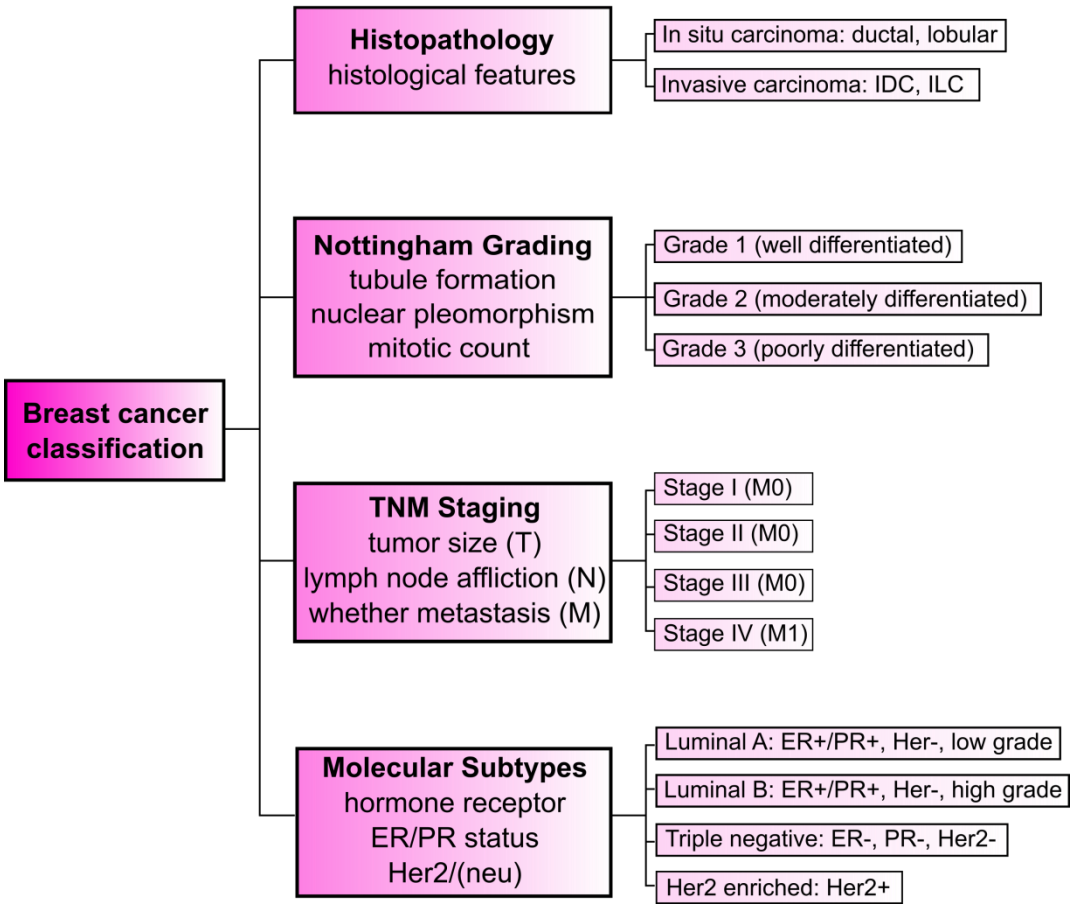


Figure 1.3: Major classification methods of breast cancer.

1.2.3 Risk Factors

Breast cancer is a result of the interplay of diverse genetic and environmental factors. Approximately 75% of breast cancer cases are sporadic, while the remaining 25% are familial/hereditary (Anderson, 1992). The risk factors of breast cancer can be classified as genetic risk factors and non-genetic risk factors.

Genetic Risk Factors

Genetic and familial factors can substantially increase the lifetime risk of developing breast cancer and are associated with the development of cancer at a young age. Women with any first-degree relative with breast cancer have about twice the risk of developing breast cancer (Pharoah *et al.*, 1997). Approximately 30% of breast cancer cases are attributed to hereditary factors (Mavaddat *et al.*, 2010; Turnbull and Rahman, 2008). *BRCA1*, a DNA-repair gene, was the first established major breast cancer susceptibility gene. *BRCA2* is another gene of the same family, and they cause most of the hereditary breast carcinoma (Hall *et al.*, 1990). Women with inherited mutations in *BRCA1* or *BRCA2* genes are at particularly high risk. Carriers of *BRCA1* mutations have a 55%-60% chance for developing breast cancer during their lifetime, whereas those with *BRCA2* mutations have a lower risk of 45% (Foulkes and Shuen, 2013; Malone *et al.*, 1998). Besides these two genes, *TP53*, *PTEN*, and *STK11/LKB1*are also essential susceptibility genes that involve in the hereditary and familial forms of breast cancer (Nathanson *et al.*, 2001). Genetic low-penetrance risk factors contribute much more modest (2 ~ 4 fold) risks of breast cancer (Thompson and Easton, 2004).

Non-genetic Risk Factors

Age is the most potent risk factor for breast cancer. The incidence of breast cancer increases with age, doubling every ten years until menopause and is usually diagnosed in the 50 ~ 60 age group (Allison, 2012; McPherson *et al.*, 2000). Reproductive factors and hormones are also established risk factors for breast cancer, such as earlier menarche, later menopause,

nulliparity, delayed childbearing, no breastfeeding, and hormone exposure (McPherson *et al.*, 2000). Also, a personal history of breast disease, benign breast diseases such as fibrosis and hyperplasia are other non-modifiable risk factors for breast cancer (McPherson *et al.*, 2000). At last, unhealthy lifestyles are also associated with breast cancer incidence, e.g., alcohol drinking, smoking, physical inactivity, excess body weight (postmenopausal females), and high-dose radiation to the chest, particularly at a young age (Verma *et al.*, 2012).

1.2.4 Metastatic Breast Cancer

Around 25%-30% of primary breast cancer (PBC) cells leave their original primary site and migrate to distant sites, which is called metastatic or stage IV breast cancer (Lorusso and Ruegg, 2012; Redig and McAllister, 2013). Approximately 10%-15% of PBC patients develop metastasis within three years. However, metastatic manifestation after ten or more years from initial diagnosis was also reported (Weigelt *et al.*, 2005). Breast cancer cells tend to metastasize to essential organs in humans, such as bone and liver, also less frequently to lung and brain (Lee, 1985). Most of the breast cancer-related deaths are due to metastasis, rather than the primary ones. Since our understanding of metastasis mechanism is inadequate, most of the metastatic cancers are still incurable.

Evolution of metastasis theories

The term metastasis was coined in 1829 (JC, 1829), and understandings of cancer metastasis evolved from time to time in nearly two hundred years (Dong *et al.*, 2009; Talmadge and Fidler, 2010). In 1889, a London surgeon Stephen Paget proposed the famous “seed and soil” hypothesis to explain the usual organ-specific metastasis pattern (Paget, 1989). Forty years later, in 1928, James Ewing suggested that metastasis occurs purely through an anatomic mechanism (Ewing J, 1928). In 1975, Irwin Bross et al. proposed the metastatic cascade for sequential events needed for disseminated cancer (Nicolson and Winkelhake, 1975). Then the clonal selection and expansion model was proposed by Peter Nowell and Isalah Fidler in 1976 (Fidler and Kripke, 1977; Nowell, 1976). Until in 2001, the role of cancer stem cells (CSC) in metastasis was proposed by Irving Weissman et al. (Reya *et al.*, 2001). Then the

next year, Rene Bernard and Robert Weinberg raised a progression puzzle to challenge the traditional view of metastasis. The dual proclivity model was subsequently proposed (Bernards and Weinberg, 2002). At the same time, Jean Paul Thiery used the epithelial-mesenchymal transition (EMT) phenomenon to explain metastatic progression (Thiery, 2002). In 2006, Kent Hunter et al. emphasized the role of genetic susceptibility for metastatic propensity (Hunter, 2006). Figure 1.4 shows a brief evolution timeline of metastasis theories.

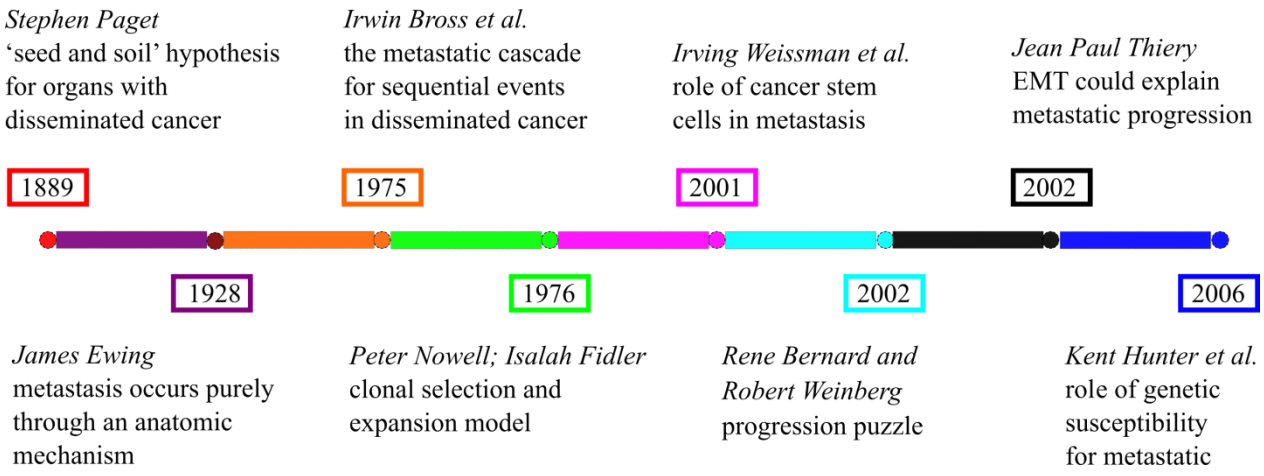


Figure 1.4: The evolution of metastasis theories. The timeline outlines the major conceptual advances in the origins and mechanisms of metastasis from Paget's seminal "seed vs. soil" hypothesis to more recent theories regarding the role of cancer stem cells and early developmental transitions. (Adapted from Dong, F. et al., clinical cancer research, 2009, Translating the metastasis paradigm from scientific theory to clinical oncology.)

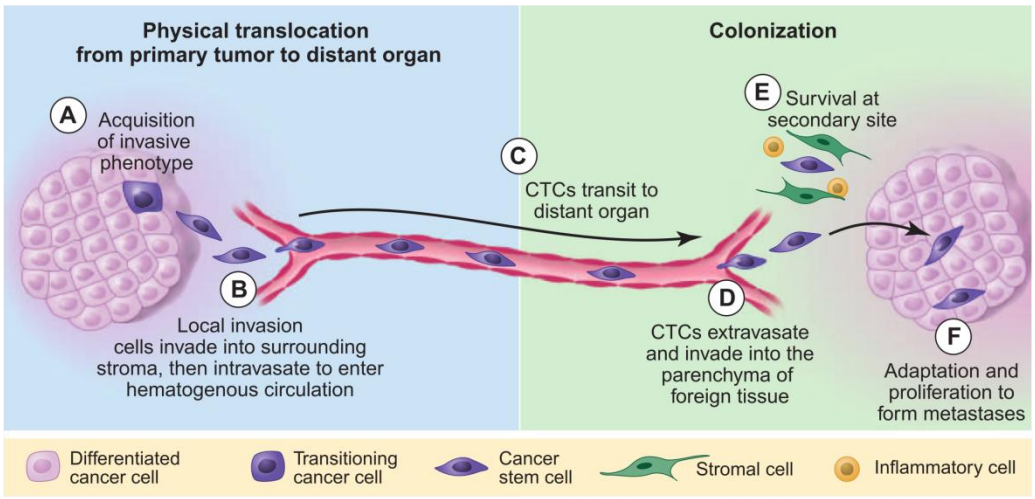
Mechanism of metastasis

Integration of the metastasis theories in nearly two hundred years, the current model for cancer metastasis proposed by Christine L. Chaffer and Robert A. Weinberg is shown in Figure 1.5 (Chaffer and Weinberg, 2011). In general, cancer metastasis is a two-phase process. The first phase is known as the metastasis cascade, which involves the physical translocation of cancer cells from the primary tumor to a distant organ and colonization of the translocated cells within the new organ. The second phase encompasses the ability of the

cancer cell to develop into a metastatic lesion at that distant site (Chaffer and Weinberg, 2011). During these processes, circulating tumor cells (CTC), tumor cells traveling through the circulating, play a vital role in establishing cancer metastasis (Reya *et al.*, 2001).

During metastasis, primary epithelial tumor cells have to leave their extracellular matrix and establish metastasis at a distant organ, during which their morphologies alter. The hypothesis of mechanism that promotes detachment and migration is the epithelial-mesenchymal transition (EMT). EMT, which plays critical roles in early embryonic morphogenesis (Thiery *et al.*, 2009), can help tumor cells transit from the epithelial state to the mesenchymal state. Once tumor cells reach the target site, the mesenchymal phenotype reverts to the epithelial state via mesenchymal-to-epithelial transitions (Chaffer and Weinberg, 2011). EMT program induction during tumorigenesis needs various signaling pathways between cancer cells and stromal cells nearby, such as transforming growth factor β (TGF-β), receptor tyrosine kinases, Wnt-β-catenin, Notch, and bone morphogenetic protein (Gonzalez and Medici, 2014).

Phase one: the metastasis cascade



Phase two: metastatic cells adapt to the distant site

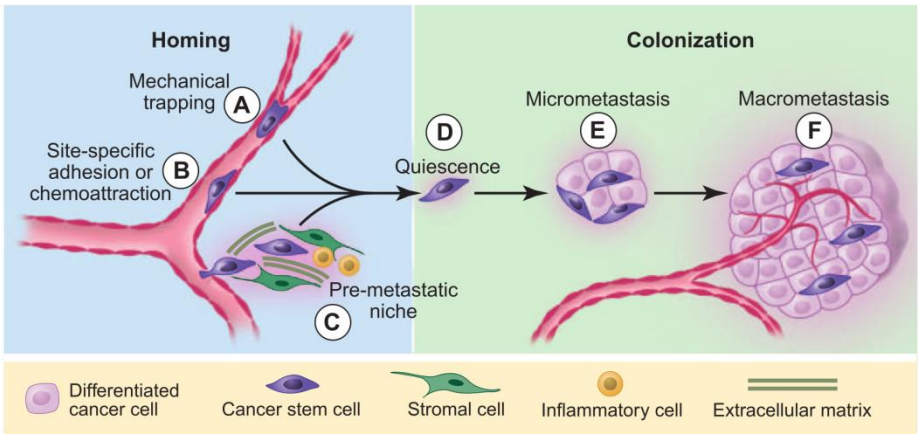


Figure 1.5: The Processes of cancer metastasis. (Adapted from Christine L. Chaffer and Robert A. Weinberg, Science, 2011, A perspective on cancer cell metastasis)

1.3 Ovarian Cancer

1.3.1 Epidemiology

Ovarian cancer (OVCA) is cancer that arises from the ovary. It is the eighth leading in cancer incidence and mortality in women worldwide, representing 3.4% of all newly diagnosed cancer cases and 4.4% of all cancer-related deaths among females in 2018 (Figure 1.2)(Bray *et al.*, 2018). Among all gynecological cancers, ovarian cancer is the third most common

cancer after cervical and uterine cancer and is associated with the second-worst prognosis and the second-highest mortality rate (Bray *et al.*, 2018). In Germany, ovarian cancer is the ninth most common cancer, but the fifth highest in mortality rate, with percentages of 2.4% and 4.9%, respectively (Figure 1.2). In most developed countries, ovarian cancer remains the leading cause of death among gynecological cancers (Siegel *et al.*, 2018). Such a high mortality rate is due to the asymptomatic and secret growth of ovarian cancer, which delays the onset of symptoms. Besides, the absence of effective screening and early detection methods result in its diagnosis in the advanced stages (Momenimovahed *et al.*, 2019). As a result, ovarian cancer is called the "silent killer" (Momenimovahed *et al.*, 2019).

1.3.2 Classification

Ovarian cancer is highly heterogeneous that encompasses a collection of neoplasms with distinct histological and molecular features and prognosis (Gilks and Prat, 2009). Ovarian cancer has a variety of subtypes. Currently, there are two conventional classification methods for ovarian cancer (Figure 1.6).

Cells of origin: Based on the cells of origin and histogenetic principles, the WHO classified ovarian cancer into three main classes of tumors: epithelial tumors, stromal tumors, and germ cell tumors (Kurman, 2014). Epithelial tumors represent around 90% of all ovarian cancer cases, which form on the outer layer of the ovaries. Epithelial ovarian tumors can be further classified into histological types as follows: serous, mucinous, endometrioid, clear cell, transitional cell tumors (Brenner tumors), carcinosarcoma, mixed epithelial tumor, undifferentiated carcinoma, and others (Kaku *et al.*, 2003; Kurman, 2014). Among these subtypes, serous tumors are the most common ones. Around 7% of ovarian cancers are stromal, which begins in the tissue that contains hormone-producing cells. Another category is the germ cell tumor that starts in egg-producing cells, which is rare and usually occurs in younger women (Kaku *et al.*, 2003; Kurman, 2014).

FIGO staging: For patient outcome prediction and treatment plan determination, the basic principle to refer to is cancer staging. The International Federation of Gynecology and

Obstetrics (FIGO) Committee on Gynecologic Oncology has a staging system for ovarian cancers since 1973 (Prat and Oncology, 2014). The FIGO stages ovarian cancer surgically and pathologically by detecting the tissue samples from ovaries, pelvis, and abdomen. Ovarian cancer is thus classified into four stages, namely stages I, II, III, and IV, as shown in Figure 1.6 (Javadi *et al.*, 2016).

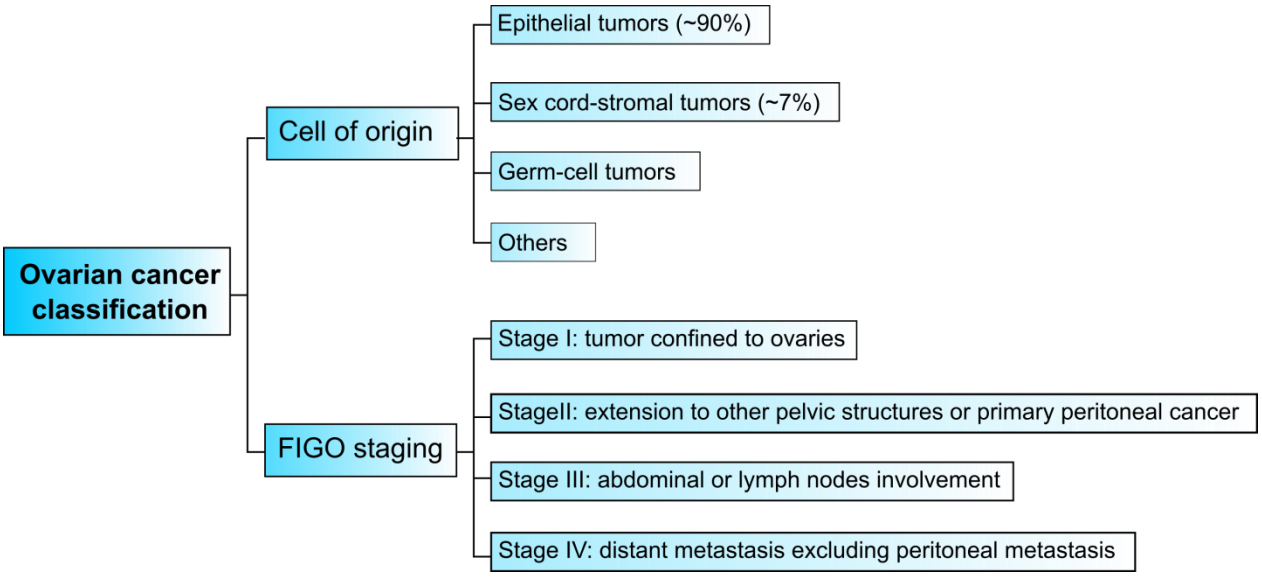


Figure 1.6: Major classification methods of ovarian cancer.

According to the data released by the Surveillance, Epidemiology, and End Results (SEER) program, the five-year survival rates for ovarian cancer vary for the different cell of origin subtypes (Table 1.1) (Howlader N *et al.*, 2019). For women with the disease is confined to the ovary at diagnosis, cure rates are as high as above 90%. Even when the tumor extends to other pelvic structures or develops into primary peritoneal cancer, the five-year survival rates are still higher than 70%. However, it decreases to lower than 20% for patients with advanced stage. One main reason for the high mortality rate of ovarian cancer is that over 70% of ovarian cancer cases are detected at the late stage when it has the worst prognosis and causes most of the deaths (Rauh-Hain *et al.*, 2011). Therefore, early detection of ovarian cancer is of significance in disease management and survival improvement.

Table 1.1: Five-year survival rates of ovarian cancer patients with different subtypes (data source: SEER, 1975-2016).

Five-year survival rates of ovarian cancer			
FIGO staging	**Epithelial tumors**	**Stromal tumors**	**Germ-cell tumors**
Stage I	90%	95%	98%
Stage II	70%	78%	94%
Stage III	39%	65%	87%
Stage IV	17%	35%	69%

1.3.3 Risk Factors

The risk factors of ovarian cancer are almost the same as that of breast cancer, which also includes genetic and non-genetic factors. The most significant genetic risk factor for ovarian cancer is the inherited mutations in *BRCA1* or *BRCA2*, which account for around 5%-15% of all ovarian cancer cases (Lynch *et al.*, 2013). Compared to women without *BRCA1/BRCA2* mutations, carriers of *BRCA1* mutations have a 65% higher ovarian cancer risk, while women with *BRCA2* mutations have a risk of up to 35% higher (Ingham *et al.*, 2013; Mavaddat *et al.*, 2013). Similar to breast cancer patients, high-risk factors of ovarian cancer patients are familial and personal history (Momenimovahed *et al.*, 2019). Older age is the leading risk factor, especially for women over 65 years old, which largely reflects the accumulation of DNA mutations (Mohammadian-Hafshejani *et al.*, 2017). Other factors like early age at menarche, late age at menopause, nulliparity, reproductive organ surgery, no contraception, hormone therapy, and unhealthy lifestyles, e.g., low physical exercise, smoking, alcohol, and high body mass index (BMI) in premenopausal women are all associated with the risk of ovarian cancer (Daniilidis and Karagiannis, 2007; Momenimovahed *et al.*, 2019).

1.4 Metabolomics

"Genomics and proteomics tell you what might happen, but metabolomics tells you what actually did happen."

— Bill Lasley, University of California, Davis.

In the "omics" family, metabolomics is the newest omics technology after genomics, epigenomics, transcriptomics, and proteomics, and is the endpoint of the pyramid (Dettmer and Hammock, 2004). Metabolomics is the technique in system biology to qualitatively or quantitatively measure and analyze the metabolome changes associated with specific physiological and pathophysiological processes (Beger, 2013; Psychogios *et al.*, 2011). The metabolome is the sum of all small-molecule (molecular weight less than 1500 Da) chemicals within cells, tissues, liquid biopsies, or the entire organism (Wikipedia). Metabolites are produced from metabolism, and they involve in many important biological processes, such as energetic regulation, enzyme activity control, structure build-up, and signaling (Vinayavekhin *et al.*, 2010). Metabolic alterations are thought to be the downstream end products of expression alterations in genes or proteins. Metabolic variations reflect not only changes in individual genetic phenotypes and molecular physiology but also environmental influences (Beger, 2013). In this respect, metabolomics has the most significant advantage over the other omics technologies, the most predictive ability for the phenotypic properties could help to elucidate the nature and identity of the biological processes (Capati *et al.*, 2017).

Metabolic alterations caused by certain diseases could be detected in biological fluids before the clinical symptoms manifest because the cascade of small changes in enzyme activities could amplify the signal, which could be detected on the metabolite level (Capati *et al.*, 2017). As a consequence, metabolomics is a promising method for disease biomarkers study. Indeed, metabolomics has attracted much attention in the discovery of biomarkers for disease diagnosis and prognosis (Wang *et al.*, 2018a).

1.4.1 Cancer Metabolism

Metabolic reprogramming, which is characterized by enhanced nutrient uptake to support biosynthetic, bioenergetics, and redox demands of malignant cells, is one of the hallmarks of cancer (Boroughs and DeBerardinis, 2015). Many metabolic processes are altered in cancer cells, tissues, or biofluids, resulting in measurable changes in metabolites that can be used as biomarkers to indicate the presence of cancer or its activity (Vazquez *et al.*, 2016). The most classical metabolic pathway in cancer is aerobic glycolysis, rather than mitochondrial oxidative phosphorylation, which generates the energy needed for cellular processes in normal cells. Aerobic glycolysis in cancer is known as the "Warburg effect" (Vander Heiden *et al.*, 2009; Warburg, 1956). In addition to glucose metabolism, glutamine metabolism is another fuel that supports cancer growth (Boroughs and DeBerardinis, 2015). Other amino acid metabolisms in cancer cells also play important roles. For instance, the oxidation of the branched-chain amino acids (BCAAs) isoleucine and valine provide an anaplerotic flux in some tissues (DeBerardinis and Chandel, 2016; Tonjes *et al.*, 2013). Fatty acid metabolism is another common metabolic change in cancer cells that participate in membrane biosynthesis, lipidation reactions, and cellular signaling, which render as potential targets (DeBerardinis and Chandel, 2016). For example, the widely expressed transmembrane protein CD36 that facilitates fatty acid uptake is one of the targets in breast cancer. What is more, the decreased levels of CD36 in stromal tissue are associated with tumorigenesis initiation (DeFilippis *et al.*, 2012).

Redox balance is also an essential feature of cancer cells. During tumorigenesis, reactive oxygen species (ROS) increased their production, which can be as high as a toxic level. Accordingly, glutathione oxidation-reduction, coupled to NADPH reduction-oxidation, is a major antioxidant pathway in cancer cells to cope with this (Boroughs and DeBerardinis, 2015; Vazquez *et al.*, 2016). The fundamental reason behind metabolic alterations of cancer cells can be attributed to the alterations in gene expression, as well as the metabolic enzyme activities (DeBerardinis and Chandel, 2016; Hart *et al.*, 2016).

1.4.2 Metabolomics Techniques

Currently, there are two standard identification techniques for metabolomics: nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). NMR spectroscopy uses isotopes processing property of magnetic spin, providing solution-state molecular structures information based on atom-centered nuclear interactions (Marion, 2013). MS involves an initial separation of metabolites from biological samples using chromatography (capillary electrophoresis (CE), gas chromatography (GC), liquid chromatography (LC) or ultra-high performance liquid chromatography (UPLC)) base on their mass to charge ratio, then identified with a tandem mass spectrometer (Feng *et al.*, 2008). However, the pre-separation stage is not always necessary.

Each technique has its strengths and weaknesses but gives complementary information (Zhang *et al.*, 2012). As summarized in Table 1.2, NMR is more reproducible, simple sample preparation, non-destructive, faster, quantitative, and less expensive, while MS has higher sensitivity, needs less sample, and gives more chemophysical information (Aboud and Weiss, 2013; Capati *et al.*, 2017; Hart *et al.*, 2016; Zhang *et al.*, 2012).

Table 1.2: Summary of the advantages and disadvantages of NMR and MS techniques.

Features	**NMR spectroscopy**	**MS**
Reproducibility	Very good	Possible variation introduced by preparation
Sensitivity	Less sensitive (LOD: 100 µM ≈100 metabolites)	Very sensitive (LOD: 100 nM >1000 metabolites)
Sample amount	More sample is required (500 µl)	Less sample is required (1-10 µl)
Sample intervention	Non-destructive	Destructive
Sample preparation	Minimal (no separation required)	Extensive (separation required, but not always)

Structural information	High	Low
Time consume	~ 10 minutes	~ 30 minutes
Quantification	Easy	Difficult
Chemophysical information	Less information	More information (time separation)
Cost	Less expensive	Expensive

So far, no technique could detect the whole metabolome. According to the aim of the investigation, one of the two approaches is used in metabolomics study: targeted or untargeted (Menni *et al.*, 2017; Suhre and Gieger, 2012). The untargeted metabolomics technique detects previously unpredicted metabolic alterations associated with a specific disease, which is vital for identifying new biomarkers or novel mechanisms. The untargeted metabolomics usually uses NMR in hypothesis-generating studies that compare different groups (Menni *et al.*, 2017; Suhre and Gieger, 2012). The targeted metabolomics technique involves hypothesis-driven experiments and is characterized by measuring sets of metabolites that are predefined and involve in a few metabolic pathways of interest. This approach mostly employs MS and has advantages of easily chemical identification, higher sensitivity, and absolute metabolite quantification (Menni *et al.*, 2017; Patti *et al.*, 2012). As a result, the typical method for metabolomics is to use NMR spectroscopy for untargeted screening and explore interesting pathways. Then employ MS to detect specific metabolites with low concentrations in a targeted way.

1.5 Tumor Markers

As defined by the National Cancer Institute, a tumor marker is “a substance found in tissue or blood or other body fluids that may be a sign of cancer or certain benign (noncancer) conditions. Most tumor markers are made by both normal cells and cancer cells, but they are made in larger amounts by cancer cells” (NCI Dictionary of Cancer Terms). Various tumor markers can be classified into the following categories according to their functions:

1) Screening markers are used to assess the cancer risk, distinguish individuals that are predisposed to particular types of cancers from the population. Typically, these markers are associated with genetic or epigenetic alterations. For instance, mutations of genes *BRCA1* or *BRCA2* indicates a higher risk of breast and ovarian cancer (Mavaddat *et al.*, 2013).
2) Diagnostic markers detect whether a specific tumor exists or not. For example, elevated levels of cancer antigen CA125 and CA19-9 could indicate ovarian cancer and pancreatic cancer, respectively (Haglund, 1986).
3) Prognostic markers can be useful in predicting the course of tumor progression or outcomes and evaluating the survival chance of cancer patients. Tumor progression can be disease-free survival (DFS), distant disease-free survival (DDFS), progression-free survival (PFS), or overall survival (OS). In this project, we will emphasize PFS, a period that the disease remains stable, and OS, time from diagnosis to death.
4) Predictive markers can predict patients' responses to a particular treatment and help to decide the treatment plans. For example, the HER2 status of breast cancer patients is needed to determine the usage and efficacy of trastuzumab treatment.

These tumor markers play essential roles in decision-making and tumor management, which have been demonstrated of significant clinical relevance.

1.5.1 Tumor Markers for Breast Cancer

Since breast cancer is a heterogeneous disease, no universal biomarker is available currently to diagnose and predict all types of breast cancer. Well-studied biomarkers for breast cancer can be categorized into several groups. Hormone receptors, including estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor family (HER2/neu), are the most important markers that assist in making treatment decisions (Harris *et al.*, 2007). However, none of them has satisfactory diagnostic ability because they are not expressing in all breast cancer patients. Proliferation markers, such as Ki67, p53, cyclin D or E, differentially expressed in breast cancer cells, are potential prognostic markers.

However, they are not sufficiently recommended for the route clinical assessment of breast cancer (Harris *et al.*, 2007; Weigel and Dowsett, 2010). The emerging multigene expression tests, such as the commercially available MammaPrint test and Onco*type* DX, as well as Rotterdam Signature, have been developed as breast cancer prognostic markers (Harris *et al.*, 2007; Paik *et al.*, 2004; van 't Veer *et al.*, 2002).

Well-established circulating markers, such as carbohydrate antigens CA15-3, CA27-29, and carcinoembryonic antigen (CEA), are important prognostic markers and are routinely used to monitor MBC relapse and treatment efficacy. However, they are not used for PBC screening or diagnosis (Bast *et al.*, 2001; Ludwig and Weinstein, 2005). The American Society of Clinical Oncology (ASCO) recommended urokinase plasminogen activator (uPA) and plasminogen activator inhibitor-1 (PAI-1) for determining prognosis in patients with newly diagnosed, lymph node-negative breast cancer (Harris *et al.*, 2007). However, they are only recommended when imaging results are available due to the limit of accuracy, so they are not recommended for MBC (Guadagni *et al.*, 2001; Lumachi *et al.*, 2000; Uehara *et al.*, 2008). Other circulating tumor markers, such as metabolites and the US Food and Drug Administration (FDA) approved prognostic marker CTCs, will be discussed in the following sections.

1.5.2 Tumor Markers for Ovarian Cancer

Carcinoembryonic antigen (CEA) was the first described single serum marker for epithelial ovarian cancer in 1976 (Khoo and Mackay, 1976). Cancer antigen 125 (CA125) is the most studied serum biomarker for serous epithelial ovarian cancer (Muinao *et al.*, 2018). It was approved by the FDA in 1981 to monitor cancer for women with a known diagnosis of ovarian cancer, rather than for preoperative use, nor associated with a survival benefit (Ueland, 2017). However, CA125 has sensitive of 47% for patients at their early stages, but a sensitivity of 80%-90% at the late stages (Colakovic *et al.*, 2000; Hogdall, 2008). Moreover, CA125 expresses not only in ovarian cancer but also in breast cancer, colon cancer, endometrial cancer, and pancreatic cancer. Around 20% of ovarian cancer patients do not

have CA125 expressed at all (Muinao *et al.*, 2018). Thus, CA125 is not a specific biomarker. As a supplement to CA125, human epididymis protein 4 (HE4) was cleared by the FDA in 2009 as a serum biomarker to monitor patients with a known diagnosed epithelial ovarian cancer and their recurrence (Ueland, 2017). In addition, HE4 could detect epithelial ovarian cancer 2 to 3 months earlier than CA125 (Ueland, 2017). However, serum HE4 levels are changing with aging, different menopausal status, and pregnancy, which lower the specificity of HE4 (Moore *et al.*, 2012).

The new multivariate index assays, OVA1, Overa, Risk of Ovarian Malignancy Algorithm (ROMA), and combining with several serum markers have been cleared by the FDA to assess the risk of ovarian malignancy (Bristow *et al.*, 2013; Coleman *et al.*, 2016; Moore *et al.*, 2009; Ueland *et al.*, 2011). The individual serum markers are CA125, HE4, transferrin, β-2 microglobulin, apolipoprotein A1, and transthyretin. These multivariate index assays significantly improved preoperative testing sensitivity compared with the single biomarker tests (Bristow *et al.*, 2013). Gene mutations, such as genes of *BRCA1/BRCA2*, *P53*, *KRAS*, *EGFR*, and microRNAs, might be promising markers for ovarian cancer screening, prognosis, and prediction (Bell, 2005; Ingham *et al.*, 2013; Milner *et al.*, 1993; Zhang *et al.*, 2008).

1.6 Circulating Tumor Cells

Circulating tumor cells (CTCs) are cancer cells derive from clones in the primary tumor and migrate around the body via the blood circulation. Thomas Ashworth first observed CTCs in the blood of a man with metastatic cancer in 1869 (Ashworth, 1869). CTCs are the intermediates in the metastatic cascade, where a subpopulation of CTCs works as seeds and is responsible for the dissemination to distant sites, namely the well-known cancer metastasis, as shown in Figure 1.5 (Chaffer and Weinberg, 2011). As revealed by experimental studies in the past decade, CTCs were present in blood of patients with different types of solid carcinomas while absent in blood of healthy individuals and patients without malignant tumors (Allard *et al.*, 2004; Cristofanilli *et al.*, 2004; Ghossein *et al.*, 1999; Racila *et al.*,

1998). Later, circulating tumor cells were proposed as an independent prognostic marker for metastasis, specifically for PFS and OS (Cohen *et al.*, 2009; Cristofanilli *et al.*, 2005; Danila *et al.*, 2007; Giuliano *et al.*, 2011). Enumeration of CTCs assists the prognosis of not only solid metastatic cancer with follow up data but also newly diagnosed cancer patients (Franken *et al.*, 2012).

A cardinal cut off of greater than or equal to five CTCs per 7.5 ml blood has been defined as CTC positive (Gregory *et al.*, 2008), and it was proposed as an indicator of poor prognosis (Cristofanilli *et al.*, 2005). Accordingly, those with less than 5 or no CTCs in 7.5 ml blood are designated as CTC negative, representing good prognosis (Cristofanilli *et al.*, 2005). Additionally, CTCs could also predict the treatment responses in MBC, metastatic prostate cancer, and metastatic colorectal cancer. Therefore CTCs were also proposed as predictive markers (Budd *et al.*, 2006; Cristofanilli *et al.*, 2007; Tol *et al.*, 2010). What is more, the expression profile of the hormone receptor status of CTCs could indicate the real-time phenotype of metastasis, which could help to make appropriate treatment plans, especially for cancer patients at the late stages (Aktas *et al.*, 2011; Fehm *et al.*, 2009).

Because of the low number of CTCs in peripheral blood, an enrichment step has to be done before detection (Zhe *et al.*, 2011). The enrichment step could concentrate CTCs from the milieu of blood cells based on their physical features such as size, density, electric charge and deformability, and biological features such as surface marker or protein expression and invasion capacity (Alix-Panabieres and Pantel, 2014; Pantel and Alix-Panabieres, 2010). Currently, the gold standard of the first and only actionable test for detecting CTCs in cancer patients is the FDA approved CellSearch™ system. This system is based on the positive expression of epithelial markers epithelial cell adhesion molecule (EpCAM) or cytokeratin 8, 18, and/or 19, and the negative expression of leukocyte CD45 on CTCs (Allard *et al.*, 2004).

