The Texas Medical Center Library
DigitalCommons@TMC

Dissertations and Theses (Open Access)

MD Anderson UTHealth Houston Graduate School

8-2014

Novel Serum Biomarkers For Lung Cancer Early Diagnosis And Clinical Outcome

Fanmao Zhang

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations

Part of the Epidemiology Commons, and the Oncology Commons

Recommended Citation
Zhang, Fanmao, "Novel Serum Biomarkers For Lung Cancer Early Diagnosis And Clinical Outcome" (2014). *Dissertations and Theses (Open Access)*. 477.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/477

This Dissertation (PhD) is brought to you for free and open access by the MD Anderson UTHealth Houston Graduate School at DigitalCommons@TMC. It has been accepted for inclusion in Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digcommons@library.tmc.edu.

The TMC LIBRARY
Health Sciences Resource Center

NOVEL SERUM BIOMARKERS FOR LUNG CANCER EARLY DIAGNOSIS AND CLINICAL OUTCOME

by

Fanmao Zhang, B.S.

APPROVED:

____
Xifeng Wu, M.D., Ph.D.,
Supervisory Professor

____
Paul Chiao, Ph.D.

____
Jian Gu, Ph.D.

____
Jonathan M. Kurie, M.D.

____
William Plunkett, Ph.D.

APPROVED:

____
Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

NOVEL SERUM BIOMARKERS FOR LUNG CANCER EARLY DIAGNOSIS AND CLINICAL OUTCOME

A

DISSERTATION

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
and
The University of Texas
MD Anderson Cancer Center
Graduate School of Biomedical Sciences

in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

by
Fanmao Zhang, B.S.
Houston, Texas

August 2014

Dedication

To all my family, my mentors and my friends,

who have stood beside me through all this time,

for their help, support and love,

I couldn't have done this without you.

Acknowledgement

It is my most sincere wish to express my deep gratitude to the many people who have helped and supported me throughout my PhD study. Without their encouragement, support and advice, the completion of this dissertation would not have been possible.

I would like to particularly thank my mentor, Dr. Xifeng Wu, for her guidance, expertise, kindness and continuous support for me over the course of my study in graduate school. She has been a role model of integrity and honesty to me. Her enthusiasm for scientific research has greatly influenced me and brought to me a great enthusiasm for success in furthering scientific advances. I was extremely fortunate to have her as my mentor and to learn from her both professionally and personally.

I would also like to thank my committee members, Drs. Paul Chiao, Jian Gu, Jonathan M. Kurie, William Plunkett and Jie Lin for their time and help. They have given me insightful suggestions and thoughtful critiques, shared with me their knowledge and expertise, and showed me their support and encouragement. My special thanks also go to the faculty, staff and my colleagues at the University of Texas Graduate School of Biomedical Science at Houston and at the MD Anderson Cancer Center.

I truly appreciate all my team members for their help and their support and care of me, especially to Drs. Jian Gu, Yuanqing Ye, Michelle A.T. Hildebrandt, David W. Chang and Jie Lin. They have been very generous sharing with me their

knowledge, expertise and experience, and I thank them for the many productive, thoughtful discussions.

Last and most importantly, I would like to thank all my friends and my family in China and in the US, to my parents, my brother and sister for their love and support for me all these years, and to my wife for bearing and staying with me, and loving me. I would not have made it without any of you.

Abstract

Novel serum biomarkers for lung cancer early diagnosis and clinical outcome

Fanmao Zhang, B.S.

Advisory Professor: Xifeng Wu, M.D., Ph.D.

The five-year survival rate for all stages of lung cancer combined is only 17%, which has changed little over the past 40 years. Despite the tremendous efforts made, serum biomarkers with clinical utility for lung cancer early detection and clinical outcome prediction are still lacking. Metabolic alterations have been recognized as an emerging hallmark of cancer. We aimed to investigate the metabolic changes associated with lung cancer and to identify novel clinically applicable serum biomarkers for lung cancer early diagnosis and clinical outcome.

Serum metabolites are potential biomarkers for lung cancer early detection. We first performed global metabolomic profiling followed by targeted validation of individual metabolites in a case-control design of 386 lung cancer cases and 193 matched controls. We then validated the most significant metabolite bilirubin as a risk marker for lung cancer incidence and mortality in a large prospective cohort comprised of 425,660 participants. In this cohort, the inverse association was only seen in male smokers. For every 0.1 mg/dL decrease of bilirubin, the risks for lung cancer incidence and mortality increased by 5% and 6%, respectively (both $P < 0.001$).

We next investigated pre-treatment laboratory tests indicative of a patient's overall metabolic status, as biomarkers for non-small cell lung cancer (NSCLC) clinical outcome. We assessed seven pre-treatment serum laboratory test levels

in 2,675 NSCLC patients, including 623 early stage and 2,052 advanced stage patients. Among 978 advanced stage NSCLC patients we studied who were treated with platinum-based chemotherapy, lower than normal levels of albumin, higher than normal levels of alkaline phosphatase and lactate dehydrogenase were all associated with worse 2-year overall survival, after adjusting for other variables. In addition, there was a cumulative effect among these three adverse laboratory test levels.

In conclusion, low serum bilirubin levels are associated with higher risks of lung cancer incidence and mortality in male smokers and may be used to identify higher risk smokers for lung cancer. In addition, pre-treatment laboratory test levels indicative of metabolic status could be utilized to enhance predictions of survival among advanced stage NSCLC patients treated with platinum-based chemotherapy. Taken together, our results suggested that metabolic alterations associated with lung cancer could serve as novel serum biomarkers with clinical significance for lung cancer early detection and clinical outcome.

Table of Contents

APPROVAL PAGE .... **i**

TITLE PAGE .... **ii**

DEDICATION .... **III**

ACKNOWLEDGEMENT .... **IV**

ABSTRACT .... **VI**

TABLE OF CONTENTS .... **VIII**

LIST OF ILLUSTRATIONS .... **XII**

LIST OF TABLES .... **XIV**

CHAPTER 1. INTRODUCTION **1**

1.1. Lung Cancer Epidemiology 2

1.1.1. Incidence and mortality *2*

1.1.2. Risk factors *3*

1.2. Lung Cancer Treatment and Survival 5

1.3. Lung Cancer Prognostic and Predictive Factors....10

1.4. Lung Cancer Screening....11

1.5. Cancer Biomarkers....12

1.5.1. Classification of biomarkers....12

1.5.2. Blood-based biomarkers....14

1.5.3. Tumor-derived biomarkers....14

1.6. Cancer Metabolism....15

1.7. Metabolomics....16

1.8. Hypothesis and Rationale....17

1.8.1. Hypothesis I: Global metabolomic profiling could identify serum metabolites that are differentially expressed during lung cancer development, which could consequently be used as novel biomarkers for the early detection of lung cancer....17

1.8.2. Hypothesis II: Pre-treatment clinical laboratory tests indicative of metabolic status are associated with survival in NSCLC patients, which could

serve as biomarkers for clinical outcome in NSCLC patients. 18

Chapter 2. Materials and Methods 20

2.1. Study Population and Data Collection 21

2.2. Global Metabolomic Profiling 26

2.3. Quantification of individual metabolites.. 27

2.4. Statistical Analyses 29

Chapter 3. Results and Discussion 32

3.1. Lung Cancer Metabolomic Profiling Study 33

3.1.1. Characteristics of the study population 33

3.1.2. Global metabolomic profiling of lung cancer 37

3.1.3. Target validation of individual metabolites.. 40

3.1.4. Validation of bilirubin as a lung cancer marker in a large cohort 42

3.1.5. The ability of bilirubin in identifying smokers with higher risk of lung cancer 52

3.2. Discussion 55

3.3. Clinical Laboratory Tests and Lung Cancer Clinical Outcome 61

3.3.1. Patient characteristics 61

3.3.2. Association of pre-treatment laboratory test levels with survival of advanced stage NSCLC patients..65

3.3.3. Cumulative effects of unfavorable laboratory test levels 69

3.3.4. Lab-test based risk score associated with survival of advanced NSCLC patients...73

3.4. Discussion 75

CHAPTER 4. CONCLUSIONS 80

CHAPTER 5. FUTURE DIRECTIONS 82

BIBLIOGRAPHY 85

APPENDIX A: PUBLICATIONS 96

APPENDIX B: SUPPLEMENTARY TABLE...98

VITA...116

List of Illustrations

Figure 1. Non-Small Cell Lung Cancer Treatment Patterns by Stage, 2008. 7

Figure 2. Five-Year Relative Survival Rates (A) and Stage Distribution (B) of Lung Cancer by Race and Stage at Diagnosis. 9

Figure 3. Study design of lung cancer metabolomic profiling and validation.......24

Figure 4. Distribution of serum total bilirubin levels among male (A) and female (B) participants in the MJ cohort. 36

Figure 5. Adjusted HRs of lung cancer incidence stratified by tertiles of serum total bilirubin level and smoking pack-years (A) or cigarettes per day (B) among male participants of the prospective cohort study. 49

Figure 6. Adjusted HRs of lung cancer mortality stratified by tertiles of serum total bilirubin level and smoking pack-years (A) or cigarettes per day (B) among male participants of the prospective cohort study. 50

Figure 7. Serum total bilirubin levels and lung cancer incidence rates (A) and mortality rates (B) in overall males of the prospective cohort study. 51

Figure 8. Serum total bilirubin levels and lung cancer incidence rates (A) and mortality rates (B) in male smokers of the prospective cohort study. 54

Figure 9. Kaplan-Meier 2-year survival curves for patients with advanced NSCLC treated with platinum-based chemotherapy grouped by individual lab test level: (A) Albumin, (B) ALP, (C) LDH. 67

Figure 10. Kaplan-Meier 5-year survival curves for patients with advanced NSCLC treated with platinum-based chemotherapy grouped by individual lab test level: (A) Albumin, (B) ALP, (C) LDH. 68

Figure 11. Kaplan-Meier 2-year survival curves for patients with advanced NSCLC treated with platinum-based chemotherapy grouped by cumulative effect of adverse lab test (A) and lab-test based risk score (B). 71

Figure 12. Kaplan-Meier 5-year survival curves for patients with advanced NSCLC treated with platinum-based chemotherapy grouped by cumulative effect of adverse lab test (A) and lab-test based risk score (B)....................72

List of Tables

Table 1. Selected host characteristics of three case-control populations............34

Table 2. Characteristics of the participants in the prospective cohort by gender and serum total bilirubin levels..35

Table 3. Metabolites with significantly differential expression between lung cancer cases and healthy controls in metabolomic profiling. 38

Table 4. Metabolites with significant trend of expression comparing normal controls, early stage and late stage cases in metabolomic profiling 39

Table 5. Levels of the two individual metabolites of interest measured with LC/MS-MS in the three case-control studies...41

Table 6. Relationship among smoking quantity, bilirubin levels and risk for lung cancer incidence by smoking status among male participants in the prospective cohort study 44

Table 7. Relationship among smoking quantity, bilirubin levels and lung cancer mortality by smoking status among male participants in the prospective

cohort study 45

Table 8. Relationship among smoking quantity, bilirubin levels and risk for lung cancer incidence by smoking status among female participants in the prospective cohort study 46

Table 9. Relationship among smoking quantity, bilirubin levels and lung cancer mortality by smoking status among female participants in the prospective cohort study 47

Table 10. Lung cancer incidence and mortality rates and adjusted HR per tertile of serum total bilirubin level among the male participants in the prospective cohort study by smoking status and smoking intensity 48

Table 11. Host characteristics of overall, early stage and late stage NSCLC patients 63

Table 12. Laboratory test levels associated with 2-year and 5-year survival in advanced NSCLC patients treated with platinum-based chemotherapy 66

Table 13. Cumulative effect of unfavorable laboratory test levels associated with

2-year and 5-year survival in advanced NSCLC patients treated with platinum-based chemotherapy..70

Table 14. Lab-test based risk score associated with 2-year and 5-year survival in advanced NSCLC patients treated with platinum-based chemotherapy......74

Chapter 1. Introduction

1.1. Lung Cancer Epidemiology

1.1.1. Incidence and mortality

Approximately 1.6 million lung cancer cases are newly diagnosed each year worldwide, accounting for about 13% of all cancer cases (1). Lung cancer is the most common cancer in males and the fourth most commonly diagnosed cancer in females worldwide (1). In the United States, approximately 224,210 cases of lung cancer are expected to be newly diagnosed in 2014 (2). It's the second most common cancer in both males and females, accounting for about 13% of all cancer cases diagnosed. In the US, the incident rates of lung cancer have been declining since the mid-1980s in males, and have just started declining in females since the mid-2000s. From 2006 to 2010, incidence rates decreased by 1.9% per year in males and by 1.2% per year in females (2). The median age at diagnosis is approximately 70 years (3). Lung cancer incidence rates are different among ethnic groups. In the US, incidence rates are highest among black males, which is 40% higher than in white males.