However, it is of importance to note that a significant fraction of patients with overt distant metastases is negative for CTCs. This could be partly explained by EMT of CTCs, in which case they can be missed by enumeration techniques that exploit the expression of epithelial markers such as EpCAM or cytokeratin (Cristofanilli *et al.*, 2004). The absence of CTCs in metastatic cancer patients gives rise to false-negative results and is the major drawback of the CellSearch™ system. Besides, leucocytes in the activated state have

epithelial markers expressed as well, thus increasing the false positive rate of CTCs (Allan and Keeney, 2010). Moreover, enrichment methods based on physical properties such as size and density have low sensitivity (Konigsberg *et al.*, 2011; Ring *et al.*, 2005). As a result, current methods for CTC detection can lead to under- or over-estimation of CTC number, so novel prognostic markers are highly needed.

1.7 Blood Metabolites as Tumor Markers

Cancer metabolomics studies with sample types of cell, tissue, urine, and blood have been reported, which have generated metabolite panels for various applications (Gunther, 2015; Wang *et al.*, 2018a; Wang *et al.*, 2016b; Xiao and Zhou, 2017). However, tissue biopsy has limitations of high invasive and heterogeneous, as well as the ineffectiveness in understanding metastatic risk, disease progression, and treatment efficiency (Marrinucci *et al.*, 2009). Urine samples can be easily affected by confounding factors and contaminated by microbiota, which makes it not the perfect sample type in biomarker study (Rodrigues *et al.*, 2016). The minimally invasive nature of blood samples (Crowley *et al.*, 2013; Haber and Velculescu, 2014), the sensitive feature of metabolite, and the changes of metabolites in breast cancer initiation and progression make blood-based metabolites attractive biomarker candidates (Gunther, 2015). Indeed, blood metabolites associated with cancer risk have been studied in most cancers including breast (Fan *et al.*, 2016b; Jasbi *et al.*, 2019; Yuan *et al.*, 2018), ovarian (Fan *et al.*, 2012; Li *et al.*, 2017a; Zhou *et al.*, 2010), prostate (Derezinski *et al.*, 2017), colorectal (Shu *et al.*, 2018a; Zaimenko *et al.*, 2019), lung (Kumar *et al.*, 2017; Xiang *et al.*, 2018), pancreatic (Di Gangi *et al.*, 2016; Kobayashi *et al.*, 2013; Shu *et al.*, 2018b), and etc.

Various studies have found promising applications for blood metabolites as detective, diagnosis, and prognostic markers for breast cancer. Some studies suggest the decreased levels of amino acids in breast cancer patients compared to healthy controls (More *et al.*, 2018; Shen *et al.*, 2013; Zhou *et al.*, 2017), and their increased levels in metastatic relative to early-stage breast cancer patients (Jobard *et al.*, 2014; Oakman *et al.*, 2011). On the contrary, higher levels of lipids were observed in breast cancer patients than in healthy controls (Shen

et al., 2013; Zhou *et al.*, 2017), whereas lower levels of lipids were investigated in metastatic breast cancer than in early-stage breast cancer (Oakman *et al.*, 2011). Apart from detection and prognosis, blood metabolite alterations have been linked to breast cancer treatment or chemotherapy response as well, which indicated the potential of metabolites as predictive markers (Stebbing *et al.*, 2012; Tenori *et al.*, 2012; Wei *et al.*, 2013). Reports also uncovered the dependence of circulating metabolites profile on subgroups by examining breast cancer and healthy control blood from individuals with different molecular subtypes (Fan *et al.*, 2016b; Hart *et al.*, 2017). However, studies about breast cancer plasma metabolites sometimes gave controversial results. For example, a lower serum level of glutamic acid in breast cancer patients was observed by Zhou et al. than in healthy controls, while Wang et al. showed an increased level (Wang *et al.*, 2018b; Zhou *et al.*, 2017). Moreover, two studies showed that lipids had higher amounts in the blood of metastatic breast cancer patients than early-stage ones (Jobard *et al.*, 2014; Tenori *et al.*, 2015), whereas another study indicated the opposite result (Oakman *et al.*, 2011).

Similar to other cancers and diseases, studies have found promising applications for blood metabolites as diagnostic, prognostic, or predictive markers in ovarian cancer, especially in epithelial ovarian cancer, the most common type of ovarian cancer (Turkoglu *et al.*, 2016). Researchers have revealed the relationship between blood metabolites and ovarian cancer risk with mass spectrometry (Bachmayr-Heyda *et al.*, 2017; Plewa *et al.*, 2019; Yang *et al.*, 2018). The pilot study to explore the relevance of plasma metabolites in ovarian cancer found an association between levels of plasma phosphatidylcholine, lysophosphatidylcholine, phosphatidylinositol, triglyceride, and sphingolipid, and ovarian cancer recurrence (Li *et al.*, 2017a). However, another study found some amino acids are correlated with ovarian cancer recurrence additional to lipids (Zhang *et al.*, 2015). Kynurenine, acetylcarnitine, PC (42:11), and LPE (22:0/0:0) were selected as potential predictive markers for ovarian cancer in another study, with plasma samples followed up for three years (Xie *et al.*, 2017). Furthermore, combining four plasma metabolites with the currently used predictive marker CA125 has been demonstrated to be more accurate than CA125 alone for ovarian cancer (Buas *et al.*, 2016).

These studies suggest the possibility of using blood-based metabolic markers for the detection, diagnosis, and prognosis of breast cancer and ovarian cancer, irrespective of full understanding of the pathophysiological mechanisms. However, evidence is still limited. Optimized marker panels are yet to be developed and tested in large populations by high-quality studies and further validated in prospective studies.

2. Aims and Objectives

2.1 Aims

Breast cancer is the most common cancer type in women worldwide and causes most of the cancer-related female deaths, which is mostly attributed to metastasis. Ovarian cancer has the second leading mortality rate among gynecological cancer. Detecting cancer at an early stage is expected to improve the survival rate. In addition, early detection and better prognostication of metastasis could improve patients' outcomes and lower the high mortality rate. Liquid biopsy has advantages over the tissue sample because it can be accessed with minimally invasive procedures and monitored continuously. Cancer can cause overt metabolic alterations in cancer cells and the surrounding milieu. Blood metabolites have shown great capability as highly sensitive and informative markers in cancer, including breast cancer and ovarian cancer. The thesis aims to explore the potential of circulating metabolites as diagnostic and prognostic markers in breast cancer and ovarian cancer using a case-control approach. Plasma metabolites with possible diagnostic and prognostic values were identified in an initial discovery cohort and further tested in a validation cohort. Besides, the prognostic capabilities of these metabolites were compared to that of circulating tumor cells, an established prognostic marker for MBC patients.

2.2 Objectives

This project includes the following objectives:

1) Identify significantly different metabolites that could distinguish:
    a. PBC patients from healthy controls
    b. MBC patients from healthy controls
    c. MBC patients from PBC patients
    d. OVCA patients from healthy controls

e. OVCA patients from PBC patients

2) Identify plasma metabolites that correlate with progression-free survival and overall survival, which is thus of prognostic significance in MBC patients.

3) Evaluate the potential of metabolites as putative biomarkers for early diagnosis and prognosis.

3. Materials and Methods

3.1 Study Samples

All studies were performed following the Declaration of Helsinki and approved by the Ethical Committee of Medical Faculty, University Heidelberg. In total, this project includes four different categories of blood samples. All subjects shared the same gender (female) and ethnic (Caucasian origin) background. They were provided written informed consent for the use of their blood samples and data.

Healthy Controls - Healthy control samples from BIOMARKER Study

The healthy controls consisted of individuals without clinically diagnosed malignancies or autoimmune diseases currently or historically, or current inflammation. Before blood withdraws, each volunteer was asked to finish a questionnaire under the supervision of study representatives regarding their lifestyle data. (Study Ethic number: S-175/2010, University of Heidelberg, Heidelberg, Germany)

PBC - Primary breast cancer patients from the GENOM Study

The PBC category included patients with sporadic and at the first diagnosis of breast cancer. Blood samples were collected before they underwent any therapeutic procedures, e.g., surgery, radiation, or systemic therapy. Histopathological and clinical-pathological features were determined by tumor tissues obtained from both the initial biopsy and surgical resection. For those who received neoadjuvant therapy, results from the former were regarded as valid only when the latter was not available. Histopathological and clinical pathology features of PBC patients were summarized in Table A1. (Study Ethic number: S-039/2008, University of Heidelberg, Heidelberg, Germany)

MBC - Metastatic breast cancer patients from CTC Study

Breast cancer patients with one or more sites of metastasis were recruited, which were confirmed by radiological investigations. All MBC patients have received or were receiving therapy for their metastatic tumors before initial blood withdraw. Histopathological and clinical-pathological features of both their primary and metastatic tumors were recorded. Tumor progression status was radiographically monitored with around three-month intervals after each round of therapy. According to the RECIST guidelines, results were classified into complete or partial response (tumor size decrease), stable disease (tumor size neither increase nor decrease), or progressive disease (tumor size increase) (Eisenhauer *et al.*, 2009). All the MBC patients were followed up after recruitment for up to 30 months. Histopathological characteristics and clinical pathology features of their metastatic tumors were summarized in Table A2. (Study Ethic number: S-295/2009, University of Heidelberg, Heidelberg, Germany)

OVCA - Ovarian cancer patients

The OVCA patients were recruited with the same standards as the PBC patients from the GENOM Study. Histopathological and clinical-pathological features of OVCA patients were summarized in Table A3. (Study Ethic number: S-052/2010, University of Heidelberg, Heidelberg, Germany).

3.2 Blood Sample Processing

Peripheral blood was drawn from all study participants into 9 ml EDTA tubes (S-Monovette®, Sarstedt, Nümbrecht, Germany) by nurses. Blood was centrifuged at 1300 g with a precooling centrifuge at 10°C for 20 min. The resulting supernatant of the fractionated blood, i.e. plasma, was transferred to 2 ml vials. The fractionated blood supernatant was transferred into 2 ml vials and was centrifuged at 15500 g with a precooling centrifuge at 10°C for 10 min, then a white pellet was observed. The supernatant plasma was pipetted out without touching the pellet and aliquoted in cryogenic vials. This step was done to ensure that the plasma was completely cell or cell debris free. All plasma aliquots were quickly frozen and stored at -

80°C until the execution of metabolic analyses. The blood processing procedures for all sample categories are summarized in Figure 3.1.

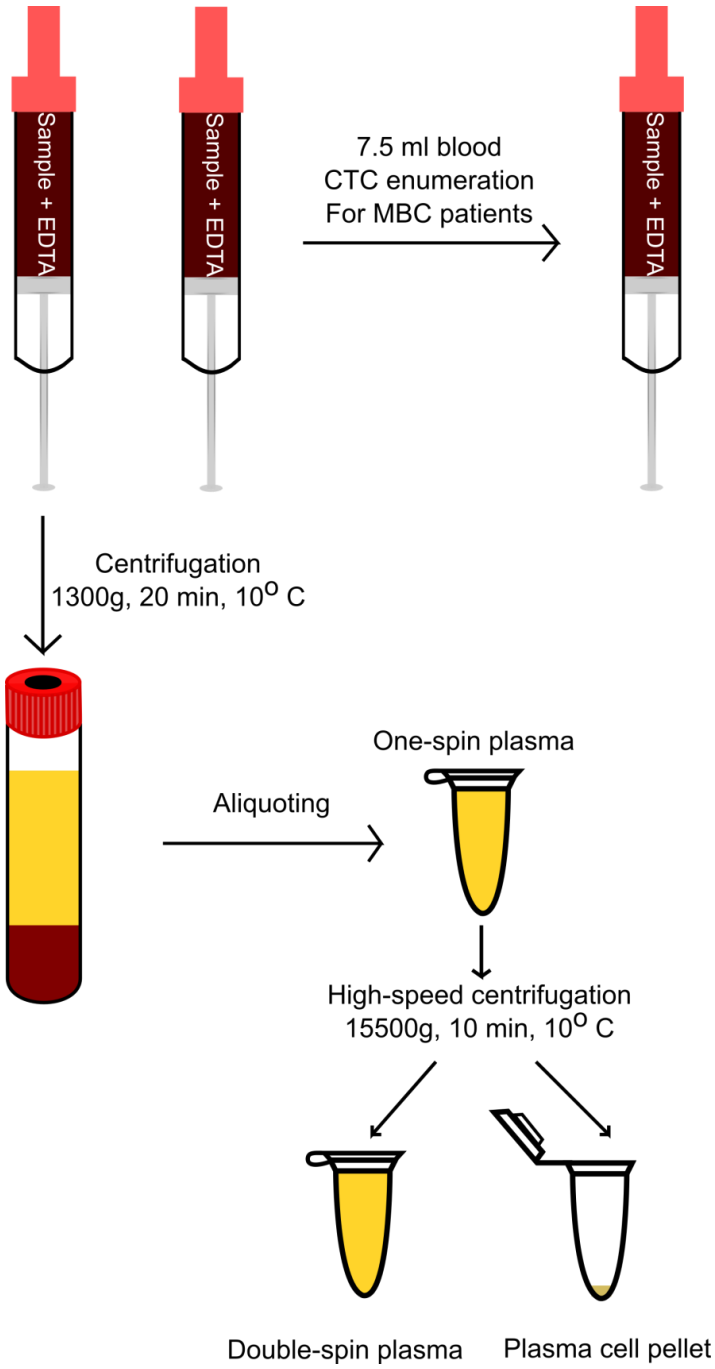


Figure 3.1: Blood sample pre-processing pipeline. Plasma is isolated by a two-step centrifugation protocol to remove cells and cell debris.

3.3 CTC Detection and Enumeration

An additional 10 ml blood from MBC patients was collected into CellSave™ tubes (Veridex, LLC, Raritan, NJ), of which 7.5 ml was used to enumerate intact, apoptotic, and enucleated CTCs. Blood samples were processed on the CellTracks™ AutoPrep system using the CellSearch™ circulating tumor cell kit (Veridex, LLC, Raritan, NJ), and CTC numbers were evaluated by the CellTracks™ AnalyserII (Veridex, LLC) (Veridex, 2010). CTC enumeration was done by our collaborators at the Department of Tumour Biology, University Hospital of Hamburg-Eppendorf. According to the number of CTCs, MBC patients were further classified as:

- CTC positive (CTCpos-MBC): more than or equal to 5 intact CTCs per 7.5 ml blood.
- CTC negative (CTCneg-MBC): less than 5 intact CTCs per 7.5 ml blood.

3.4 Metabolites Quantification by Targeted Metabolomics

3.4.1 Sample Preparation and the Absolute*IDQ*® p180 Kit

Plasma samples were thawed on ice, centrifuged, and aliquots of 10 µL were used for targeted quantitative metabolite analysis. Samples were prepared and analyzed using the Absolute*IDQ*® p180 Kit (Biocrates Life Sciences AG, Innsbruck, Austria) in strict accordance with the user manual. This kit was designated to target 188 common metabolites in human samples, including 21 amino acids (AAs), 21 biogenic amines (BAs), 40 acylcarnitines (ACs), 76 phosphatidylcholines (PCs), 14 lysophosphatidylcholines (lysoPCs), 15 sphingomyelins (SMs), and the sum of hexoses (H1) (Table A4; Table 3.1). The kit is high throughput, reproducible, translatable, and needs a small sample volume.

Table 3.1: Metabolite classes in the Biocrates Absolute*IDQ*® p180 kit (adapted from the Biocrates product information sheet).

Metabolite classes	Numbers	Biological relevances (representatives)
Amino Acids	21	• Immune regulation • Nutritional status • Energy homeostasis
Biogenic Amines	21	• Detoxification • Cell cycle control • Neurotransmission
Acylcarnitines	40	• Mitochondrial function • Fatty acid oxidation
(Lyso-)Phosphatidylcholines	90	• Lipoprotein structure • Inflammation
Sphingolipids	15	• Membrane composition • Neuroprotection
Hexoses	1	• Insulin resistance • Glycolysis and gluconeogenesis

3.4.2 Metabolites Quantification

In brief, added 10 µL of the internal standard solution, which serves as a reference for quantification, to each well on a filtering spot of the 96-well extraction plate. Then 10 µL of each plasma sample, low/medium/high-quality control (QC) samples, blank, zero samples, or calibration standard were added to the appropriate wells (St John-Williams *et al.*, 2017). Nitrogen of a gentle stream was used to dry the plate. Samples were derivatized for AAs and BAs with phenyl isothiocyanate. Then with 5 mM ammonium acetate in methanol, sample extracts were eluted. Furthermore, sample extract dilution was performed with either kit running solvent for flow injection analysis tandem mass spectrometry (FIA-MS/MS) (5:1) or 40% methanol in water for the UPLC-MS/MS analysis (2:1). Absolute quantification of AAs

and BAs was performed by UPLC-MS/MS analysis. ACs, SMs, and (lyso)PCs were analyzed by FIA-MS/MS. Peak integration, calibration, and concentration calculations of the UPLC-MS/MS data were done by TargetLynx™ (Waters Corporation). Then Biocrates' Met*IDQ*™ software was used to analyze both the UPLC-MS/MS data processed by TargetLynx™ and the data generated from FIA-MS/MS. All metabolite concentrations are reported in micromolar.

3.4.3 Quality Controls and Technical Validation

All samples were randomized and evenly distributed on seven kit plates. On each plate, one low-level QC, five medium levels QC, and one high-level QC samples were measured. To prove the validity of the run, and to verify the authentic performance of the applied quantitative procedure, technical validation of each analyzed kit plate was done automatically. Met*IDQ*™ software was used to perform inter-plate technical validation (St John-Williams *et al.*, 2017). All measured plates passed the technical validation, which confirmed the high quality and accuracy of the quantitative metabolomics data obtained. In order to minimize batch effects, normalization using medium level QC measurement was applied. Normalization procedure was performed by Met*IDQ*™ software too.

3.5 Statistical Analysis

For each sample category, two independent study cohorts were included, the discovery cohort of 314 individuals included 100 MBC, 80 PBC, and 34 OVCA patients, and 100 healthy controls to identify potential candidate metabolites, while the validation cohort of 237 samples consisted of 78 MBC, 109 PBC, and 35 OVCA patients, and 50 healthy controls. Among the 80 PBC samples in the discovery cohort, three with confirmed late-stage breast cancer and one with breast angiosarcoma, so these four samples were removed from the downstream analysis. Among the 78 MBC patients in the validation cohort, CTC status information was lacking for two, and they were excluded when identifying prognostic

markers for MBC. Additionally, four samples did not feature any FIGO staging information within the 35 OVCA samples in the validation cohort. The types and number of samples were listed in Table 3.2. All statistical analysis was performed with R3.5.1 (Team, 2015).

Table 3.2: Overall of the type and number of samples used in this thesis work.

Plasma samples	Discovery cohort	Validation cohort	Clinical features
Healthy Controls	**100**	**50**	
Primary Breast Cancer (PBC)	**80**	**109**	**Table A1**
Stage 0&I	25	38	
Stage II	48	62	
Stage III	3	9	
Metastatic Breast Cancer (MBC)	**100**	**78**	**Table A2**
Good prognosis (PFS and OS ≥ 16 months)	8	23	
Poor prognosis (PFS or OS < 3 months)	26	17	
CTC positive (≥ 5 intact CTC/7.5ml blood)	44	21	
CTC negative (< 5 intact CTC/7.5ml blood)	56	55	
Ovarian Cancer (OVCA)	**34**	**35**	**Table A3**
Stage I	2	5	
Stage II	1	5	
Stage III	29	15	
Stage IV	2	6	

3.5.1 Quality Control

Metabolites were excluded when the percentage of concentration values, missing or below the limit of detection (LOD), was higher than 20%, and when batch coefficients of variation were larger than 0.3. Five healthy and five OVCA plasma samples from the discovery cohort were measured in triplicates to test the reliability of the mass spectrometry generated metabolite data. Metabolite SM C22:3 had coefficients of variation larger than 0.3 and, therefore, it was excluded from the following statistical analysis (Figure A1). After filtering

out metabolites with percentages of LOD or missing values higher than 20% (n = 50), the remaining values that below LOD or missing were replaced by imputed values between LOD and LOD/2 following equal distribution. At last, the remaining 138 unique and variably expressed metabolites were used for subsequent comparisons. All metabolite concentration values were log2 transformed and then normalized by the metabolite-wise batch standardization.

3.5.2 The Difference in Metabolites between Groups

Age was the only clinical characteristic available in all study groups and was unequally distributed among different groups, especially in the validation cohort (Figure A2). Human serum metabolic profiles are age-dependent (Yu *et al.*, 2012), which was also verified in our study (Figure A3). Therefore, age was always adjusted in the univariate logistic regression analysis, which aimed to identify significantly altered metabolites between study groups or subgroups in the discovery cohort. The resulting *p*-values were adjusted for multiple testing by controlling for the false discovery rate (FDR; Benjamini-Hochberg method). All statistical tests were two-sided; FDR below 0.05 were regarded as statistically significant.

The identified metabolites were further validated by age-adjusted univariate logistic regression analysis in an independent validation cohort. Only metabolites with FDR less than 0.05 were chosen for downstream analysis. To compute the least redundant and most informative panel of metabolites that could discriminate different groups, penalized LASSO logistic regression models (ten-fold cross-validation to tune penalty parameter) were built (Tibshirani, 1996). This process was implemented in the R package “glmnet”. The corresponding area under the curve (AUC) was calculated for each multivariable model, as well as the sensitivities at predefined specificities with corresponding 95% confidence intervals (CIs). The performance of each metabolite was evaluated by using a ten-fold cross-validated receiver operating characteristic (ROC) curve based on the logistic regression model.

3.5.3 Correlation between Variables

Relationships between investigated metabolites were evaluated by partial Spearman correlation, which could adjust the remaining variables when calculating the correlation between any two metabolites. Relationships between metabolite concentrations and CTC numbers were analyzed by Spearman rank correlation.

3.5.4 Correlation with Survival

Period (in months) from blood withdrawn to the progression of the disease or last radiological examination was defined as PFS, while to death or last visit was regarded as OS. Each metabolite was dichotomized as the lowest quartile and the residual quartiles based on its concentration. Then the log-rank test was applied on the stratified data to compare the probability of survival between these two groups. Kaplan-Meier curve was constructed at the same time to demonstrate the relationship between individual metabolite levels and survival time. Metabolites that showed significant associations with survival in the univariate log-rank tests were imported into a multivariate LASSO cox model. Relevant metabolite variables were selected automatically with the LASSO penalty (ten-fold cross-validation to tune the penalty parameter) (Tibshirani, 1996).

Comparison of survival models

Survival models, including metabolite variables alone, CTC status alone, or metabolite variables with CTC status, were built. The 0.632+ bootstrap was used to evaluate the prognostic values of models by estimating the prediction error of LASSO cox models. The resulting values were summarized as the integrated prediction error (IPE) curve up to 30 months (Gerds and Schumacher, 2007). Then IPEs of the models were compared to determine the best model. The lower of the IPE value of a model indicates the higher model accurate, which fits the data better.

3.5.5 Correlation with Clinical Characteristics

When calculating the relationships between metabolite levels and clinical characteristics, different tests were used according to the data type. For categorical and binary data, Wilcoxon rank-sum tests were used. Spearman correlation permutation tests fitted the quantitative and continuous data better, while Jonckheere-Terpstra tests were the best for ordinal data. For clinical features that are distributed differently between CTCpos-MBC and CTCneg-MBC sub-groups, logistic regression analysis with covariates adjusted was done. In this way, authentic significance contributed by metabolites in differentiating between groups or the actual association between metabolites and survival could be evaluated.

3.6 Pathway Analysis

Associated metabolic pathways that these significantly different metabolites between PBC and healthy controls participated in were enriched by MetaboAnalyst 3.0 (Xia *et al.*, 2015). This online software MetaboAnalyst 3.0 imported the concentration data of metabolites, and the standard metabolite names were human metabolome database (HMDB) IDs. The pathway library of “Homo sapiens” was selected. Pathway analysis algorithm of “Global test” was used for enrichment analysis, while the algorithm of “Out-degree Centrality” was used for topology analysis.

4. Results

4.1 Early Detective Value of Plasma Metabolites in PBC Patients

4.1.1 Global plasma metabolite profiles of PBC patients and healthy controls

Estrogen receptor (ER) status and age were not equally distributed between discovery and validation cohorts, with the *p*-values for chi-square test were 0.047 and 0.001, respectively (Table A2). Fold changes (FCs) were calculated to show the distribution trend of metabolites and univariate logistic regression models were constructed to distinguish statistically different metabolites ($p < 0.05$) between PBC patients and healthy controls, with age was adjusted for its effect. Resulting *p*-values were adjusted for multiple comparisons. In the discovery cohort, between PBC patients and healthy controls, a total of 37 metabolites were significantly different, with six up-regulated metabolites and 31 down-regulated metabolites observed in PBC patients. In the validation cohort, 24 metabolites were significantly different, of which 21 metabolites were found to be down-regulated while three metabolites up-regulated in PBC patients (Figure 4.1) (Yuan *et al.*, 2018).

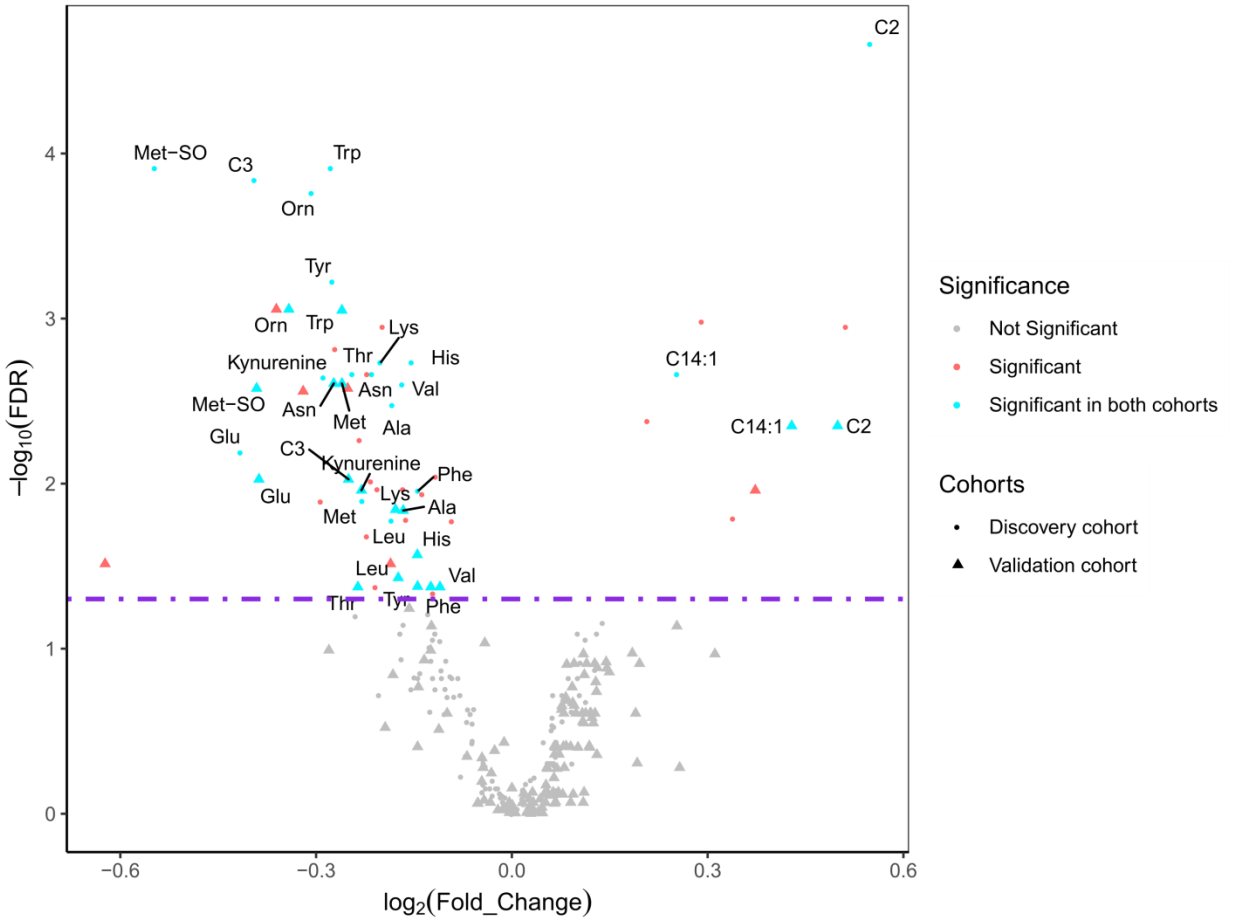


Figure 4.1: The volcano plot shows the comparison between PBC patients and healthy controls. The Y-axis represents -log10 (FDR: adjusted *p*-value with Benjamini-Hochberg), while the horizontal axis represents log2 (fold change: PBC patient versus healthy control), each spot symbolizes one metabolite. The dotted horizontal purple line highlights the statistical significance threshold (FDR < 0.05). Round spots represent metabolites in the discovery cohort, while triangular spots represent metabolites in the validation cohort. Metabolites that are not significantly altered are shown in grey, while significant metabolites are shown in red. Those that are prominent in both cohorts are shown in cyan and marked with text.

Among these significantly altered metabolites, 18 metabolites were present in both discovery and validation cohorts (Table 4.1). Compared to healthy controls, 13 amino acids, two biogenic amines, and one acylcarnitine of downregulation, two acylcarnitines of

upregulation were detected in PBC patients. A heatmap visualized sample clustering based on the expression levels of these 18 metabolites with Euclidean distance metrics showed different clustering patterns of up-regulated and down-regulated metabolites (Figure 4.2) (Yuan *et al.*, 2018).

Table 4.1: List of significantly different levels of metabolite between PBC patients and healthy controls in the discovery and validation cohort.

Metabolites (18)	Discovery cohort			Validation cohort		
	***p*-value**	**FDR**	**FC**	***p*-value**	**FDR**	**FC**
Ala ↓	4,59E-04	3,33E-03	-1,13	1,59E-03	1,45E-02	-1,12
Asn ↓	2,45E-04	2,16E-03	-1,16	7,46E-05	2,48E-03	-1,21
Glu ↓	1,03E-03	6,45E-03	-1,33	7,74E-04	9,41E-03	-1,31
His ↓	1,60E-04	1,83E-03	-1,11	3,14E-03	2,70E-02	-1,11
Leu ↓	3,99E-03	1,67E-02	-1,14	5,10E-03	3,72E-02	-1,13
Lys ↓	1,52E-04	1,83E-03	-1,15	1,48E-03	1,44E-02	-1,13
Met ↓	2,67E-03	1,27E-02	-1,17	8,48E-05	2,48E-03	-1,20
Orn ↓	6,26E-06	1,73E-04	-1,24	1,20E-05	8,79E-04	-1,27
Phe ↓	2,15E-03	1,10E-02	-1,10	6,95E-03	4,23E-02	-1,09
Thr ↓	2,51E-04	2,16E-03	-1,18	6,72E-03	4,23E-02	-1,18
Trp ↓	2,04E-06	1,22E-04	-1,21	1,83E-05	8,91E-04	-1,20
Tyr ↓	2,59E-05	5,95E-04	-1,21	6,04E-03	4,20E-02	-1,11
Val ↓	3,26E-04	2,50E-03	-1,12	6,79E-03	4,23E-02	-1,08
Kynurenine ↓	2,79E-04	2,27E-03	-1,22	1,05E-03	1,10E-02	-1,17
Met-SO ↓	2,65E-06	1,22E-04	-1,46	1,27E-04	2,65E-03	-1,31
C3 ↓	4,18E-06	1,44E-04	-1,31	7,59E-04	9,41E-03	-1,19
C2 ↑	1,56E-07	2,16E-05	1,46	3,06E-04	4,47E-03	1,41
C14:1 ↑	2,40E-04	2,16E-03	1,19	2,91E-04	4,47E-03	1,35

Note: The arrows indicate the relative increase/decrease of metabolite levels in PBC patients compared to healthy controls. p-values were adjusted for Benjamini-Hochberg correction. The criteria for selecting significantly altered metabolites were: FDR <0.05.

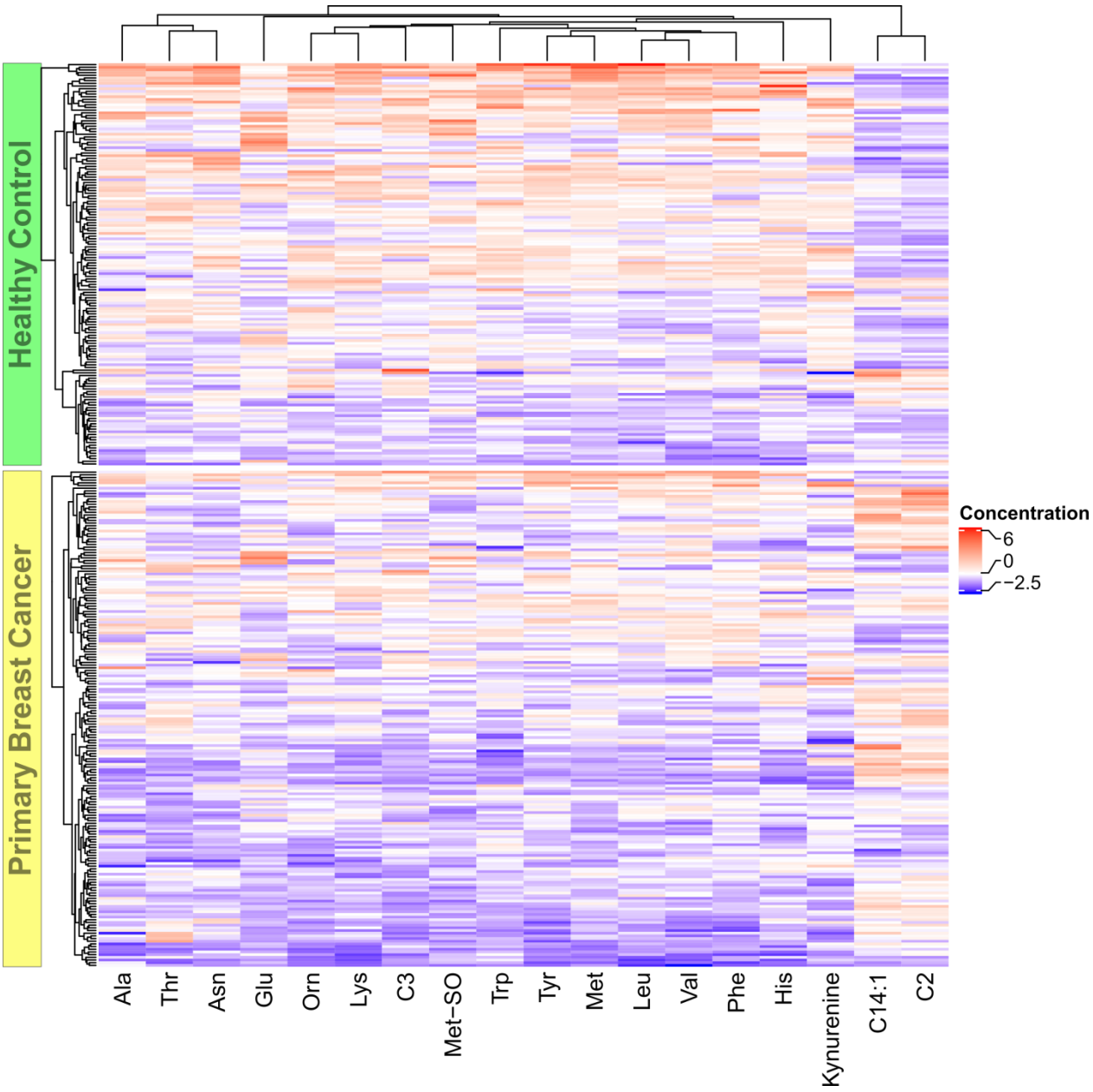


Figure 4.2: Supervised hierarchical complete linkage clustering with Euclidean distance metric, depending on the 18 significantly different metabolites between PBC patients and healthy controls.

4.1.2 A panel of metabolites associated with breast cancer early diagnosis and detection

The diagnostic potential of metabolite panel was assessed through more advanced statistical tools. A multi-parameter ROC curve that established through a penalized LASSO logistic regression model (adjusting penalty parameters through 10-fold cross-validation) was used to build the least redundant and most informative metabolite panel. This panel was able to distinguish between PBC patients and healthy controls. The discovery cohort was used to build a robust statistical model, while the validation cohort was used for model selection and final classification model testing. The corresponding AUC was calculated for the LASSO logistic regression model. Finally, a multivariate model containing seven metabolites (Glu, Orn, Thr, Trp, Met-SO, C2, C3) was predicted, which produced high discrimination accuracy with AUCs of 0.87 (95% CI: 0.81 ~ 0.92) and 0.80 (95% CI: 0.71 ~ 0.87) for the discovery and validation cohorts, respectively (Figure 4.3) (Yuan *et al.*, 2018).