Worldwide, lung cancer accounts for nearly 1.4 million deaths each year – 18% of overall cancer-related mortality (3). It is the leading cause of cancer-related mortality in males, and the second leading cause of cancer-related mortality in females worldwide (1). In the US, approximately 159,260 patients are expected to die from lung cancer in 2014. It's the leading cause of cancer-related mortality in both males and females, accounting for about 27% of all

cancer-related mortality in the US, more than the mortality attributed to the next four most deadly cancers combined (breast, prostate, colon and pancreatic cancers) (2). Mortality rates started declining in the early 1990s in males, and mid-2000s in females. From 2006 to 2010, mortality rates decreased by 2.9% per year in males and by 1.4% per year in females. On the other hand, lung cancer incidence and mortality rates are still on the rise in many other countries (3).

Worldwide, the highest lung cancer incidence rates in males are observed in Central and Eastern Europe, Northern America, while the lowest rates are observed in sub-Saharan Africa. In females, the highest lung cancer incidence rates are observed in Northern America and Northern Europe, and the lowest rates are observed in Africa as well (1).

1.1.2. Risk factors

Tobacco smoking has been recognized as the most predominant risk factor for lung cancer since the 1950s. Tobacco smoking is estimated to account for 80% of the worldwide lung cancer burden among males and over 50% of the burden in females (1). However, less than 10% of ever-smokers develop lung cancer. Both smoking intensity (i.e. number of cigarettes smoked per day) and duration of smoking increase lung cancer risks. Other factors of smoking influencing lung cancer risk include age starting to smoke, time since quitting and types of tobacco products smoked. Lung cancer risks are at least ten-times higher among smokers compared to those who never smoked, in both males and females (4). Former smokers have a lower risk of lung cancer, although risk is still higher among

former smokers compared to never-smokers. The observed differences in lung cancer incidence rates across countries and between genders are primarily due to differences in the smoking epidemic (1). Besides cigarettes, cigar or pipe smoking also increases lung cancer risk.

Exposure to second-hand smoking is a major risk factor for those who do not smoke. It is estimated that 40% of non-smokers in the US are exposed to second-hand smoking, which increases lung cancer risk by up to 30% and contributes to around 3,000 lung cancer deaths each year.

Exposure to radon gas, which is commonly released from construction materials, is responsible for over 20,000 lung cancer cases each year in the United States, making it the second leading risk factor for lung cancer in North America and Europe (2). Other known risk factors include environmental or occupational exposure asbestos, air pollution, diesel exhaust, radiation, polycyclic aromatic hydrocarbons (PAHs) as well as certain metals (arsenic, chromium and cadmium). There is also interaction among these risk factors, e.g. asbestos exposure and cigarette smoking have been shown to jointly increase lung cancer risk (5).

Of note, although smoking is less prevalent in Chinese females (less than 4% adults) compared to those in certain European countries such as Germany and Italy (about 20% adults), lung cancer incidence rates are higher among Chinese females, presumably reflecting indoor air pollution from stoves fueled by coal and without ventilation and from cooking fumes in China (1). Other risk factors include a family or personal history of cancer as well as a medical history

of tuberculosis. Genetics also plays a contributing role in lung cancer carcinogenesis, particularly among patients with early age-onset (3).

1.2. Lung Cancer Treatment and Survival

Lung cancer can be broadly categorized into two major histopathologic groups: small-cell lung cancer (SCLC, 14% of cases) and non-small cell lung cancer (NSCLC, 84% of cases) for the purposes of treatment (6). NSCLCs are of epithelial cells origin, while SCLCs are of endocrine cells origin. NSCLC can be further classified into several subtypes based on different characteristics of tumor cells, which includes adenocarcinoma (~40%), squamous cell carcinoma (25% - 30%), large cell carcinoma (10% - 15%) as well as bronchioloalveolar carcinoma (3).

Treatment regimens include surgery, radiation therapy, chemotherapy and targeted therapy, depending on type and stage of disease, as well as molecular characteristics of the tumors (6). For SCLC, radiation therapy alone is the standard treatment for limited disease (LD), while radiation therapy in combination with chemotherapy is the standard treatment for extensive disease (ED) (6). A majority of patients experience at least temporary remission under this regimen, although they often experience recurrence later.

For NSCLC patients with localized diseases, surgery is usually the treatment of choice. Survival is improved when chemotherapy is administered after surgery for most of these patients. About 30% of early stage patients will develop recurrence or progress to metastatic disease. Therefore, besides overall

survival, prevention of recurrence and progression is another major concern for early stage NSCLC patients. For advanced stage NSCLC patients, chemotherapy, radiation therapy or a combination of both are used for treatment (6) (Figure 1). Currently, the standard treatment for advanced stage NSCLC patients is platinum-based chemotherapy, with an overall response rate of 17% to 32% and a moderately improved survival among patients. Radiation therapy is usually administered either concurrently or sequentially in combination with chemotherapy. However, many patients develop severe toxic effects from the treatment. Therefore, patients with advanced stage NSCLC are often treated with palliative purposes to reduce symptoms and to improve quality of life. Targeted therapies are used to treat advanced stage NSCLC as well (3).

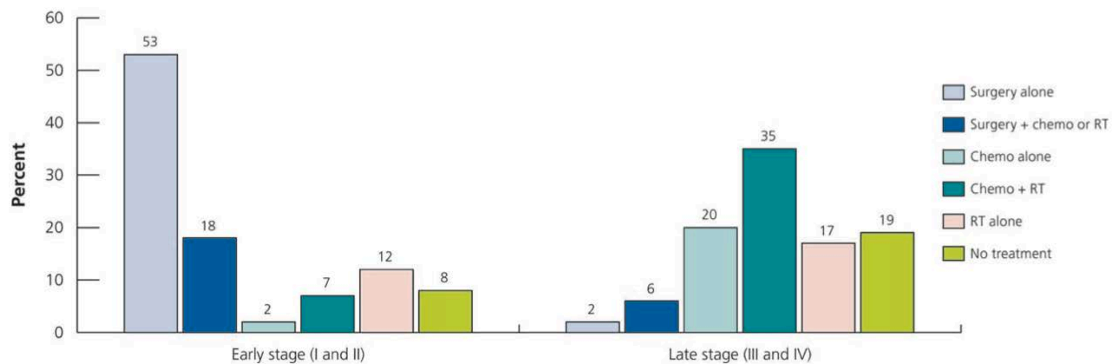


Figure 1. Non-Small Cell Lung Cancer Treatment Patterns by Stage, 2008.

Chemo indicates chemotherapy (may include common targeted therapies); RT, radiation therapy. Percentages do not sum to 100% due to rounding. Data source: NCDB (6). Reprinted by permission from CA CANCER J CLIN, copyright (2012).

Overall outcomes for all stages of lung cancer have improved in recent years. The 1-year survival rate has increased during the past several decades primarily due to improvements in surgical techniques and chemotherapy for lung cancer. However, the overall 5-year survival rate for lung cancer has remained relatively unchanged at approximately 15% for the past two decades (2). The 5-year survival rates for local, regional and distant diseases are 54%, 26% and 4% (Figure 2), respectively. Therefore, early detection is the best way to reduce lung cancer mortality. However, only 15% of lung cancer cases are diagnosed at the local stage. Most lung cancer patients are diagnosed at inoperable advanced stages when the prognosis is particularly dismal, as the early stage disease is typically asymptomatic (3). The overall 5-year survival rate is 6% for SCLC and 18% for NSCLC (6). Among NSCLCs, patients with adenocarcinoma usually have better prognosis compared to those with other subtypes, while patients with large cell carcinoma usually have worse prognosis since tumors are often poorly differentiated and metastasize early.

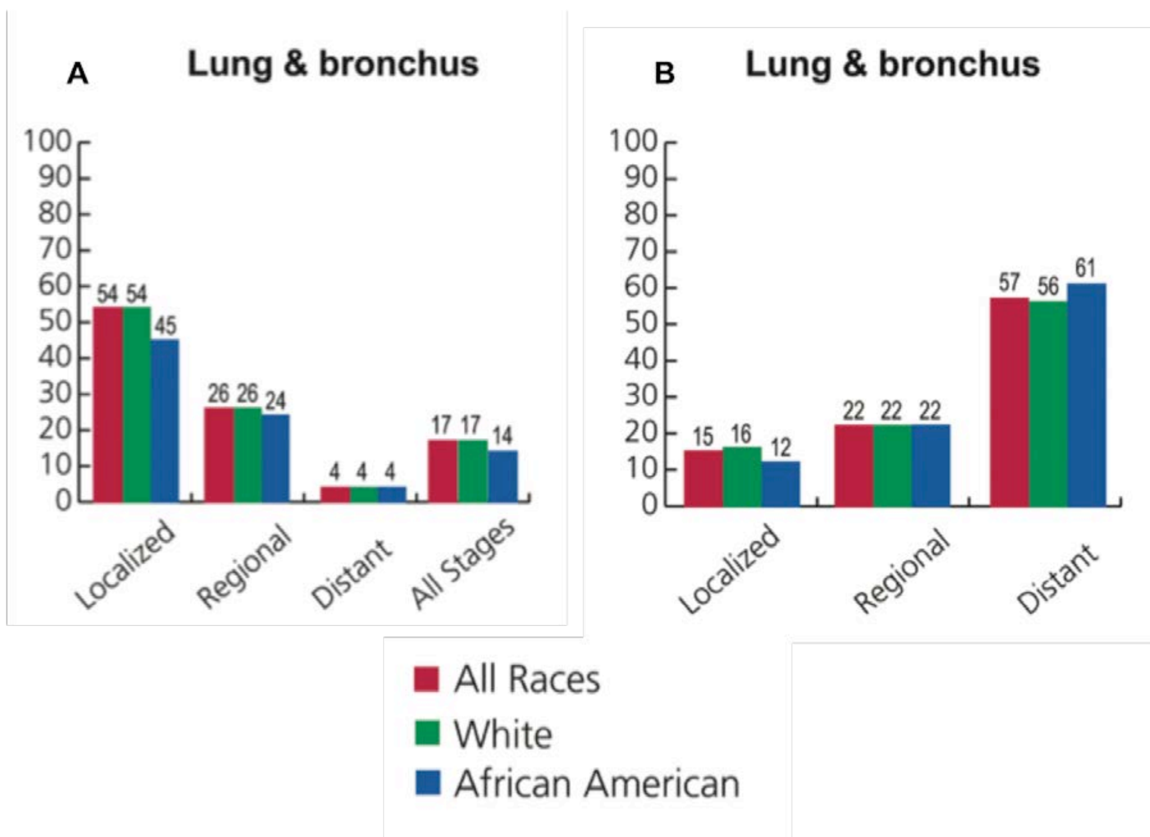


Figure 2. Five-Year Relative Survival Rates (A) and Stage Distribution (B) of Lung Cancer by Race and Stage at Diagnosis.