The leave-one-out cross-validated ROC curve built with the logistic regression model revealed that each metabolite in the panel could distinguish PBC patients from healthy controls (Figure 4.4). However, the combination of these seven metabolites performed better than individual metabolites (Figure 4.3). The AUC of each metabolite in the validation group was as follows: Glu (AUC 0.66), Orn (AUC 0.68), Thr (AUC 0.66), Trp (AUC 0.73), Met-SO (AUC 0.68), C2 (AUC 0.72), and C3 (AUC 0.64) (Yuan *et al.*, 2018).

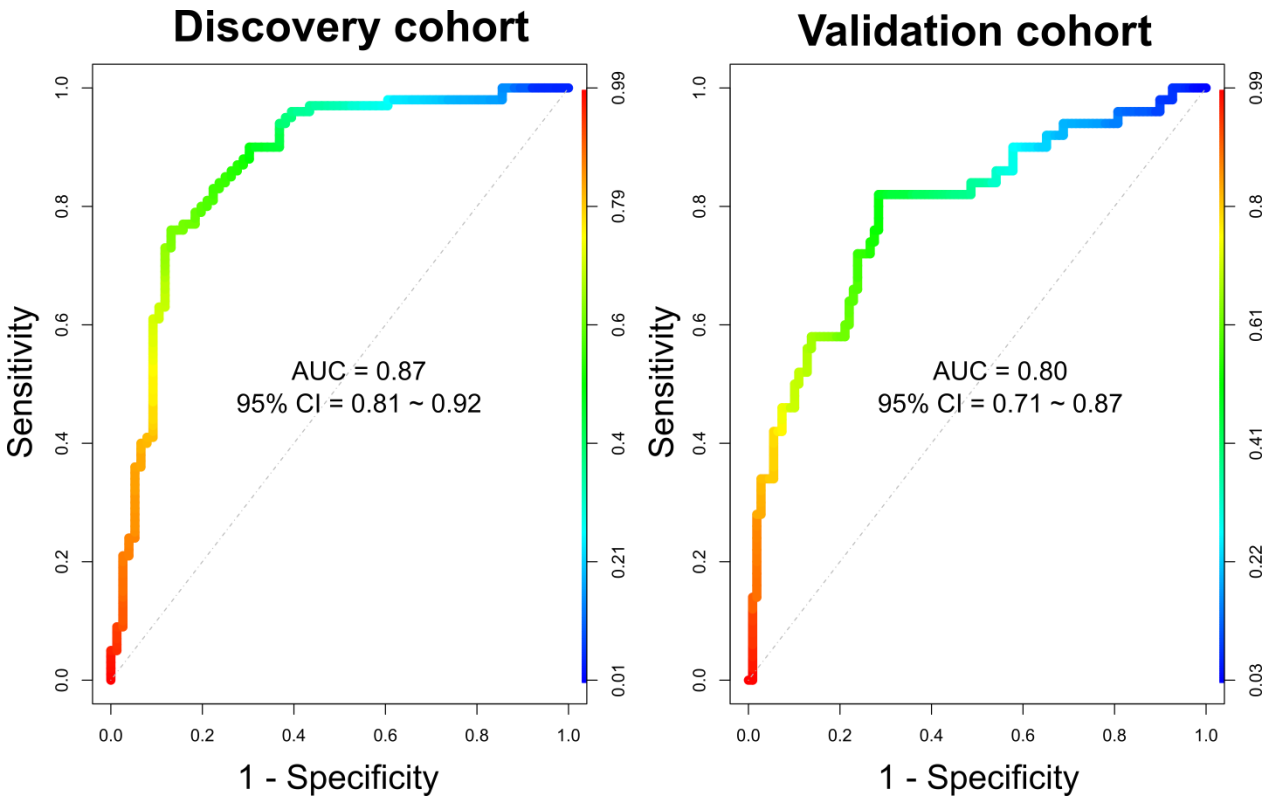


Figure 4.3: Multi-parameter ROC classifier curves for distinguishing between PBC patients and healthy controls. The multi-parameter panel with seven variables selected by the penalized LASSO logistic regression model shows the best performance for the discovery and validation cohorts. The solid colored line typifies the discriminative power of the classifier and corresponding AUC value and 95% CI are displayed. ROC curve is colored according to cutoff values, and scale is shown on the right side of the graph. The x-axis and the y-axis represent the false positive rate (1-specificity) and the true positive rate (sensitivity), respectively. The diagonal dashed gray line means no separation ability, i.e. random classification.

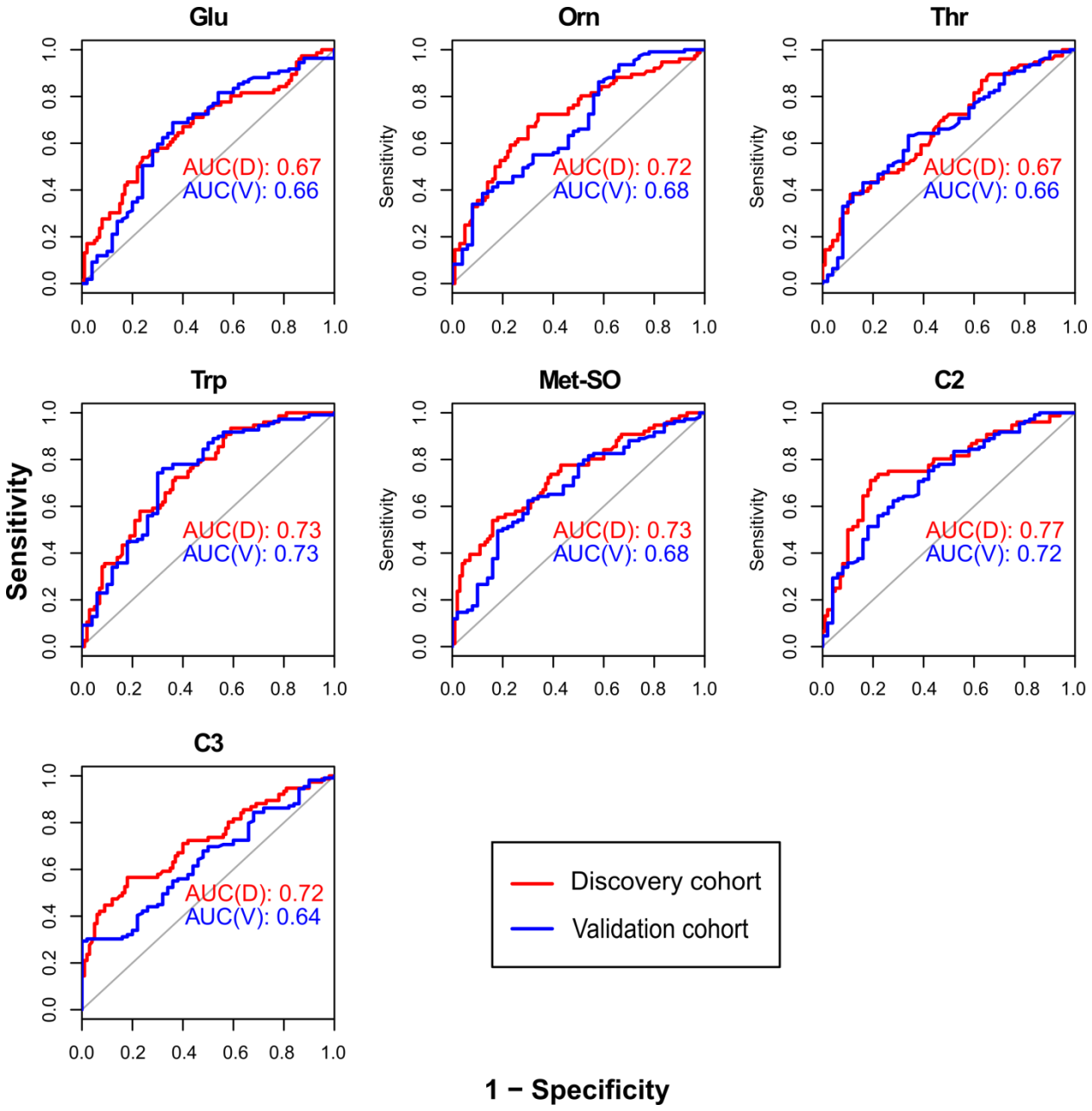


Figure 4.4: Univariate ROC curves for leave-one-out cross-validated logistic regression models based on the ability of individual metabolites to distinguish between PBC patients and healthy controls. The ROC curves for the discovery and validation cohorts are displayed in red and blue lines, respectively. The diagonal gray line means no separation ability, i.e. random classification.

In this thesis project, most of the PBC patients were at stage I or stage II (71 of 80 in the discovery cohort and 99 of 109 in the validation cohort). The logistic regression model, with only stage I and II breast cancer patients, was established to validate the potential classification functions of these seven metabolites further. The multi-parameter ROC curve had an AUC of 0.80 (95% CI: 0.71 ~ 0.87), which was equivalent to the AUC of all PBC patients. AUCs of individual metabolites from stage I and II breast cancer patients were similar to those from all PBC patients. The metabolite panel had better breast cancer detection capabilities than each metabolite in the panel as well (Yuan *et al.*, 2018).

4.1.3 Pathway analysis

With metabolite concentration data, pathway enrichment analysis was performed to compare the metabolism of PBC patients to healthy controls. As a result, significant differences were observed between the two groups (Figure 4.5). In total, a list of 28 metabolic pathways was connected with significantly differentiated metabolites, among which 12 pathways were significant ($p < 0.05$). These pathways mostly involve acid metabolism, aminoacyl-tRNA biosynthesis, and nitrogen metabolism (Yuan et al., 2018). These pathways mostly involve acid metabolism, aminoacyl-tRNA biosynthesis, and nitrogen metabolism (Yuan *et al.*, 2018).

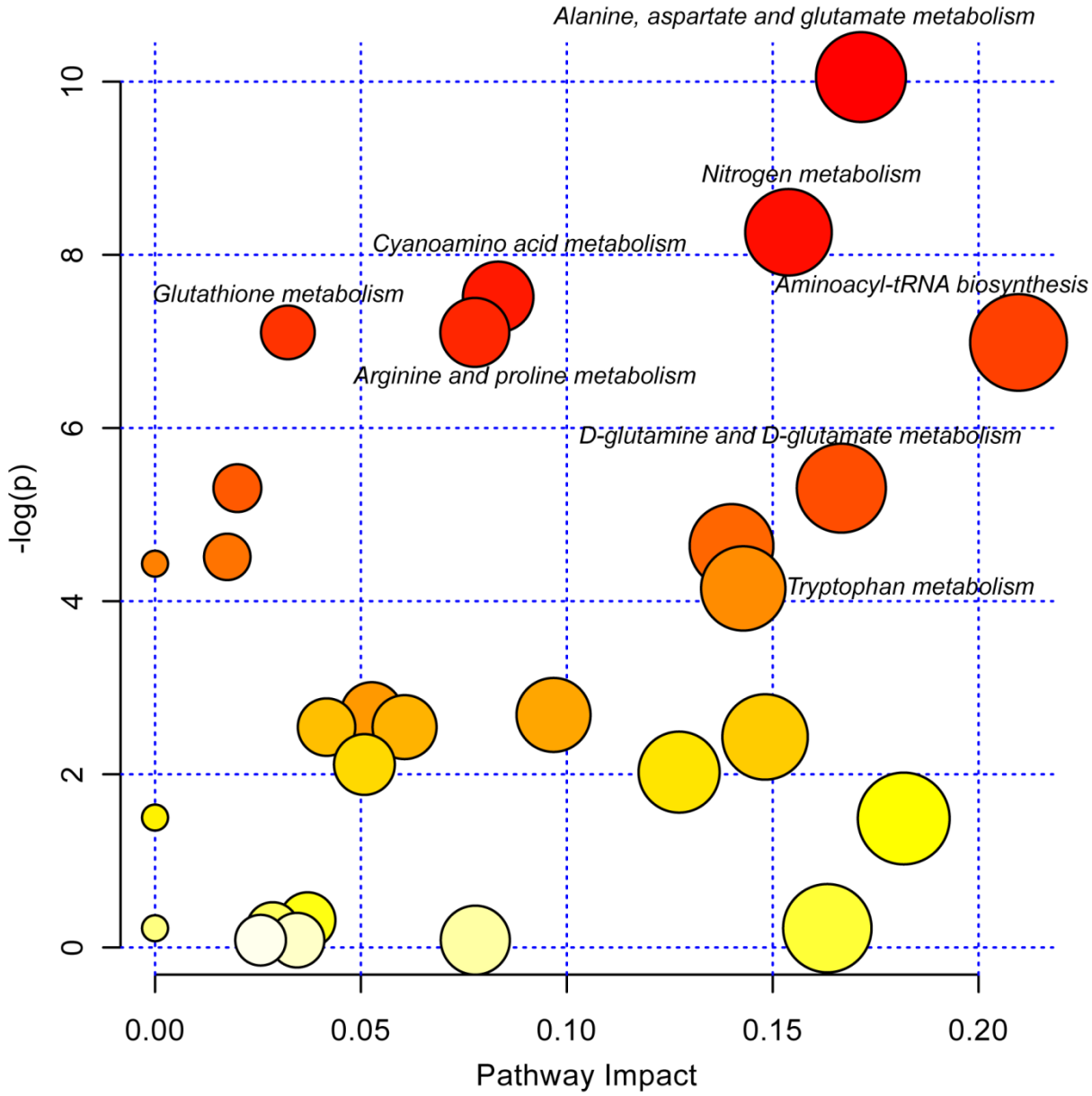


Figure 4.5: All matched pathways, according to the *p*-values from pathway enrichment analysis (y-axis) and pathway impact values obtained from pathway topology analysis (x-axis), for metabolites that could distinguish PBC patients from healthy controls. The color and size of each circle represent the *p*-value and pathway impact value, respectively. Pathway analysis was performed using the MetaboAnalyst 3.0 online tool.

4.2 Diagnostic Value of Plasma Metabolites in MBC Patients

4.2.1 Plasma metabolite profiles of MBC, PBC patients, and healthy controls

In the discovery cohort, plasma samples consist of 100 MBC patients (44 CTC-positive and 56 CTC-negative), 76 PBC patients, and 100 healthy controls. In CTC-positive MBC patients, 82 metabolites were found to be differentially expressed in comparison to healthy controls, whereas, only 17 metabolites were differentially expressed in CTC-negative MBC patients when compared to healthy controls. Similar results were obtained when comparing MBC to PBC patients. Indeed, 62 significantly different metabolite levels were observed after comparing CTC-positive MBC to PBC patients, while only 19 metabolite levels were significantly different between CTC-positive MBC and PBC patients.

Significant differentially expressed metabolites identified in the discovery cohort were further tested in an expanded independent sample set with 76 MBC patients (21 CTC-positive, 55 CTC-negative), 109 PBC patients, and 50 healthy controls. Univariate logistic regression analysis was used to validate the metabolites that could separate different groups. We confirmed 19 metabolite levels (five increased and 14 decreased) for CTC-positive MBC patients, and 12 metabolite levels (seven elevated and five reduced) for CTC-negative MBC patients were significantly different from healthy controls (Table 4.2). Likewise, 25 metabolite levels (six increased and 19 decreased) for CTC-positive patients and nine metabolite levels (seven elevated and two reduced) for CTC-negative patients were found to be significantly different from PBC patients (Table 4.3). More metabolites were identified for patients with CTC-positive MBC than CTC-negative MBC in comparison to healthy controls or PBC patients. So the differences between CTC-positive MBC patients and controls or PBC patients were larger than that for CTC-negative MBC patients (Figure 4.6). These 35 metabolites, including amino acids, biogenic amines, short- and long-chain acylcarnitines, lysophosphatidylcholines, and phosphatidylcholines, differed significantly between MBC patients and PBC patients or healthy controls.

Table 4.2: Metabolite levels are significantly different between CTC-positive or CTC-negative MBC patients and healthy controls.

CTC-positive MBC vs. Control				CTC-negative MBC vs. Control			
Metabolites	**FC**	**FDR**	**AUC**	**Metabolites**	**FC**	**FDR**	**AUC**
Asn ↓	-1.25	2.71E-02	0.71	Asn ↓	-1.24	7.26E-04	0.72
Glu ↑	1.69	1.29E-02	0.82	Glu ↑	1.60	7.26E-04	0.72
His ↓	-1.16	3.06E-02	0.79	His ↓	-1.16	9.21E-04	0.73
Trp ↓	-1.17	3.92E-02	0.75	Thr ↓	-1.17	1.73E-02	0.69
C4:0 ↑	1.99	9.59E-03	0.79	Trp ↓	-1.24	2.44E-04	0.71
C16:0 ↑	1.29	3.05E-02	0.77	C2:0 ↑	1.41	9.21E-04	0.68
C18:1 ↑	1.29	4.19E-02	0.80	C16:0 ↑	1.41	3.28E-04	0.74
lysoPC a C18:2 ↓	-1.65	9.56E-03	0.88	C18:0 ↑	1.34	9.65E-04	0.66
PC aa C40:6 ↑	1.52	9.59E-03	0.74	C18:1 ↑	1.50	2.44E-04	0.78
PC ae C34:2 ↓	-1.25	3.42E-02	0.75	C18:2 ↑	1.35	9.21E-04	0.71
PC ae C34:3 ↓	-1.43	9.59E-03	0.76	lysoPC a C18:2 ↓	-1.33	9.21E-04	0.72
PC ae C36:3 ↓	-1.32	9.59E-03	0.78	PC aa C40:6 ↑	1.29	1.91E-03	0.68
PC ae C40:4 ↓	-1.25	1.41E-02	0.68				
PC ae C42:3 ↓	-1.29	1.29E-02	0.83				
PC ae C42:4 ↓	-1.53	2.81E-03	0.77				
PC ae C44:3 ↓	-1.25	2.29E-02	0.77				
PC ae C44:4 ↓	-1.51	2.81E-03	0.80				
PC ae C44:5 ↓	-1.51	2.81E-03	0.72				
PC ae C44:6 ↓	-1.36	9.56E-03	0.69				

Note: Results of fold change (FC), univariate logistic regression with corresponding false discovery rate (FDR), and ten-fold cross-validated area under the curve (AUC) estimate for significantly different metabolites in both cohorts are shown.

Table 4.3: Metabolite levels are significantly different between CTC-positive or CTC-negative MBC and PBC patients.

CTC-positive MBC vs. PBC				CTC-negative MBC vs. PBC			
Metabolites	**FC**	**FDR**	**AUC**	**Metabolites**	**FC**	**FDR**	**AUC**
Arg ↓	-1.25	2.03E-02	0.80	Arg ↓	-1.17	1.67E-02	0.67
Glu ↑	2.22	1.58E-04	0.90	Glu ↑	2.10	1.77E-07	0.85
Orn ↑	1.29	1.81E-03	0.70	Orn ↑	1.49	2.57E-08	0.83
Kynurenine ↑	1.28	9.84E-03	0.66	Kynurenine ↑	1.25	1.63E-03	0.68
Met-SO ↑	1.36	1.37E-02	0.68	Met-SO ↑	1.36	6.11E-04	0.69
C4:0 ↑	2.19	4.23E-04	0.85	Spermidine ↑	1.13	1.67E-02	0.63
lysoPC a C18:2 ↓	-1.49	2.07E-03	0.83	C4:0 ↑	1.49	3.20E-03	0.69
lysoPC a C20:3 ↓	-1.31	1.37E-02	0.69	C18:2 ↑	1.33	1.91E-04	0.69
lysoPC a C20:4 ↓	-1.22	3.52E-02	0.71	lysoPC a C18:2 ↓	-1.20	8.31E-03	0.64
PC aa C40:6 ↑	1.43	1.11E-03	0.73				
PC aa C42:1 ↓	-1.21	3.52E-02	0.67				
PC ae C34:2 ↓	-1.28	4.58E-03	0.73				
PC ae C34:3 ↓	-1.38	1.63E-03	0.73				
PC ae C36:3 ↓	-1.37	4.23E-04	0.76				
PC ae C40:1 ↓	-1.19	1.35E-02	0.75				
PC ae C40:3 ↓	-1.19	7.75E-03	0.68				
PC ae C40:4 ↓	-1.33	4.18E-04	0.71				
PC ae C42:1 ↓	-1.19	7.75E-03	0.73				
PC ae C42:3 ↓	-1.31	2.07E-03	0.81				
PC ae C42:4 ↓	-1.62	1.15E-04	0.81				
PC ae C42:5 ↓	-1.42	1.45E-04	0.75				
PC ae C44:3 ↓	-1.31	2.72E-03	0.78				
PC ae C44:4 ↓	-1.61	1.15E-04	0.82				
PC ae C44:5 ↓	-1.64	1.15E-04	0.77				
PC ae C44:6 ↓	-1.46	4.23E-04	0.73				

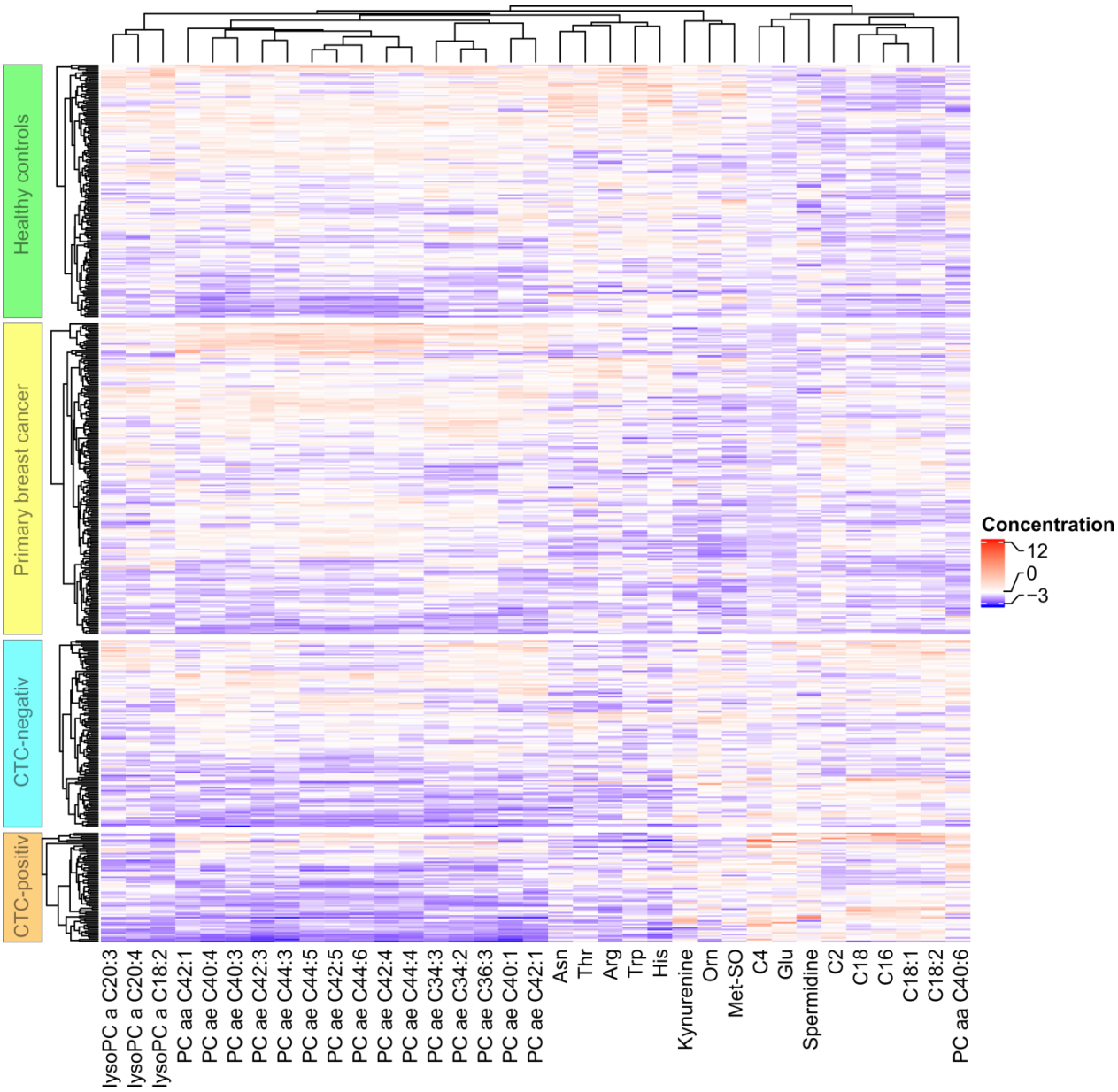


Figure 4.6: Supervised hierarchical complete linkage clustering with Euclidean distance metric, depending on the 35 significantly different metabolites between MBC patients and PBC patients or healthy controls.

4.2.2 Plasma metabolite panels associated with metastatic breast cancer early detection

A multivariable penalized LASSO logistic regression model (ten-fold cross-validated) comprising six metabolites (His, C4:0, C18:1, lysoPC a C18:2, PC aa C40:6, and PC ae C42:3) was predicted. It yielded a high discriminatory accuracy with an AUC of 0.92 (95% CI: 0.86 ~ 0.99) for CTC-positive MBC patients versus healthy controls. For CTC-negative MBC patients versus healthy controls, the predicted model consisting of 11 metabolites (Asn, Glu, His, Thr, Trp, C16:0, C18:0, C18:1, C18:2, lysoPC a C18:2, and PC aa C40:6) had the best distinguishing performance with an AUC of 0.89 (95% CI: 0.83 ~ 0.95). Another multivariable model was built to compare CTC-positive MBC and PBC patients, which contained 15 selected metabolites (Arg, Glu, Orn, C4:0, lysoPC a C18:2, lysoPC a C20:4, PC aa C40:6, PC aa C42:1, PC ae C34:2, PC ae C34:3, PC ae C36:3, PC ae C42:1, PC ae C42:3, PC ae C42:5, and PC ae C44:4) and generated an AUC of 0.95 (95% CI: 0.9 ~ 1.0). Though individual metabolite could not distinguish CTC-negative MBC patients from PBC patients with high accuracy, the model with a combination of seven metabolites (Arg, Glu, Orn, Met-SO, spermidine, C4:0, and lysoPC a C18:2) predicted an appreciable AUC of 0.90 (95% CI: 0.86 ~ 0.95) (Figure 4.7).

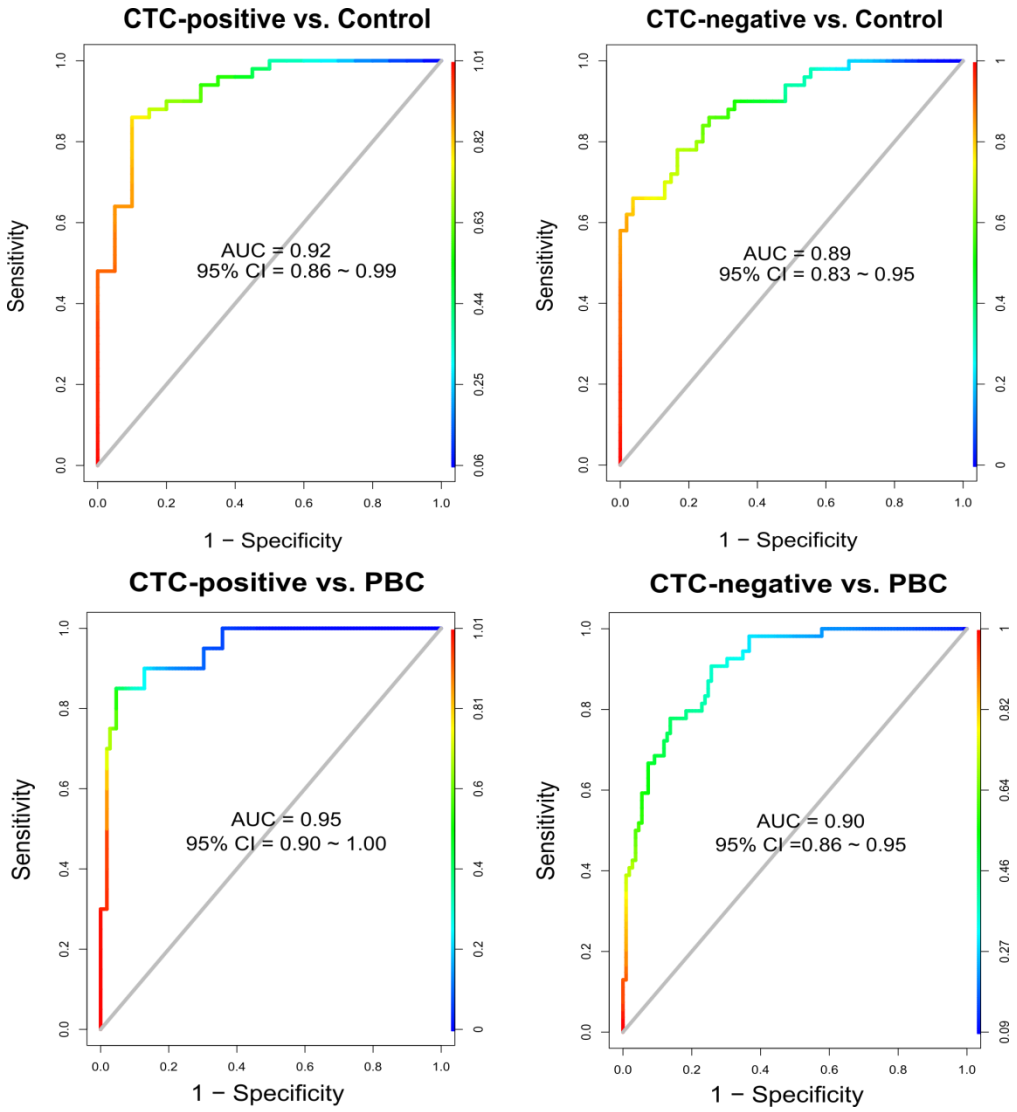


Figure 4.7: ROC curves for multivariable panels based on penalized LASSO logistic regression models to discriminate MBC patients from PBC patients or healthy controls. CTC-positive MBC patients versus control: His, C4:0, C18:1, lysoPC a C18:2, PC aa C40:6, and PC ae C42:3; CTC-negative MBC patients versus control: Asn, Glu, His, Thr, Trp, C16:0, C18:0, C18:1, C18:2, lysoPC a C18:2, and PC aa C40:6; CTC-positive MBC versus PBC patients: Arg, Glu, Orn, C4:0, lysoPC a C18:2, lysoPC a C20:4, PC aa C40:6, PC aa C42:1, PC ae C34:2, PC ae C34:3, PC ae C36:3, PC ae C42:1, PC ae C42:3, PC ae C42:5, and PC ae C44:4; CTC-negative MBC versus PBC patients: Arg, Glu, Orn, Met-SO, spermidine, C4:0, and lysoPC a C18:2.

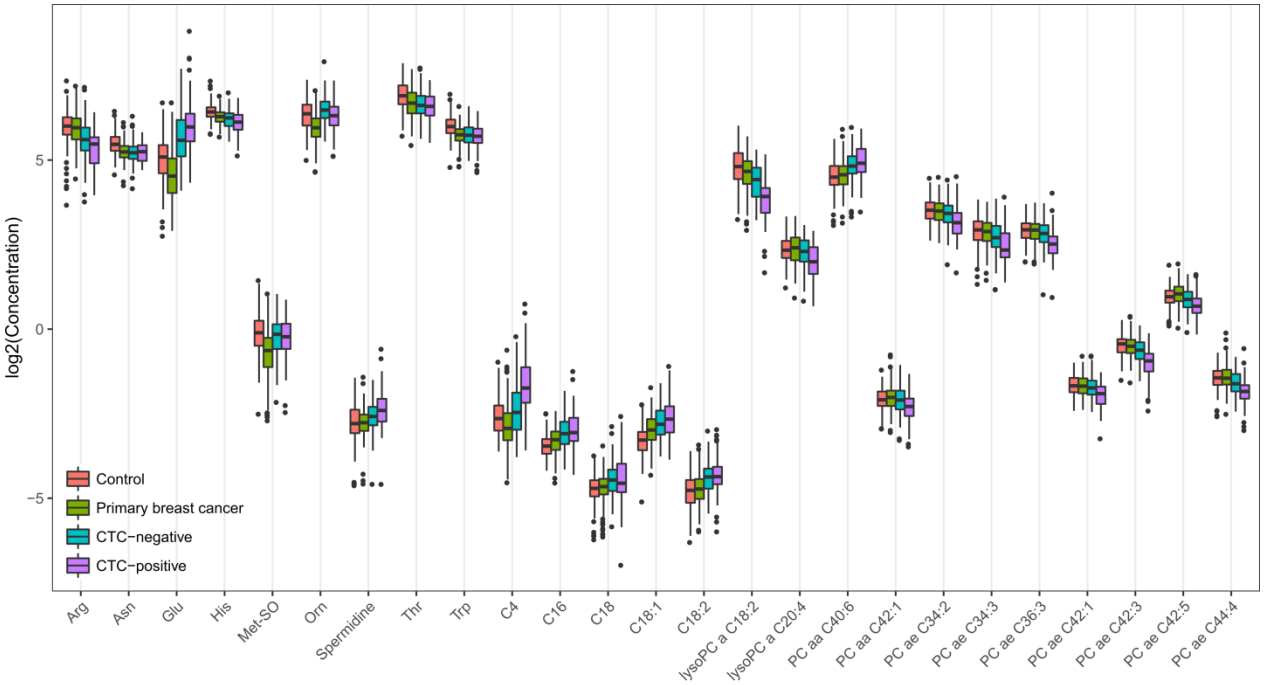


Figure 4.8: Distribution of the LASSO selected metabolite concentrations in CTC-positive MBC, CTC-negative MBC, PBC patients, and healthy controls. Box plots of metabolites are represented as the logarithm of concentration to the base 2 values.

Boxplots for the 25 metabolites selected by the four multivariate penalized LASSO logistic regression models were constructed to show the distributions of metabolite concentrations in each group (Figure 4.8). Ten-fold cross-validated univariate ROC analysis for each candidate metabolite in the set could predict the ability to distinguish CTC-positive or CTC-negative MBC patients from PBC patients or healthy controls. Even though individual metabolites could distinguish different groups by themselves, multivariable models with selected panels performed better (Figure 4.7; Figure 4.9). Among these 25 metabolites, glutamate showed the best ability to discriminate different groups (Table 4.2; Table 4.3; Figure 4.9). Remarkably, ornithine and Met-SO could differentiate MBC patients from PBC patients better than MBC patients from healthy controls (Figure 4.9).

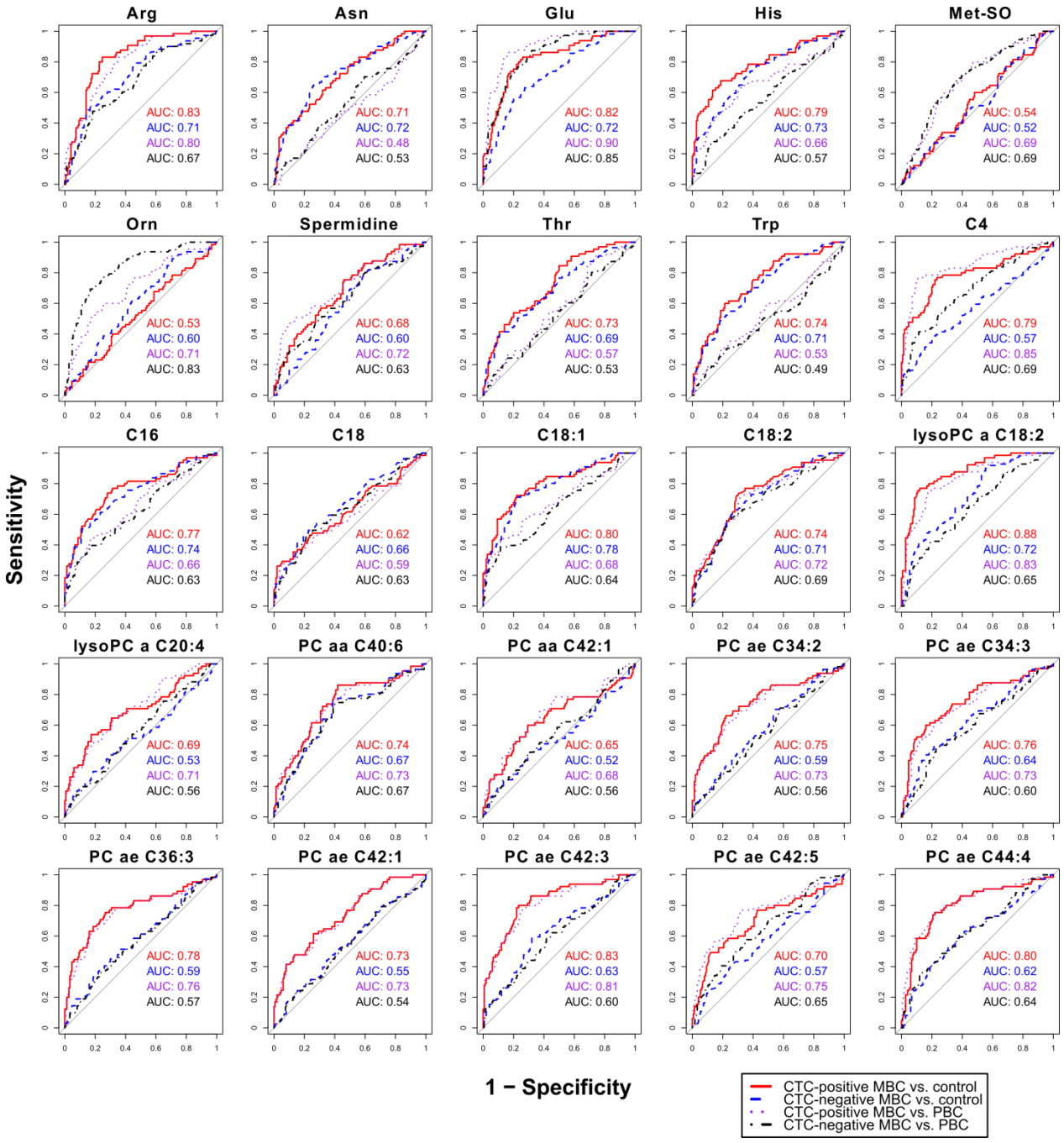


Figure 4.9: Ten-fold cross-validated ROC curves based on univariate logistic regression models for the selected individual metabolites. The ROC curves for CTC-positive MBC patients versus healthy controls, CTC-negative MBC patients versus healthy controls, CTC-positive MBC versus PBC patients, and CTC-negative MBC versus PBC patients are displayed in red solid, blue dashed, purple dotted, and black dot-dashed lines, respectively. The diagonal gray line means no separation ability, i.e. random classification.