United States, 2003 to 2009 (2). Reprinted by permission from CA CANCER J CLIN copyright (2014).

1.3. Lung Cancer Prognostic and Predictive Factors

Survival time varies significantly among lung cancer patients, with some patients surviving years, and others less than a few months. Besides the well-known TNM stage, several other factors have been shown to affect clinical outcomes among NSCLC patients. These factors can be classified as prognostic or predictive markers. A prognostic factor is a characteristic that indicates the course of disease and clinical outcomes irrespective of treatments – it indicates the effect of the tumor on the patient. A predictive marker is a characteristic that indicates better clinical outcomes from a specific treatment, which determines the effect of treatment on the tumor (7, 8). Prognostic factors may help identify patients who are more likely to experience recurrence or progression to advanced disease and facilitate physicians making treatment plans accordingly, while predictive factors could be used to predict treatment responses, thus maximizing responses from effective treatments, minimizing toxicity associated with ineffective cytotoxic treatments and therefore improve overall survival and quality of life with personalized treatments. Several variables have been identified as prognostic/predictive markers for NSCLC clinical outcomes as our understanding in the molecular mechanisms underlying NSCLC tumorigenesis evolves (8).

Prognostic factors for NSCLC can generally be grouped into three categories: patient-related factors (e.g. age, gender, performance status and comorbidity, etc.), tumor-related factors (e.g. TNM stage, tumor histology and grade, etc.) and environmental factors (e.g. nutritional status, etc.). Several

molecular markers have been well studied as prognostic factors for NSCLC, e.g., excision repair cross-complementation group 1 protein (ERCC1), P53, KRAS and EGFR (7). Several studies have also investigated clinical and laboratory variables as prognostic factors for lung cancer clinical outcomes, and to guide selection of treatment plan. Among those, only a few are validated and used clinically as prognostic factors, e.g. TNM stage and performance status of patients. The clinical utility of other factors, e.g. clinical laboratory tests, although suggested by some studies, are still controversial and requires further large studies, preferably in large randomized clinical trials (RCTs) (9).

1.4. Lung Cancer Screening

In 2011, the National Lung Screening Trial (NLST) concluded that low-dose helical computed tomography (LDCT) screening of high-risk individuals could reduce lung cancer mortality by 20% among current and former heavy smokers compared with standard chest x-ray based on findings from 53,454 participants enrolled (10). Based on these findings, LDCT screening according to NLST selection criteria, i.e., current or former smokers aged 55-74 years with at least 30 pack-years of smoking history and no more than 15 years since quitting, has been recommended by the majority of professional organizations in the US (11-14). Moreover, it has recently been reported that participants with the highest risk for lung cancer deaths accounted for the most screening-prevented lung-cancer deaths and benefitted most from LDCT, while very few deaths were prevented among those at lowest risk (15). However, the feasibility of large-scale screening

is limited by a false positive rate of greater than 95% (10, 11). Biomarkers are urgently needed for improving risk prediction for lung cancer beyond smoking variables alone to reduce false positives and shift the balance towards higher cost-effectiveness for screening and early detection of lung cancer.

1.5. Cancer Biomarkers

According to the definition of National Cancer Institute, a biomarker is “a biological molecule found in blood, other body fluids or in tissues that is a sign of a normal or abnormal process, or of a condition or disease” (16). Accordingly, sources of biomarkers include whole blood/plasma/serum, urine, sputum, tissue or peripheral blood lymphocyte, etc. Cancer biomarkers can be used to assess exposure or disease risk, screen the general population, aid in the diagnosis and clinical staging of cancer, predict response or resistance to a specific treatment, monitor disease course to assess recurrence or progression, or serve as prognostic factors for disease, etc. (16). Cancer biomarkers can be applied across the cancer continuum, offering one of the best ways for risk prediction, early diagnosis, therapeutic response prediction and prognosis.

1.5.1. Classification of biomarkers

Generally, cancer biomarkers can be produced from either the tumor itself or other tissues in response to the presence of tumor or other conditions associated with the tumor. There are four categories of cancer molecular biomarkers: (i) risk biomarkers, which are factors associated with or contributing

to the carcinogenesis process, examples of which include genetic risk factors, high estradiol levels for breast cancer patients as well as HPV infection for cervical cancer patients; (ii) released biomarkers, which are factors released in abnormal levels in cancer patients due to anatomical or metabolic abnormality associated with tumor, examples of which include high PSA levels in prostate cancer patients or blood in stools in colorectal cancer patients; (iii) response biomarkers, e.g. antibodies and acute-phase reactants (17), which are factors generated by the body in response to the tumor; (iv) carcinogenesis biomarkers, such as DNA mutation or hyper-methylation, which are shed by the tumor and products of the carcinogenic process (18).

There are potential limitations associated with each class of biomarkers. Risk biomarkers could be used to predict individual's cancer risk. However, for a given risk biomarker, only a small percentage of individuals with the risk factor would develop cancer. Therefore, the integration of several risk biomarkers into a risk prediction model is necessary to improve the prediction. Release biomarkers are present in abnormal levels in cancer patients due to the presence of the tumor. However, other conditions could also lead to abnormal levels of these factors since they are often non-specific molecular biomarkers for cancer. Similarly, many other pathological conditions could lead to generation of response biomarkers by the body. Conversely, carcinogenesis biomarkers are a group of markers that are specific for patients with different types of cancer. Depending on the purpose and design of the study, carcinogenesis biomarkers may be identified for early

diagnosis of cancer, although they are more likely to be present in metastatic or invasive tumors.

1.5.2. Blood-based biomarkers

Blood/serum/plasma is the most common source for biomarker studies and the most used biological material in the clinic. It offers several advantages over other biological samples: (i) it is minimally invasive to obtain blood samples; (ii) it's abundant and could be obtained serially to allow for disease follow-up; (iii) blood serves as a metabolic source/sink and reflects the cumulative impact of all organ systems including perturbations relating to disease. However, one of the major disadvantages with blood for biomarker studies is that analytes in blood, including RNAs, proteins and metabolites, usually present at levels spanning a large range, making it difficult to measure all analytes accurately (19). Nevertheless, blood/serum/plasma is still the most cost-effective and widely used source for biomarker studies. Unfortunately, there are not many validated serum biomarkers of sufficient sensitivity and specificity that can be applied clinically.

1.5.3. Tumor-derived biomarkers

Tissue is another important source for cancer biomarker studies. Tissue-based biomarkers from the tumor are more specific to the underlying cancer and reflect more directly the pathogenic process compared to biomarkers from other sources. However, the sampling procedure is usually invasive, and therefore more difficult and costly to obtain. Moreover, tumor samples are usually

obtained at a single time point (at the time of surgery) and are less amenable for tracking disease status or progression.

Biomarkers with clinical significance can be applied along the cancer continuum to substantially reduce cancer burden through risk prediction, prevention, early detection, personalized treatment and disease monitoring after treatment. During the past decade, there have been substantial new discoveries on cancer biomarkers with the application of large-scale high-throughput technologies. However, thus far, very few biomarkers have been validated and applied in the clinic. One major problem is the lack of validation for the identified biomarkers. Many studies lacked a replication phase, and few biomarkers have been prospectively validated, which minimizes biases and reverse causality. Another major concern is the relatively low sensitivity or specificity of the candidate biomarkers. Many potential biomarkers don't have sufficient sensitivity and specificity to be useful in the clinic for decision-making purposes. Nevertheless, those markers that are not specific to cancer could serve as markers of other diseases and pathologies that are associated with cancer, including inflammation, anemia, malnutritrion, cachexia, etc.

1.6. Cancer Metabolism

It is well established that cancer cells undergo profound changes in cellular metabolism to sustain the additional demands for energy and synthesis of essential biochemical precursors required for uncontrolled proliferation (20). Metabolic reprogramming has been increasingly recognized as an emerging

hallmark of cancer, providing cancer cells with energy and biosynthetic materials to support continuous cell growth and proliferation in cancer (21). Alterations in several pathways, including glycolysis, glutaminolyis and mitochondrial biogenesis are among the most significant alterations of cancer cell metabolism, supporting both cancer initiation and progression. They provide cancer cells both the energy required as well as metabolites for synthesizing macromolecules and organelles for cell proliferation. For example, increased glycolysis in cancer cells diverts glycolytic intermediates to molecules used for generating nucleosides and amino acids, which facilitates macromolecules biosynthesis (22). Therefore, a deep understanding of the fundamental metabolic changes that occur during lung cancer development could lead to the identification of novel biomarkers for lung cancer early diagnosis.

1.7. Metabolomics

Metabolomics is the systematic, unbiased study of the unique chemical fingerprints generated by metabolic processes that can inform the cellular processes of an organism (23). In the first issue of the new decade, *Nature* asked a selection of leading researchers and policy-makers in five areas where their fields would be ten years from then, and one of the fields chosen was metabolomics (23). In contrast to the widely studied genomics and transcriptomics, metabolomics is an emerging science. Metabolites include the end products of the cellular processes, which represent the distal read-out of the cellular state. It is therefore thought that metabolomic profiling reflects

physiological functions and pathological characteristics in greatest detail. As part of a system biology perspective, metabolomic profiling has become an important complement to the other “omic” approaches including genomics, epigenomics, transcriptomics, and proteomics and is emerging as an important tool to identify biomarkers for the early detection, diagnosis, and prognosis of cancers (24).

Multiple technologies have been applied for metabolomic profiling, including nuclear magnetic resonance (NMR), high pressure liquid chromatography (HPLC), gas chromatography/mass spectrometry and liquid chromatography/mass spectrometry (GC/MS and LC/MS). Among these, NMR has the limitation of defining only named compounds, although it is sensitive and high-throughput. HPLC requires an external standard and identifies compounds based solely on their chromatographic retention time. On the other hand, GC/MS and LC/MS are sensitive and allow for both identification of known metabolites and characterization of unknown compounds (24).

Several studies have utilized metabolomic profiling to reveal metabolic alterations associated with various malignancies, including breast (25), colorectal (26), esophageal (27), gastric (28), liver (29), kidney (30), oral (31), pancreatic (32) and prostate (24) cancers. Recently, a few studies have investigated metabolic profiles of lung cancer patients (33-35). However, they only targeted and measured a small, selected number of metabolites.

1.8. Hypothesis and Rationale

1.8.1. Hypothesis I: Global metabolomic profiling could identify serum metabolites

that are differentially expressed during lung cancer development, which could consequently be used as novel biomarkers for the early detection of lung cancer.

Cancer cells have been known to acquire altered cellular metabolism to sustain the additional demands for energy and other biochemical precursors for uncontrolled proliferation. Metabolomic profiling offers a functional readout of the physiological state and a comprehensive picture of the metabolic changes associated with cancer development. We therefore hypothesized that global metabolomics profiling followed by target validation of individual metabolites can identify serum metabolites as potential biomarkers for the early detection of lung cancer. In this study, we propose to use a global metabolomic profiling platform, which can measure hundreds of metabolites simultaneously, to provide a comprehensive assessment of metabolites from serum of NSCLC patients compared to paired healthy control subjects. These metabolomic profiles can guide the identification of novel biomarkers for early detection of NSCLC.

1.8.2. Hypothesis II: Pre-treatment clinical laboratory tests indicative of metabolic status are associated with survival in NSCLC patients, which could serve as biomarkers for clinical outcome in NSCLC patients.

Survival time varies significantly among early stage or late stage NSCLC patients. Clinically relevant biomarkers that predict prognosis and survival time are urgently needed to personalize treatment regimens and manage expectations. Patients' pre-treatment clinical laboratory test levels are reflective of

patient-related and environmental factors. Serum pre-treatment laboratory test levels have long been reported to be prognostic factors for SCLC (36). Therefore, we hypothesized that pre-treatment laboratory test levels that indicate metabolic status are associated with NSCLC patients' survival, and could be utilized as biomarkers for NSCLC clinical outcome prediction.