4.3 Prognostic Value of Plasma Metabolites in MBC Patients

4.3.1 Plasma metabolites correlated with PFS and OS

Survival data were available for all MBC patients in both discovery and validation cohorts. The log-rank test was employed to identify metabolites that could predict PFS and OS. Each metabolite was categorized into two classes according to its concentration value: the lowest quartile and the residual quartiles. In the discovery cohort, 16 and 51 metabolites were significantly correlated with PFS and OS ($p < 0.05$), respectively. These metabolites were further tested in the validation cohort. Ultimately, four metabolites, kynurenine, PC aa C36:3, PC ae C36:1, and PC ae C38:3, were correlated with PFS significantly. Here, *p*-values were not adjusted for multiple testing. Higher levels of these metabolites were found to be associated with higher probabilities of PFS (Table 4.4; Figure 4.10A). Twelve metabolites, lysoPC a C18:1, lysoPC a C20:3, lysoPC a C20:4, PC aa C36:3, PC aa C36:4, PC aa C38:5, PC ae C36:1, PC ae C38:3, PC ae C38:4, PC ae C40:2, SM C18:1, and SM (OH) C22:2, were also positively related to OS significantly (Table 4.4; Figure 4.10B).

Table 4.4: Associations between plasma metabolites or CTC counts and PFS and OS assessed by the log-rank model.

Metabolites	PFS	OS
Kynurenine	**3.88E-04**	3.94E-01
lysoPC a C18:1	3.08E-01	**1.18E-03**
lysoPC a C20:3	7.78E-02	**6.63E-05**
lysoPC a C20:4	7.47E-01	**5.09E-04**
PC aa C36:3	**2.98E-03**	**1.81E-07**
PC aa C36:4	2.20E-01	**1.15E-05**
PC aa C38:5	2.30E-01	**9.59E-05**
PC ae C36:1	**2.77E-03**	**6.33E-04**
PC ae C38:3	**1.79E-04**	**1.12E-08**
PC ae C38:4	7.98E-02	**1.30E-03**
PC ae C40:2	4.35E-01	**5.40E-03**
SM C18:1	1.73E-01	**1.23E-06**
SM (OH) C22:2	3.30E-01	**6.28E-05**
CTC	**3.05E-02**	**6.37E-10**

Note: significant results are in bold.

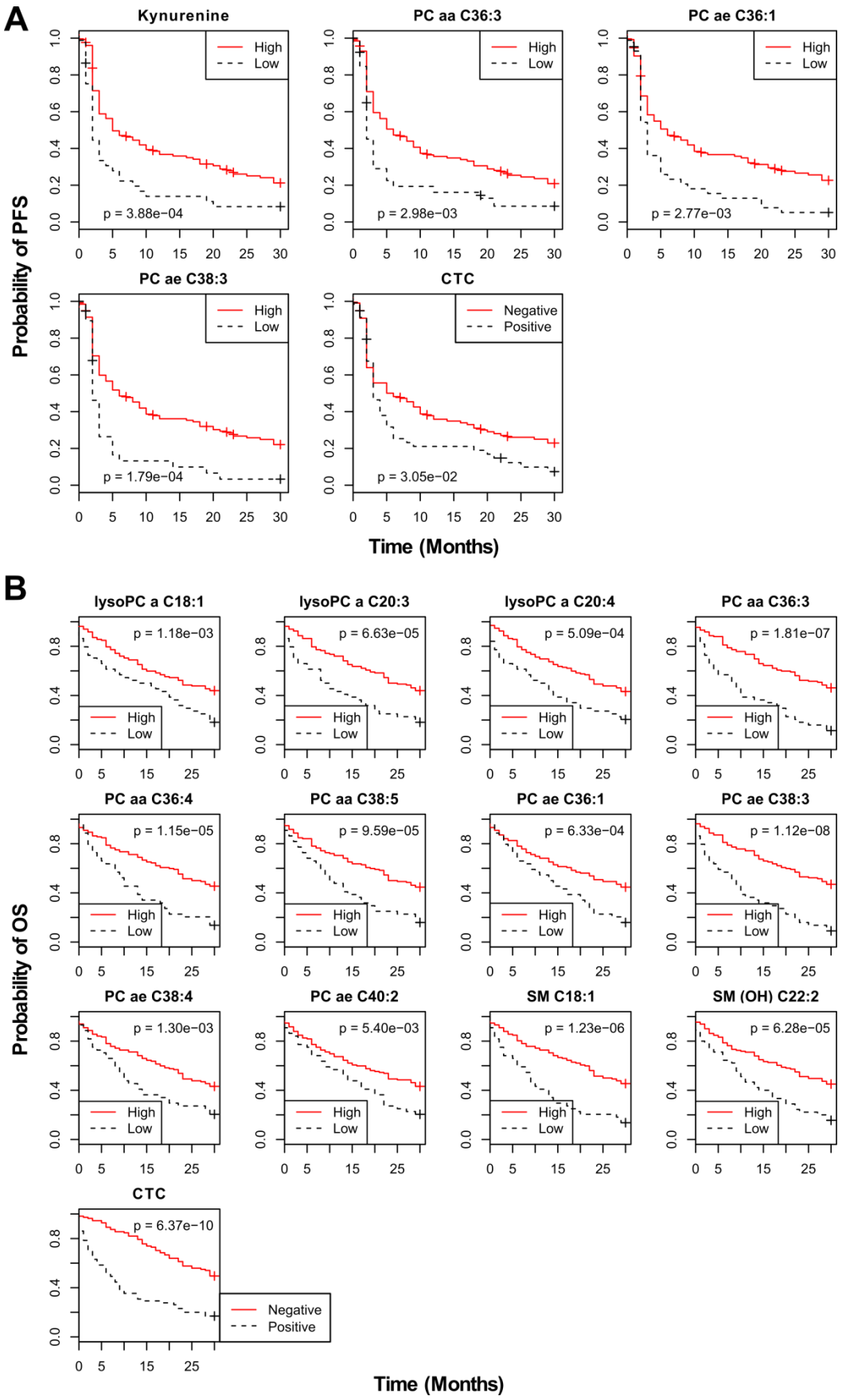

A
Kynurenine
p = 3.88e−04
PC aa C36:3
p = 2.98e−03
PC ae C36:1
p = 2.77e−03
PC ae C38:3
p = 1.79e−04
CTC
p = 3.05e−02
High
Low
Negative
Positive
Probability of PFS
Time (Months)
B
lysoPC a C18:1
p = 1.18e−03
lysoPC a C20:3
p = 6.63e−05
lysoPC a C20:4
p = 5.09e−04
PC aa C36:3
p = 1.81e−07
PC aa C36:4
p = 1.15e−05
PC aa C38:5
p = 9.59e−05
PC ae C36:1
p = 6.33e−04
PC ae C38:3
p = 1.12e−08
PC ae C38:4
p = 1.30e−03
PC ae C40:2
p = 5.40e−03
SM C18:1
p = 1.23e−06
SM (OH) C22:2
p = 6.28e−05
CTC
p = 6.37e−10
Probability of OS
Time (Months)

Figure 4.10: Kaplan-Meier curves of the CTC status and metabolites that significantly correlated with progression-free survival (PFS, A) and overall survival (OS, B). Metabolite levels were stratified according to their concentrations as the lowest quartile ("Low") and residual quartiles ("High"). CTCs were stratified as CTC-positive and CTC-negative.

4.3.2 Correlations of plasma metabolites with CTC counts

For these 12 metabolites that related to OS, spearman correlation analysis was performed to assess the correlations between metabolite concentrations and CTC counts. Results revealed that kynurenine, PC ae C36:1, PC ae C40:2, and SM C18:1 were not significantly correlated with CTC counts, while the other metabolites were negatively correlated with CTC counts significantly. The lower metabolite concentrations, the higher CTC numbers (Figure 4.11).

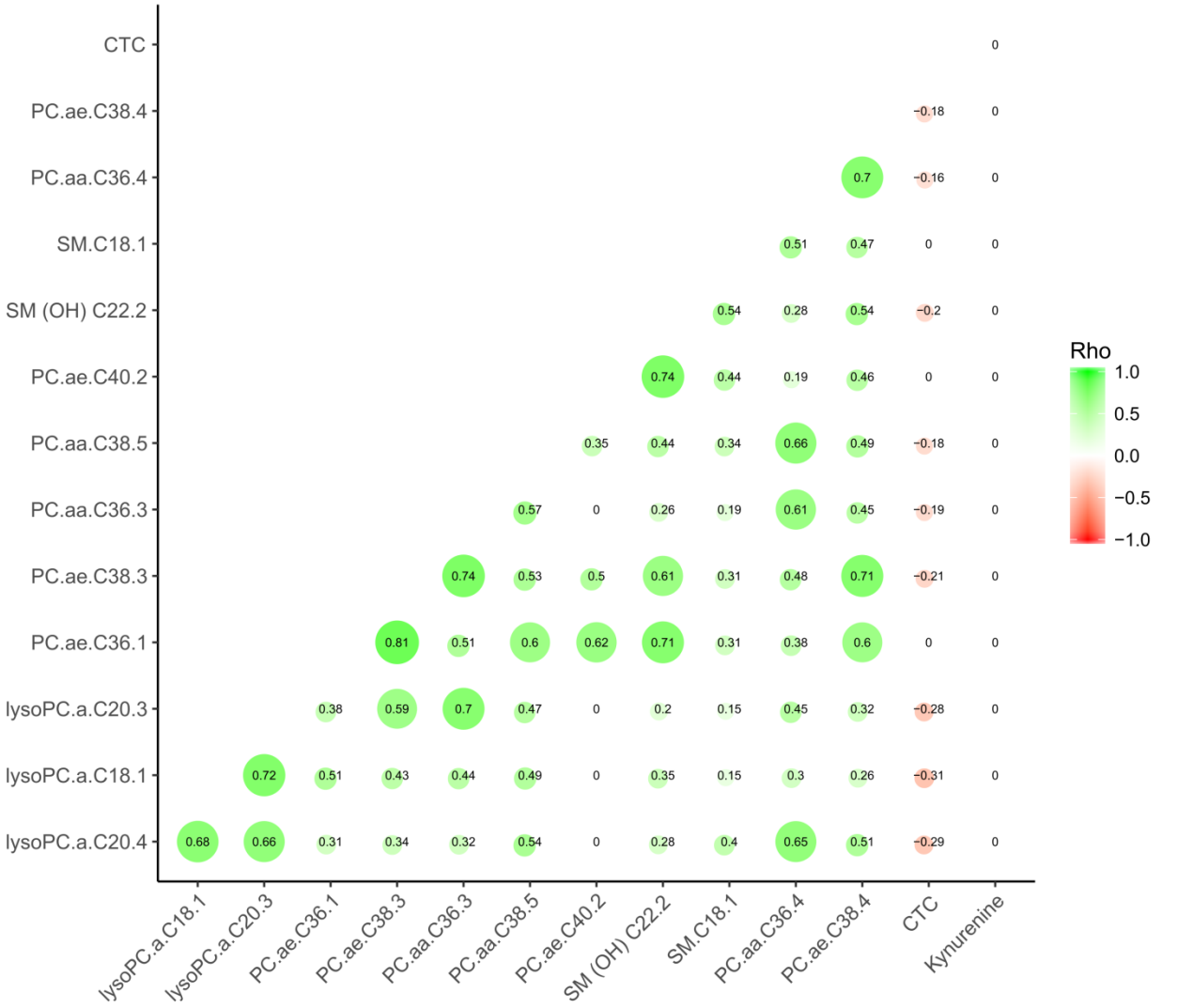


Figure 4.11: Spearman rank correlation of metabolite concentrations and CTC numbers. Metabolites significantly associated with CTC counts are marked with texts of correlation coefficients. The green dots represent positive correlations, while red dots indicate negative correlations.

4.3.3 Evaluation and comparison of putative prognostic models

The LASSO cox model predicted that combination of PC ae C36:1 and PC ae C38:3 for PFS (IPE = 5.669 compared to IPE0 = 5.689 for the null model without covariate information) and combination of lysoPC a C20:3, lysoPC a C20:4, PC aa C38:5, PC ae C38:3, and SM (OH)

C22:2 for OS (IPE = 5.795 compared to IPE0 = 6.130 for the null model without covariate information) were the best-fitting multivariate models, as shown in Table 4.5. When CTCs were introduced into these multivariable metabolite models, the IPEs were somewhat lower but essentially unchanged (PFS: 5.610, OS: 5.254; Figure 4.12). For PFS, PC ae C36:1 outperformed the multivariable model (IPE = 5.617), while lysoPC a C20:3 outperformed the multivariable for OS (IPE = 5.722). In comparison to CTCs (IPE of 5.628 for PFS and 5.268 for OS), PC ae C36:1 performed slightly better for PFS, but lysoPC a C20:3 performed worse for OS. However, combinations of PC ae C36:1 with CTCs and lysoPC a C20:3 with CTCs showed better prediction accuracies than CTCs alone for PFS (IPE = 5.535) and OS (IPE = 5.082), respectively (Figure 4.13).

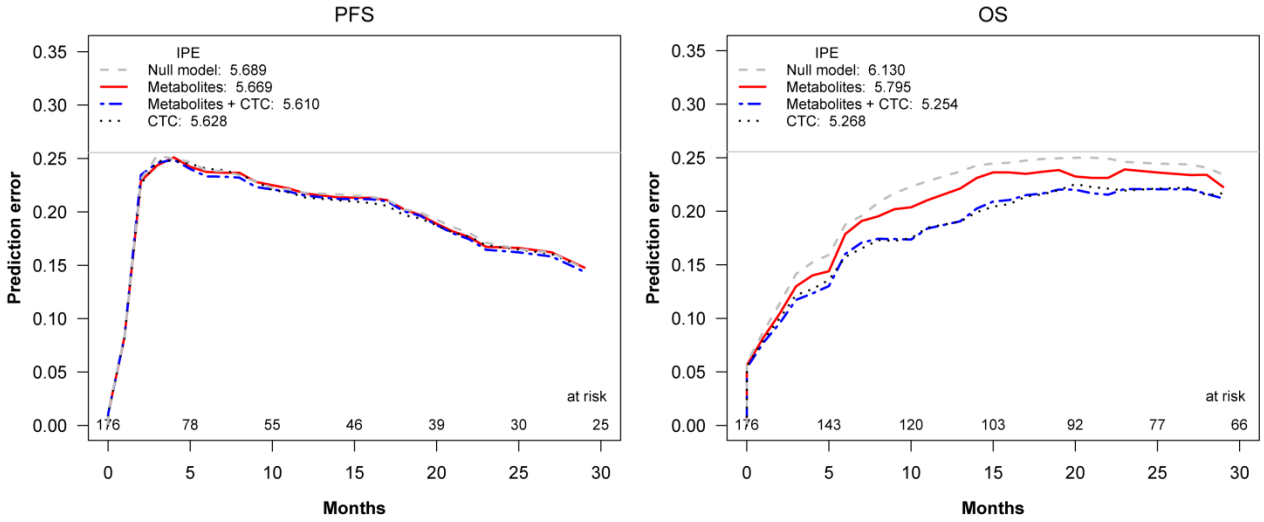


Figure 4.12: Integrated prediction error (IPE) curves up to 30 months for the null model (Kaplan-Meier model without any covariate information), CTC, multivariate metabolite model, and metabolite model + CTC for PFS and OS. The number of individuals at risk at different time points is indicated along the x-axis. The y-axis represents prediction errors of the LASSO cox model.

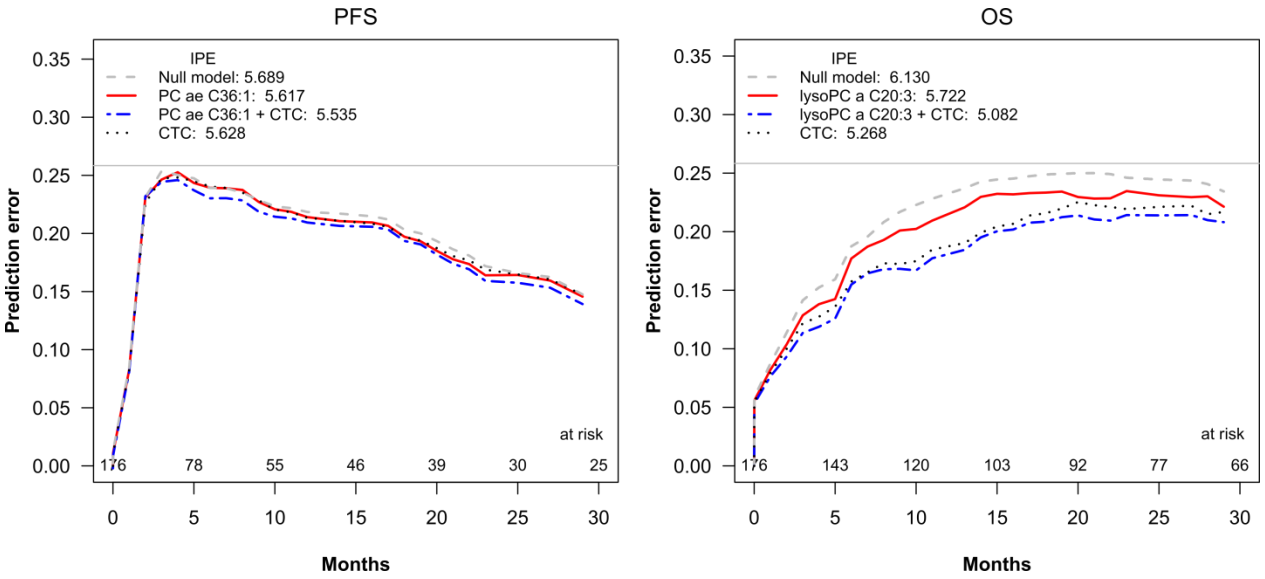


Figure 4.13: Integrated prediction error (IPE) curves for PFS (A) and OS (B). The curves were up to 30 months for the null model (Kaplan-Meier model without any covariate information), CTCs, PC ae C36:1 (PFS)/lysoPC a C20:3 (OS), and PC ae C36:1 (PFS)/lysoPC a C20:3 (OS) + CTCs. The number of individuals at risk at different time points is indicated along the x-axis. The y-axis represents prediction errors of the LASSO cox model.

Table 4.5: Different LASSO cox models for PFS and OS in MBC patients, corresponding integrated prediction errors (IPE) were shown.

Type	Model	PFS		OS	
		Variables	**IPE**	**Variables**	**IPE**
One-metabolite	Null model	None	5.689	None	6.130
	Metabolite model	PC ae C36:1	5.617	lysoPC a C20:3	5.722
	CTC model	CTC status	5.628	CTC status	5.268
	Metabolite + CTC model	PC ae C36:1, CTC status	5.535	lysoPC a C20:3, CTC status	5.082
Multi-metabolites	Null model	None	5.689	None	6.130
	Metabolites model	PC ae C36:1, PC ae C38:3	5.669	lysoPC a C20:3, lysoPC a C20:4, PC aa C38:5, PC ae C38:3, SM (OH) C22:2	5.795
	CTC model	CTC status	5.628	CTC status	5.268
	Metabolites + CTC model	PC ae C36:1, PC ae C38:3, CTC status	5.610	lysoPC a C20:3, lysoPC a C20:4, PC aa C38:5, PC ae C38:3, SM (OH) C22:2, CTC status	5.254

4.4 Diagnostic Value of Plasma Metabolites in OVCA Patients

4.4.1 Plasma metabolite profiles of OVCA, PBC patients, and healthy controls

With the remaining 138 metabolites after removing outliers, principal component analysis (PCA) was performed for OVCA patients and healthy controls in the discovery and validation cohorts. As shown in Figure 4.14, the separation of these two groups was based on the first principal component, and the separation of samples within the groups was mostly depending on the second principal component. The variance of ovarian cancer patients was generally higher than healthy controls, which may be related to the heterogeneous nature of ovarian cancers. The highest contribution to the overall variance originates from phosphatidylcholines and acylcarnitines.

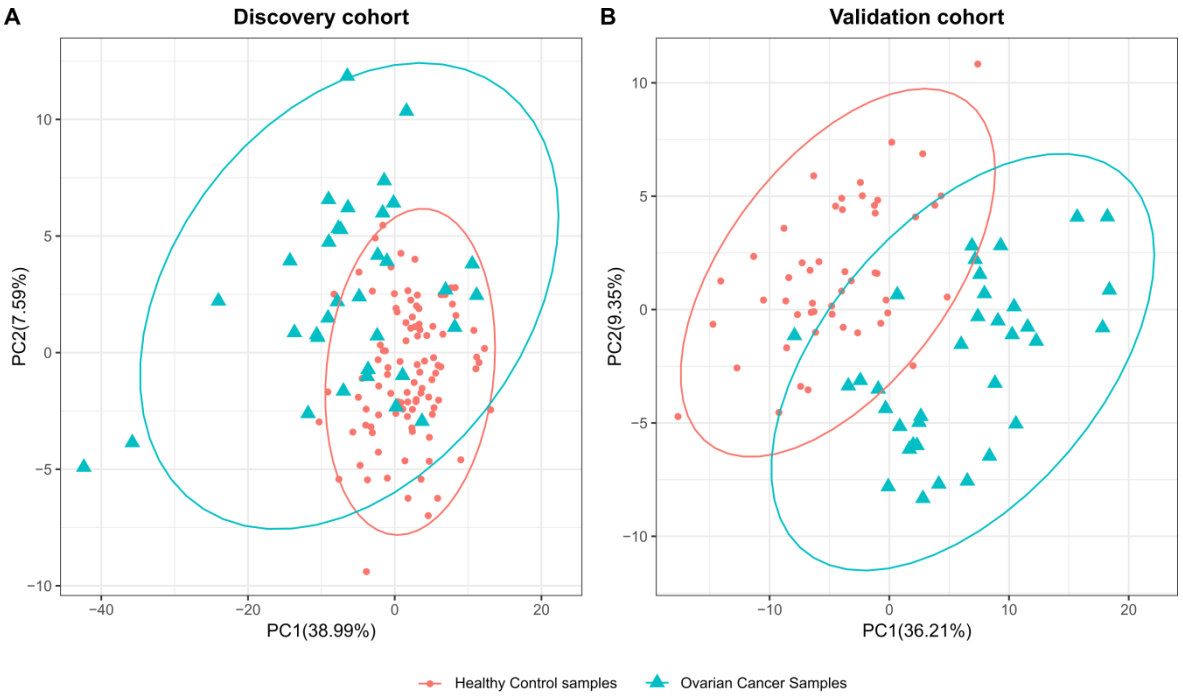


Figure 4.14: PCA scoring plot on the metabolite dataset of ovarian cancer patients and healthy controls for the discovery cohort (A) and validation cohort (B) with outliers excluded,

95% confidence ellipses are drawn for each group. Round red spots represent samples from healthy controls, while triangular aqua spots represent samples from ovarian cancer.

The data set of the discovery cohort consisted of 34 OVCA patients, 80 PBC patients, and 100 healthy controls. A total of 94 metabolites were differentially expressed between OVCA patients and healthy controls (FDR < 0.05). The data set for validating the potential metabolites consisted of 35 OVCA patients, 109 BC patients, and 50 healthy controls. Among these significant differentially expressed metabolites identified in the discovery cohort, 71 were verified independently to separate OVCA samples from healthy controls, including seven amino acids, four acylcarnitines, five lysophosphatidylcholines, 54 phosphatidylcholines, and one sphingomyelin. Except for the increased levels of Glu, C2, C18:1, and C18:2 in ovarian cancer patients relative to healthy controls, all the other identified differentially expressed metabolites were down-regulated.

4.4.2 A panel of plasma metabolites associated with OVCA diagnosis

The resulting 71 features were reduced by LASSO regression analysis as described in the Methods. Videlicet, models were trained on the discovery cohort with ten-fold cross-validation, which generated an optimal parameterized model consisted of a subset of seven metabolites: His, Trp, C18:1, lysoPC a C18:2, PC aa C32:2, PC aa C34:4, and PC ae C34:3. Following training, this panel of metabolites was tested on an independent validation dataset consisted of 35 OVCA patients, 109 PBC patients, and 50 healthy controls. The performance of this classifier was as high as AUC = 0.95 (95% CI: 0.89 ~ 1.00) (Figure 4.15, left).

FIGO stages I and II are designated as early-stage OVCA, while stages III and IV are designated as late-stage OVCA. Correspondingly, the discovery cohort included three early-stage and 31 late-stage OVCA samples, while the validation cohort contained ten early-stage and 21 late-stage OVCA samples. There were only ten and three early-stage samples in the discovery and validation cohort, respectively. The model with seven metabolites that could discriminate OVCA patients from healthy controls achieved an AUC of 0.87 (95% CI: 0.76 ~

0.97) for early-stage OVCA samples (Figure 4.15, middle) and an AUC of 0.96 (95% CI: 0.92 ~ 1.00) for late-stage OVCA samples to differentiate from healthy controls (Figure 4.15, right).

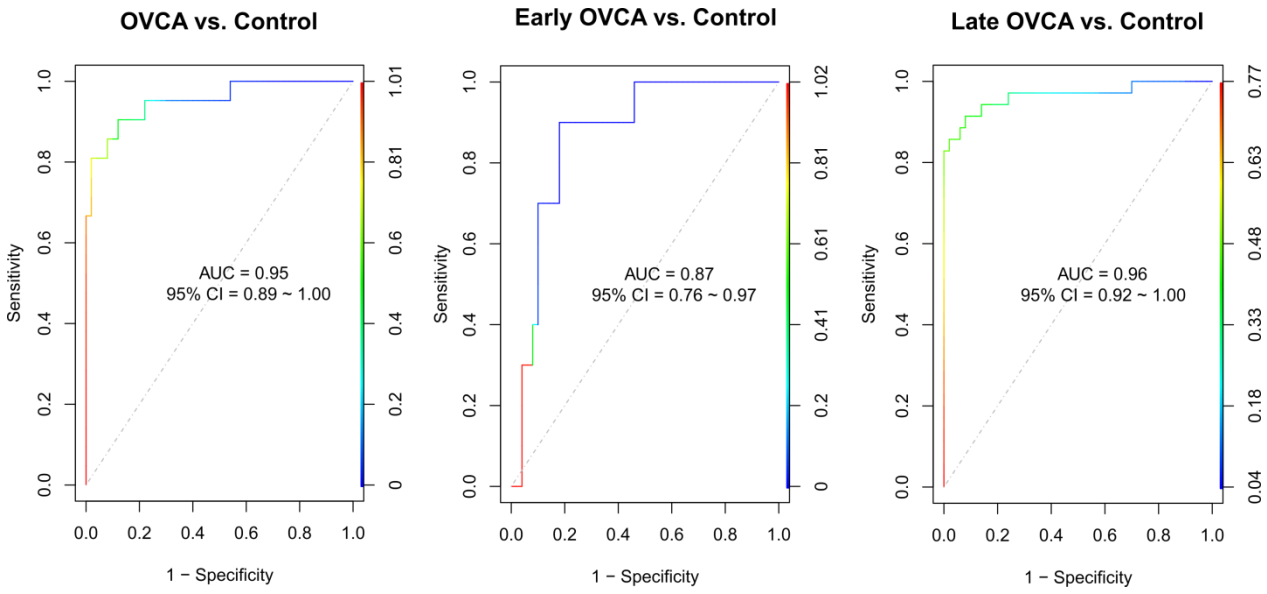


Figure 4.15: ROC curves for a seven metabolites panel to distinguish between OVCA patients and healthy controls (left), early-stage OVCA patients and healthy controls (middle), late-stage OVCA patients and healthy controls (right). This classifier panel includes His, Trp, C18:1, lysoPC a C18:2, PC aa C32:2, PC aa C34:4, and PC ae C34:3. The diagonal dot-dashed gray line means no separation ability, i.e. random classification.

The performance of this seven metabolites classification model in discriminating OVCA patients from PBC patients was also evaluated. The ROC curve analysis for OVCA and PBC patients separation achieved an AUC of 0.93 (95% CI: 0.88 ~ 0.98) (Figure 4.16, left). At the same time, OVCA patients at the early and late stages were also investigated. Early-stage OVCA and PBC patients distinction had a relatively low AUC of 0.70 (95% CI: 0.57 ~ 0.83) (Figure 4.16, middle). However, the discriminatory power between late-stage OVCA and PBC patients could be as high as AUC = 0.943 (95% CI: 0.88 ~ 1.00) (Figure 4.16, right). The low number of early-stage OVCA samples might be the reason for their low performance,

and the reason for the similar discriminative ability between late-stage and overall OVCA patients.

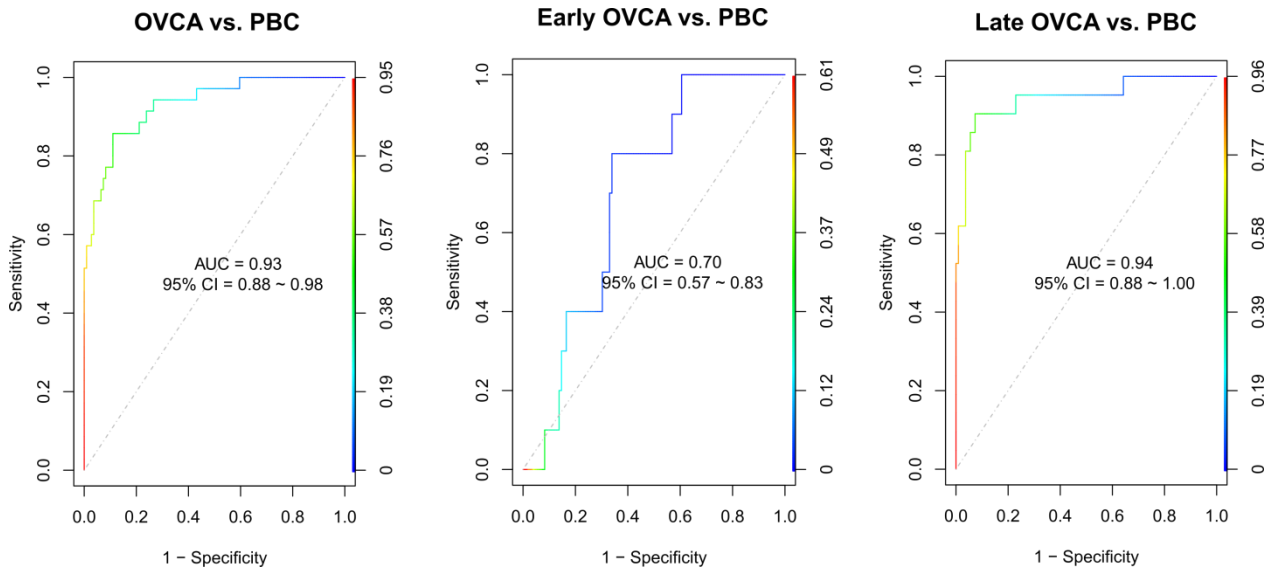


Figure 4.16: ROC curve for a seven metabolites panel to distinguish between OVCA and PBC patients (left), early-stage OVCA and PBC patients (middle), late-stage OVCA and PBC patients (right). This classifier panel includes His, Trp, C18:1, lysoPC a C18:2, PC aa C32:2, PC aa C34:4, and PC ae C34:3. The diagonal dot-dashed gray line means no separation ability, i.e. random classification.

To evaluate the power of each metabolite in discriminating OVCA patients from healthy controls or PBC patients, leave-one-out cross-validated ROC analysis for the logistic regression model was built in the discovery and validation cohorts (Table 4.5). AUCs of early-stage OVCA patients differed a lot from the overall OVCA patients because of the small sample size. Therefore, results for early-stage OVCA patients could not reflect the authentic ability of the metabolite panel in differentiating them from healthy controls or PBC patients. Each metabolite in the panel performed with high AUC in separating OVCA patients from healthy controls or PBC patients (Figure 4.17). However, the multivariate model containing these seven metabolites had the highest differentiation power (Figure 4.15; Figure 4.16).

Table 4.6: AUCs of individual metabolites that separate OVCA patients from healthy controls or PBC patients in the discovery cohort (D) and validation cohort (V).

Metabolites	OVCA vs. Control		Early OVCA vs. Control		Late OVCA vs. Control		OVCA vs. PBC		Early OVCA vs. PBC		Late OVCA vs. PBC	
	D	V	D	V	D	V	D	V	D	V	D	V
His	0.846	0.784	0.877	0.802	0.843	0.753	0.685	0.668	0.680	0.670	0.686	0.639
Trp	0.833	0.976	0.703	1.000	0.846	0.974	0.713	0.931	0.606	0.972	0.723	0.931
C18:1	0.733	0.849	0.553	0.892	0.761	0.829	0.614	0.726	0.627	0.778	0.637	0.709
lysoPC a C18:2	0.812	0.946	0.880	0.946	0.806	0.953	0.783	0.927	0.860	0.931	0.775	0.936
PC aa C32:2	0.862	0.920	0.703	0.915	0.877	0.919	0.771	0.914	0.623	0.905	0.785	0.916
PC aa C34:4	0.830	0.888	0.737	0.853	0.839	0.889	0.754	0.907	0.640	0.869	0.765	0.910
PC ae C34:3	0.821	0.848	0.767	0.753	0.827	0.896	0.811	0.844	0.763	0.729	0.816	0.904

Note: D represents discovery cohort. V means validation cohort.