Chapter 2. Materials and Methods

2.1. Study Population and Data Collection

MD Anderson Case-Control Population in Metabolomics Study

The lung cancer cases and control subjects were from an ongoing lung cancer case-control study at the University of Texas MD Anderson Cancer Center (37). Cases were newly diagnosed, histologically confirmed non-small cell lung cancer (NSCLC) patients previously untreated with chemotherapy or radiotherapy at MD Anderson Cancer Center. There were no restrictions on age, sex, or ethnicity at study recruitment. Early stage NSCLC included stages I and II, while late stage NSCLC included stages III and IV. The controls were recruited in the Kelsey Seybold Clinics, Houston's largest private multispecialty group practice with 18 clinics in the Houston metropolitan area. The control subjects were healthy individuals without prior history of cancer (except for non-melanoma skin cancer). To control for the confounding effect of ethnicity, we only included Caucasians in this study. Twenty each of controls, early-stage, and late-stage NSCLC cases (hereafter referred to as "trio") were used for metabolomic profiling. Promising metabolites identified from this profiling were examined in two additional case-control samples, consisting of 50 trios and 123 trios, respectively (Figure 3). All participants completed an in-person interview administered by MD Anderson staff interviewers using a structured questionnaire. Demographic characteristics, smoking history, family history of cancer, environmental exposures and other epidemiologic data were collected and recorded. At the end of the interview, each participant donated 40ml blood sample for molecular analysis.

All patients and controls gave written informed consent before participation, and the study was approved by the Institutional Review Boards of the MD Anderson Cancer Center and the Kelsey Seybold Clinics.

The Taiwanese Prospective Cohort Population for Metabolomic Validation

The Taiwanese prospective cohort population was obtained from a standard medical screening program conducted by the MJ Health Management Institution at Taiwan (hereon referred to as “MJ”). From 1994 to 2008, a total of 435,985 subjects aged 20 years and older, free of cancer at baseline, were recruited. Median follow-up time for the cohort is 8 years (interquartile range: 5-11 years) for male participants and 9 years (interquartile range: 5-11 years) for female participants. All participants completed a self-administered questionnaire covering demographic characteristics and health history. All subjects underwent a series of medical tests for blood, urine, physical examination, body measurements and functional tests, including testing for anthropometric measurements (e.g., height, weight, waist circumference, hip circumference, body fat percentage, etc.), blood pressure, pulse rate and respiration rate. Overnight fasting blood was analyzed for a standard panel, including hemogram and testing for blood sugar, liver function, renal function, thyroid function, blood lipids and blood grouping. The cohort members were followed through 2008 for cancer and vital status, which were assessed by linkage of the individual ID to the National Cancer Registry and National Death file.

All participants signed an informed consent. The studies were approved by the Institutional Review Boards of the University of Texas MD Anderson Cancer Center and National Health Research Institutes, Taiwan.

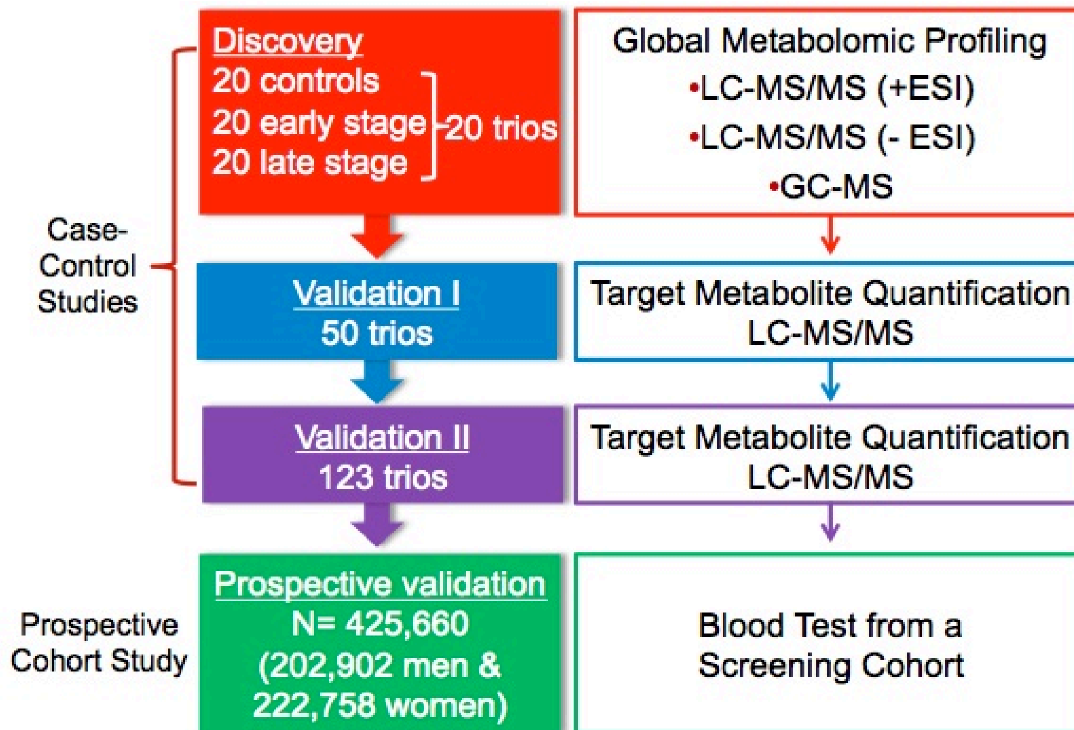


Figure 3. Study design of lung cancer metabolomic profiling and validation

MD Anderson Lung Cancer Patient Population for Lung Cancer Clinical Outcome

This study was conducted in 2,806 patients with newly diagnosed, histologically confirmed NSCLC recruited at the University of Texas MD Anderson Cancer Center. None of the study subjects had undergone chemotherapy or radiotherapy treatment before study enrollment. There were no restrictions on age, sex, or ethnicity at recruitment. Excluding 131 patients with unknown clinical stage information, there were 2,675 patients remaining for analysis. Early stage NSCLC included stages I and II patients, while advanced stage NSCLC included stages III and IV patients. All participants completed an in-person interview by trained MD Anderson Cancer Center staff using a structured questionnaire. Demographic characteristics, smoking history, family history of cancer, and exposure data were collected. After the interview, each participant donated 40 ml blood sample for molecular analysis. Clinical and follow-up data were extracted from medical charts by trained medical staff. Pretreatment serum levels of albumin, alkaline phosphatase, alanine aminotransferase, total bilirubin, glucose, lactate dehydrogenase and total protein were measured as part of a standard battery of tests evaluating patients' overall metabolic status. The normal ranges of the tests were based on standard laboratory norms. All study subjects had their blood drawn and laboratory tests evaluated no earlier than 30 days before diagnosis and before implementation of any treatment.

All patients signed written informed consent before participation, and the study was approved by the Institutional Review Board of the University of Texas MD Anderson Cancer Center.

2.2. Global Metabolomic Profiling

The metabolomic profiling analysis of all samples was carried out by Metabolon, Inc (Durham, NC) using the following general protocol described previously (38). Briefly, this process involved: sample extraction, separation, detection, spectral analysis, data normalization, delineation of class-specific metabolites and pathway mapping. All samples were randomized prior to mass spectrometric analyses to avoid any experimental drifts. A number of internal standards, including injection standards, process standards, and alignment standards were used to assure QA/QC targets were met and to control for experimental variability. Samples were kept frozen until assays were ready to be performed. The samples were extracted and divided into three equal portions for analysis on one of the three platforms: ultrahigh performance liquid chromatography/tandem mass spectrometry (UHLC/MS/MS2) optimized for basic species, UHLC/MS/MS2 optimized for acidic species, and gas chromatography/mass spectrometry (GC/MS). For UHLC/MS/MS2 analysis, two aliquots were processed with one using acidic positive ion optimized conditions and the other using basic negative ion optimized conditions in two independent injections using separate dedicated columns. The platform includes a Waters ACQUITY ultra-performance liquid chromatography (UPLC) system (Waters, Millford, MA) and an LTQ mass spectrometer (Thermo Fisher Scientific, Inc., Waltham, MA) comprised of electrospray ionization (ESI) source and linear ion-trap (LIT) mass analyzer. The MS instrument scans 99-1000 m/z and alternates between MS and MS2 scans using dynamic exclusion with

approximately 6 scans per second. For GC/MS analysis, one aliquot was used and separated on a 5% phenyldimethyl silicone column with helium as the carrier gas and a temperature ramp from 60°C to 340°C. The platform includes a Thermo-Finnigan Trace DSQ fast-scanning single-quadrupole mass spectrometer that uses electron impact ionization (Thermo Fisher Scientific, Inc.) and has a 50-750 atomic mass unit scan range. Metabolites were identified by automated comparison of the ion features in the experimental samples to the metabolomic library of chemical standard entries using software developed at Metabolon (39). Known chemical entities were mapped to the metabolomic library entries of purified standards. Identification of additional entities was based on the virtue of their recurrent nature (both chromatographic and mass spectral).

2.3. Quantification of individual metabolites

Gamma-glutamylalanine and bilirubin standard powders were purchased from Sigma-Aldrich (St. Louis, MO). Individual metabolite quantification was done at Texas Southern University (Houston, TX). Quantification of individual metabolite in serum was determined by LC-MS/MS methods using a 3200 QTRAP® MS/MS coupled by an Agilent 1200 Series HPLC system. Standard curves for each compound were constructed by spiking known amount of the standard to a series of blank plasma (Gulf Coast Blood Bank, Houston, TX). For the determination of λ-glutamylalanine, aliquot of serum samples (50 μL) was mixed with 500 μL of a 2:1 ratio (v/v) solution of methanol/acetone containing etravirine as an internal standard. After a brief vortex and centrifugation, the supernatant was transferred to a 12x75 mm glass tube where it was dried under

pure nitrogen. The residual was then reconstituted with 100 µL of a 1:1 (v/v) solution of methanol/water before HPLC injection. Gamma-glutamylalanine and the internal standard were separated on a reverse phase XTerra MS-C18 (2.1 x 50mm, 3.5 µm) column. The mobile phases consisted of a 0.5 mM ammonium acetate aqueous solution (mobile phase A) and a 100% acetonitrile solution (mobile phase B). A gradient elution starting with 95% mobile phase A at a flow rate of 200 µL/min was applied to achieve retention time of 0.69 min and 4.3 min for λ-glutamylalanine and the internal standard, respectively. For the determination of bilirubin, aliquot of serum samples (50 µL) was mixed with 250 µL of methanol containing etravirine as an internal standard. After a brief vortex and centrifugation, the supernatant was injected directly onto the HPLC system. Bilirubin and the internal standard were separated on a reverse phase XTerra MS-C18 (2.1 x 50mm, 3.5 µm) column. The mobile phases consisted of a 2 mM ammonium acetate aqueous solution adjusted to pH 7.8 by ammonium hydroxide (mobile phase A) and a 98% methanol solution containing 2% (v/v) of a 2 mM ammonium acetate (mobile phase B). A gradient elution starting with 70% mobile phase A at a flow rate of 200 µL/min was applied to achieve retention time of 5.6 min and 5.8 min for bilirubin and the internal standard, respectively.

Mass spectrometric detection was performed on a 3200 QTRAP® (AB Sciex, Foster City, CA, USA), a hybrid triple quadrupole linear ion trap equipped with a TurboIonspray ion source. A Parker Balston Source 5000Tri Gas generator was used to generate pure nitrogen. The mass spectrophotometer was set at the negative mode. The transition ions were detected using multiple reaction

monitoring from a specific parent ion transition to product ion for λ-glutamylalanine (m/z 217→128), bilirubin (583.4→285), and etravirine (432.9→141.6). Peak areas and other compound parameters were determined by Analyst R software, version 1.5. Standard curve ranges are 25 – 500 ng/mL and 500 – 10,000 ng/mL for λ-glutamylalanine and bilirubin, respectively.

2.4. Statistical Analyses

Lung Cancer Metabolomic Profiling Study

In the case-control analysis, the Pearson χ^2 test was used to examine the differences in sex and smoking status between cases and controls. Student's t test was used to test for differences in age and pack-years of smoking as continuous variables. For the metabolomic profiling assay, missing metabolite measurements, which were due to levels below detection limits, were imputed with the compound minimum (minimum value imputation). Only metabolites with detectable expression in at least 80% of the samples were analyzed. Nonparametric two-sided Wilcoxon rank-sum tests were used to analyze the differences of metabolite levels between cases and controls for both metabolomic profiling and individual metabolite validation assay. Nonparametric trend test was used to analyze the trend across normal controls, early stage lung cancer and late stage lung cancer cases for both metabolomic profiling and individual metabolite validation assay. Spearman's correlation test was used to assess the correlation between the two values measured with metabolomic profiling and individual metabolite assay with LC-MS/MS.

For the MJ cohort, serum total bilirubin levels were divided into three groups with equal tertile in men or women. Cox proportional hazard models were used to assess the association of serum total bilirubin levels with lung cancer incidence or mortality rate using the highest tertile category of bilirubin levels (>1 mg/dL for men and >0.82 mg/dL for women) as reference. For lung cancer incidence, the event time was from the date of recruitment to the end of follow-up, or the date of lung cancer identification if earlier. For lung cancer mortality, the event time was from the date of recruitment to the end of follow-up, or the date of death due to lung cancer if earlier. Hazard ratios were adjusted for age, educational level (middle school or lower, high school, junior college, or college or higher), body mass index (BMI) and pack-years of smoking in a multivariable model with continuous variables whenever appropriate. The proportional hazards assumption was assessed by plotting Schoenfeld residuals versus time and examining their correlation. Two-way interactions between smoking status and serum total bilirubin levels were assessed for lung cancer incidence and mortality in men and women. Non-parametric trend tests were used to analyze the trend of serum total bilirubin levels across each characteristic group of the cohort (e.g., age, BMI, etc). All statistical tests were two-sided with the threshold for significance set at 0.05. Statistical analyses were performed using STATA 10.0 (StataCorp, College Station, TX).

Clinical Laboratory Tests and Lung Cancer Clinical Outcome Study

The primary endpoint for this study was overall survival. Secondary endpoints for early stage NSCLC patients were recurrence and progression.

Survival time was calculated from the date of diagnosis to the date of death or last patient follow-up. Chi-square test was used to assess differences in patient characteristics. Multivariable Cox proportional hazards model was used to assess the effect of each pretreatment serum laboratory test level on 2-year or 5-year survival, adjusted for age, sex, smoking status, clinical stage, and treatment regimen. Kaplan-Meier survival analysis was used to estimate survival, and log-rank tests were used to assess the differences in overall survival for each laboratory test. The cumulative effects of multiple unfavorable laboratory test levels were evaluated for the three tests that showed statistical significance in the main analysis (i.e., $P < 0.05$). The lab-test based risk score for each patient was derived by linear combination of the product of reference-normalized expression level of each laboratory test by its Cox regression corresponding coefficient (21). All patients were dichotomized by the median risk score, and individuals with a risk score higher or lower than the median were classified as high or low risk groups, respectively. A P value less than 0.05 was considered significant in all statistical analyses. All statistical analyses were conducted using STATA software, version 10 (StataCorp, College Station).

Chapter 3. Results and Discussion

3.1. Lung Cancer Metabolomic Profiling Study

3.1.1. Characteristics of the study population

In the case-control studies, the lung cancer cases and healthy controls were all Caucasians, matched on age and gender (Table 1). The Taiwanese MJ cohort consisted of 425,660 subjects (202,902 men and 222,758 women) aged 20 years and older (Table 2). Median follow-up time for this cohort was 8 years (interquartile range [IQR], 5-11 years) for men and 9 years (IQR, 5-11 years) for women, which yielded a total of more than 3.4 million person-years of follow-up. The average age at testing was 41 for both men and women. Median total bilirubin level was 0.87 mg/dL (IQR, 0.68-1.11 mg/dL) in men and 0.7 mg/dL (IQR, 0.56-0.9 mg/dL) in women. Serum total bilirubin levels were higher in men than in women, which were consistent with previous studies (40, 41). Selected demographic characteristics and exposures of the cohort participants are shown in Table 2, presented by gender and tertiles of bilirubin level (<0.75, 0.75-1 and >1 mg/dL for men and <0.61, 0.61-0.82 and >0.82 mg/dL for women). Distribution of serum total bilirubin levels among the participants in the cohort is shown in Figure 3. Among male participants in the cohort, over half (52.1%) were smokers, with 25% of them being heavy smokers of ≥30 pack-years. In contrast, only 17,123 (8.3%) female participants were smokers, with 1,327 (8.3%) of them being heavy smokers. During the follow-up, there were 809 incident lung cancer cases and 614 lung cancer deaths among the males, and 524 lung cancer cases and 330 deaths among the females.

Table 1. Selected host characteristics of three case-control populations

Characteristics	Phase I (N=60)			Phase II (N=150)			Phase III (N=369)		
	Cases (N=40)	**Controls (N=20)**	***P***	**Cases (N=100)**	**Controls (N=50)**	***P***	**Cases (N=246)**	**Controls (N=123)**	***P***
Age									
Mean (SD), y	62 (11)	59 (10)	0.38	61 (12)	59 (12)	0.43	64 (10)	64 (10)	0.82
Sex, No. (%)									
Male	13 (32.5)	8 (40.0)		40 (40.0)	20 (40.0)		174 (70.7)	87 (70.7)	
Female	27 (67.5)	12 (60.0)	0.57	60 (60.0)	30 (60.0)	1	72 (29.3)	36 (29.3)	1
Smoking status, No. (%)									
Non-smoker	6 (15.0)	7 (35.0)		21 (21.0)	16 (32.0)		37 (15.0)	34 (27.6)	
Smoker	34 (85.0)	13 (65.0)	0.19	79 (79.0)	34 (68.0)	0.14	209 (85.0)	89 (72.4)	0.009
Pack year, Mean (SD)	48.0 (30.7)	54.1 (42.3)	0.59	44.7 (31.5)	48.5 (32.9)	0.57	52.1 (32.1)	33.9 (25.9)	<0.001

Table 2. Characteristics of the participants in the prospective cohort by gender and serum total bilirubin levels [a]

Characteristics	Men (N = 202,902), N (%)				Women (N = 222,758), N (%)			
	Total	Total bilirubin level (mg/dL)			Total	Total bilirubin level (mg/dL)		
		>1	0.75-1	<0.75		>0.82	0.61-0.82	<0.61
Total	202,902	67,841 (33.4)	65,540 (32.3)	69,521 (34.3)	222,758	75,189 (33.8)	72,207 (32.4)	75,362 (33.8)
Age (y), mean (SD)	41 (14)	40 (14)	42 (14)	41 (14)	41 (14)	41 (14)	42 (14)	41 (13)
20-39	112,584 (55.5)	39,399 (35.0)	35,270 (31.3)	37,915 (33.7)	119,946 (53.9)	41,854 (34.9)	37,510 (31.3)	40,582 (33.8)
40-59	63,447 (31.3)	19,927 (31.4)	21,201 (33.4)	22,319 (35.2)	76,087 (34.2)	23,908 (31.4)	25,500 (33.5)	26,679 (35.1)
≥60	26,871 (13.2)	8,515 (31.7)	9,069 (33.8)	9,287 (34.6)	26,725 (12)	9,427 (35.3)	9,197 (34.4)	8,101 (30.3)
BMI (kg/m^2), mean (SD)	23.9 (3.4)	23.5 (3.3)	23.9 (3.3)	24.2 (3.4)	22.3 (3.6)	21.8 (3.5)	22.3 (3.6)	22.8 (3.7)
<25	134,591 (66.4)	47,499 (35.3)	43,264 (32.1)	43,828 (32.6)	176,567 (79.3)	62,056 (35.2)	57,139 (32.4)	57,372 (32.5)
25-29.9	59,734 (29.5)	18,062 (30.2)	19,565 (32.8)	22,107 (37.0)	38,454 (17.3)	11,040 (28.7)	12,660 (32.9)	14,754 (38.4)
≥30	8,516 (4.2)	2,264 (26.6)	2,696 (31.7)	3,556 (41.8)	7,689 (3.5)	2,080 (27.1)	2,394 (31.1)	3,215 (41.8)
Educational levels								
Middle school or lower	40,499 (20.6)	12,823 (31.7)	13,379 (33.0)	14,297 (35.3)	70,385 (32.6)	23,109 (32.8)	23,546 (33.5)	23,730 (33.7)
High school	45,601 (23.2)	14,665 (32.2)	14,693 (32.2)	16,243 (35.6)	54,124 (25.1)	17,238 (31.9)	17,186 (31.8)	19,700 (36.4)
Junior college	45,367 (23.1)	15,705 (34.6)	14,526 (32.0)	15,136 (33.4)	42,941 (19.9)	15,153 (35.3)	13,525 (31.5)	14,263 (33.2)
College or higher	64,987 (33.1)	22,603 (34.8)	20,844 (32.1)	21,540 (33.2)	48,400 (22.4)	17,428 (36.0)	15,692 (32.4)	15,280 (31.6)
Smoking status								
Non-smoker	92,864 (47.9)	35,175 (37.9)	30,188 (32.5)	27,501 (29.6)	188,685 (91.7)	64,488 (34.2)	61,534 (32.6)	62,663 (33.2)
Smoker	101,092 (52.1)	29,632 (29.3)	32,451 (32.1)	39,009 (38.6)	17,123 (8.3)	4,891 (28.6)	5,228 (30.5)	7,004 (40.9)
<30 pack-years	72,153 (74.9)	21,843 (30.3)	23,084 (32.0)	27,226 (37.7)	14,662 (91.7)	4,303 (29.4)	4,403 (30.0)	5,956 (40.6)
≥30 pack-years	24,146 (25.1)	6,269 (26.0)	7,777 (32.2)	10,100 (41.8)	1,327 (8.3)	279 (21)	434 (32.7)	614 (46.3)
Lung cancer incidence	809 (0.4)	215 (26.6)	270 (33.4)	324 (40.1)	524 (0.2)	155 (29.6)	187 (35.7)	182 (34.7)
Lung cancer mortality	614 (0.3)	147 (23.9)	214 (34.9)	253 (41.2)	330 (0.2)	107 (32.4)	115 (34.9)	108 (32.7)

[a] Percentage may not total 100 because of rounding.