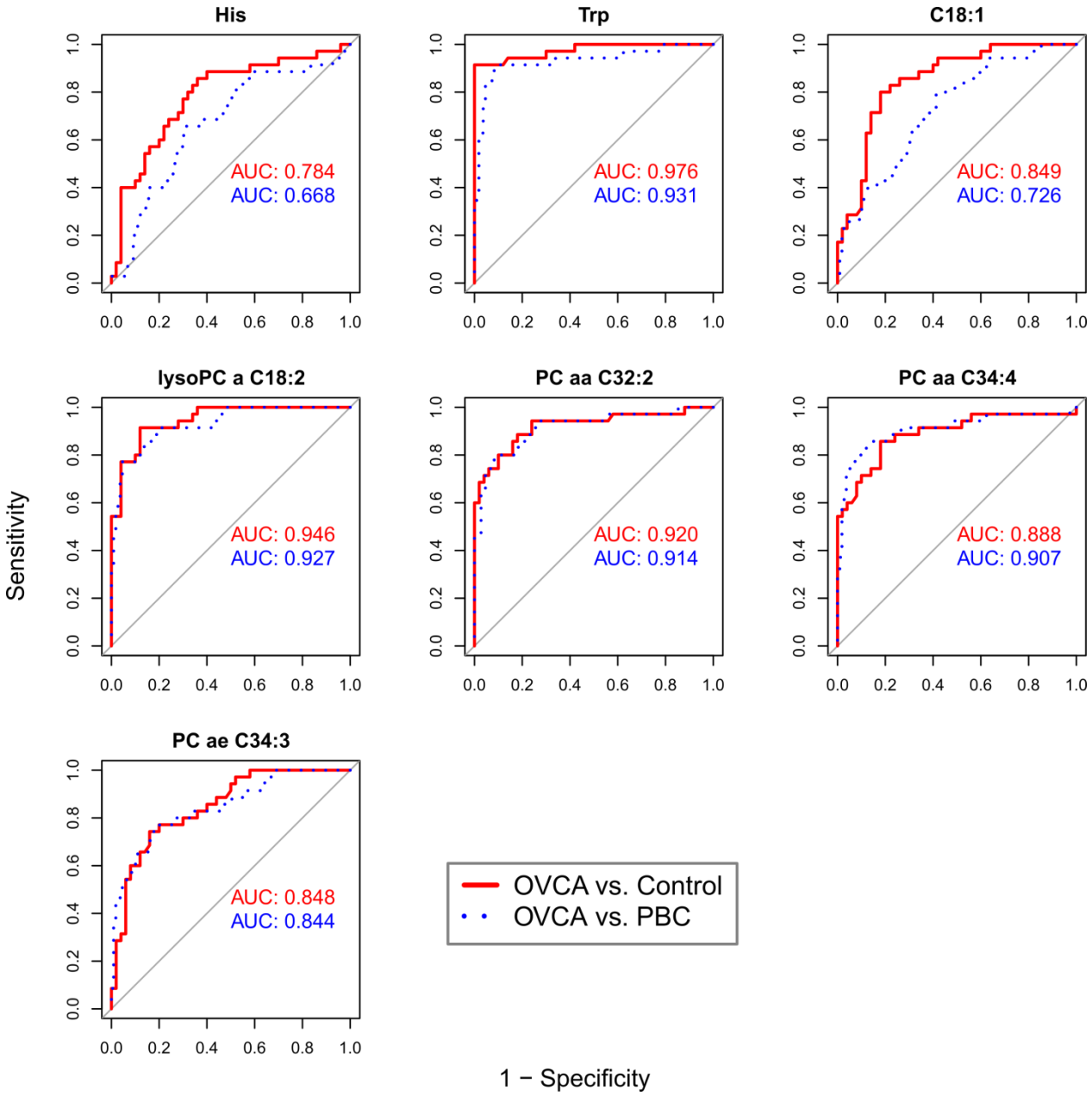


Figure 4.17: Leave-one-out cross-validated ROC analysis for individual metabolites to differentiate OVCA patients from healthy controls or PBC patients. The continuous red line and dotted blue line represent the ROC curves for OVCA patients versus healthy controls, OVCA patients versus and PBC patients, respectively. The diagonal gray line means no separation ability, i.e. random classification.

4.4.3 Plasma metabolites associated with OVCA diagnosis identified by Biocrates Absolute*IDQ*® p400 HR Kit

During the process of data analysis, Biocrates released the Absolute*IDQ*® p400 HR Kit. This new kit could measure 408 metabolites simultaneously, including 21 amino acids, 21 biogenic amines, one hexose, 172 phosphatidylcholines, 24 lysophosphatidylcholines, 31 sphingomyelins, nine ceramides, 55 acylcarnitines, 14 cholesteryl esters, 18 diglycerides, and 42 triglycerides (Table A5). With the same patient samples, plasma metabolites were again measured by the p400 HR kit. The profiling procedure was the same as for the p180 kit, which was described in the Methods.

A total of 248 metabolites were left for downstream analysis after filtering with the same criteria. Metabolites concentration data were log2 transformed, and univariate logistic regression analysis identified the significantly different metabolites between groups. In the end, 133 differentially expressed metabolites between ovarian cancer patients and healthy controls were identified in the discovery cohort, and 77 metabolites were verified in the validation cohort. These 77 metabolites consisted of eight amino acids, four biogenic amines, 35 phosphatidylcholines, nine lysophosphatidylcholines, 12 sphingomyelins, two ceramides, four acylcarnitines, one cholesteryl ester, one diglyceride, and one triglyceride. Moreover, all the metabolites in the classifier panel that measured with the p180 kit to distinguish between OVCA patients and healthy controls were also differentially expressed when measured with the p400 kit.

Further, a panel of 17 metabolites was selected by multivariate penalized LASSO logistic regression analysis, as with the p180 kit. This discriminative panel contained Ala, Asn, Cit, Glu, Trp, ADMA, c4-OH-Pro, SDMA, AC(4:0), AC(16:1), LPC(18:2), PC(34:5), PC(35:1), PC(42:4), PC-O(30:0), SM(32:2), and SM(42:1) achieved a high AUC of 0.99 (95% CI: 0.97 ~ 1.00) (Figure 4.18). The sensitivity and specificity were higher than the panel selected with the p180 kit. Trp and LPC(18:2) were the overlapped metabolites in the classier panels that selected with the p180 kit and p400 kit. Likewise, the combination of the classier panel performed better than individual metabolites in separating OVCA patients from healthy controls (Figure 4.18; Figure 4.19).

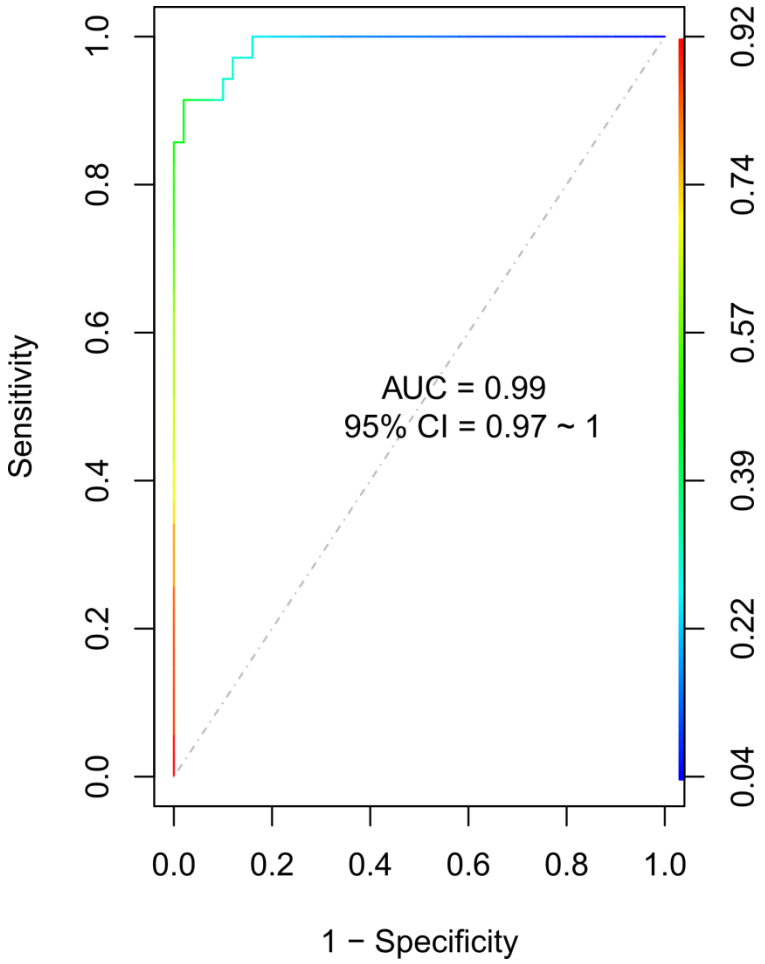


Figure 4.18: ROC curve for multiparametric panel based on penalized LASSO logistic regression model to distinguish ovarian cancer patients from healthy controls. The diagonal dot-dashed gray line means no separation ability, i.e. random classification.

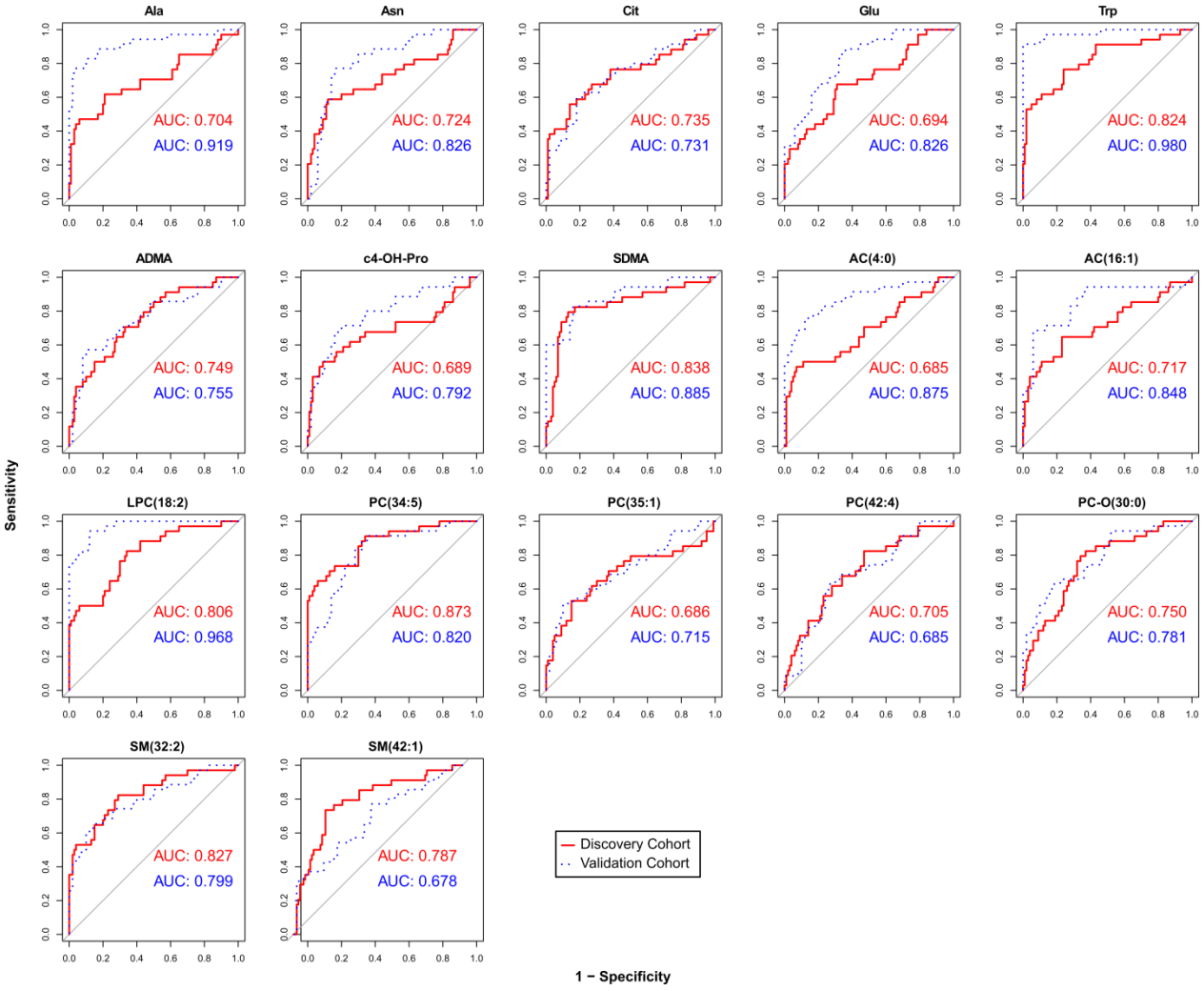


Figure 4.19: ROC curves of individual metabolites from the classifier panel that measured with p400 kit to separate OVCA patients from healthy controls. ROC curves for the discovery and validation cohorts are shown with red and blue lines, respectively. The diagonal gray line means no separation ability, i.e. random classification.

5. Discussion

Cancer management and treatment have benefited a lot from biomarkers, which have improved disease outcomes and the overall life quality of cancer patients. Cancer biomarkers have increased the survival rates, for instance, amongst breast cancer patients (Cardoso *et al.*, 2009; Kataja *et al.*, 2009) and ovarian cancer patients (Ledermann *et al.*, 2018). In addition, the prediction of the metastasis risk of cancer patients might facilitate to generate tailored treatment (Weigelt *et al.*, 2005). The better prognostic prediction will promote a more personalized medicine approach, which is the future of cancer treatment. However, well-established circulating biomarkers have limited applications. For example, CA15-3, CA27-29, and CEA are only appropriate for metastatic breast cancer relapse monitoring and prognosis, rather than diagnosis (Bast *et al.*, 2001; Ludwig and Weinstein, 2005). Similarly, FDA approved CEA, CA125 and HE4 are markers for ovarian cancer diagnosis and monitoring, but lacking high specificity and sensitivity, especially for patients at an early stage (Hogdall, 2008; Moore *et al.*, 2012; Muinao *et al.*, 2018; Ueland, 2017). Though CTCs are very promising prognostic markers of metastatic breast cancer, their detection methods increase the false positive and false negative rate (Allan and Keeney, 2010; Cristofanilli *et al.*, 2004). Therefore, it is urgent to develop better and more accurate markers for breast cancer and ovarian cancer early detection, as well as for metastatic breast cancer early detection and prognosis.

Circulating biomarkers, especially biomarkers in blood, are suitable and practical for routine clinical use. Compared to tumor tissues, blood samples can be easily accessed with minimally invasive procedures, which is highly appropriate for repeated sampling and continuously monitoring. Blood metabolic profile alterations, which have been proposed as diagnostic, prognostic, and predictive markers for several cancers including breast cancer and ovarian cancer, were investigated (Gunther, 2015; Hart *et al.*, 2017; Loke and Lee, 2018; Plewa *et al.*, 2019). Metabolites represent the ultimate changes of the organism and could directly reflect the phenotypic properties, which help to elucidate the nature and identity of biological processes (Beger, 2013; Capati *et al.*, 2017). This is the most significant advantage

of metabolites over other signatures such as somatic DNA mutations, DNA methylations, microRNAs, or proteins. Measurement of metabolites by mass spectrometry is fast, sensitive, robust, and needs only a very small sample volume. These properties indicate the potential of metabolites as biomarkers for breast cancer and ovarian cancer.

5.1 Metabolites Associated with PBC

Comparative profiling of plasma metabolite levels generated a list of 18 metabolites that were significantly different (FDR < 0.05) between PBC patients and healthy controls. In detail, increased levels of C2 and C14:1, and decreased levels of Ala, Asn, Glu, His, Leu, Lys, Met, Orn, Phe, Thr, Trp, Tyr, Val, Kynurenine, Met-SO and C3 in PBC samples were observed.

A multi-marker panel is superior to a single marker, because the multi-marker panel is more informative, accurate, and stable. The panel of metabolites differentiating PBC patients from healthy controls showed that a combination of the discriminatory power of individual metabolites resulted in higher performance than each metabolite. An AUC of 0.80 was achieved, wherein the models consisted of Glu, Orn, Thr, Trp, Met-SO, C2, and C3 that may serve as potential early diagnostic biomarkers of breast cancer. The inclusion of metabolites in the final model was based on a penalized automatic variable selection with ten-fold cross-validation regression analysis. The cross-validation method randomly divides the data into training and validation tests, so multiple rounds of cross-validation are performed, reducing data variability and over-fitting. This is statistically equivalent to independent validation, so it is crucial here because these results are not independently validated. Pathway enrichment analysis revealed significant enrichment of 12 pathways, including disturbances in amino acid metabolism, aminoacyl-tRNA biosynthesis, and nitrogen metabolism.

As one of the most prominent discriminatory features, the majority of amino acids in plasma of PBC patients were lower than in healthy controls, which is consistent with the results of most previous reports of breast cancer patients with lower amino acid amounts than healthy controls (Jove *et al.*, 2017; Shen *et al.*, 2013; Wang *et al.*, 2016a; Zhou *et al.*, 2017). Lower levels of amino acids might associate with higher protein requests in the cancer cell to

support its rapid growth and proliferation (Lai *et al.*, 2005). Another feature of this study is the significant increase in fatty acid and lipid levels in PBC patients, relative to those in healthy controls, which have been mentioned in other reports (Lv and Yang, 2012; Wang *et al.*, 2016a; Zhou *et al.*, 2017).

5.2 Metabolites Associated with MBC

MBC subgroups defined by CTC status had distinct plasma metabolite profiles, as evident from the clustering of samples. Abundances of 19 metabolites for CTC-positive MBC samples and 12 metabolites for CTC-negative MBC samples were validated to significantly (FDR < 0.05) different from healthy controls. Increased levels of Glu, C4:0, C16:0, C18:1, PC aa C40:6 in CTC-positive MBC samples, and Glu, C2:0, C16:0, C18:0, C18:1, C18:2, PC aa C40:6 in CTC-negative MBC were observed when compared to healthy controls. On the other hand, Asn, His, Trp, and lysoPC a C18:2 were decreased in all MBC samples in comparison to healthy controls.

Multivariate ten-fold cross-validated LASSO regression analysis generated two panels of metabolites to distinguish between MBC subgroups and healthy controls. All metabolites were allowed to enter the model as penalized variables and retained as long as they have unique contributions to the discriminative power of the model. For discriminating CTC-positive MBC from healthy samples, a combination of His, C4:0, C18:1, lysoPC a C18:2, PC aa C40:6, and PC ae C42:3 was postulated with AUC = 0.92. Although the difference between CTC-negative MBC and healthy controls were not as high as the CTC-positive MBC, combining Asn, Glu, His, Thr, Trp, C16:0, C18:0, C18:1, C18:2, lysoPC a C18:2, and PC aa C40:6 could differentiate CTC-negative MBC and control with an AUC of 0.89.

Metabolites that were different between MBC subgroups and PBC patients exhibited many overlaps with those between MBC subgroups and healthy controls. Among the 25 significantly different metabolites, a combination of Arg, Glu, Orn, C4:0, lysoPC a C18:2, lysoPC a C20:4, PC aa C40:6, PC aa C42:1, PC ae C34:2, PC ae C34:3, PC ae C36:3, PC ae C42:1, PC ae C42:3, PC ae C42:5, and PC ae C44:4 was postulated with AUC = 0.95 for

CTC-positive MBC samples, whereas within the nine significantly different metabolites, a panel of Arg, Glu, Orn, Met-SO, spermidine, C4:0, and lysoPC a C18:2 with AUC = 0.90 was selected for CTC-negative MBC samples compared to PBC samples. Nowadays, MBC patient detection depends on the clinical manifestations of disease transmission to distant organs, a biopsy of affected organs, imaging methods, radiological assessment, and serum tumor markers (Scully *et al.*, 2012). However, these methods are less effective in detecting metastasis, especially at the early stage, when the patient could have a better outcome (Scully *et al.*, 2012). Serum markers such as CA 15-3 and CEA are widely used to detect the development of breast cancer metastasis, however, only useful in 54%-80% and 30%-50% of MBC patients, respectively (Lee *et al.*, 2013). In comparison to these markers, the performance of our panel of metabolites was much higher for distinguishing subjects with MBC from healthy controls or PBC. Thus, the panels of metabolites identified here, if verified in large multi-centric prospective studies, might be potential diagnostic markers, which would have important applications in identifying early-stage MBC patients or micro-metastatic patients without clinically overt metastasis, which are usually missed by imaging techniques. In turn, early capture of the disease might increase the survival rate of patients with MBC.

Phosphatidylcholines were the mostly identified metabolites associated with MBC, and they were widely studied in many cancer types (Bertini *et al.*, 2012; Mir *et al.*, 2015). However, for breast cancer, they were mostly studied in breast cancer tissues (Brockmoller *et al.*, 2012; Hilvo *et al.*, 2011) or blood samples from mice with MBC (Kus *et al.*, 2018), and they were present at lower levels in all these studies. lysoPC a C18:2, a lysophosphatidylcholine, decreases with the progression of breast cancer. Previous studies have claimed that most of the amino acid levels in patients with metastatic breast cancer are high (De Luca *et al.*, 2010; Hart *et al.*, 2017; Seidlitz *et al.*, 2009). Dysregulation of acylcarnitines has been widely reported in other cancers types, but rarely in MBC blood samples (Lu *et al.*, 2016; Shen *et al.*, 2013; Xie *et al.*, 2017). The involvement of here-identified metabolites in other epithelial cancers raises the questions about their specificity for breast cancer and we hypothesize that they might be generally elevated in epithelial cancer.

5.3 Putative Prognostic Metabolites for MBC

The prognostic significance of plasma metabolites in MBC patients was investigated and characterized by analyzing their correlations with PFS and OS. Disease progression quantified as PFS is one of the original four categorical outcomes developed by the WHO to describe changes in tumor burden (Miller *et al.*, 1981). Determination of whether the disease has progressed or not was established based on the RECIST guideline in the MBC cohort (Eisenhauer *et al.*, 2009). PFS and OS are important outcomes in advanced cancers wherein PFS and OS are related to quality and quantity of life for cancer patients, respectively (Booth and Eisenhauer, 2012; Siena *et al.*, 2007).

Within MBC patients, metabolites that were differentially expressed between patients with poor and good prognosis correlated with PFS, while metabolites that were differentially expressed between deceased and alive patients correlated with OS. Log-rank tests were applied to metabolites data stratified into lower quartile and the residual quartiles based on their concentrations. A list of four metabolites, kynurenine, PC aa C36:3, PC ae C36:1, and PC ae C38:3, correlated with PFS and 12 metabolites, lysoPC a C18:1, lysoPC a C20:3, lysoPC a C20:4, PC aa C36:3, PC aa C36:4, PC aa C38:5, PC ae C36:1, PC ae C38:3, PC ae C38:4, PC ae C40:2, SM C18:1, and SM (OH) C22:2, correlated with OS. Kaplan-Meier curves indicated that lower levels of the metabolites correlated with the lower probability of PFS and OS, namely, a shorter progress-free survival time and a shorter overall survival time.

Penalized LASSO Cox regression analysis was carried out to predict putative metabolite combinations from these prognostic metabolites with optimal performance. Signatures of PC ae C36:1 and PC ae C38:3 for PFS, lysoPC a C20:3, lysoPC a C20:4, PC aa C38:5, PC ae C38:3, and SM (OH) C22:2 for OS were put forth by these analyses. To understand the performance of metabolite panels associated with currently available markers, they were compared to CTC status, which is an established and FDA approved prognostic marker for MBC (Allard *et al.*, 2004). Evaluating the IPE of the model with metabolites or CTC status proved that the metabolite signatures (IPE = 5.669) performed not as well as CTC status (IPE = 5.628) for PFS, the difference in IPE between the metabolite panel (IPE = 5.795) and the CTC status (IPE = 5.268) was much wider for OS. Next, a panel combining both metabolite

variables and CTC status was built. Here, as we found no significant interaction between metabolites and CTC status, interaction between the metabolites and CTC status was ignored. The results showed that the combination of the CTC status and the metabolite panel largely diminish the prediction error in comparison to the CTC-status alone or metabolite profile alone. It is interesting to note that the best-fitting multivariable models for PFS were PC ae C36:1 and CTCs (5.535), for OS were lysoPC a C20:3 and CTCs (5.082). This study is the first to explore the associations of metabolite and CTC status combination with breast cancer prognosis.

Prognostic significance of plasma phospholipids and sphingolipids has been demonstrated in breast cancer and other cancers (Bertini *et al.*, 2012; Bilal *et al.*, 2019; Hilvo *et al.*, 2011; Ruckhaberle *et al.*, 2008; Zaimenko *et al.*, 2019). Serum phosphatidylcholine levels decreased in breast cancer patients after systemic chemotherapy (Li *et al.*, 2017b). The lower stearic acid level in tumor membrane phosphatidylcholine, which related to the fatty acid composition of phosphatidylcholine, indicates poor breast cancer prognosis (Bougnoux *et al.*, 1992). Besides, a study revealed that the survival of patients with metastatic lung and gastrointestinal cancers correlates with low levels of total plasma phospholipids (Murphy *et al.*, 2010), which supports our hypothesis. This result is consistent with previous studies showing that higher plasma levels of lysophosphatidylcholines are related to a lower risk of several cancer types, including breast cancer (Kuhn *et al.*, 2016; Zhao *et al.*, 2007). In addition, higher levels of sphingomyelins were reported to be significantly associated with better DFS in triple-negative breast cancer patients (Purwaha *et al.*, 2018), which is in concordance with our results. In future, large multi-centric cohorts will be needed to test the reproducibility of our results.

5.4 Metabolites Associated with OVCA

Next, we identified metabolites associated with ovarian cancer. A panel of seven metabolites, including His, Trp, C18:1, lysoPC a C18:2, PC aa C32:2, PC aa C34:4, PC ae C34:3, could distinguish ovarian cancer patients from healthy controls with an AUC of 0.95. Such a high

performance raises the hope that the identified metabolites, if they can be validated in large prospective studies, can prospectively indicate the existence of ovarian cancer at an earlier stage than ovarian cancers are diagnosed today. Moreover, this panel of metabolites could distinguish OVCA patients from PBC patients as well, which indicates their OVCA-specific discriminative feature. In addition, the measurement of plasma metabolites with p400 kit has extended the library of potential candidate markers. Currently, no sufficient screening strategy has been established for OVCA. CA125 and HE4 are the two blood-based protein markers for OVCA monitoring approved by FDA (Ueland, 2017). However, screening with CA125 or vaginal ultrasound did not show a significant mortality reduction (Jacobs *et al.*, 2016), and nowadays, CA125 is only used as an indicator of disease recurrence due to its limitation in specificity and sensitivity (Suh *et al.*, 2010). Serum HE4 levels are changing with aging, different menopausal status, and pregnancy, which lower the specificity of HE4 (Moore *et al.*, 2012). The multivariate index assay demonstrated lower specificity compared with physician assessment and CA 125 in detecting ovarian malignancies (Ueland *et al.*, 2011).

Levels of lipid alterations are usually associated with cancer progression (Chen *et al.*, 2016; Zhang *et al.*, 2016). Lipid metabolism is increasingly required for cell membranes and cell signaling construction during rapid cell proliferation and apoptosis (Zhang *et al.*, 2018). In our study, the decreased levels of lipids in OVCA patients, especially phospholipids, are associated with an increased risk for ovarian cancer. Similar results have been partially presented before (Bachmayr-Heyda *et al.*, 2017; Fan *et al.*, 2016a; Gaul *et al.*, 2015; Plewa *et al.*, 2019). Lysophosphatidylcholine is an important signaling molecule for cellular proliferation, inflammation, and cell invasion regulation (Li *et al.*, 2017a). Phosphatidylcholine, another critical signaling molecule, regulates cell signaling and structural integrity of cell membranes (Li *et al.*, 2017a). Previous studies provided evidence of the associations between amino acid levels and OVCA metabolism, which demonstrated that blood amino acid profiles might be useful for ovarian cancer detection (Plewa *et al.*, 2017; Zhou *et al.*, 2010). Our present study also observed correlations between decreased Ala, Asn, Cit, His, Trp, and Tyr levels and increased ovarian cancer risk, except for Glu.

5.5 Functional Relevance of Identified Metabolites

Amino acids, being both metabolic fuel and building blocks of protein synthesis, have been proposed as predominant participants in the development of cancer (Lai *et al.*, 2005). In general, amino acids decreased in PBC patients but increased in MBC and OVCA patients in our study. Hart et al. investigated higher levels of most amino acids in metastatic breast cancer patients than the ones at the early stage, which supports our results (Hart *et al.*, 2017). Bone metastasis of breast cancer exhibits increased secretion of glutamate into the extracellular environment, which contributes to the disruption of normal bone homeostasis (Seidlitz *et al.*, 2009).

The increased activity of arginase and ornithine decarboxylase might explain the lower amount of arginine and the higher level of spermidine (Miolo *et al.*, 2016; Perez *et al.*, 2012). Methionine is a major target of reactive oxygen species (ROS), leading to the formation of Met-SO (Moskovitz, 2005). During breast cancer metastasis, a decreased amount of methionine sulfoxide reductases causes Met-SO to accumulate (De Luca *et al.*, 2010). Tryptophan may originate from an increased metabolization via the indoleamine-2,3-dioxygenase to kynurenine or via tryptophan-hydroxylase to serotonin (O'Mahony *et al.*, 2015). Tryptophan is particularly unexpected in early-stage cancer patients, as it is associated with cancer progression (Miyagi *et al.*, 2011). Kynurenine not only promotes epithelial-to-mesenchymal transition via activating aryl hydrocarbon receptor but also promotes the degradation of E-cadherin by enhancing the formation of the E-cadherin/AhR/Skp2 complex. These processes are believed to play vital roles in cancer invasiveness and metastasis (Chen *et al.*, 2014; Duan *et al.*, 2018).

Glutamate, the interchangeable substrate of glutamine, was significantly lower in PBC patients, whereas significantly higher in metastatic breast cancer and ovarian cancer patients was observed by us. Glutaminolysis is one of the distinctive central features of cancer metabolism (Cluntun *et al.*, 2017). Thereby glutamine is converted via glutamate to α-ketoglutarate, which enters the tricarboxylic cycle to produce energy and anabolic carbons for the synthesis of amino acids, nucleotides, and lipids (Jin *et al.*, 2016). These biological processes may partially explain the change of glutamate levels in our results.

The putrescine is a polyamine product from ornithine that catalyzed by ornithine decarboxylase, which exists in breast cancer cells and other cancer types (Afshar *et al.*, 2017; Kim *et al.*, 2017). Polyamine is indispensable for normal cell growth and development, and its elevated level has been observed to associate with several cancer types(Luk and Casero, 1987). We observed a lower level of ornithine in PBC patients, which possibly due to its conversion to polyamine.

Compared to healthy controls, PBC patients have significantly lower threonine concentrations in our study. We speculate that this may be a sign of increased production of pyruvate from threonine, which also associated with impaired glutamate uptake in a previous study (Poschke *et al.*, 2013). On the other hand, threonine also has a role in immunoglobulin production (Wils-Plotz *et al.*, 2013). The gut may absorb a large proportion of dietary threonine for the production of immunoglobulins because of disease or treatment. This is an interesting perspective and future research is needed to assess this possibility.

Acylcarnitines play an essential role in regulating the balance between intracellular carbohydrate and lipid metabolism (Li *et al.*, 2019). Increased concentrations of acylcarnitines hint an excess of acetyl-CoA in cancer cells, which are converted to C2 by the carnitine O-acetyltransferase with subsequent excretion (Pietrocola *et al.*, 2015). As acetyl-CoA inhibits pyruvate dehydrogenase and induces gluconeogenesis, so its accumulation is undesirable in cancer cells (Li *et al.*, 2014; Schug *et al.*, 2015). In order to cross the mitochondrial matrix for fatty acid β-oxidation, which produces energy to maintain cancer progression, long-chain fatty acids must be conjugated to carnitine to form long-chain acylcarnitines (Schug *et al.*, 2015). We speculate that this process leads to the accumulation of long-chain acylcarnitines observed in cancer plasma. Compared to healthy controls, less free carnitine and odd number short-chain acylcarnitines C3 and C5 were observed by us in the plasma of PBC patients because the conjugation of medium- and long-chain acyl groups to free carnitine (C0). Similar results had demonstrated that several acylcarnitines associated with β-oxidation showed higher plasma concentrations in breast cancer patients than in healthy controls (Lv and Yang, 2012; Shen *et al.*, 2013).

In our study, the majority of the identified metabolites associated with MBC and OVCA were phosphatidylcholines. OVCA and CTC-positive patients showed lower levels of phosphatidylcholines, so we speculate that phosphatidylcholines participate in the metastasis of cancer, especially in the late stage. Indeed, a study showing that the survival of patients with metastatic lung and gastrointestinal cancers correlates with low levels of total plasma phospholipids supports our hypothesis (Murphy *et al.*, 2010). Currently, few phosphatidylcholines have been investigated in human breast cancer blood samples. Most of them were studied either in breast cancer tissues (Brockmoller *et al.*, 2012; Hilvo *et al.*, 2011) or blood samples derived from mice with MBC (Kus *et al.*, 2018). Phosphatidylcholines were present at lower levels in all these studies. Lower levels of phosphatidylcholines were also observed in OVCA patients (Fan *et al.*, 2016a). Phosphatidylcholines are one of the most abundant glycerophospholipids in human cells, which are essential components of the cell membrane, maintaining its integrity and functions. Phosphatidylcholines play a crucial role in cell proliferation and apoptosis (Ridgway, 2013). Also, we found lysoPC a C18:2, one of the lysophosphatidylcholines, decreases as breast cancer progressed. This result is consistent with previous studies showing that higher plasma levels of lysophosphatidylcholines are related to a lower risk of several cancer types, including breast cancer (Kuhn *et al.*, 2016; Zhao *et al.*, 2007).

5.6 Strengths and Limitations of the Study

The significant strength of this project is that the samples were carefully collected and processed. Pre-analytical conditions play a vital role in investigating blood metabolites and are one of the main reasons for the lack of reproducibility in results among studies (Santos Ferreira *et al.*, 2019). According to a previous report and our experience, the processing of blood samples within two hours after blood collection and two-step centrifugation before snap-freezing are critical to avoid hemolysis and metabolite contamination from cells or cell debris (Yin *et al.*, 2013). This protocol is designed to make sure that the metabolites originated from the cell-free portion of the blood exclusively. Sample preparation is of translational importance as not all samples are prepared ideally in a clinical setting, so the

newly speculated markers need to be able to buffer such differences. The choice of plasma as the sample type is also one of the advantages. Serum is more sensitive than plasma in metabolomics study because most metabolites have higher concentrations in serum (Liu *et al.*, 2010; Yu *et al.*, 2011). However, partly this higher metabolite levels are suspected to originate from metabolites that have been released from cells or cell debris into the serum during sampling, as plasma is more reproducible than serum, which is quite popular in the biomarker study (Yu *et al.*, 2011).

The other significant advantage of this study is the systematic study design consisting of the discovery and validation phase. Plasma metabolites were profiled in both phases with identical experimental procedures independently to demonstrate metabolites that have potential diagnostic and prognostic values. Sample randomization through all steps of the experimental setup was done to avoid batch variations and, thus, any possible bias. The large sample size of breast cancer patients is another positive aspect. Additionally, the prognostic metabolites were validated in an independent sample group for their association with PFS and OS, which consolidate the validity of our results.

One major drawback of the study is that, with this experiment, the length of side chains, distribution and position of double bonds of metabolites cannot be discerned. Thus, the exact conformations of metabolites are unknown. In addition, this kit identifies many metabolites with functions not well known currently. As a result, the reason why they appear in the blood or plasma is hard to explain.

Another limitation of the study is the lack of lifestyle and physiological information of cancer patients. Thus we cannot adjust all the confounding factors with the univariate and multivariate logistic regression analysis. In the study, we have identified panels of metabolites as supposed diagnostic and prognostic markers with high performance. However, whether they are specific for their corresponding cancer type is unknown. Furthermore, no animal model experiment was conducted to explain or investigate our results concerning the underlying mechanism.

Last but not least, the sample size of OVCA patients, especially early-stage OVCA samples, is relatively low. Therefore, the metabolite panel identified here fits the late-stage

OVCA patients better. This limitation reflects the general diagnosis of ovarian cancer at a relatively late stage. In our study, we focused on OVCA patients with rather an early-stage diagnosis. But as the sample size would have been too small, we had to include also even more advanced-stage samples. Large studies including more early-stage OVCA patients as well as large prospective studies are needed to identify and verify metabolites for early-stage detection. The fact that no significant metabolites were detected between breast tumor molecular subgroups may be due to the uneven distribution of samples between subgroups, resulting in relatively small sample sizes of rare subgroups. Large cohorts in the future including enough samples of even rare breast cancer subgroups are needed.

6. Perspectives

Identified metabolite panels associated with PBC, MBC, and OVCA patients and MBC prognosis are needed to be further validated with independent and more extensive multi-centric prospective studies. Metabolites connected with breast cancer and ovarian cancer from different layers could extend to other epithelial cancers. Therefore, population-based metabolomics studies are essential to be conducted. Especially prognostic metabolites for metastatic breast cancer should be analyzed in epithelial cancers that also use CTC status as prognostic factors for metastasis, such as prostate cancer and colorectal cancer.

In this project, metabolites of different combinations are speculated as promising biomarkers for breast cancer and ovarian cancer. However, whether these combinatorial panels are specific for the corresponding cancer types needs to be further verified. Plasma metabolite profiles are also affected by patient lifestyles, physiological status, and environmental factors. Unfortunately, this information was not available for most of the cancer patients in our study. Investigations in the future should be applied to adjust all meaningful confounding factors. Moreover, higher mass resolution methods need to be employed to measure the exact composition of lipids. Metabolites that measured with Biocrates Absolute*IDQ*® p400 HR kit for breast cancer patients will be analyzed in the future to expand the library of breast cancer diagnostic and prognostic biomarkers.

It might be promising and illuminating to combine metabolomics data and transcriptomics data, together with genomics data, to explore the underlying mechanisms of metabolic changes. The multi-omics analysis will not only deepen our understanding of the metabolic processes of breast cancer and ovarian cancer but also likely help to develop new therapeutic targets for cancer treatment. The functional characterization of the metabolites identified in this project may promote understanding of the biological processes in cancer occurrence and development. Further, the combination of metabolites with other promising blood-based molecular markers, such as DNA methylation, microRNA, cell-free DNA mutation markers, will be an attractive option. Multimarkers from different layers could be

potentially useful in developing blood-based diagnostic tests for breast cancer and ovarian cancer.

Bibliography

Aboud, O. A. and Weiss, R. H. (2013). **New opportunities from the cancer metabolome**. Clin Chem *59*, 138-146, doi: 10.1373/clinchem.2012.184598.

Afshar, A. S., Nematpour, F. S., Meshkani, M. and Khafi, A. (2017). **Growth inhibition of human breast cancer cells and down-regulation of ODC1 and ADA genes by Nepeta binaloudensis**. Rev Bras Farmacogn *27*, 84-90, doi: 10.1016/j.bjp.2016.07.005.