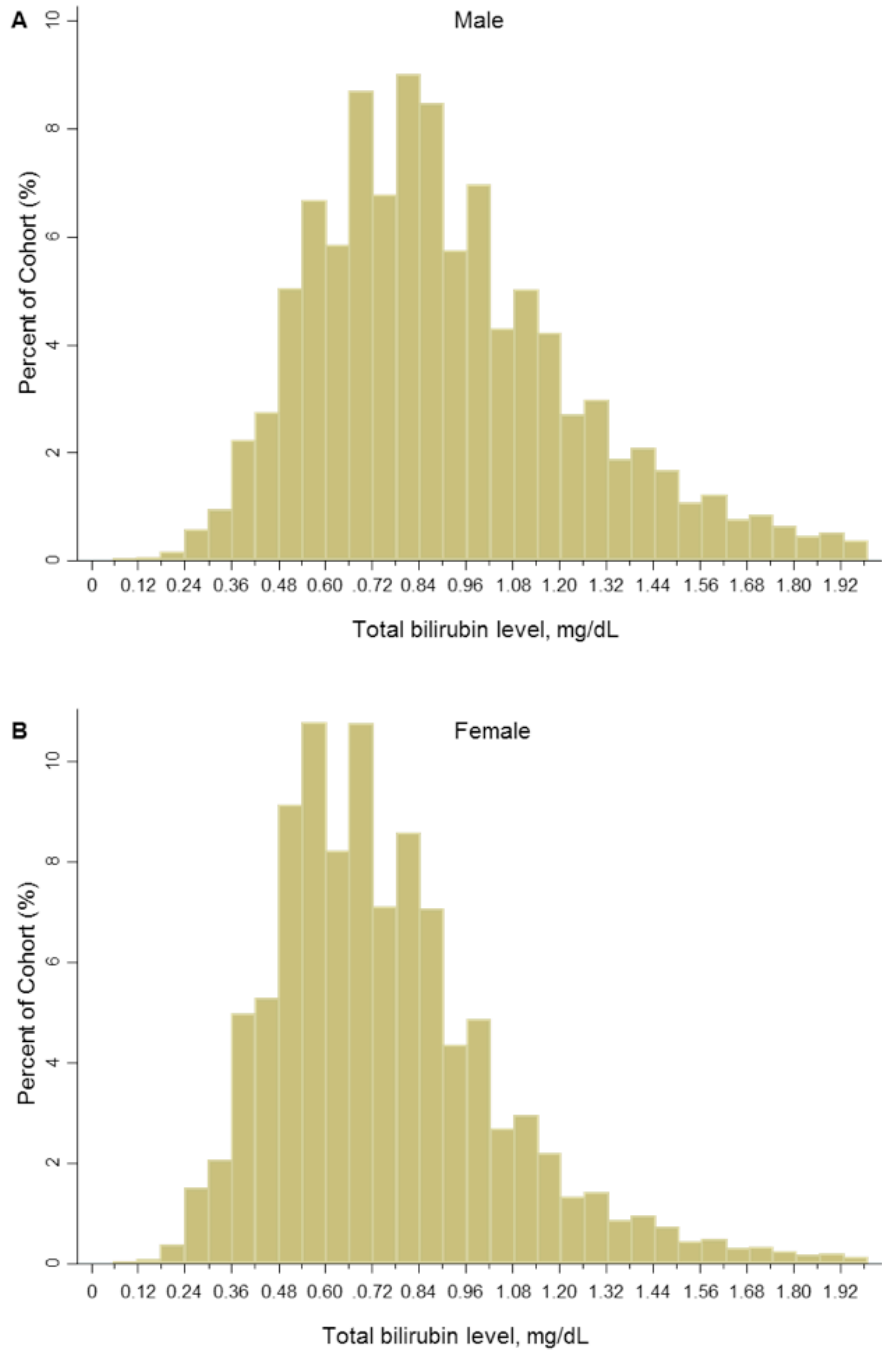


Figure 4. Distribution of serum total bilirubin levels among male (A) and female (B) participants in the MJ cohort.

3.1.2. Global metabolomic profiling of lung cancer

Serum global metabolomic profiles of 40 lung cancer cases and 20 healthy controls (20 trios) were assessed in the initial case-control study and a total of 403 named metabolites were identified. After exclusion of metabolites detected in less than 80% of samples, 306 (76%) metabolites remained. These metabolites were mapped to 8 super-pathways and 61 sub-pathways (Appendix B). Among these, 24 metabolites exhibited significantly differential expression between lung cancer cases and healthy controls (Table 3). Furthermore, 29 metabolites exhibited a significant trend of expression when comparing normal controls, early and late stage cases, 12 of which had *P* for trend values < 0.01 (Table 4).

Table 3. Metabolites with significantly differential expression between lung cancer cases and healthy controls in metabolomic profiling.

Metabolite	Relative Expression Value		*P* Value
	Controls, Mean (SD) N=20	Cases, Mean (SD) N=40	
2-pyrrolidinone	184,592 (67,739)	120,309 (38,326)	<0.001
λ-glutamylalanine	73,251 (15,477)	55,538 (14,767)	<0.001
prolylhydroxyproline	99,837 (32,892)	147,019 (85,297)	0.003
bradykinin, des-arg(9)	149,185 (191,016)	52,631 (57,472)	0.004
citrate	4,436,480 (1,350,699)	3,526,780 (910,400)	0.008
inosine	6,089 (7,742)	15,731 (25,092)	0.008
octadecanedioate (C18)	18,830 (13,613)	25,914 (12,996)	0.009
ascorbate (Vitamin C)	379,215 (354,715)	186,720 (262,879)	0.01
xylonate	34,788 (19,194)	48,479 (21,691)	0.01
ribose	65,583 (26,471)	92,345 (56,000)	0.02
pyroglutamine*	83,523 (47,468)	120,072 (72,884)	0.02
2-hydroxyglutarate	73,643 (22,606)	59,895 (21,528)	0.02
bilirubin	42,818 (37,297)	28,382 (23,597)	0.03
1-docosahexaenoyl-GPC* (22:6)*	326,254 (110,182)	460,063 (260,002)	0.03
serotonin (5HT)	41,042 (26,983)	58,634 (33,030)	0.03
alpha-tocopherol	1,275,449 (447,775)	16,23,417 (872,291)	0.03
acetylcarnitine (C2)	1,127,480 (371,971)	1,292,182 (307,364)	0.03
caproate (6:0)	55,317 (13,834)	49,014 (15,451)	0.03
gamma-CEHC	9,740 (5,548)	6,694 (3,971)	0.04
tryptophan	7,266,894 (1,210,393)	6,764,267 (1,438,869)	0.04
nonadecanoate (19:0)	34,914 (20,482)	42,556 (19,167)	0.04
citrulline	192,653 (61,750)	157,052 (45,065)	0.05
3-hydroxybutyrate (BHBA)	3,537,910 (6,648,727)	5,629,617 (7,296,060)	0.05
lathosterol	83,975 (37,169)	65,360 (37,853)	0.05

Table 4. Metabolites with significant trend of expression comparing normal controls, early stage and late stage cases in metabolomic profiling

Metabolite	Relative Expression Value			*P* Value
	Controls Mean (SD) N=20	Early stage cases Mean (SD) N=20	Late stage cases Mean (SD) N=20	
λ-glutamylalanine	73,251 (15,477)	58,721 (15,063)	52,355 (14,121)	<0.0001
bradykinin,,des-arg(9)	149,185 (191,016)	72,002 (72,510)	33,260 (27,061)	0.0004
2-pyrrolidinone	184,592 (67,739)	119,052 (40,019)	121,566 (37,554)	0.0007
ribose	65,583 (26,471)	85,819 (74,815)	98,871 (27,388)	0.0008
prolylhydroxyproline	99,837 (32,892)	145,811 (109,956)	148,228 (53,298)	0.001
ascorbate (Vitamin,C)	379,215 (354,715)	241,922 (323,738)	131,519 (175,005)	0.003
pyroglutamine	83,523 (47,468)	94,411 (38,913)	145,732 (89,461)	0.004
bilirubin	42,818 (37,297)	34,885 (27,919)	21,879 (16,568)	0.005
citrate	4,436,480 (1,350,699)	3,703,495 (826,483)	3,350,065 (975,942)	0.006
dimethylglycine	374,999 (117,999)	394,627 (129,304)	519,430 (162,458)	0.007
λ-glutamylglutamine	158,358 (41,747)	147,878 (48,266)	115,265 (50,958)	0.009
allantoin	45,083 (18,638)	62,444 (60,083)	105,248 (75,072)	0.009
xylonate	34,788 (19,194)	44,653 (14,651)	52,306 (26,838)	0.013
citrulline	192,653 (61,750)	168,397 (36,799)	145,707 (50,432)	0.015
taurochenodeoxycholate	35,192 (25,112)	36,905 (29,972)	107,807 (157,690)	0.015
betaine	801,195 (206,034)	880,737 (261,529)	1,014,750 (243,831)	0.015
octadecanedioate (C18)	18,830 (13,613)	24,281 (10,670)	27,548 (15,075)	0.019
C-glycosyltryptophan*	30,314 (5,611)	31,348 (7,797)	42,725 (18,271)	0.021
caproate (6:0)	55,317 (13,834)	51,113 (14,652)	46,915 (16,312)	0.023
2-hydroxypalmitate	273,748 (54,880)	284,755 (59,536)	315,180 (67,137)	0.024
N-acetylalanine	11,056 (1,936)	11,228 (1,837)	13,237 (3,363)	0.024
N-acetylneuraminate	93,377 (30,573)	93,898 (46,055)	184,215 (142,088)	0.025
beta-hydroxyisovalerate	23,341 (16,904)	20,560 (13,029)	32,634 (23,122)	0.027
isoleucine	7,826,930 (1,625,275)	8,346,346 (2,541,086)	9,321,622 (2,069,754)	0.027
2-ethylhexanoic,acid	837,760 (370,924)	808,623 (287,933)	622,298 (400,523)	0.027
bilirubin (E,E)*	545,806 (488,517)	595,041 (394,994)	333,095 (247,461)	0.036
benzoate	901,738 (163,812)	929,460 (118,775)	830,883 (136,694)	0.043
taurolithocholate,3-sulfate	22,282 (14,217)	27,824 (26,186)	41,376 (28,524)	0.044
glycocholate	74,886 (58,775)	81,622 (98,815)	144,855 (155,358)	0.048

3.1.3. Target validation of individual metabolites

Metabolites exhibiting a significant trend in levels from normal individuals to early and late stage patients are potential biomarkers for the detection and prognosis of lung cancer. Of the 29 metabolites with significant trends, bilirubin caught our most interest given its potent endogenous cytoprotective properties and more importantly, its inverse association with cardiovascular disease and respiratory disease in previous reports (41-44). Therefore, we selected bilirubin and λ-glutamylalanine, which showed the most significant trend from metabolomic profiling for further validation. We developed standard LC/MS-MS assays for these metabolites and used these assays to measure their levels in the 20 trios of cases and controls from phase I of the case-control study; we found excellent correlation with metabolomic profiling data. We further examined levels of bilirubin and λ-glutamylalanine in additional 50 trios of serum samples (phase II) and 123 trios of serum samples (phase III) from controls and early and late-stage patients (Table 5). Through this process, bilirubin emerged as a metabolite that consistently showed a statistically significant trend in all three sets of trio data.

Table 5. Levels of the two individual metabolites of interest measured with LC/MS-MS in the three case-control studies

Metabolite	Expression Level, Mean (SD)			*P* Value
	Controls	Early stage	Late stage	
	Phase I (N=60)			
	(N = 20)	(N = 20)	(N = 20)	
λ-glutamylalanine (ng/mL)	98.28 (35.01)	68.06 (42.58)	73.96 (32.82)	0.03
Bilirubin (μg/mL)	2.81 (1.43)	2.12 (1.29)	1.54 (0.84)	0.004
	Phase II (N=150)			
	(N = 50)	(N = 50)	(N = 50)	
λ-glutamylalanine (ng/mL)	113.57 (43.01)	107.55 (42.68)	101.28 (36.3)	0.1
Bilirubin (μg/mL)	3.10 (1.81)	2.78 (1.54)	1.91 (1.03)	0.0001
	Phase III (N=369)			
	(N = 123)	(N = 123)	(N = 123)	
λ-glutamylalanine (ng/mL)	183.17 (81.69)	181.17 (104.05)	144.04 (78.6)	0.0002
Bilirubin (μg/mL)	2.69 (1.77)	2.56 (1.73)	2.25 (1.47)	0.02

3.1.4. Validation of bilirubin as a lung cancer marker in a large cohort

Since bilirubin is a routine blood test in health examination, we next assessed the association of blood test serum total bilirubin levels with lung cancer incidence and mortality using a large prospective cohort in Taiwan. In this cohort, younger age, lower BMI, never smoking status, lower systolic blood pressure and higher education status were all significantly associated with higher bilirubin levels in men (Supplementary Table 5). Similar associations were also observed in women. Total incidence rate for lung cancer was 5.64 (95 % CI, 5.26-6.05) per 10,000 person-years in men and 3.17 (95 % CI, 2.90-3.46) in women. Total lung cancer specific mortality rate was 3.88 (95% CI, 3.58-4.20) per 10,000 person-years in men and 1.85 (95% CI, 1.66-2.06) in women. Lower bilirubin level was associated with significantly higher rates of both lung cancer incidence and lung cancer specific mortality in men but not women. The incidence rate of lung cancer per 10,000 person-years in men was 4.27 (95% CI, 3.71-4.90) in the highest tertile of bilirubin level (>1 mg/dL) compared to 6.93 (95% CI, 6.20-7.75) in the lowest tertile (<0.75 mg/dL), which translated to a 52% increase in lung cancer incidence for the low bilirubin group ($P < 0.001$). The corresponding lung cancer specific mortality rate in men was 2.70 (95% CI, 2.30-3.17) in the highest tertile compared to 4.88 (95% CI, 4.32-5.52) in the lowest tertile, a 71% increase in lung cancer specific mortality for the low bilirubin group ($P<0.001$) (Table 10).