Aktas, B., Muller, V., Tewes, M., Zeitz, J., Kasimir-Bauer, S., Loehberg, C. R., Rack, B., Schneeweiss, A., Fehm, T. and Grp, D. S. (2011). **Comparison of estrogen and progesterone receptor status of circulating tumor cells and the primary tumor in metastatic breast cancer patients**. Gynecologic Oncology *122*, 356-360, doi: 10.1016/j.ygyno.2011.04.039.

Alix-Panabieres, C. and Pantel, K. (2014). **Technologies for detection of circulating tumor cells: facts and vision**. Lab Chip *14*, 57-62, doi: 10.1039/c3lc50644d.

Allan, A. L. and Keeney, M. (2010). **Circulating tumor cell analysis: technical and statistical considerations for application to the clinic**. J Oncol *2010*, 426218, doi: 10.1155/2010/426218.

Allard, W. J., Matera, J., Miller, M. C., Repollet, M., Connelly, M. C., Rao, C., Tibbe, A. G. J., Uhr, J. W. and Terstappen, L. W. M. M. (2004). **Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases**. Clinical Cancer Research *10*, 6897-6904, doi: 10.1158/1078-0432.Ccr-04-0378.

Allison, K. H. (2012). **Molecular pathology of breast cancer: what a pathologist needs to know**. Am J Clin Pathol *138*, 770-780, doi: 10.1309/AJCPIV9IQ1MRQMOO.

Anderson, D. E. (1992). **Familial versus sporadic breast cancer**. Cancer *70*, 1740-1746.

Ashworth, T. R. (1869). **A case of cancer in which cells similar to those in the tumours were seen in the blood after death**. Australian Medical Journal *14*, 146–147.

Bachmayr-Heyda, A., Aust, S., Auer, K., Meier, S. M., Schmetterer, K. G., Dekan, S., Gerner, C. and Pils, D. (2017). **Integrative Systemic and Local Metabolomics with Impact on Survival in High-Grade Serous Ovarian Cancer**. Clinical Cancer Research *23*, 2081-2092, doi: 10.1158/1078-0432.Ccr-16-1647.

Bast, R. C., Ravdin, P., Hayes, D. F., Bates, S., Fritsche, H., Jessup, J. M., Kemeny, N., Locker, G. Y., Mennel, R. G., Somerfield, M. R. and M, A. S. C. O. T. (2001). **2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: Clinical practice guidelines of the American Society of Clinical Oncology**. J Clin Oncol *19*, 1865-1878, doi: 10.1200/Jco.2001.19.6.1865.

Beger, R. D. (2013). **A review of applications of metabolomics in cancer**. Metabolites *3*, 552-574, doi: 10.3390/metabo3030552.

Bell, D. A. (2005). **Origins and molecular pathology of ovarian cancer**. Mod Pathol *18 Suppl 2*, S19-32, doi: 10.1038/modpathol.3800306.

Bernards, R. and Weinberg, R. A. (2002). **A progression puzzle**. Nature *418*, 823, doi: 10.1038/418823a.

Bertini, I., Cacciatore, S., Jensen, B. V., Schou, J. V., Johansen, J. S., Kruhoffer, M., Luchinat, C., Nielsen, D. L. and Turano, P. (2012). **Metabolomic NMR fingerprinting to identify and predict survival of patients with metastatic colorectal cancer**. Cancer Res *72*, 356-364, doi: 10.1158/0008-5472.CAN-11-1543.

Bilal, F., Montfort, A., Gilhodes, J., Garcia, V., Riond, J., Carpentier, S., Filleron, T., Colacios, C., Levade, T., Daher, A., Meyer, N., Andrieu-Abadie, N. and Segui, B. (2019). **Sphingomyelin Synthase 1 (SMS1) Downregulation Is Associated With Sphingolipid Reprogramming and a Worse Prognosis in Melanoma**. Frontiers in Pharmacology *10*, 443, doi: 10.3389/fphar.2019.00443.

Bleyer, A. and Welch, H. G. (2012). **Effect of three decades of screening mammography on breast-cancer incidence**. N Engl J Med *367*, 1998-2005, doi: 10.1056/NEJMoa1206809.

Bloom, H. J. and Richardson, W. W. (1957). **Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years**. Br J Cancer *11*, 359-377, doi: 10.1038/bjc.1957.43.

Booth, C. M. and Eisenhauer, E. A. (2012). **Progression-Free Survival: Meaningful or Simply Measurable?** Journal of Clinical Oncology *30*, 1030-1033, doi: 10.1200/Jco.2011.38.7571.

Boroughs, L. K. and DeBerardinis, R. J. (2015). **Metabolic pathways promoting cancer cell survival and growth**. Nat Cell Biol *17*, 351-359, doi: 10.1038/ncb3124.

Bougnoux, P., Chajes, V., Lanson, M., Hacene, K., Body, G., Couet, C. and Le Floch, O. (1992). **Prognostic significance of tumor phosphatidylcholine stearic acid level in breast carcinoma**. Breast Cancer Res Treat *20*, 185-194.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. and Jemal, A. (2018). **Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries**. CA Cancer J Clin, doi: 10.3322/caac.21492.

Bristow, R. E., Smith, A., Zhang, Z., Chan, D. W., Crutcher, G., Fung, E. T. and Munroe, D. G. (2013). **Ovarian malignancy risk stratification of the adnexal mass using a multivariate index assay**. Gynecologic Oncology *128*, 252-259, doi: 10.1016/j.ygyno.2012.11.022.

Brockmoller, S. F., Bucher, E., Muller, B. M., Budczies, J., Hilvo, M., Griffin, J. L., Oresic, M., Kallioniemi, O., Iljin, K., Loibl, S., Darb-Esfahani, S., Sinn, B. V., Klauschen, F., Prinzler, J., Bangemann, N., Ismaeel, F., Fiehn, O., Dietel, M. and Denkert, C. (2012). **Integration of metabolomics and expression of glycerol-3-phosphate acyltransferase (GPAM) in breast cancer-link to patient survival, hormone receptor status, and metabolic profiling**. J Proteome Res *11*, 850-860, doi: 10.1021/pr200685r.

Broeders, M., Moss, S., Nystrom, L., Njor, S., Jonsson, H., Paap, E., Massat, N., Duffy, S., Lynge, E., Paci, E. and Group, E. W. (2012). **The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies**. J Med Screen *19 Suppl 1*, 14-25, doi: 10.1258/jms.2012.012078.

Buas, M. F., Gu, H. W., Djukovic, D., Zhu, J. J., Drescher, C. W., Urban, N., Raftery, D. and Li, C. I. (2016). **Identification of novel candidate plasma metabolite biomarkers for distinguishing serous ovarian carcinoma and benign serous ovarian tumors**. Gynecologic Oncology *140*, 138-144, doi: 10.1016/j.ygyno.2015.10.021.

Budd, G. T., Cristofanilli, M., Ellis, M. J., Stopeck, A., Borden, E., Miller, M. C., Matera, J., Repollet, M., Doyle, G. V., Terstappen, L. W. M. M. and Hayes, D. F. (2006). **Circulating tumor cells versus imaging - Predicting overall survival in metastatic breast cancer**. Clinical Cancer Research *12*, 6403-6409, doi: 10.1158/1078-0432.Ccr-05-1769.

Buist, D. S., Porter, P. L., Lehman, C., Taplin, S. H. and White, E. (2004). **Factors contributing to mammography failure in women aged 40-49 years**. J Natl Cancer Inst *96*, 1432-1440, doi: 10.1093/jnci/djh269.

Capati, A., Ijare, O. B. and Bezabeh, T. (2017). **Diagnostic Applications of Nuclear Magnetic Resonance-Based Urinary Metabolomics**. Magn Reson Insights *10*, 1178623X17694346, doi: 10.1177/1178623X17694346.

Cardoso, F., Castiglione, M. and Group, E. G. W. (2009). **Locally recurrent or metastatic breast cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up**. Ann Oncol *20 Suppl 4*, 15-18, doi: 10.1093/annonc/mdp115.

Chaffer, C. L. and Weinberg, R. A. (2011). **A perspective on cancer cell metastasis**. Science *331*, 1559-1564, doi: 10.1126/science.1203543.

Chen, J. Y., Li, C. F., Kuo, C. C., Tsai, K. K., Hou, M. F. and Hung, W. C. (2014). **Cancer/stroma interplay via cyclooxygenase-2 and indoleamine 2,3-dioxygenase promotes breast cancer progression**. Breast Cancer Res *16*, 410, doi: 10.1186/s13058-014-0410-1.

Chen, X., Chen, H., Dai, M., Ai, J., Li, Y., Mahon, B., Dai, S. and Deng, Y. (2016). **Plasma lipidomics profiling identified lipid biomarkers in distinguishing early-stage breast cancer from benign lesions**. Oncotarget *7*, 36622-36631, doi: 10.18632/oncotarget.9124.

Cluntun, A. A., Lukey, M. J., Cerione, R. A. and Locasale, J. W. (2017). **Glutamine Metabolism in Cancer: Understanding the Heterogeneity**. Trends Cancer *3*, 169-180, doi: 10.1016/j.trecan.2017.01.005.

Cohen, S. J., Punt, C. J., Iannotti, N., Saidman, B. H., Sabbath, K. D., Gabrail, N. Y., Picus, J., Morse, M. A., Mitchell, E., Miller, M. C., Doyle, G. V., Tissing, H., Terstappen, L. W. and Meropol, N. J. (2009). **Prognostic significance of circulating tumor cells in patients with metastatic colorectal cancer**. Ann Oncol *20*, 1223-1229, doi: 10.1093/annonc/mdn786.

Colakovic, S., Lukic, V., Mitrovic, L., Jelic, S., Susnjar, S. and Marinkovic, J. (2000). **Prognostic value of CA125 kinetics and half-life in advanced ovarian cancer**. International Journal of Biological Markers *15*, 147-152.

Coleman, R. L., Herzog, T. J., Chan, D. W., Munroe, D. G., Pappas, T. C., Smith, A., Zhang, Z. and Wolf, J. (2016). **Validation of a second-generation multivariate index assay for malignancy risk of adnexal masses**. American Journal of Obstetrics and Gynecology *215*, e1-e11, doi: 10.1016/j.ajog.2016.03.003.

Cristofanilli, M., Broglio, K. R., Guarneri, V., Jackson, S., Fritsche, H. A., Islam, R., Dawood, S., Reuben, J. M., Kau, S. W., Lara, J. M., Krishnamurthy, S., Ueno, N. T., Hortobagyi, G. N. and Valero, V. (2007). **Circulating tumor cells in metastatic breast cancer: biologic staging beyond tumor burden**. Clin Breast Cancer *7*, 471-479, doi: 10.3816/CBC.2007.n.004.

Cristofanilli, M., Budd, G. T., Ellis, M. J., Stopeck, A., Matera, J., Miller, M. C., Reuben, J. M., Doyle, G. V., Allard, W. J., Terstappen, L. W. M. M. and Hayes, D. F. (2004). **Circulating tumor cells, disease progression, and survival in metastatic breast cancer**. New England Journal of Medicine *351*, 781-791, doi: 10.1056/NEJMoa040766.

Cristofanilli, M., Hayes, D. F., Budd, G. T., Ellis, M. J., Stopeck, A., Reuben, J. M., Doyle, G. V., Matera, J., Allard, W. J., Miller, M. C., Fritsche, H. A., Hortobagyi, G. N. and Terstappen, L. W. M. M. (2005). **Circulating tumor cells: A novel prognostic factor for newly diagnosed metastatic breast cancer**. Journal of Clinical Oncology *23*, 1420-1430, doi: 10.1200/Jco.2005.08.140.

Crowley, E., Di Nicolantonio, F., Loupakis, F. and Bardelli, A. (2013). **Liquid biopsy: monitoring cancer-genetics in the blood**. Nat Rev Clin Oncol *10*, 472-484, doi: 10.1038/nrclinonc.2013.110.

Daniilidis, A. and Karagiannis, V. (2007). **Epithelial ovarian cancer. Risk factors, screening and the role of prophylactic oophorectomy**. Hippokratia *11*, 63-66.

Danila, D. C., Heller, G., Gignac, G. A., Gonzalez-Espinoza, R., Anand, A., Tanaka, E., Lilja, H., Schwartz, L., Larson, S., Fleisher, M. and Scher, H. I. (2007). **Circulating tumor cell number and prognosis in progressive castration-resistant prostate cancer**. Clin Cancer Res *13*, 7053-7058, doi: 10.1158/1078-0432.CCR-07-1506.

De Luca, A., Sanna, F., Sallese, M., Ruggiero, C., Grossi, M., Sacchetta, P., Rossi, C., De Laurenzi, V., Di Ilio, C. and Favaloro, B. (2010). **Methionine sulfoxide reductase A down-regulation in human breast cancer cells results in a more aggressive phenotype**. Proc Natl Acad Sci U S A *107*, 18628-18633, doi: 10.1073/pnas.1010171107.

DeBerardinis, R. J. and Chandel, N. S. (2016). **Fundamentals of cancer metabolism**. Sci Adv *2*, e1600200, doi: 10.1126/sciadv.1600200.

DeFilippis, R. A., Chang, H., Dumont, N., Rabban, J. T., Chen, Y. Y., Fontenay, G. V., Berman, H. K., Gauthier, M. L., Zhao, J., Hu, D., Marx, J. J., Tjoe, J. A., Ziv, E., Febbraio, M., Kerlikowske, K., Parvin, B. and Tlsty, T. D. (2012). **CD36 repression activates a multicellular stromal program shared by high mammographic density and tumor tissues**. Cancer Discov *2*, 826-839, doi: 10.1158/2159-8290.CD-12-0107.

Derezinski, P., Klupczynska, A., Sawicki, W., Palka, J. A. and Kokot, Z. J. (2017). **Amino Acid Profiles of Serum and Urine in Search for Prostate Cancer Biomarkers: a Pilot Study**. Int J Med Sci *14*, 1-12, doi: 10.7150/ijms.15783.

Dettmer, K. and Hammock, B. D. (2004). **Metabolomics--a new exciting field within the "omics" sciences**. Environ Health Perspect *112*, A396-397, doi: 10.1289/ehp.112-1241997.

Di Gangi, I. M., Mazza, T., Fontana, A., Copetti, M., Fusilli, C., Ippolito, A., Mattivi, F., Latiano, A., Andriulli, A., Vrhovsek, U. and Pazienza, V. (2016). **Metabolomic profile in pancreatic cancer patients: a consensus-based approach to identify highly discriminating metabolites**. Oncotarget *7*, 5815-5829, doi: 10.18632/oncotarget.6808.

Dong, F., Budhu, A. S. and Wang, X. W. (2009). **Translating the metastasis paradigm from scientific theory to clinical oncology**. Clin Cancer Res *15*, 2588-2593, doi: 10.1158/1078-0432.CCR-08-2356.

Duan, Z., Li, Y. and Li, L. (2018). **Promoting epithelial-to-mesenchymal transition by D-kynurenine via activating aryl hydrocarbon receptor**. Mol Cell Biochem *448*, 165-173, doi: 10.1007/s11010-018-3323-y.

Edge S, Byrd DR, Compton CC, Fritz AG, Greene FL and A, T. (2010). **AJCC Cancer Staging Manual, 7th Edition**, Springer.

Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., Rubinstein, L., Shankar, L., Dodd, L., Kaplan, R., Lacombe, D. and Verweij, J. (2009). **New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)**. Eur J Cancer *45*, 228-247, doi: 10.1016/j.ejca.2008.10.026.

Elmore, J. G., Armstrong, K., Lehman, C. D. and Fletcher, S. W. (2005). **Screening for breast cancer**. JAMA *293*, 1245-1256, doi: 10.1001/jama.293.10.1245.

Ewing J (1928). **Neoplastic diseases: a treatise on tumors**, 3. edn, Philadelphia (PA): WB Saunders.

Fan, L. J., Yin, M. Z., Ke, C. F., Ge, T. T., Zhang, G. M., Zhang, W., Zhou, X. H., Lou, G. and Li, K. (2016a). **Use of Plasma Metabolomics to Identify Diagnostic Biomarkers for Early Stage Epithelial Ovarian Cancer**. Journal of Cancer *7*, 1265-1272, doi: 10.7150/jca.15074.

Fan, L. J., Zhang, W., Yin, M. Z., Zhang, T., Wu, X. Y., Zhang, H. Y., Sun, M., Li, Z. Z., Hou, Y., Zhou, X. H., Lou, G. and Li, K. (2012). **Identification of metabolic biomarkers to diagnose epithelial ovarian cancer using a UPLC/QTOF/MS platform**. Acta Oncologica *51*, 473-479, doi: 10.3109/0284186x.2011.648338.

Fan, Y., Zhou, X., Xia, T. S., Chen, Z., Li, J., Liu, Q., Alolga, R. N., Chen, Y., Lai, M. D., Li, P., Zhu, W. and Qi, L. W. (2016b). **Human plasma metabolomics for identifying differential metabolites and predicting molecular subtypes of breast cancer**. Oncotarget *7*, 9925-9938, doi: 10.18632/oncotarget.7155.

Fehm, T., Hoffmann, O., Aktas, B., Becker, S., Solomayer, E. F., Wallwiener, D., Kimmig, R. and Kasimir-Bauer, S. (2009). **Detection and characterization of circulating tumor cells in blood of primary breast cancer patients by RT-PCR and comparison to status of bone marrow disseminated cells**. Breast Cancer Research *11*, doi: 10.1186/bcr2349.

Feng, X., Liu, X., Luo, Q. and Liu, B. F. (2008). **Mass spectrometry in systems biology: an overview**. Mass Spectrom Rev *27*, 635-660, doi: 10.1002/mas.20182.

Fidler, I. J. and Kripke, M. L. (1977). **Metastasis Results from Preexisting Variant Cells within a Malignant-Tumor**. Science *197*, 893-895, doi: 10.1126/science.887927.

Foulkes, W. D. and Shuen, A. Y. (2013). **In brief: BRCA1 and BRCA2**. J Pathol *230*, 347-349, doi: 10.1002/path.4205.

Franken, B., de Groot, M. R., Mastboom, W. J., Vermes, I., van der Palen, J., Tibbe, A. G. and Terstappen, L. W. (2012). **Circulating tumor cells, disease recurrence and survival in newly diagnosed breast cancer**. Breast Cancer Res *14*, R133, doi: 10.1186/bcr3333.

Gaul, D. A., Mezencev, R., Long, T. Q., Jones, C. M., Benigno, B. B., Gray, A., Fernandez, F. M. and McDonald, J. F. (2015). **Highly-accurate metabolomic detection of early-stage ovarian cancer**. Sci Rep *5*, 16351, doi: 10.1038/srep16351.

Gerds, T. A. and Schumacher, M. (2007). **Efron-type measures of prediction error for survival analysis**. Biometrics *63*, 1283-1287, doi: 10.1111/j.1541-0420.2007.00832.x.

Ghossein, R. A., Bhattacharya, S. and Rosai, J. (1999). **Molecular detection of micrometastases and circulating tumor cells in solid tumors**. Clin Cancer Res *5*, 1950-1960.

Gilks, C. B. and Prat, J. (2009). **Ovarian carcinoma pathology and genetics: recent advances**. Human Pathology *40*, 1213-1223, doi: 10.1016/j.humpath.2009.04.017.

Giuliano, M., Giordano, A., Jackson, S., Hess, K. R., De Giorgi, U., Mego, M., Handy, B. C., Ueno, N. T., Alvarez, R. H., De Laurentiis, M., De Placido, S., Valero, V., Hortobagyi, G. N., Reuben, J. M. and Cristofanilli, M. (2011). **Circulating tumor cells as prognostic and predictive markers in metastatic breast cancer patients receiving first-line systemic treatment**. Breast Cancer Research *13*, doi: 10.1186/bcr2907.

Gonzalez, D. M. and Medici, D. (2014). **Signaling mechanisms of the epithelial-mesenchymal transition**. Science Signaling *7*, doi: 10.1126/scisignal.2005189.

Gralow, J., Ozols, R. F., Bajorin, D. F., Cheson, B. D., Sandler, H. M., Winer, E. P., Bonner, J., Demetri, G. D., Curran, W., Jr., Ganz, P. A., Kramer, B. S., Kris, M. G., Markman, M., Mayer, R. J., Raghavan, D., Ramsey, S., Reaman, G. H., Sawaya, R., Schuchter, L. M., Sweetenham, J. W., Vahdat, L. T., Davidson, N. E., Schilsky, R. L., Lichter, A. S. and American Society of Clinical, O. (2008). **Clinical cancer advances 2007: major research advances in cancer treatment, prevention, and screening--a report from the American Society of Clinical Oncology**. J Clin Oncol *26*, 313-325, doi: 10.1200/JCO.2007.15.4088.

Gregory, P. A., Bert, A. G., Paterson, E. L., Barry, S. C., Tsykin, A., Farshid, G., Vadas, M. A., Khew-Goodall, Y. and Goodall, G. J. (2008). **The mir-200 family and mir-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1**. Nature Cell Biology *10*, 593-601, doi: 10.1038/ncb1722.

Guadagni, F., Ferroni, P., Carlini, S., Mariotti, S., Spila, A., Aloe, S., D'Alessandro, R., Carone, M. D., Cicchetti, A., Ricciotti, A., Venturo, I., Perri, P., Di Filippo, F., Cognetti, F., Botti, C. and Roselli, M. (2001). **A re-evaluation of carcinoembryonic antigen (CEA) as a serum marker for breast cancer: A prospective longitudinal study**. Clinical Cancer Research *7*, 2357-2362.

Gunther, U. L. (2015). **Metabolomics Biomarkers for Breast Cancer**. Pathobiology *82*, 153-165, doi: 10.1159/000430844.

Haber, D. A. and Velculescu, V. E. (2014). **Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA**. Cancer Discov *4*, 650-661, doi: 10.1158/2159-8290.CD-13-1014.

Haglund, C. (1986). **Tumour marker antigen CA125 in pancreatic cancer: a comparison with CA19-9 and CEA**. Br J Cancer *54*, 897-901, doi: 10.1038/bjc.1986.259.

Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B. and King, M. C. (1990). **Linkage of early-onset familial breast cancer to chromosome 17q21**. Science *250*, 1684-1689, doi: 10.1126/science.2270482.

Hanahan, D. and Weinberg, R. A. (2000). **The hallmarks of cancer**. Cell *100*, 57-70, doi: 10.1016/s0092-8674(00)81683-9.

Hanahan, D. and Weinberg, R. A. (2011). **Hallmarks of cancer: the next generation**. Cell *144*, 646-674, doi: 10.1016/j.cell.2011.02.013.

Harris, L., Fritsche, H., Mennel, R., Norton, L., Ravdin, P., Taube, S., Somerfield, M. R., Hayes, D. F. and Bast, R. C. (2007). **American society of clinical oncology 2007 update of recommendations for the use of tumor markers in breast cancer**. Journal of Clinical Oncology *25*, 5287-5312, doi: 10.1200/Jco.2007.14.2364.

Hart, C. D., Tenori, L., Luchinat, C. and Di Leo, A. (2016). **Metabolomics in Breast Cancer: Current Status and Perspectives**. Adv Exp Med Biol *882*, 217-234, doi: 10.1007/978-3-319-22909-6_9.

Hart, C. D., Vignoli, A., Tenori, L., Uy, G. L., To, T. V., Adebamowo, C., Hossain, S. M., Biganzoli, L., Risi, E., Love, R. R., Luchinat, C. and Di Leo, A. (2017). **Serum Metabolomic Profiles Identify ER-Positive Early Breast Cancer Patients at Increased Risk of Disease Recurrence in a Multicenter Population**. Clinical Cancer Research *23*, 1422-1431, doi: 10.1158/1078-0432.Ccr-16-1153.

Hilvo, M., Denkert, C., Lehtinen, L., Muller, B., Brockmoller, S., Seppanen-Laakso, T., Budczies, J., Bucher, E., Yetukuri, L., Castillo, S., Berg, E., Nygren, H., Sysi-Aho, M., Griffin, J. L., Fiehn, O., Loibl, S., Richter-Ehrenstein, C., Radke, C., Hyotylainen, T., Kallioniemi, O., Iljin, K. and Oresic, M. (2011). **Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression**. Cancer Res *71*, 3236-3245, doi: 10.1158/0008-5472.CAN-10-3894.

Hogdall, E. (2008). **Cancer antigen 125 and prognosis**. Curr Opin Obstet Gynecol *20*, 4-8, doi: 10.1097/GCO.0b013e3282f2b124.

Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ and (eds), C. K. (2019). **SEER Cancer Statistics Review, 1975-2016, National Cancer Institute**. URL: https://seer.cancer.gov/csr/1975_2016/, based on November 2018 SEER data submission, posted to the SEER web site, April 2019. [as of 17.06.2019].

Hunter, K. (2006). **Host genetics influence tumour metastasis**. Nat Rev Cancer *6*, 141-146, doi: 10.1038/nrc1803.

Ingham, S. L., Warwick, J., Buchan, I., Sahin, S., O'Hara, C., Moran, A., Howell, A. and Evans, D. G. (2013). **Ovarian cancer among 8005 women from a breast cancer family history clinic: no increased risk of invasive ovarian cancer in families testing negative for BRCA1 and BRCA2**. Journal of Medical Genetics *50*, 368-372, doi: 10.1136/jmedgenet-2013-101607.

Jacobs, I. J., Menon, U., Ryan, A., Gentry-Maharaj, A., Burnell, M., Kalsi, J. K., Amso, N. N., Apostolidou, S., Benjamin, E., Cruickshank, D., Crump, D. N., Davies, S. K., Dawnay, A., Dobbs, S., Fletcher, G., Ford, J., Godfrey, K., Gunu, R., Habib, M., Hallett, R., Herod, J., Jenkins, H., Karpinskyj, C., Leeson, S., Lewis, S. J., Liston, W. R., Lopes, A., Mould, T., Murdoch, J., Oram, D., Rabideau, D. J., Reynolds, K., Scott, I., Seif, M. W., Sharma, A., Singh, N., Taylor, J., Warburton, F., Widschwendter, M., Williamson, K., Woolas, R., Fallowfield, L., McGuire, A. J.,

Campbell, S., Parmar, M. and Skates, S. J. (2016). **Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial**. Lancet *387*, 945-956, doi: 10.1016/S0140-6736(15)01224-6.

Jasbi, P., Wang, D., Cheng, S. L., Fei, Q., Cui, J. Y., Liu, L., Wei, Y., Raftery, D. and Gu, H. (2019). **Breast cancer detection using targeted plasma metabolomics**. J Chromatogr B Analyt Technol Biomed Life Sci *1105*, 26-37, doi: 10.1016/j.jchromb.2018.11.029.

Javadi, S., Ganeshan, D. M., Qayyum, A., Iyer, R. B. and Bhosale, P. (2016). **Ovarian Cancer, the Revised FIGO Staging System, and the Role of Imaging**. AJR Am J Roentgenol *206*, 1351-1360, doi: 10.2214/AJR.15.15199.

JC, R. (1829). **Recherches sur le traitement du cancer sur la compression methodique simple ou combinee et sur l'histoire generale de la meme maladie,2nd**.

Jiang, H., Walter, S. D., Brown, P. E. and Chiarelli, A. M. (2016). **Estimation of screening sensitivity and sojourn time from an organized screening program**. Cancer Epidemiol *44*, 178-185, doi: 10.1016/j.canep.2016.08.021.

Jin, L., Alesi, G. N. and Kang, S. (2016). **Glutaminolysis as a target for cancer therapy**. Oncogene *35*, 3619-3625, doi: 10.1038/onc.2015.447.

Jobard, E., Pontoizeau, C., Blaise, B. J., Bachelot, T., Elena-Herrmann, B. and Tredan, O. (2014). **A serum nuclear magnetic resonance-based metabolomic signature of advanced metastatic human breast cancer**. Cancer Lett *343*, 33-41, doi: 10.1016/j.canlet.2013.09.011.

Jove, M., Collado, R., Quiles, J. L., Ramirez-Tortosa, M. C., Sol, J., Ruiz-Sanjuan, M., Fernandez, M., de la Torre Cabrera, C., Ramirez-Tortosa, C., Granados-Principal, S., Sanchez-Rovira, P. and Pamplona, R. (2017). **A plasma metabolomic signature discloses human breast cancer**. Oncotarget *8*, 19522-19533, doi: 10.18632/oncotarget.14521.

Kaku, T., Ogawa, S., Kawano, Y., Ohishi, Y., Kobayashi, H., Hirakawa, T. and Nakano, H. (2003). **Histological classification of ovarian cancer**. Med Electron Microsc *36*, 9-17, doi: 10.1007/s007950300002.

Kataja, V., Castiglione, M. and Group, E. G. W. (2009). **Primary breast cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up**. Ann Oncol *20 Suppl 4*, 10-14, doi: 10.1093/annonc/mdp114.

Khoo, S. K. and Mackay, E. V. (1976). **Carcinoembryonic Antigen (Cea) in Ovarian Cancer - Factors Influencing Its Incidence and Changes Which Occur in Response to Cytotoxic Drugs**. British Journal of Obstetrics and Gynaecology *83*, 753-759.

Kim, H. I., Schultz, C. R., Buras, A. L., Friedman, E., Fedorko, A., Seamon, L., Chandramouli, G. V. R., Maxwell, G. L., Bachmann, A. S. and Risinger, J. I. (2017). **Ornithine decarboxylase as a therapeutic target for endometrial cancer**. Plos One *12*, doi: 10.1371/journal.pone.0189044.

Kobayashi, T., Nishiumi, S., Ikeda, A., Yoshie, T., Sakai, A., Matsubara, A., Izumi, Y., Tsumura, H., Tsuda, M., Nishisaki, H., Hayashi, N., Kawano, S., Fujiwara, Y., Minami, H., Takenawa, T., Azuma, T. and Yoshida, M. (2013). **A novel serum metabolomics-based diagnostic approach to pancreatic cancer**. Cancer Epidemiol Biomarkers Prev *22*, 571-579, doi: 10.1158/1055-9965.EPI-12-1033.

Konigsberg, R., Obermayr, E., Bises, G., Pfeiler, G., Gneist, M., Wrba, F., de Santis, M., Zeillinger, R., Hudec, M. and Dittrich, C. (2011). **Detection of EpCAM positive and negative circulating tumor cells in metastatic breast cancer patients**. Acta Oncol *50*, 700-710, doi: 10.3109/0284186X.2010.549151.

Kuhn, T., Floegel, A., Sookthai, D., Johnson, T., Rolle-Kampczyk, U., Otto, W., von Bergen, M., Boeing, H. and Kaaks, R. (2016). **Higher plasma levels of lysophosphatidylcholine 18:0 are related to**

a lower risk of common cancers in a prospective metabolomics study. BMC Med *14*, 13, doi: 10.1186/s12916-016-0552-3.

Kumar, N., Shahjaman, M., Mollah, M. N. H., Islam, S. M. S. and Hoque, M. A. (2017). **Serum and Plasma Metabolomic Biomarkers for Lung Cancer**. Bioinformation *13*, 202-208, doi: 10.6026/97320630013202.

Kurman, R. J. (2014). **WHO classification of tumours of female reproductive organs**, Lyon : International Agency for Research on Cancer, 2014.

Kus, K., Kij, A., Zakrzewska, A., Jasztal, A., Stojak, M., Walczak, M. and Chlopicki, S. (2018). **Alterations in arginine and energy metabolism, structural and signalling lipids in metastatic breast cancer in mice detected in plasma by targeted metabolomics and lipidomics**. Breast Cancer Res *20*, 148, doi: 10.1186/s13058-018-1075-y.

Lai, H. S., Lee, J. C., Lee, P. H., Wang, S. T. and Chen, W. J. (2005). **Plasma free amino acid profile in cancer patients**. Semin Cancer Biol *15*, 267-276, doi: 10.1016/j.semcancer.2005.04.003.

Ledermann, J. A., Raja, F. A., Fotopoulou, C., Gonzalez-Martin, A., Colombo, N., Sessa, C. and Group, E. G. W. (2018). **Newly diagnosed and relapsed epithelial ovarian carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up**. Ann Oncol *29*, iv259, doi: 10.1093/annonc/mdy157.

Lee, J. S., Park, S., Park, J. M., Cho, J. H., Kim, S. I. and Park, B. W. (2013). **Elevated levels of serum tumor markers CA 15-3 and CEA are prognostic factors for diagnosis of metastatic breast cancers**. Breast Cancer Res Treat *141*, 477-484, doi: 10.1007/s10549-013-2695-7.

Lee, Y. T. (1985). **Patterns of metastasis and natural courses of breast carcinoma**. Cancer Metastasis Rev *4*, 153-172.

Li, C. I., Uribe, D. J. and Daling, J. R. (2005). **Clinical characteristics of different histologic types of breast cancer**. British Journal of Cancer *93*, 1046-1052, doi: 10.1038/sj.bjc.6602787.

Li, J., Xie, H., Li, A., Cheng, J., Yang, K., Wang, J., Wang, W., Zhang, F., Li, Z., Dhillon, H. S., Openkova, M. S., Zhou, X., Li, K. and Hou, Y. (2017a). **Distinct plasma lipids profiles of recurrent ovarian cancer by liquid chromatography-mass spectrometry**. Oncotarget *8*, 46834-46845, doi: 10.18632/oncotarget.11603.

Li, P., Tian, W. and Ma, X. (2014). **Alpha-mangostin inhibits intracellular fatty acid synthase and induces apoptosis in breast cancer cells**. Mol Cancer *13*, 138, doi: 10.1186/1476-4598-13-138.

Li, S., Gao, D. and Jiang, Y. (2019). **Function, Detection and Alteration of Acylcarnitine Metabolism in Hepatocellular Carcinoma**. Metabolites *9*, doi: 10.3390/metabo9020036.

Li, X., Lim, J., Kolatkar, A., Welter, L., Waitman, K., Hwang, E. S. S., Lu, J. M., Kuhn, P. and Nieva, J. J. (2017b). **Serum phosphatidylcholine is lower among breast cancer patients on systemic chemotherapy**. Journal of Clinical Oncology *35*, doi: 10.1200/JCO.2017.35.15_suppl.e12571.

Liu, L. S., Aa, J. Y., Wang, G. J., Yan, B., Zhang, Y., Wang, X. W., Zhao, C. Y., Cao, B., Shi, J. A., Li, M. J., Zheng, T. A., Zheng, Y. T., Hao, G., Zhou, F., Sun, J. G. and Wu, Z. M. (2010). **Differences in metabolite profile between blood plasma and serum**. Analytical Biochemistry *406*, 105-112, doi: 10.1016/j.ab.2010.07.015.

Loeb, K. R. and Loeb, L. A. (2000). **Significance of multiple mutations in cancer**. Carcinogenesis *21*, 379-385, doi: 10.1093/carcin/21.3.379.

Loke, S. Y. and Lee, A. S. G. (2018). **The future of blood-based biomarkers for the early detection of breast cancer**. European Journal of Cancer *92*, 54-68, doi: 10.1016/j.ejca.2017.12.025.

Lorusso, G. and Ruegg, C. (2012). **New insights into the mechanisms of organ-specific breast cancer metastasis**. Seminars in Cancer Biology *22*, 226-233, doi: 10.1016/j.semcancer.2012.03.007.

Lu, Y. H., Li, N., Gao, L., Xu, Y. J., Huang, C., Yu, K. K., Ling, Q. X., Cheng, Q., Chen, S. S., Zhu, M. Q., Fang, J. L., Chen, M. Q. and Ong, C. N. (2016). **Acetylcarnitine Is a Candidate Diagnostic and Prognostic Biomarker of Hepatocellular Carcinoma**. Cancer Research *76*, 2912-2920, doi: 10.1158/0008-5472.Can-15-3199.

Ludwig, J. A. and Weinstein, J. N. (2005). **Biomarkers in cancer staging, prognosis and treatment selection**. Nat Rev Cancer *5*, 845-856, doi: 10.1038/nrc1739.

Luk, G. D. and Casero, R. A., Jr. (1987). **Polyamines in normal and cancer cells**. Adv Enzyme Regul *26*, 91-105.

Lumachi, F., Brandes, A. A., Ermani, M., Bruno, G. and Boccagni, P. (2000). **Sensitivity of serum tumor markers CEA and CA 15-3 in breast cancer recurrences and correlation with different prognostic factors**. Anticancer Res *20*, 4751-4755.

Lv, W. and Yang, T. (2012). **Identification of possible biomarkers for breast cancer from free fatty acid profiles determined by GC-MS and multivariate statistical analysis**. Clin Biochem *45*, 127-133, doi: 10.1016/j.clinbiochem.2011.10.011.

Lynch, H. T., Snyder, C. and Casey, M. J. (2013). **Hereditary ovarian and breast cancer: what have we learned?** Annals of Oncology *24*, 83-95, doi: 10.1093/annonc/mdt313.

Malhotra, G. K., Zhao, X., Band, H. and Band, V. (2010). **Histological, molecular and functional subtypes of breast cancers**. Cancer Biol Ther *10*, 955-960, doi: 10.4161/cbt.10.10.13879.

Malone, K. E., Daling, J. R., Thompson, J. D., O'Brien, C. A., Francisco, L. V. and Ostrander, E. A. (1998). **BRCA1 mutations and breast cancer in the general population - Analyses in women before age 35 years and in women before age 45 years with first-degree family history**. Jama-Journal of the American Medical Association *279*, 922-929, doi: 10.1001/jama.279.12.922.

Marion, D. (2013). **An introduction to biological NMR spectroscopy**. Mol Cell Proteomics *12*, 3006-3025, doi: 10.1074/mcp.O113.030239.

Marrinucci, D., Bethel, K., Luttgen, M., Bruce, R. H., Nieva, J. and Kuhn, P. (2009). **Circulating tumor cells from well-differentiated lung adenocarcinoma retain cytomorphologic features of primary tumor type**. Arch Pathol Lab Med *133*, 1468-1471, doi: 10.1043/1543-2165-133.9.1468.

Mavaddat, N., Antoniou, A. C., Easton, D. F. and Garcia-Closas, M. (2010). **Genetic susceptibility to breast cancer**. Mol Oncol *4*, 174-191, doi: 10.1016/j.molonc.2010.04.011.

Mavaddat, N., Peock, S., Frost, D., Ellis, S., Platte, R., Fineberg, E., Evans, D. G., Izatt, L., Eeles, R. A., Adlard, J., Davidson, R., Eccles, D., Cole, T., Cook, J., Brewer, C., Tischkowitz, M., Douglas, F., Hodgson, S., Walker, L., Porteous, M. E., Morrison, P. J., Side, L. E., Kennedy, M. J., Houghton, C., Donaldson, A., Rogers, M. T., Dorkins, H., Miedzybrodzka, Z., Gregory, H., Eason, J., Barwell, J., McCann, E., Murray, A., Antoniou, A. C., Easton, D. F. and Embrace (2013). **Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE**. J Natl Cancer Inst *105*, 812-822, doi: 10.1093/jnci/djt095.

McPherson, K., Steel, C. M. and Dixon, J. M. (2000). **ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics**. BMJ *321*, 624-628, doi: 10.1136/bmj.321.7261.624.

Menni, C., Zierer, J., Valdes, A. M. and Spector, T. D. (2017). **Mixing omics: combining genetics and metabolomics to study rheumatic diseases**. Nat Rev Rheumatol *13*, 174-181, doi: 10.1038/nrrheum.2017.5.

Miglioretti, D. L., Lange, J., van den Broek, J. J., Lee, C. I., van Ravesteyn, N. T., Ritley, D., Kerlikowske, K., Fenton, J. J., Melnikow, J., de Koning, H. J. and Hubbard, R. A. (2016). **Radiation-Induced Breast Cancer Incidence and Mortality From Digital Mammography Screening: A Modeling Study**. Ann Intern Med *164*, 205-214, doi: 10.7326/M15-1241.

Miller, A. B., Hoogstraten, B., Staquet, M. and Winkler, A. (1981). **Reporting results of cancer treatment**. Cancer *47*, 207-214.

Milner, B. J., Allan, L. A., Eccles, D. M., Kitchener, H. C., Leonard, R. C. F., Kelly, K. F., Parkin, D. E. and Haites, N. E. (1993). **P53 Mutation Is a Common Genetic Event in Ovarian-Carcinoma**. Cancer Research *53*, 2128-2132.

Miolo, G., Muraro, E., Caruso, D., Crivellari, D., Ash, A., Scalone, S., Lombardi, D., Rizzolio, F., Giordano, A. and Corona, G. (2016). **Pharmacometabolomics study identifies circulating spermidine and tryptophan as potential biomarkers associated with the complete pathological response to trastuzumab-paclitaxel neoadjuvant therapy in HER-2 positive breast cancer**. Oncotarget *7*, 39809-39822, doi: 10.18632/oncotarget.9489.

Mir, S. A., Rajagopalan, P., Jain, A. P., Khan, A. A., Datta, K. K., Mohan, S. V., Lateef, S. S., Sahasrabuddhe, N., Somani, B. L., Prasad, T. S. K., Chatterjee, A., Kumar, K. V. V., VijayaKumar, M., Kumar, R. V., Gundimeda, S., Pandey, A. and Gowda, H. (2015). **LC-MS-based serum metabolomic analysis reveals dysregulation of phosphatidylcholines in esophageal squamous cell carcinoma**. Journal of Proteomics *127*, 96-102, doi: 10.1016/j.jprot.2015.05.013.

Miyagi, Y., Higashiyama, M., Gochi, A., Akaike, M., Ishikawa, T., Miura, T., Saruki, N., Bando, E., Kimura, H., Imamura, F., Moriyama, M., Ikeda, I., Chiba, A., Oshita, F., Imaizumi, A., Yamamoto, H., Miyano, H., Horimoto, K., Tochikubo, O., Mitsushima, T., Yamakado, M. and Okamoto, N. (2011). **Plasma free amino acid profiling of five types of cancer patients and its application for early detection**. PLoS One *6*, e24143, doi: 10.1371/journal.pone.0024143.

Mohammadian-Hafshejani, A., Bakeshei, F. A., Aryaie, M., Salehiniya, H., Khosravi, B., Ghafari, M. and Mohammadian, M. (2017). **Variations in the Incidence and Mortality of Ovarian Cancer and Their Relationship with the Human Development Index in European Countries in 2012**. Biomedical Research and Therapy *4*, doi: 10.15419/bmrat.v4i08.228.

Momenimovahed, Z., Tiznobaik, A., Taheri, S. and Salehiniya, H. (2019). **Ovarian cancer in the world: epidemiology and risk factors**. Int J Womens Health *11*, 287-299, doi: 10.2147/IJWH.S197604.

Moore, R. G., McMeekin, D. S., Brown, A. K., DiSilvestro, P., Miller, M. C., Allard, W. J., Gajewski, W., Kurman, R., Bast, R. C., Jr. and Skates, S. J. (2009). **A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass**. Gynecol Oncol *112*, 40-46, doi: 10.1016/j.ygyno.2008.08.031.

Moore, R. G., Miller, M. C., Eklund, E. E., Lu, K. H., Bast, R. C., Jr. and Lambert-Messerlian, G. (2012). **Serum levels of the ovarian cancer biomarker HE4 are decreased in pregnancy and increase with age**. Am J Obstet Gynecol *206*, 349 e341-347, doi: 10.1016/j.ajog.2011.12.028.

More, T. H., RoyChoudhury, S., Christie, J., Taunk, K., Mane, A., Santra, M. K., Chaudhury, K. and Rapole, S. (2018). **Metabolomic alterations in invasive ductal carcinoma of breast: A comprehensive metabolomic study using tissue and serum samples**. Oncotarget *9*, 2678-2696, doi: 10.18632/oncotarget.23626.

Moskovitz, J. (2005). **Methionine sulfoxide reductases: ubiquitous enzymes involved in antioxidant defense, protein regulation, and prevention of aging-associated diseases**. BBA-Proteins Proteom *1703*, 213-219, doi: 10.1016/j.bbapap.2004.09.003.

Muinao, T., Deka Boruah, H. P. and Pal, M. (2018). **Diagnostic and Prognostic Biomarkers in ovarian cancer and the potential roles of cancer stem cells - An updated review**. Exp Cell Res *362*, 1-10, doi: 10.1016/j.yexcr.2017.10.018.

Murphy, R. A., Wilke, M. S., Perrine, M., Pawlowicz, M., Mourtzakis, M., Lieffers, J. R., Maneshgar, M., Bruera, E., Clandinin, M. T., Baracos, V. E. and Mazurak, V. C. (2010). **Loss of adipose tissue**

and plasma phospholipids: Relationship to survival in advanced cancer patients. Clinical Nutrition *29*, 482-487, doi: 10.1016/j.clnu.2009.11.006.

Nathanson, K. N., Wooster, R. and Weber, B. L. (2001). **Breast cancer genetics: What we know and what we need**. Nature Medicine *7*, 552-556, doi: 10.1038/87876.

NCI Dictionary of Cancer Terms **Tumor marker**. URL: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/tumor-marker [as of 16.06.2019].

Nicolson, G. L. and Winkelhake, J. L. (1975). **Organ specificity of blood-borne tumour metastasis determined by cell adhesion?** Nature *255*, 230-232, doi: 10.1038/255230a0.

Nowell, P. C. (1976). **The clonal evolution of tumor cell populations**. Science *194*, 23-28, doi: 10.1126/science.959840.

O'Mahony, S. M., Clarke, G., Borre, Y. E., Dinan, T. G. and Cryan, J. F. (2015). **Serotonin, tryptophan metabolism and the brain-gut-microbiome axis**. Behav Brain Res *277*, 32-48, doi: 10.1016/j.bbr.2014.07.027.

Oakman, C., Tenori, L., Claudino, W. M., Cappadona, S., Nepi, S., Battaglia, A., Bernini, P., Zafarana, E., Saccenti, E., Fornier, M., Morris, P. G., Biganzoli, L., Luchinat, C., Bertini, I. and Di Leo, A. (2011). **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**. Ann Oncol *22*, 1295-1301, doi: 10.1093/annonc/mdq606.

Paget, S. (1989). **The distribution of secondary growths in cancer of the breast. 1889**. Cancer Metastasis Rev *8*, 98-101.

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F. L., Walker, M. G., Watson, D., Park, T., Hiller, W., Fisher, E. R., Wickerham, D. L., Bryant, J. and Wolmark, N. (2004). **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer**. N Engl J Med *351*, 2817-2826, doi: 10.1056/NEJMoa041588.

Pantel, K. and Alix-Panabieres, C. (2010). **Circulating tumour cells in cancer patients: challenges and perspectives**. Trends Mol Med *16*, 398-406, doi: 10.1016/j.molmed.2010.07.001.

Patti, G. J., Yanes, O. and Siuzdak, G. (2012). **Innovation: Metabolomics: the apogee of the omics trilogy**. Nat Rev Mol Cell Biol *13*, 263-269, doi: 10.1038/nrm3314.

Perez, G., Olivares, I. M., Rodriguez, M. G., Ceballos, G. M. and Garcia Sanchez, J. R. (2012). **Arginase activity in patients with breast cancer: an analysis of plasma, tumors, and its relationship with the presence of the estrogen receptor**. Onkologie *35*, 570-574, doi: 10.1159/000343005.

Pharoah, P. D., Day, N. E., Duffy, S., Easton, D. F. and Ponder, B. A. (1997). **Family history and the risk of breast cancer: a systematic review and meta-analysis**. Int J Cancer *71*, 800-809.

Pietrocola, F., Galluzzi, L., Bravo-San Pedro, J. M., Madeo, F. and Kroemer, G. (2015). **Acetyl coenzyme A: a central metabolite and second messenger**. Cell Metab *21*, 805-821, doi: 10.1016/j.cmet.2015.05.014.

Plewa, S., Horala, A., Derezinski, P., Klupczynska, A., Nowak-Markwitz, E., Matysiak, J. and Kokot, Z. J. (2017). **Usefulness of Amino Acid Profiling in Ovarian Cancer Screening with Special Emphasis on Their Role in Cancerogenesis**. International Journal of Molecular Sciences *18*, doi: 10.3390/ijms18122727.

Plewa, S., Horala, A., Derezinski, P., Nowak-Markwitz, E., Matysiak, J. and Kokot, Z. J. (2019). **Wide spectrum targeted metabolomics identifies potential ovarian cancer biomarkers**. Life Sciences *222*, 235-244, doi: 10.1016/j.lfs.2019.03.004.

Polyak, K. (2011). **Heterogeneity in breast cancer**. J Clin Invest *121*, 3786-3788, doi: 10.1172/JCI60534.

Poschke, I., Mao, Y. M., Kiessling, R. and de Boniface, J. (2013). **Tumor-dependent increase of serum amino acid levels in breast cancer patients has diagnostic potential and correlates with molecular tumor subtypes**. Journal of Translational Medicine *11*, doi: 10.1186/1479-5876-11-290.

Prat, J. and Oncology, F. C. o. G. (2014). **Staging classification for cancer of the ovary, fallopian tube, and peritoneum**. Int J Gynaecol Obstet *124*, 1-5, doi: 10.1016/j.ijgo.2013.10.001.

Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T. L., Smith, S. R., Bamforth, F., Greiner, R., McManus, B., Newman, J. W., Goodfriend, T. and Wishart, D. S. (2011). **The human serum metabolome**. PLoS One *6*, e16957, doi: 10.1371/journal.pone.0016957.

Purwaha, P., Gu, F., Piyarathna, D. W. B., Rajendiran, T., Ravindran, A., Omilian, A. R., Jiralerspong, S., Das, G., Morrison, C., Ambrosone, C., Coarfa, C., Putluri, N. and Sreekumar, A. (2018). **Unbiased Lipidomic Profiling of Triple-Negative Breast Cancer Tissues Reveals the Association of Sphingomyelin Levels with Patient Disease-Free Survival**. Metabolites *8*, doi: 10.3390/metabo8030041.

Qureshi-Baig, K., Ullmann, P., Haan, S. and Letellier, E. (2017). **Tumor-Initiating Cells: a criTICal review of isolation approaches and new challenges in targeting strategies**. Mol Cancer *16*, 40, doi: 10.1186/s12943-017-0602-2.

Racila, E., Euhus, D., Weiss, A. J., Rao, C., McConnell, J., Terstappen, L. W. and Uhr, J. W. (1998). **Detection and characterization of carcinoma cells in the blood**. Proc Natl Acad Sci U S A *95*, 4589-4594, doi: 10.1073/pnas.95.8.4589.

Rauh-Hain, J. A., Krivak, T. C., Del Carmen, M. G. and Olawaiye, A. B. (2011). **Ovarian cancer screening and early detection in the general population**. Rev Obstet Gynecol *4*, 15-21.

Redig, A. J. and McAllister, S. S. (2013). **Breast cancer as a systemic disease: a view of metastasis**. Journal of Internal Medicine *274*, 113-126, doi: 10.1111/joim.12084.

Reya, T., Morrison, S. J., Clarke, M. F. and Weissman, I. L. (2001). **Stem cells, cancer, and cancer stem cells**. Nature *414*, 105-111, doi: 10.1038/35102167.

Ridgway, N. D. (2013). **The role of phosphatidylcholine and choline metabolites to cell proliferation and survival**. Crit Rev Biochem Mol Biol *48*, 20-38, doi: 10.3109/10409238.2012.735643.

Ring, A. E., Zabaglo, L., Ormerod, M. G., Smith, I. E. and Dowsett, M. (2005). **Detection of circulating epithelial cells in the blood of patients with breast cancer: comparison of three techniques**. Br J Cancer *92*, 906-912, doi: 10.1038/sj.bjc.6602418.

Rodrigues, D., Jeronimo, C., Henrique, R., Belo, L., Bastos, M. D., de Pinho, P. G. and Carvalho, M. (2016). **Biomarkers in bladder cancer: A metabolomic approach using in vitro and ex vivo model systems**. International Journal of Cancer *139*, 256-268, doi: 10.1002/ijc.30016.

Ruckhaberle, E., Rody, A., Engels, K., Gaetje, R., von Minckwitz, G., Schiffmann, S., Grosch, S., Geisslinger, G., Holtrich, U., Karn, T. and Kaufmann, M. (2008). **Microarray analysis of altered sphingolipid metabolism reveals prognostic significance of sphingosine kinase 1 in breast cancer**. Breast Cancer Res Treat *112*, 41-52, doi: 10.1007/s10549-007-9836-9.

Santos Ferreira, D. L., Maple, H. J., Goodwin, M., Brand, J. S., Yip, V., Min, J. L., Groom, A., Lawlor, D. A. and Ring, S. (2019). **The Effect of Pre-Analytical Conditions on Blood Metabolomics in Epidemiological Studies**. Metabolites *9*, doi: 10.3390/metabo9040064.

Schnitt, S. J. (2010). **Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy**. Mod Pathol *23 Suppl 2*, S60-64, doi: 10.1038/modpathol.2010.33.

Schug, Z. T., Peck, B., Jones, D. T., Zhang, Q., Grosskurth, S., Alam, I. S., Goodwin, L. M., Smethurst, E., Mason, S., Blyth, K., McGarry, L., James, D., Shanks, E., Kalna, G., Saunders, R. E., Jiang, M., Howell, M., Lassailly, F., Thin, M. Z., Spencer-Dene, B., Stamp, G., van den Broek, N. J., Mackay, G., Bulusu, V., Kamphorst, J. J., Tardito, S., Strachan, D., Harris, A. L., Aboagye, E. O., Critchlow, S. E., Wakelam, M. J., Schulze, A. and Gottlieb, E. (2015). **Acetyl-CoA synthetase 2 promotes acetate utilization and maintains cancer cell growth under metabolic stress**. Cancer Cell *27*, 57-71, doi: 10.1016/j.ccell.2014.12.002.

Scully, O. J., Bay, B. H., Yip, G. and Yu, Y. (2012). **Breast cancer metastasis**. Cancer Genomics Proteomics *9*, 311-320.

Seidlitz, E. P., Sharma, M. K., Saikali, Z., Ghert, M. and Singh, G. (2009). **Cancer cell lines release glutamate into the extracellular environment**. Clinical & Experimental Metastasis *26*, 781-787, doi: 10.1007/s10585-009-9277-4.

Shen, J., Yan, L., Liu, S., Ambrosone, C. B. and Zhao, H. (2013). **Plasma metabolomic profiles in breast cancer patients and healthy controls: by race and tumor receptor subtypes**. Transl Oncol *6*, 757-765.

Shu, X., Xiang, Y. B., Rothman, N., Yu, D. X., Li, H. L., Yang, G., Cai, H., Ma, X., Lan, Q., Gao, Y. T., Jia, W., Shu, X. O. and Zheng, W. (2018a). **Prospective study of blood metabolites associated with colorectal cancer risk**. International Journal of Cancer *143*, 527-534, doi: 10.1002/ijc.31341.

Shu, X., Zheng, W., Yu, D., Lan Li, H., Lan, Q., Yang, G., Cai, H., Ma, X., Rothman, N., Gao, Y. T., Jia, W., Xiang, Y. B. and Shu, X. O. (2018b). **Prospective metabolomics study identifies potential novel blood metabolites associated with pancreatic cancer risk**. Int J Cancer, doi: 10.1002/ijc.31574.

Siegel, R. L., Miller, K. D. and Jemal, A. (2018). **Cancer statistics, 2018**. CA Cancer J Clin *68*, 7-30, doi: 10.3322/caac.21442.

Siena, S., Peeters, M., Van Cutsem, E., Humblet, Y., Conte, P., Bajetta, E., Comandini, D., Bodoky, G., Van Hazel, G., Salek, T., Wolf, M., Devercelli, G., Woolley, M. and Amado, R. G. (2007). **Association of progression-free survival with patient-reported outcomes and survival: results from a randomised phase 3 trial of panitumumab**. Br J Cancer *97*, 1469-1474, doi: 10.1038/sj.bjc.6604053.

Sopik, V. and Narod, S. A. (2018). **The relationship between tumour size, nodal status and distant metastases: on the origins of breast cancer**. Breast Cancer Research and Treatment *170*, 647-656, doi: 10.1007/s10549-018-4796-9.

St John-Williams, L., Blach, C., Toledo, J. B., Rotroff, D. M., Kim, S., Klavins, K., Baillie, R., Han, X., Mahmoudiandehkordi, S., Jack, J., Massaro, T. J., Lucas, J. E., Louie, G., Motsinger-Reif, A. A., Risacher, S. L., Alzheimer's Disease Neuroimaging, I., Alzheimer's Disease Metabolomics, C., Saykin, A. J., Kastenmuller, G., Arnold, M., Koal, T., Moseley, M. A., Mangravite, L. M., Peters, M. A., Tenenbaum, J. D., Thompson, J. W. and Kaddurah-Daouk, R. (2017). **Targeted metabolomics and medication classification data from participants in the ADNI1 cohort**. Sci Data *4*, 170140, doi: 10.1038/sdata.2017.140.

Stebbing, J., Sharma, A., North, B., Athersuch, T. J., Zebrowski, A., Pchejetski, D., Coombes, R. C., Nicholson, J. K. and Keun, H. C. (2012). **A metabolic phenotyping approach to understanding relationships between metabolic syndrome and breast tumour responses to chemotherapy**. Annals of Oncology *23*, 860-U862, doi: 10.1093/annonc/mdr347.

Suh, K. S., Park, S. W., Castro, A., Patel, H., Blake, P., Liang, M. and Goy, A. (2010). **Ovarian cancer biomarkers for molecular biosensors and translational medicine**. Expert Rev Mol Diagn *10*, 1069-1083, doi: 10.1586/erm.10.87.

Suhre, K. and Gieger, C. (2012). **Genetic variation in metabolic phenotypes: study designs and applications**. Nat Rev Genet *13*, 759-769, doi: 10.1038/nrg3314.

Tabar, L., Fagerberg, G., Chen, H. H., Duffy, S. W., Smart, C. R., Gad, A. and Smith, R. A. (1995). **Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial**. Cancer *75*, 2507-2517.

Taherian-Fard, A., Srihari, S. and Ragan, M. A. (2015). **Breast cancer classification: linking molecular mechanisms to disease prognosis**. Brief Bioinform *16*, 461-474, doi: 10.1093/bib/bbu020.

Talmadge, J. E. and Fidler, I. J. (2010). **AACR Centennial Series: The Biology of Cancer Metastasis: Historical Perspective**. Cancer Research *70*, 5649-5669, doi: 10.1158/0008-5472.Can-10-1040.

Tavassoéli, F. A., Devilee, P. (2003). **Pathology and Genetics Tumours of the Breast and Female Genital Organs**, Vol 4, 3. edn, International Agency for Research on Cancer.

Team, R. C. (2015). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Tenori, L., Oakman, C., Claudino, W. M., Bernini, P., Cappadona, S., Nepi, S., Biganzoli, L., Arbushites, M. C., Luchinat, C., Bertini, I. and Di Leo, A. (2012). **Exploration of serum metabolomic profiles and outcomes in women with metastatic breast cancer: a pilot study**. Mol Oncol *6*, 437-444, doi: 10.1016/j.molonc.2012.05.003.

Tenori, L., Oakman, C., Morris, P. G., Gralka, E., Turner, N., Cappadona, S., Fornier, M., Hudis, C., Norton, L., Luchinat, C. and Di Leo, A. (2015). **Serum metabolomic profiles evaluated after surgery may identify patients with oestrogen receptor negative early breast cancer at increased risk of disease recurrence. Results from a retrospective study**. Mol Oncol *9*, 128-139, doi: 10.1016/j.molonc.2014.07.012.

Thiery, J. P. (2002). **Epithelial-mesenchymal transitions in tumour progression**. Nat Rev Cancer *2*, 442-454, doi: 10.1038/nrc822.

Thiery, J. P., Acloque, H., Huang, R. Y. and Nieto, M. A. (2009). **Epithelial-mesenchymal transitions in development and disease**. Cell *139*, 871-890, doi: 10.1016/j.cell.2009.11.007.

Thompson, D. and Easton, D. (2004). **The genetic epidemiology of breast cancer genes**. J Mammary Gland Biol Neoplasia *9*, 221-236, doi: 10.1023/B:JOMG.0000048770.90334.3b.

Tibshirani, R. (1996). **Regression shrinkage and selection via the Lasso**. Journal of the Royal Statistical Society Series B-Methodological *58*, 267-288.

Tol, J., Koopman, M., Miller, M. C., Tibbe, A., Cats, A., Creemers, G. J. M., Vos, A. H., Nagtegaal, I. D., Terstappen, L. W. M. M. and Punt, C. J. A. (2010). **Circulating tumour cells early predict progression-free and overall survival in advanced colorectal cancer patients treated with chemotherapy and targeted agents**. Annals of Oncology *21*, 1006-1012, doi: 10.1093/annonc/mdp463.

Tonjes, M., Barbus, S., Park, Y. J., Wang, W., Schlotter, M., Lindroth, A. M., Pleier, S. V., Bai, A. H. C., Karra, D., Piro, R. M., Felsberg, J., Addington, A., Lemke, D., Weibrecht, I., Hovestadt, V., Rolli, C. G., Campos, B., Turcan, S., Sturm, D., Witt, H., Chan, T. A., Herold-Mende, C., Kemkemer, R., Konig, R., Schmidt, K., Hull, W. E., Pfister, S. M., Jugold, M., Hutson, S. M., Plass, C., Okun, J. G., Reifenberger, G., Lichter, P. and Radlwimmer, B. (2013). **BCAT1 promotes cell proliferation through amino acid catabolism in gliomas carrying wild-type IDH1**. Nature Medicine *19*, 901-908, doi: 10.1038/nm.3217.

Torre, L. A., Islami, F., Siegel, R. L., Ward, E. M. and Jemal, A. (2017). **Global Cancer in Women: Burden and Trends**. Cancer Epidemiol Biomarkers Prev *26*, 444-457, doi: 10.1158/1055-9965.EPI-16-0858.

Turkoglu, O., Zeb, A., Graham, S., Szyperski, T., Szender, J. B., Odunsi, K. and Bahado-Singh, R. (2016). **Metabolomics of biomarker discovery in ovarian cancer: a systematic review of the current literature**. Metabolomics *12*, doi: 10.1007/s11306-016-0990-0.

Turnbull, C. and Rahman, N. (2008). **Genetic predisposition to breast cancer: past, present, and future**. Annu Rev Genomics Hum Genet *9*, 321-345, doi: 10.1146/annurev.genom.9.081307.164339.

Uehara, M., Kinoshita, T., Hojo, T., Akashi-Tanaka, S., Iwamoto, E. and Fukutomi, T. (2008). **Long-term prognostic study of carcinoembryonic antigen (CEA) and carbohydrate antigen 15-3 (CA 15-3) in breast cancer**. Int J Clin Oncol *13*, 447-451, doi: 10.1007/s10147-008-0773-3.

Ueland, F. R. (2017). **A Perspective on Ovarian Cancer Biomarkers: Past, Present and Yet-To-Come**. Diagnostics (Basel) *7*, doi: 10.3390/diagnostics7010014.

Ueland, F. R., Desimone, C. P., Seamon, L. G., Miller, R. A., Goodrich, S., Podzielinski, I., Sokoll, L., Smith, A., van Nagell, J. R., Jr. and Zhang, Z. (2011). **Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors**. Obstet Gynecol *117*, 1289-1297, doi: 10.1097/AOG.0b013e31821b5118.

van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R. and Friend, S. H. (2002). **Gene expression profiling predicts clinical outcome of breast cancer**. Nature *415*, 530-536, doi: 10.1038/415530a.

Vander Heiden, M. G., Cantley, L. C. and Thompson, C. B. (2009). **Understanding the Warburg effect: the metabolic requirements of cell proliferation**. Science *324*, 1029-1033, doi: 10.1126/science.1160809.

Vazquez, A., Kamphorst, J. J., Markert, E., Schug, Z. T., Tardito, S. and Gottlieb, E. (2016). **Cancer metabolism at a glance**. Journal of Cell Science *129*, 3367-3373, doi: 10.1242/jcs.181016.

Veridex, L., Johnson & Johnson (2010). **CellSearch circulating tumour cell kit**.

Verma, R., Bowen, R. L., Slater, S. E., Mihaimeed, F. and Jones, J. L. (2012). **Pathological and epidemiological factors associated with advanced stage at diagnosis of breast cancer**. British Medical Bulletin *103*, 129-145, doi: 10.1093/bmb/lds018.

Viale, G. (2012). **The current state of breast cancer classification**. Ann Oncol *23 Suppl 10*, x207-210, doi: 10.1093/annonc/mds326.

Vinayavekhin, N., Homan, E. A. and Saghatelian, A. (2010). **Exploring Disease through Metabolomics**. Acs Chemical Biology *5*, 91-103, doi: 10.1021/cb900271r.

Wang, L., Liu, X. and Yang, Q. (2018a). **Application of Metabolomics in Cancer Research: As a Powerful Tool to Screen Biomarker for Diagnosis, Monitoring and Prognosis of Cancer**. Biomarkers Journal *01*, doi: 10.21767/2472-1646.100050.

Wang, M., Zhao, J., Zhang, L., Wei, F., Lian, Y., Wu, Y., Gong, Z., Zhang, S., Zhou, J., Cao, K., Li, X., Xiong, W., Li, G., Zeng, Z. and Guo, C. (2017). **Role of tumor microenvironment in tumorigenesis**. J Cancer *8*, 761-773, doi: 10.7150/jca.17648.

Wang, Q., Sun, T., Cao, Y., Gao, P., Dong, J., Fang, Y., Fang, Z., Sun, X. and Zhu, Z. (2016a). **A dried blood spot mass spectrometry metabolomic approach for rapid breast cancer detection**. Onco Targets Ther *9*, 1389-1398, doi: 10.2147/OTT.S95862.

Wang, X., Chen, S. and Jia, W. (2016b). **Metabolomics in Cancer Biomarker Research**. Current Pharmacology Reports *2*, 293-298, doi: 10.1007/s40495-016-0074-x.

Wang, X., Zhao, X., Chou, J., Yu, J., Yang, T., Liu, L. and Zhang, F. (2018b). **Taurine, glutamic acid and ethylmalonic acid as important metabolites for detecting human breast cancer based on the targeted metabolomics**. Cancer Biomark, doi: 10.3233/CBM-181500.

Warburg, O. (1956). **On the origin of cancer cells**. Science *123*, 309-314.

Wei, S., Liu, L., Zhang, J., Bowers, J., Gowda, G. A., Seeger, H., Fehm, T., Neubauer, H. J., Vogel, U., Clare, S. E. and Raftery, D. (2013). **Metabolomics approach for predicting response to neoadjuvant chemotherapy for breast cancer**. Mol Oncol *7*, 297-307, doi: 10.1016/j.molonc.2012.10.003.

Weigel, M. T. and Dowsett, M. (2010). **Current and emerging biomarkers in breast cancer: prognosis and prediction**. Endocr Relat Cancer *17*, R245-262, doi: 10.1677/ERC-10-0136.

Weigelt, B., Peterse, J. L. and van't Veer, L. J. (2005). **Breast cancer metastasis: Markers and models**. Nature Reviews Cancer *5*, 591-602, doi: 10.1038/nrc1670.

Weinberg, R. A. (2014). **The biology of cancer (2nd Edition)**, 2. edn, Garland Science, Taylor & Francis Group, LLC

Wikipedia **Metabolome**. URL: https://en.wikipedia.org/wiki/Metabolome [as of 26.07.2019].

Wils-Plotz, E. L., Jenkins, M. C. and Dilger, R. N. (2013). **Modulation of the intestinal environment, innate immune response, and barrier function by dietary threonine and purified fiber during a coccidiosis challenge in broiler chicks**. Poult Sci *92*, 735-745, doi: 10.3382/ps.2012-02755.

Xia, J. G., Sinelnikov, I. V., Han, B. and Wishart, D. S. (2015). **MetaboAnalyst 3.0-making metabolomics more meaningful**. Nucleic Acids Research *43*, W251-W257, doi: 10.1093/nar/gkv380.

Xiang, C. C., Jin, S. D., Zhang, J., Chen, M. J., Xia, Y. K., Shu, Y. Q. and Guo, R. H. (2018). **Cortisol, cortisone, and 4-methoxyphenylacetic acid as potential plasma biomarkers for early detection of non-small cell lung cancer**. International Journal of Biological Markers *33*, 314-320, doi: 10.1177/1724600818778754.

Xiao, S. and Zhou, L. (2017). **Gastric cancer: Metabolic and metabolomics perspectives (Review)**. Int J Oncol, doi: 10.3892/ijo.2017.4000.

Xie, H., Hou, Y., Cheng, J., Openkova, M. S., Xia, B., Wang, W., Li, A., Yang, K., Li, J., Xu, H., Yang, C., Ma, L., Li, Z., Fan, X., Li, K. and Lou, G. (2017). **Metabolic profiling and novel plasma biomarkers for predicting survival in epithelial ovarian cancer**. Oncotarget *8*, 32134-32146, doi: 10.18632/oncotarget.16739.

Yang, W., Mu, T., Jiang, J., Sun, Q., Hou, X., Sun, Y., Zhong, L., Wang, C. and Sun, C. (2018). **Identification of Potential Biomarkers and Metabolic Profiling of Serum in Ovarian Cancer Patients Using UPLC/Q-TOF MS**. Cell Physiol Biochem *51*, 1134-1148, doi: 10.1159/000495492.

Yin, P., Peter, A., Franken, H., Zhao, X., Neukamm, S. S., Rosenbaum, L., Lucio, M., Zell, A., Haring, H. U., Xu, G. and Lehmann, R. (2013). **Preanalytical aspects and sample quality assessment in metabolomics studies of human blood**. Clin Chem *59*, 833-845, doi: 10.1373/clinchem.2012.199257.

Yu, Z. H., Kastenmuller, G., He, Y., Belcredi, P., Moller, G., Prehn, C., Mendes, J., Wahl, S., Roemisch-Margl, W., Ceglarek, U., Polonikov, A., Dahmen, N., Prokisch, H., Xie, L., Li, Y. X., Wichmann, H. E., Peters, A., Kronenberg, F., Suhre, K., Adamski, J., Illig, T. and Wang-Sattler, R. (2011).

Differences between Human Plasma and Serum Metabolite Profiles. Plos One *6*, e21230, doi: 10.1371/journal.pone.0021230.

Yu, Z. H., Zhai, G. J., Singmann, P., He, Y., Xu, T., Prehn, C., Roemisch-Margl, W., Lattka, E., Gieger, C., Soranzo, N., Heinrich, J., Standl, M., Thiering, E., Mittelstrass, K., Wichmann, H. E., Peters, A., Suhre, K., Li, Y. X., Adamski, J., Spector, T. D., Illig, T. and Wang-Sattler, R. (2012). **Human serum metabolic profiles are age dependent**. Aging Cell *11*, 960-967, doi: 10.1111/j.1474-9726.2012.00865.x.

Yuan, B., Schafferer, S., Tang, Q., Scheffler, M., Nees, J., Heil, J., Schott, S., Golatta, M., Wallwiener, M., Sohn, C., Koal, T., Wolf, B., Schneeweiss, A. and Burwinkel, B. (2018). **A plasma metabolite panel as biomarkers for early primary breast cancer detection**. Int J Cancer, doi: 10.1002/ijc.31996.

Zaimenko, I., Jaeger, C., Brenner, H., Chang-Claude, J., Hoffmeister, M., Grotzinger, C., Detjen, K., Burock, S., Schmitt, C. A., Stein, U. and Lisec, J. (2019). **Non-invasive metastasis prognosis from plasma metabolites in stage II colorectal cancer patients: The DACHS study**. Int J Cancer *145*, 221-231, doi: 10.1002/ijc.32076.

Zhang, A., Sun, H., Wang, P., Han, Y. and Wang, X. (2012). **Modern analytical techniques in metabolomics analysis**. Analyst *137*, 293-300, doi: 10.1039/c1an15605e.

Zhang, C., Wang, K., Yang, L., Liu, R., Chu, Y., Qin, X., Yang, P. and Yu, H. (2018). **Lipid metabolism in inflammation-related diseases**. Analyst *143*, 4526-4536, doi: 10.1039/c8an01046c.

Zhang, H., Ge, T., Cui, X., Hou, Y., Ke, C., Yang, M., Yang, K., Wang, J., Guo, B., Zhang, F., Lou, G. and Li, K. (2015). **Prediction of advanced ovarian cancer recurrence by plasma metabolic profiling**. Mol Biosyst *11*, 516-521, doi: 10.1039/c4mb00407h.

Zhang, L., Volinia, S., Bonome, T., Calin, G. A., Greshock, J., Yang, N., Liu, C. G., Giannakakis, A., Alexiou, P., Hasegawa, K., Johnstone, C. N., Megraw, M. S., Adams, S., Lassus, H., Huang, J., Kaur, S., Liang, S., Sethupathy, P., Leminen, A., Simossis, V. A., Sandaltzopoulos, R., Naomoto, Y., Katsaros, D., Gimotty, P. A., DeMichele, A., Huang, Q., Butzow, R., Rustgi, A. K., Weber, B. L., Birrer, M. J., Hatzigeorgiou, A. G., Croce, C. M. and Coukos, G. (2008). **Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer**. Proc Natl Acad Sci U S A *105*, 7004-7009, doi: 10.1073/pnas.0801615105.

Zhang, Y., Liu, Y., Li, L., Wei, J., Xiong, S. and Zhao, Z. (2016). **High resolution mass spectrometry coupled with multivariate data analysis revealing plasma lipidomic alteration in ovarian cancer in Asian women**. Talanta *150*, 88-96, doi: 10.1016/j.talanta.2015.12.021.

Zhao, Z. W., Xiao, Y. J., Elson, P., Tan, H. Y., Plummer, S. J., Berk, M., Aung, P. P., Lavery, I. C., Achkar, J. P., Li, L., Casey, G. and Xu, Y. (2007). **Plasma lysophosphatidylcholine levels: Potential biomarkers for colorectal cancer**. Journal of Clinical Oncology *25*, 2696-2701, doi: 10.1200/Jco.2006.08.5571.