As expected, there was a strong dose response relationship between lung cancer risk/mortality and pack-years of smoking or smoking intensity in this cohort (Tables 6 and 7). Furthermore, among males, using non-smokers with the highest

tertile of bilirubin levels (>1 mg/dL) as reference, smokers in the lowest tertile of bilirubin levels (<0.75 mg/dL) had a 2.86-fold increased risk of developing lung cancer (Table 6). Smokers with <30 and ≥30 pack-years of smoking had HRs of 1.40 and 4.14 respectively (Table 6 and Fig. 4A). Similarly, smokers in the lowest tertile of bilirubin levels who smoked <10, 10-19 and ≥20 cigarettes per day had HRs of 1.85, 2.70 and 4.32, respectively (Table 6 and Fig. 4B). Similar results were found for lung cancer mortality (Table 7, Fig. 5A and 5B). In contrast, among females, lower serum bilirubin levels were not significantly associated with lung cancer incidence or mortality overall, in female smokers or in female non-smokers (Tables 8 and 9). Table 10 presents the rates of lung cancer incidence and mortality stratified by tertiles of serum bilirubin levels and corresponding risk estimates. We plotted the lung cancer incidence rates against subgroups of bilirubin levels and introduced a best-fit model. Those with bilirubin levels <0.42 mg/dL showed more than 80% increase in lung cancer incidence rate (6.1 vs 3.27 per 100,000 person-years, Figure 4A) and over two folds increase in mortality rate (4.09 vs 1.94 per 100,000 person-years, Figure 4B) compared to the subgroup with bilirubin levels >1.62 mg/dL.

Table 6. Relationship among smoking quantity, bilirubin levels and risk for lung cancer incidence by smoking status among male participants in the prospective cohort study

	Men (N=202,902)											
	Total			Total bilirubin level (mg/dL)								
				>1			0.75-1			<0.75		
	No.	No. of incidence	HR[a] (95% CI)	No.	No. of incidence	HR[a] (95% CI)	No.	No. of incidence	HR[a] (95% CI)	No.	No. of incidence	HR[a] (95% CI)
Non-smoker	92,864	156	1 (Ref)	35,175	64	1 (Ref)	30,188	50	0.87 (0.59-1.27)	27,501	42	0.85 (0.56-1.27)
Total smokers	101,092	603	2.64 (2.19-3.18)	29,632	139	1.84 (1.35-2.51)	32,451	202	2.38 (1.77-3.19)	39,009	262	2.86 (2.15-3.81)
Pack-year												
<30 pack-years	72,153	123	1.31 (1.02-1.69)	21,843	27	0.80 (0.50-1.29)	23,084	47	1.37 (0.92-2.03)	27,226	49	1.40 (0.94-2.07)
≥30 pack-years	24,146	454	4.01 (3.27-4.91)	6,269	108	3.14 (2.25-4.36)	7,777	145	3.48 (2.54-4.77)	10,100	201	4.14 (3.06-5.60)
# of Cigarettes per day												
<10	31,520	106	1.55 (1.20-2.01)	10,602	23	0.96 (0.58-1.57)	10,270	37	1.39 (0.91-2.13)	10,648	46	1.85 (1.24-2.74)
10-19	38,866	261	2.71 (2.19-3.34)	11,031	62	1.97 (1.36-2.84)	12,557	91	2.58 (1.84-3.60)	15,278	108	2.70 (1.95-3.72)
≥20	26,879	221	4.29 (3.46-5.33)	6,759	52	3.39 (2.31-4.96)	8,374	70	3.72 (2.61-5.30)	11,746	99	4.32 (3.11-5.99)

[a]Adjusted for age, educational level and body mass index

Table 7. Relationship among smoking quantity, bilirubin levels and lung cancer mortality by smoking status among male participants in the prospective cohort study

	Men (N = 202,902)											
	Total			Total bilirubin level (mg/dL)								
				>1			0.75-1			<0.75		
	No.	No. of mortality	HR[a] (95% CI)	No.	No. of mortality	HR[a] (95% CI)	No.	No. of mortality	HR[a] (95% CI)	No.	No. of mortality	HR[a] (95% CI)
Non-smoker	92,864	98	1 (Ref)	35,175	36	1 (Ref)	30,188	34	1.04 (0.65-1.66)	27,501	28	0.99 (0.60-1.63)
Total smokers	101,092	478	3.24 (2.60-4.05)	29,632	104	2.39 (1.63-3.50)	32,451	165	3.28 (2.28-4.72)	39,009	209	3.96 (2.77-5.65)
Pack-year												
<30 pack-years	72,153	90	1.62 (1.20-2.18)	21,843	14	0.80 (0.43-1.48)	23,084	39	2.08 (1.31-3.30)	27,226	37	2.01 (1.26-3.20)
≥30 pack-years	24,146	370	4.78 (3.77-6.05)	6,269	89	4.18 (2.81-6.22)	7,777	119	4.56 (3.11-6.69)	10,100	162	5.52 (3.81-7.99)
# of Cigarettes per day												
<10	31,520	84	1.95 (1.45-2.62)	10,602	14	1.05 (0.56-1.94)	10,270	29	1.91 (1.17-3.14)	10,648	41	2.88 (1.83-4.52)
10-19	38,866	211	3.38 (2.65-4.32)	11,031	48	2.65 (1.71-4.09)	12,557	77	3.65 (2.45-5.45)	15,278	86	3.75 (2.53-5.55)
≥20	26,879	172	5.16 (4.01-6.65)	6,759	42	4.64 (2.96-7.27)	8,374	55	4.95 (3.24-7.57)	11,746	75	5.74 (3.85-8.56)

[a]Adjusted for age, educational level and body mass index

Table 8. Relationship among smoking quantity, bilirubin levels and risk for lung cancer incidence by smoking status among female participants in the prospective cohort study

	Women (N = 222,758)											
	Total			Total bilirubin level (mg/dL)								
				>0.82			0.61-0.82			<0.61		
	No.	No. of incidence	HR[a] (95% CI)	No.	No. of incidence	HR[a] (95% CI)	No.	No. of incidence	HR[a] (95% CI)	No.	No. of incidence	HR[a] (95% CI)
Non-smoker	188,685	435	1 (Ref)	64,488	127	1 (Ref)	61,534	157	1.35 (1.06-1.73)	62,663	151	1.36 (1.06-1.74)
Total smokers	17,123	37	1.23 (0.87-1.74)	4,891	11	1.55 (0.81-2.98)	5,228	10	1.23 (0.64-2.35)	7,004	16	1.68 (0.99-2.86)
Pack-year												
<30 pack-years	14,662	13	0.61 (0.34-1.08)	4,303	5	0.86 (0.32-2.34)	4,403	2	0.38 (0.09-1.56)	5,956	6	0.97 (0.43-2.22)
≥30 pack-years	1,327	21	2.84 (1.82-4.43)	279	4	3.16 (1.16-8.57)	434	7	3.25 (1.51-6.99)	614	10	3.88 (2.02-7.43)
# of Cigarettes per day												
<10	10,403	16	0.89 (0.53-1.49)	3,237	8	1.75 (0.81-3.75)	3,201	2	0.44 (0.11-1.77)	3,965	6	1.18 (0.52-2.70)
10-19	4,054	12	1.54 (0.86-2.73)	1,004	3	2.03 (0.64-6.40)	1,210	2	0.93 (0.23-3.76)	1,840	7	2.61 (1.21-5.61)
≥20	1,693	8	3.02 (1.50-6.08)	382	0	N/A	475	6	8.91 (3.91-20.29)	836	2	2.03 (0.50-8.21)

[a]Adjusted for age, educational level and body mass index

Table 9. Relationship among smoking quantity, bilirubin levels and lung cancer mortality by smoking status among female participants in the prospective cohort study

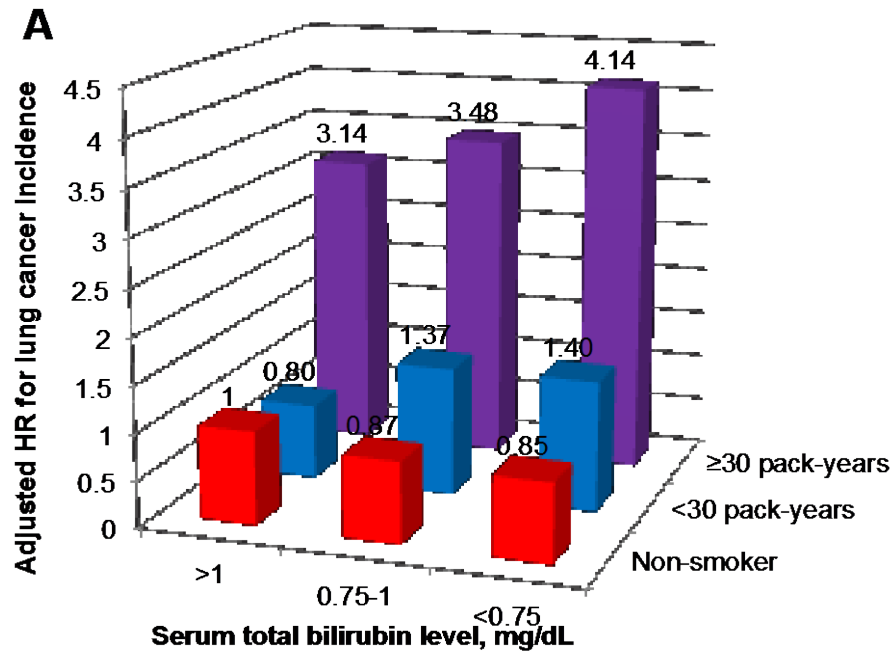
	Women (N = 222,758)											
	Total			Total bilirubin level (mg/dL)								
				>0.82			0.61-0.82			<0.61		
	No.	No. of mortality	HR[a] (95% CI)	No.	No. of mortality	HR[a] (95% CI)	No.	No. of mortality	HR[a] (95% CI)	No.	No. of mortality	HR[a] (95% CI)
Non-smoker	188,685	261	1 (Ref)	64,488	85	1 (Ref)	61,534	90	1.10 (0.82-1.48)	62,663	86	1.10 (0.81-1.49)
Total smokers	17,123	31	1.69 (1.16-2.46)	4,891	9	1.89 (0.94-3.78)	5,228	9	1.50 (0.75-3.00)	7,004	13	1.88 (1.04-3.39)
Pack-year												
<30 pack-years	14,662	8	0.67 (0.33-1.35)	4,303	3	0.91 (0.29-2.88)	4,403	1	0.27 (0.04-1.96)	5,956	4	0.93 (0.34-2.55)
≥30 pack-years	1,327	20	4.00 (2.52-6.33)	279	4	4.05 (1.48-11.05)	434	7	4.15 (1.91-8.99)	614	9	4.44 (2.23-8.85)
# of Cigarettes per day												
<10	10,403	12	1.14 (0.64-2.03)	3,237	6	2.05 (0.89-4.70)	3,201	1	0.30 (0.04-2.14)	3,965	5	1.39 (0.56-3.43)
10-19	4,054	11	2.25 (1.23-4.12)	1,004	3	2.86 (0.90-9.05)	1,210	2	1.28 (0.31-5.20)	1,840	6	3.05 (1.33-7.00)
≥20	1,693	7	4.23 (1.99-8.98)	382	0	N/A	475	5	10.49 (4.25-25.9)	836	2	2.76 (0.68-11.23)

[a]Adjusted for age, educational level and body mass index

Table 10. Lung cancer incidence and mortality rates and adjusted HR per tertile of serum total bilirubin level among the male participants in the prospective cohort study by smoking status and smoking intensity

Characteristics	Men (N=202,902)								
	No. of lung cancer incidence			Adjusted HR[a] (95% CI)			Incidence Rate Per 10 000 Person-Years (95% CI)		
	Total bilirubin level (mg/dL)			Total bilirubin level (mg/dL)			Total bilirubin level (mg/dL)		
	>1	0.75-1	<0.75	>1	0.75-1	<0.75	>1	0.75-1	<0.75
Total	215	270	324	1 (Ref)	1.24 (1.03-1.51)	1.52 (1.26-1.82)	4.27 (3.71-4.90)	5.79 (5.12-6.54)	6.93 (6.20-7.75)
Non-smoker	64	50	42	1 (Ref)	0.86 (0.59-1.27)	0.84 (0.56-1.26)	2.56 (1.98-3.30)	2.46 (1.86-3.27)	2.35 (1.72-3.22)
Total smokers	139	202	262	1 (Ref)	1.29 (1.03-1.62)	1.55 (1.25-1.92)	6.05 (5.09-7.18)	8.37 (7.27-9.64)	9.75 (8.61-11.03)
<30 pack-years	27	47	49	1 (Ref)	1.71 (1.04-2.79)	1.77 (1.09-2.89)	1.57 (1.06-2.33)	2.74 (2.05-3.67)	2.61 (1.95-3.48)
≥30 pack-years	108	145	201	1 (Ref)	1.10 (0.85-1.43)	1.31 (1.03-1.67)	22.41 (18.46-27.21)	24.85 (21.05-29.35)	27.79 (24.11-32.02)
	No. of lung cancer mortality			Adjusted HR[a] (95% CI)			Mortality Rate Per 10 000 Person-Years (95% CI)		
	Total bilirubin level (mg/dL)			Total bilirubin level (mg/dL)			Total bilirubin level (mg/dL)		
	>1	0.75-1	<0.75	>1	0.75-1	<0.75	>1	0.75-1	<0.75
Total	147	214	253	1 (Ref)	1.39 (1.12-1.72)	1.71 (1.39-2.10)	2.70 (2.30-3.17)	4.11 (3.59-4.69)	4.88 (4.32-5.52)
Non-smoker	36	34	28	1 (Ref)	1.03 (0.65-1.65)	0.98 (0.59-1.61)	1.34 (0.96-1.85)	1.49 (1.06-2.09)	1.43 (0.99-2.07)
Total smokers	104	165	209	1 (Ref)	1.37 (1.07-1.76)	1.66 (1.31-2.10)	4.18 (3.45-5.07)	6.17 (5.30-7.19)	7.02 (6.13-8.04)
<30 pack-years	14	39	37	1 (Ref)	2.60 (1.41-4.80)	2.56 (1.38-4.74)	0.77 (0.45-1.30)	2.06 (1.51-2.82)	1.80 (1.31-2.49)
≥30 pack-years	89	119	162	1 (Ref)	1.09 (0.82-1.44)	1.32 (1.01-1.71)	16.54 (13.44-20.36)	18.02 (15.06-21.57)	20.13 (17.26-23.48)

*Adjusted for age, educational level and body mass index



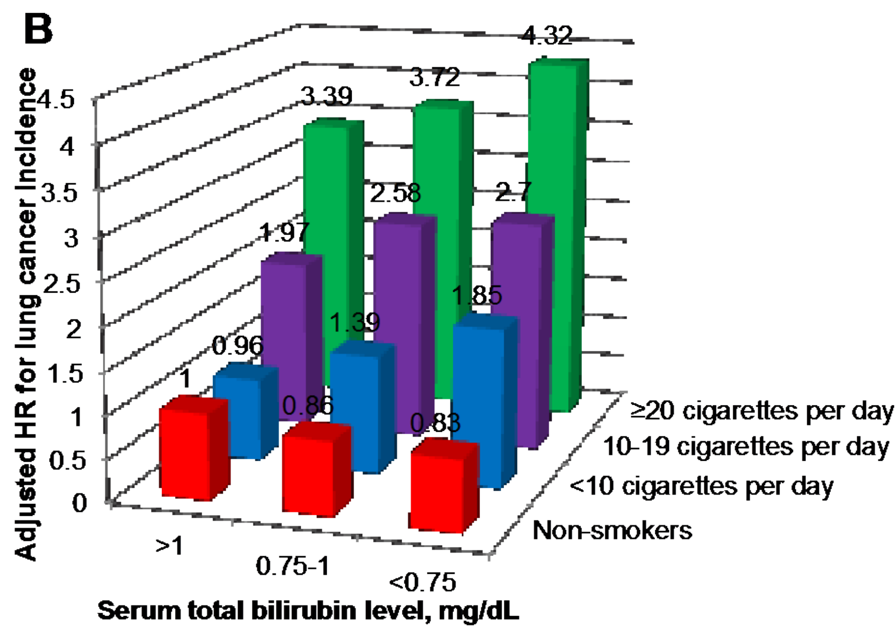
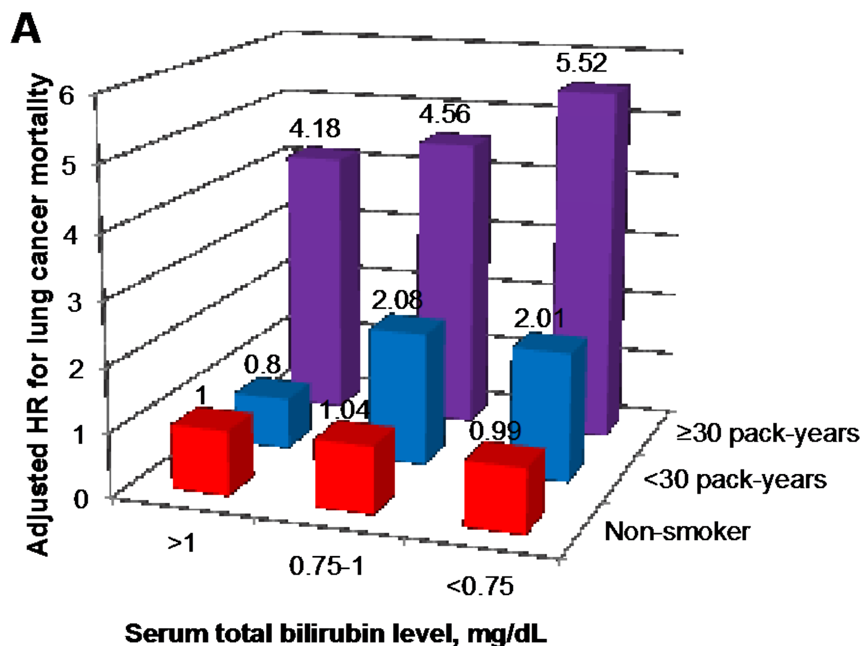


Figure 5. Adjusted HRs of lung cancer incidence stratified by tertiles of serum total bilirubin level and smoking pack-years (A) or cigarettes per day (B) among male participants of the prospective cohort study.



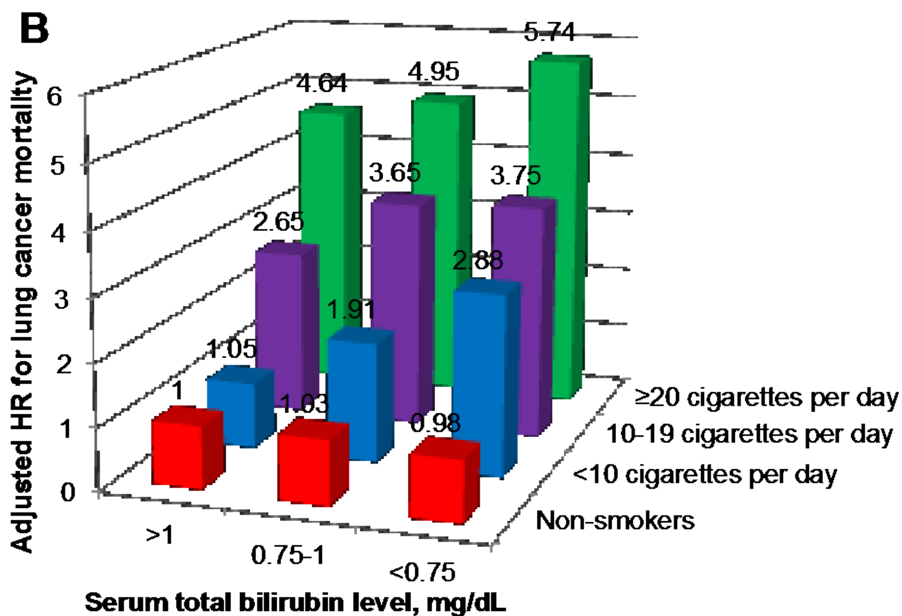
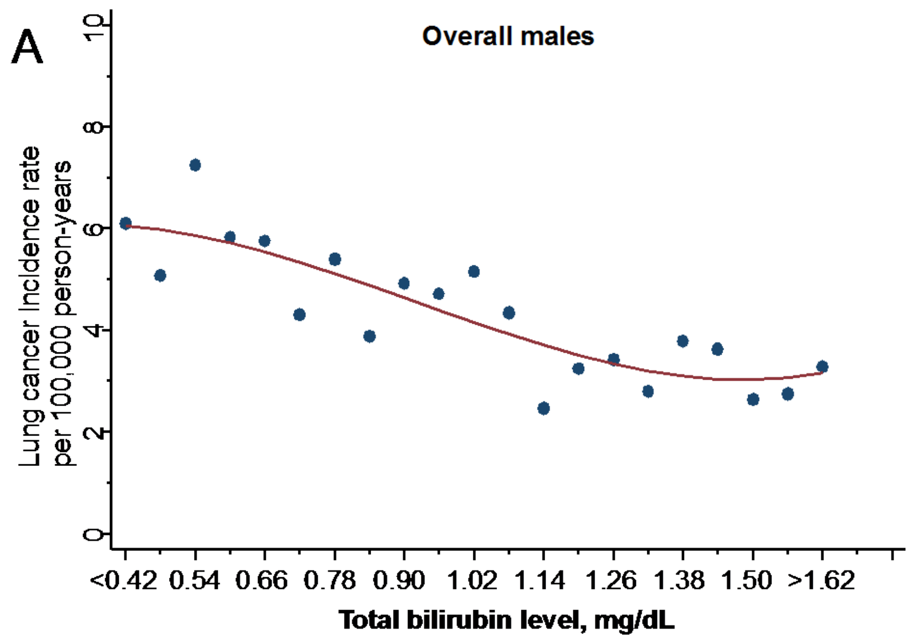


Figure 6. Adjusted HRs of lung cancer mortality stratified by tertiles of serum total bilirubin level and smoking pack-years (A) or cigarettes per day (B) among male participants of the prospective cohort study.