Zhe, X., Cher, M. L. and Bonfil, R. D. (2011). **Circulating tumor cells: finding the needle in the haystack**. Am J Cancer Res *1*, 740-751.

Zhou, J., Wang, Y. and Zhang, X. (2017). **Metabonomics studies on serum and urine of patients with breast cancer using 1H-NMR spectroscopy**. Oncotarget, doi: 10.18632/oncotarget.16210.

Zhou, M., Guan, W., Walker, L. D., Mezencev, R., Benigno, B. B., Gray, A., Fernandez, F. M. and McDonald, J. F. (2010). **Rapid mass spectrometric metabolic profiling of blood sera detects ovarian cancer with high accuracy**. Cancer Epidemiol Biomarkers Prev *19*, 2262-2271, doi: 10.1158/1055-9965.EPI-10-0126.

Publications

Original articles (from the thesis)

Yuan BW, Schafferer S, Tang QQ, Scheffler M, Nees J, Heil J, Schott S, Golatta M, Wallwiener M, Sohn C, Koal T, Wolf B, Schneeweiss A, Burwinkel B. *A plasma metabolite panel as biomarkers for early primary breast cancer detection*. International Journal of Cancer 2019;144: 2833-42.

Yuan BW, Schafferer S, Deutsch T, Tang QQ, Scheffler M, Nees J, Koal T, Wolf B, Wallwiener M, Schneeweiss A, Burwinkel B. *Panels of Plasma Metabolites as Diagnostic and Prognostic Markers in Metastatic Breast Cancer*. (Under review)

Yuan BW, Schafferer S, Scheffler M, Nees J, Koal T, Wolf B, Wallwiener M, Schneeweiss A, Burwinkel B. *A comprehensive analysis of metabolomic and transcriptomic data reveals metabolic pathway alteration in ovarian cancer*. (Ready for submission)

Yuan BW, Burwinkel B. *Blood metabolites as biomarkers for breast cancer and ovarian cancer: a systematic review*. (Manuscript in preparation)

The published paper is based on the results of the dissertation chapters 4.1. The significance of metabolites as primary breast cancer early detective markers (chapter 5.1) was discussed based on this publication. My own contribution to the publication extends to the analysis and evaluation of data, writing, and revising of the whole manuscript draft.

The manuscript under review is based on the results of the dissertation chapter 4.2 and 4.3. The discussions of chapter 5.2 and 5.3 are based on this manuscript. My own contribution to the publication extends to the analysis and evaluation of data, writing, and revising of the whole manuscript draft.

Patents

The patent entitled "*Metabolite-based breast cancer detection and diagnosis*" filed with the European Patent and Trademark Office, patent number: PCT/EP2019/057688.

The patent entitled "*Metabolites used for ovarian cancer detection and diagnosis*" filed with the European Patent and Trademark Office.

Conference Abstracts

Yuan BW, Schafferer S, Tang QQ, Scheffler M, Nees J, Heil J, Schott S, Golatta M, Wallwiener M, Sohn C, Koal T, Wolf B, et al. *A plasma metabolite panel as biomarkers for early primary breast cancer detection*. 13th International Ph.D. Cancer Conference; 12-14.06.2019, Amsterdam, The Netherlands. (Gave talk).

Yuan BW, Schafferer S, Tang QQ, Scheffler M, Nees J, Heil J, Schott S, Golatta M, Wallwiener M, Sohn C, Koal T, Wolf B, et al. *A plasma metabolite panel as biomarkers for early primary breast cancer detection*. DKFZ Poster presentation. 16.11.2018. Heidelberg.

Other publication

YUAN Bao-Wen, WANG Xiu-Jie. *Chromatin higher-order structure: an important form of genome regulation*. Chinese Bulletin of Life Sciences. 2015, 27(3): 336-343

Appendix

Detailed Clinical Features of Samples

Table A1: Clinicopathological characteristics of the PBC patients in discovery and validation cohorts. [a]Chi-squared test, [b]Wilcoxon rank-sum test, [c]tests with $p < 0.05$.

Clinical Characteristics		**Discovery Cohort ($n = 80$)**	**Validation Cohort ($n = 109$)**	***p*-value**
Age[b]	Mean	55.6	50	
	Median	53.0	46.2	0.001[c]
	Range	30.6 ~ 84.4	26.8 ~ 77.4	
Menopause[a]	Premenopause	29 (36.3%)	61 (56.0%)	
	Perimenopause	8 (10.0%)	7 (6.4%)	
	Postmenopause	42 (52.5%)	39 (35.8%)	0.055
	NA	1 (1.3%)	2 (1.8%)	
Grading[a]	G1	6 (7.5%)	12 (11.0%)	
	G2	38 (47.5%)	53 (48.6%)	
	G3	36 (45.0%)	43 (39.4%)	0.641
	NA	0	1 (0.9%)	
Tumor size[a]	Tis	2 (2.5%)	1 (0.9%)	
	T1	29 (36.3%)	45 (41.3%)	
	T2	44 (55.0%)	55 (50.5%)	0.184
	T3	5 (6.3%)	3 (2.8%)	
	T4	0	5 (4.6%)	
Lymph node spread[a]	N0	57 (71.3%)	72 (66.1%)	
	N1	22 (27.5%)	35 (32.1%)	0.738
	N2	1 (1.3%)	2 (1.8%)	

AJCC stage[a]	0	2 (2.5%)	1 (0.9%)	0.159
	I	23 (30.0%)	37 (33.9%)	
	II	48 (60.0%)	62 (56.9%)	
	III	3 (3.8%)	9 (8.3%)	
	IV	3 (3.8%)	0	
	NA	1 (1.25%)	0	
ER status[a]	Positive	46 (57.5%)	80 (73.4%)	0.047[c]
	Negative	33 (41.3%)	29 (26.6%)	
	NA	1 (1.3%)	0	
PR status[a]	Positive	43 (53.8%)	74 (67.9%)	0.087
	Negative	36 (45.0%)	35 (32.1%)	
	NA	1 (1.3%)	0	
HER2 status[a]	Positive	16 (20.0%)	30 (27.5%)	0.262
	Negative	63 (78.8%)	79 (72.5%)	
	NA	1 (1.3%)	0	
Triple negative[a]	Yes	20 (25%)	25 (22.9%)	0.470
	No	59 (73.8%)	84 (77.1%)	
	NA	1 (1.3%)	0	

Table A2: Clinicopathological characteristics of the MBC patients in discovery and validation cohorts. [a]Chi-squared test, [b]Wilcoxon rank-sum test, [c]tests with $p < 0.05$.

Clinical characteristics		Discovery Cohort (*n*=100)			Validation Cohort (*n*=76)		
		CTC-positive (*n*=44)	CTC-negative (*n*=56)	*p*-value	CTC-positive (*n*=21)	CTC-negative (*n*=55)	*p*-value
Age[b]	Mean	57.25	60.63	0.17	59.86	58.96	0.74
	Median	56	62		57	59	
	Range	33 ~ 81	31 ~ 89		42 ~ 77	36 ~ 78	
Bone metastasis[a]	Yes	34	33	0.09	16	32	0.23
	No	10	23		5	23	
Lung metastasis[a]	Yes	13	29	0.04[c]	12	31	1
	No	31	27		9	24	
Liver metastasis[a]	Yes	22	20	0.22	10	18	0.35
	No	22	36		11	37	
Brain metastasis[a]	Yes	4	7	0.83	0	5	0.36
	No	40	49		21	50	
Number of sites of metastasis[a]	1-2	25	33	0.97	10	32	0.57
	3-5	18	22		11	23	
	> 5	1	1		0	0	
Disseminated metastasis[a]	Yes	34	22	0.0003[c]	16	43	0.78
	No	10	34		5	11	
	NA	0	0		0	1	
Metastatic ER status[a]	Positive	17	26	0.64	10	24	0.6
	Negative	6	5		1	7	
	NA	21	25		10	24	
Metastatic PR status[a]	Positive	14	22	0.71	8	17	0.57
	Negative	9	9		3	14	
	NA	21	25		10	24	
Metastatic HER2 status[a]	Positive	3	8	0.49	1	8	0.5
	Negative	21	25		10	23	
	NA	20	23		10	24	
PFS status[a]	Progression	27	34	1	18	48	1
	No Progression	17	22		3	7	
PFS (Month)[b]	Mean	14.2	6.4	5.45E-04[c]	8.2	4.5	0.24
OS status[a]	Dead	35	25	8.66E-04[c]	19	24	6.14E-04[c]
	Alive	9	31		2	31	
OS (Month)[b]	Median	23.3	11.3	2.69E-07[c]	21.5	11.4	1.3E-04[c]

Prognosis status [a]	Good Bad Middle	8 26 10	23 17 16	0.048 [c]	2 17 2	13 34 8	0.39
Histology [a]	IDC ILC Others NA	19 8 5 12	31 4 0 21	0.02[c]	13 1 1 6	31 7 4 13	0.73
Grading [a]	G1 G2 G3 NA	1 26 14 3	4 21 23 8	0.14	0 9 7 5	2 25 21 7	0.56
Tumor size [a]	Tis T1 T2 T3 T4 NA	1 12 21 6 4 0	0 25 22 4 4 1	0.34	0 7 9 0 4 1	0 10 29 5 7 4	0.36
Lymph node spread [a]	N0 N1 N2 N3 NA	20 11 3 8 2	21 22 5 5 3	0.45	6 7 3 4 1	21 16 8 6 4	0.84
Distant metastasis [a]	M0 M1 NA	19 8 17	29 9 18	0.69	11 6 4	29 15 11	1
Primary ER status [a]	Positive Negative NA	31 13 0	45 9 2	0.14	18 2 1	38 15 2	0.25
Primary PR status [a]	Positive Negative NA	25 19 0	42 11 3	0.02[c]	14 5 2	33 19 3	0.59
Primary HER2 status [a]	Positive Negative NA	4 37 3	9 40 7	0.33	0 16 5	11 38 6	0.05
DFS (Month) [b]	Median	28	46	0.06	51	45	0.66
DDFS (Month) [b]	Median	29	70.5	0.006[c]	79	50	0.83

Table A3: Clinicopathological characteristics of the ovarian cancer patients in discovery and validation cohorts. [a]Chi-squared test, [b]Wilcoxon rank-sum test, [c]tests with $p < 0.05$.

Clinical Characteristics		Discovery Cohort (n=34)	Validation Cohort (n = 35)	*p*-value
Age[b]	Mean	61.2	61.6	0.889
	Median	62.6	64.5	
	Range	37.5 ~ 77	25.5 ~ 83.6	
FIGO staging[a]	I	2 (5.9%)	5 (14.3%)	0.006[c]
	II	1 (2.9%)	5 (14.3%)	
	III	29 (85.3%)	15 (42.9%)	
	IV	2 (5.9%)	6 (17.1%)	
	NA	0	4 (11.4%)	
Grading[a]	G1	1 (2.9%)	3 (8.6%)	0.111
	G2	7 (20.6%)	1 (2.9%)	
	G3	21 (61.8%)	24 (68.6%)	
	NA	5 (14.7%)	7 (20%)	
Diabetes[a]	Yes	5 (14.7%)	1 (2.9%)	0.214
	No	26 (76.5%)	30 (85.7%)	
	NA	3 (8.8%)	4 (11.4%)	
BMI[b]	Mean	25.7	26	0.842
	Median	24.8	25	
	Range	19.2 ~ 42	19.5 ~ 42.5	
	NA	9 (26.5%)	10 (28.6%)	
Menopause[a]	Premenopause	3 (8.8%)	6 (17.1%)	0.382
	Perimenopause	1 (2.9%)	0	
	Postmenopause	24 (70.6%)	26 (74.3%)	
	NA	6 (17.6%)	3 (8.6%)	

Metabolites List

Table A4: List of metabolites measured with Biocrates Absolute*IDQ*® p180 kit.

Category	N	Metabolites
Amino Acids	21	Ala, Arg, Asn, Asp, Cit, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Orn, Phe, Pro, Ser, Thr, Trp, Tyr, Val
Biogenic Amines	21	alpha-AAA, ADMA, Ac-Orn, carnosine, creatinine, DOPA, dopamine, histamine, kynurenine, Met-SO, Nitro-Tyr, *cis*-OH-Pro, *trans*-OH-Pro, PEA, putrescine, SDMA, sarcosine, serotonin, spermidine, spermine, taurine
Carnitine	1	C0
Acylcarnitines (C*x:y*)	26	C2, C3, C3:1, C4, C4:1, C5, C5:1, C6 (or C4:1-DC), C6:1, C8, C8:1, C9, C10, C10:1, C10:2, C12, C12:1,C14, C14:1, C14:2, C16, C16:1, C16:2, C18, C18:1,C18:2
Hydroxy- and Dicarboxy-Acylcarnitines (C*x:y*-DC/OH)	13	C3-OH, C4-OH (or C3-DC), C5-DC (or C6-OH), C5-OH (or C3-DC-M), C5:1-DC, C5-M-DC, C7-DC, C12-DC, C14:1-OH, C14:2-OH, C16:1-OH, C16:2-OH, C16-OH, C18:1-OH
Sum of Hexoses (H1)	1	H1
Sphingomyelins (SM C*x:y*)	10	SM C16:0/ C16:1/ C18:0/C18:1/ C20:2/ C22:3/ C24:0/ C24:1/ C26:0/C26:1
Hydroxy-Sphingomyelins (SM (OH) C*x:y*)	5	SM (OH) C14:1/ C16:1/ C22:1/ C22:2/ C24:1
Diacyl-Phosphatidylcholines (PC aa C*x:y*)	38	PC aa C24:0/C26:0/C28:1/C30:0/C30:2/C32:0/C32:1/C32:2/C32:3/C34:1/C34:2/C34:3/C34:4/C36:0/C36:1/C36:2/C36:3/C36:4/C36:5/C36:6/C38:0/C38:1/C38:3/C38:4/C38:5/C38:6/C40:1/C40:2/C40:3/C40:4/C40:5/C40:6/C42:0/C42:1/C42:2/C42:4/C42:5/C42:6
Acyl-alkyl-Phosphatidylcholines (PC ae C*x:y*)	38	PC ae C30:0/C30:1/C30:2/C32:1/C32:2/C34:0/C34:1/C34:2/C34:3/C36:0/C36:1/C36:2/C36:3/C36:4/C36:5/C38:0/C38:1/C38:2/C38:3/C38:4/C38:5/C38:6/C40:1/C40:2/C40:3/C40:4/C40:5/C40:6/C42:0/C42:1/C42:2/C42:3/C42:4/C42:5/C44:3/C44:4/C44:5/C44:6
Lyso-phosphatidylcholines (lysoPC a C*x:y*)	14	lysoPC a C14:0/C16:0/C16:1/C17:0/C18:0/C18:1/C18:2/C20:3/C20:4/C24:0/C26:0/C26:1/C28:0/C28:1

Abbreviations:
a, acyl; aa, acyl-acyl; ae, acyl-alkyl;

DC, decarboxyl groups; M, methyl; OH, hydroxyl groups;
PC, phosphatidylcholine; SM, sphingomyelin;
x:y, x is the number of carbons and y is the number of double bonds in the fatty acid side chain

Table A5: List of metabolites measured with Biocrates Absolute*IDQ*® p400 HR kit.

Category	**N**	**Metabolites**
Amino Acids	21	Ala, Arg, Asn, Asp, Cit, Gln, Glu, Gly, His, Ile, xLeu, Lys, Met, Orn, Phe, Pro, Ser, Thr, Trp, Tyr, Val
Biogenic Amines	21	alpha-AAA, ADMA, Ac-Orn, carnosine, creatinine, DOPA, dopamine, histamine, kynurenine, Met-SO, Nitro-Tyr, *cis*-OH-Pro, *trans*-OH-Pro, PEA, putrescine, SDMA, sarcosine, serotonin, spermidine, spermine, taurine
Sum of Hexoses (H1)	1	H1
Diacyl-Phosphatidylcholines PC(*x:y*)	108	PC (24:0)/(25:0)/(26:0)/(27:0)/(27:1)/(28:1)/(29:0)/(29:1)/(29:2)/(30:0)/(30:1)/(30:2)/(30:3)/(31:0)/(31:1)/(31:2)/(31:3)/(32:0)/(32:1)/(32:2)/(32:3)/(32:4)/(32:5)/(32:6)/(33:0)/(33:1)/(33:2)/(33:3)/(33:4)/(33:5)/(34:0)/(34:1)/(34:2)/(34:3)/(34:4)/(34:5)/(35:0)/(35:1)/(35:2)/(35:3)/(35:4)/(35:5)/(36:0)/(36:1)/(36:2)/(36:3)/(36:4)/(36:5)/(36:6)/(37:0)/(37:1)/(37:2)/(37:3)/(37:4)/(37:5)/(37:6)/(37:7)/(38:0)/(38:1)/(38:2)/(38:3)/(38:4)/(38:5)/(38:6)/(38:7)/(39:0)/(39:1)/(39:2)/(39:3)/(39:4)/(39:5)/(39:6)/(39:7)/(40:1)/(40:2)/(40:3)/(40:4)/(40:5)/(40:6)/(40:7)/(40:8)/(40:9)/(41:1)/(41:2)/(41:3)/(41:4)/(41:5)/(41:8)/(42:0)/(42:1)/(42:2)/(42:3)/(42:4)/(42:5)/(42:6)/(42:7)/(42:10)/(43:2)/(43:6)/(44:1)/(44:3)/(44:5)/(44:6)/(44:7)/(44:10)/(44:12)/(46:1)/(46:2)
Acyl-alkyl-Phosphatidylcholines PC-O-(*x:y*)	64	PC-O-(26:0)/(26:1)/(28:0)/(28:1)/(29:0)/(30:0)/(30:1)/(30:2)/(31:0)/(31:1)/(31:3)/(32:0)/(32:1)/(32:2)/(32:3)/(33:0)/(33:1)/(33:2)/(33:3)/(33:4)/33:6)/(34:0)/(34:1)/(34:2)/(34:3)/(34:4)/(35:3)/(35:4)/(36:0)/(36:1)/(36:2)/(36:3)/(36:4)/(36:5)/(36:6)/(37:6)/(37:7)/(38:0)/(38:1)/(38:2)/(38:3)/(38:4)/(38:5)/(38:6)/(40:0)/(40:1)/(40:2)/(40:3)/(40:4)/(40:5)/(40:6)/(40:7)/(40:8)/(42:0)/(42:1)/(42:2)/(42:3)/(42:4)/(42:5)/(42:6)/(44:3)/(44:4)/(44:5)/(44:6)
Lyso-phosphatidylcholines LPC(*x:y*)	24	LPC (12:0)/(14:0)/(15:0)/(16:0)/(16:1)/(17:0)/(17:1)/(18:0)/(18:1)/(18:2)/(20:0)/(20:1)/(20:2)/(20:3)/(20:4)/(22:5)/(22:6)/(24:0)/(24:1)/-O(16:1)/ -O(17:1)/ -O(18:0)/ -O(18:1)/ -O(18:2)

Sphingomyelins SM(*x:y*)	31	SM (30:1)/(31:0)/(31:1)/(32:1)/(32:2)/(33:1)/(33:2)/(34:1)/(34:2)/ (35:1)/(36:0)/(36:1)/(36:2)/(37:1)/(38:1)/(38:2)/(38:3)/(39:1)/ (39:2)/(40:1)/(40:2)/(40:4)/(41:1)/(41:2)/(42:1)/(42:2)/(42:3)/ (43:1)/(43:2)/(44:1)/(44:2)
Ceramides Cer(*x:y*)	9	Cer (34:0)/(34:1)/(38:1)/(40:1)/(41:1)/(42:1)/(42:2)/(43:1)/(44:0)/
Acylcarnitines AC(*x:y*)	34	AC (0:0)/(2:0)/(3:0)/(3:1)/(4:0)/(4:1)/(5:0)/(5:1)/(6:0)/(6:1)/(7:0)/ (8:0)/(8:1)/(9:0)/(10:0)/(10:1)/(10:2)/(10:3)/(11:0)/(12:0)/(12: 1)/(13:0)/(14:0)/(14:1)/(14:2)/(15:0)/(16:0)/(16:1)/(16:2)/(17: 0)/(18:0)/(18:1)/(18:2)/(19:0)
Hydroxy- and Dicarboxy- Acylcarnitines AC(*x:y*-DC/OH)	21	AC (3:0-DC)/(3:0-OH)/(4:0-DC)/(4:0-OH)/(4:1-DC)/(5:0-DC)/ (5:0-OH)/ (5:1-DC)/ (6:0-DC)/(6:0-OH)/(7:0-DC)/(8:1-OH)/ (12:0-DC)/(14:0-OH)/(14:1-DC)/(14:1-OH)/(14:2- OH)/(16:0-OH)/(16:1-OH)/(16:2-OH)/(18:1-OH)
Diglycerides DG(*x:y*)	18	DG (32:1)/(32:2)/(34:1)/(34:3)/(36:2)/(36:3)/(36:4)/(38:0)/(38:5)/ (39:0)/(41:1)/(42:0)/(42:1)/(42:2)/(42:3)/-O(32:2)/-O(34:1)/ - O(36:4)
Triglycerides TG(*x:y*)	42	TG (44:1)/(44:2)/(44:4)/(46:2)/(48:1)/(48:2)/(48:3)/(49:1)/(49:2)/ (50:1)/(50:2)/(50:3)/(50:4)/(51:1)/(51:2)/(51:3)/(51:4)/(51:5)/ (52:2)/(52:3)/(52:4)/(52:5)/(52:6)/(52:7)/(53:3)/(53:4)/(53:5)/ (53:6)/ (54:2)/(54:3)/(54:4)/(54:5)/(54:6)/(54:7)/(55:6)/(55:7)/ (55:8)/(55:9)/(56:6)/(56:7)/ (56:8)/(56:9)
Cholesteryl Esters CE(*x:y*)	14	CE (16:0)/(16:1)/(17:0)/(17:1)/(17:2)/(18:1)/(18:2)/(18:3)/(19:2)/ (19:3)/(20:4)/(20:5)/(22:5)/(22:6)

Abbreviations:

-O-, acyl-alkyl;

PC, phosphatidylcholine; LPC: lyso- phosphatidylcholine; SM, sphingomyelin;

Cer: ceramide; AC: acylcarnitine; DG: diglyceride; TG: triglyceride; CE: cholesteryl ester;

DC, decarboxyl groups; OH, hydroxyl groups;

x:y, x is the number of carbons and y is the number of double bonds in the fatty acid side chain

R Packages

Table A6: R packages used in this thesis work. R package source: Comprehensive R Archive Network.

R3.5.1 was used in this project	
Boruta v6.0.0	caret v6.0.80
CoxBoost v1.4	dplyr v0.7.8
e1071 v1.7-0	ggcorrplot v0.1.2
ggplot2 v3.1.0	ggpubr v0.2
ggrepel v0.8.0	ggsignif v0.4.0
glmnet v2.0.16	gmodels v2.18.1
gridExtra v2.3	gtools v3.8.1
Hmisc v4.2.0	penalized v0.9.51
peperr v1.1.7.1	pROC v1.13.0
randomForest v4.6-14	ROCR v1.0.7
survival v2.43.1	

Supplementary Figures

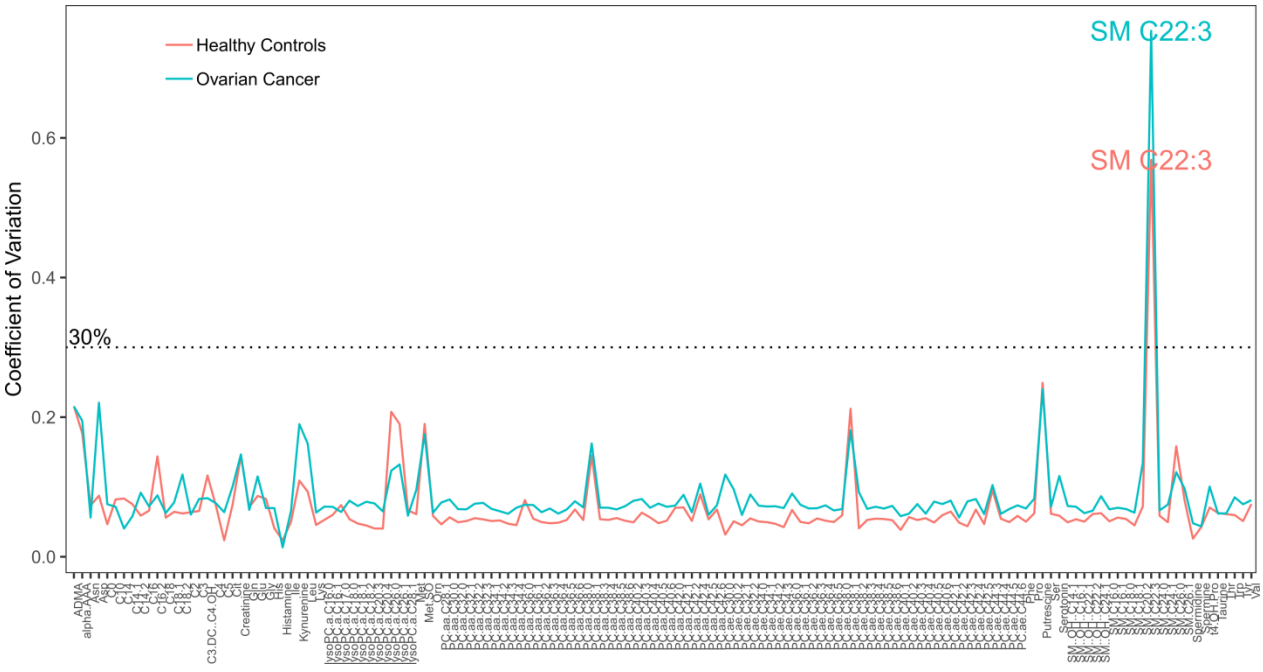


Figure A1: Coefficient of variation of triplicate samples from five healthy controls and five ovarian cancer patients. Metabolite with a coefficient of variation larger than 0.3 is marked (SM C22:3).

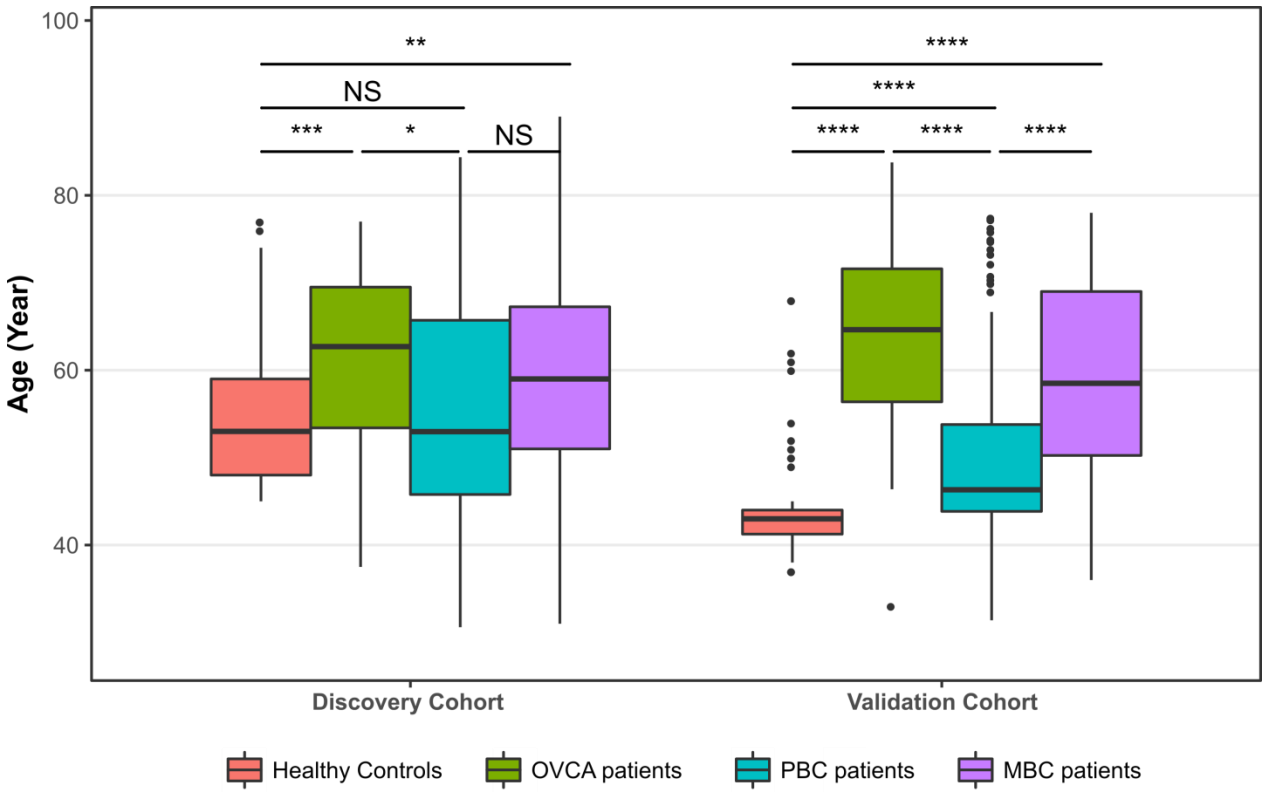


Figure A2: Age is unequally distributed among MBC patients, PBC patients, OVCA patients, and healthy controls, especially in the validation cohort.

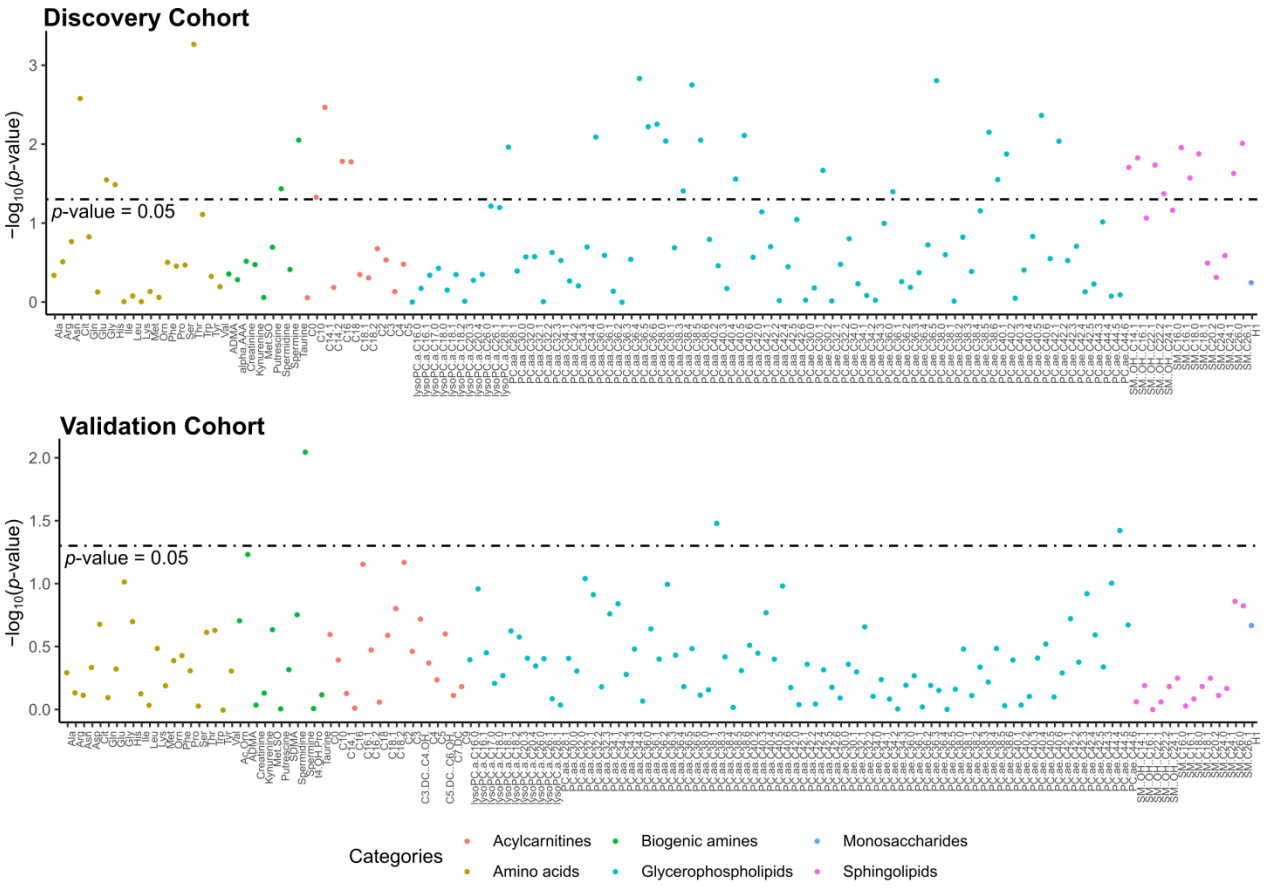


Figure A3: Associations between age (continuous) and metabolite concentrations plotted as $-\log_{10}$(*p*-values) for healthy control individuals in discovery and validation cohorts. The dashed line shows a statistical significance level ($p = 0.05$). The *p*-values were derived from partial spearman correlation adjusted for BMI, alcohol, sports, and smoking status.

Curriculum Vitae

Personal Details

Name:	Baowen Yuan
Gender:	Female
Date of birth:	21st May 1990
Nationality:	Chinese
Contact:	Im Neuenheimer Feld 133, 4-5-1 69120, Heidelberg, Germany Tel: +49 160 5964385 E-Mail: b.yuan@dkfz.de

Academic Profile

2016 - Present

Ph.D. Student at the Helmholtz Group Molecular Epidemiology, German Cancer Research Center (DKFZ)/Molecular Biology of Breast Cancer, Women's Clinic, University Heidelberg, Heidelberg, Germany.

Doctoral thesis: "*Identification of Plasma Metabolites Associated with Breast and Ovarian Cancer and Breast Cancer Prognosis*"

Supervisor: Prof. Dr. Barbara Burwinkel

2013 - 2016

Master of Science, Bioinformatics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Master Thesis: "*A Study on Regulatory Principles of Human Genomes at the Chromosome Structure Level*"

Supervisor: Prof. Xiu-Jie Wang

2013 - 2016

Master of Science, Bioinformatics, Department of Health Technology, Technical University of Denmark, Lyngby, Denmark

Master Thesis: "*A Study on Regulatory Principles of Mammalian Genomes at the Chromosome Structure Level*" (Full marks for thesis defense: Excellent)

Supervisor: Prof. Ramneek Gupta

2009 - 2013

Bachelor of Science, Biotechnology, College of Biological Engineering, Henan University of Technology, Zhengzhou, Henan, China

Bachelor Thesis: "*Research on RT-PCR Technology in Identification of Potato Virus Y (PVY) Serotype*"

2006 - 2009

High School: Nanyang 5th high school, Nanyang, Henan, China.

Acknowledgments

At this point, I would like to thank all people who contributed to my Ph.D. project and who supported me during my time as a Ph.D. student.

First and foremost, I would like to express my most profound appreciation and sincere gratitude to Prof. Barbara Burwinkel for the opportunity to work on this project. The support and guidance she gave me and her confidence in me throughout my Ph.D. work have given me the confidence and skills. Her constant encouragement and the freedom she gave have helped me grow as an independent scientist. This thesis would not have competed without her help and assistance.

I want to thank my Thesis Advisory Committee (TAC) members, Prof. Stefan Wiemann, and Dr. Jürgen G. Okun, for their mentorship, constant support, insightful discussions, and valuable advice throughout my Ph.D. program. Besides, I would thank our collaborators from the clinic for providing the samples and collaborators from the Biocrates for measuring and pre-processing all the metabolites data used in this project.

I also extend my gratitude to the German Cancer Research Center (DKFZ) and especially Helmholtz International Graduate School for providing the valuable scientific training courses and career-related seminars, the opportunities to attend international conferences, and the strict and interactive research atmospheres. I specially thanks to the China Scholarship Council (CSC) for financial support during my Ph.D. period.

I am very grateful to my friends and colleagues from DKFZ for their collaboration, support, encouragement, and friendship. My thanks go to Dr. Andrey Turchinovich, Dr. Qiuqiong Tang, and Dr. Melanie Maierthaler for their help and introduction to my project and new life in Heidelberg. My thanks also go to people, Dr. Elena Bausch, Dr. Anna Cazzola, Dr. Jing Xu, Dr. Viani and Gianmatteo, who share the same office with me in DKFZ. They helped me a lot and created an enjoyable work atmosphere. I also thank my family for their patience, encouraging words, and their unstinting support.

EIDESSTATTLICHE VERSICHERUNG

1. Bei der eingereichten Dissertation zu dem Thema

……

handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.*

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Ort und Datum Unterschrift

* Wenn dies nicht zutrifft, machen Sie folgende Angaben:

3. Die Arbeit oder Teile davon habe ich wie folgt an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt:

Titel der Arbeit:

Hochschule und Jahr:

Art der Prüfungs- oder Qualifikationsleistung: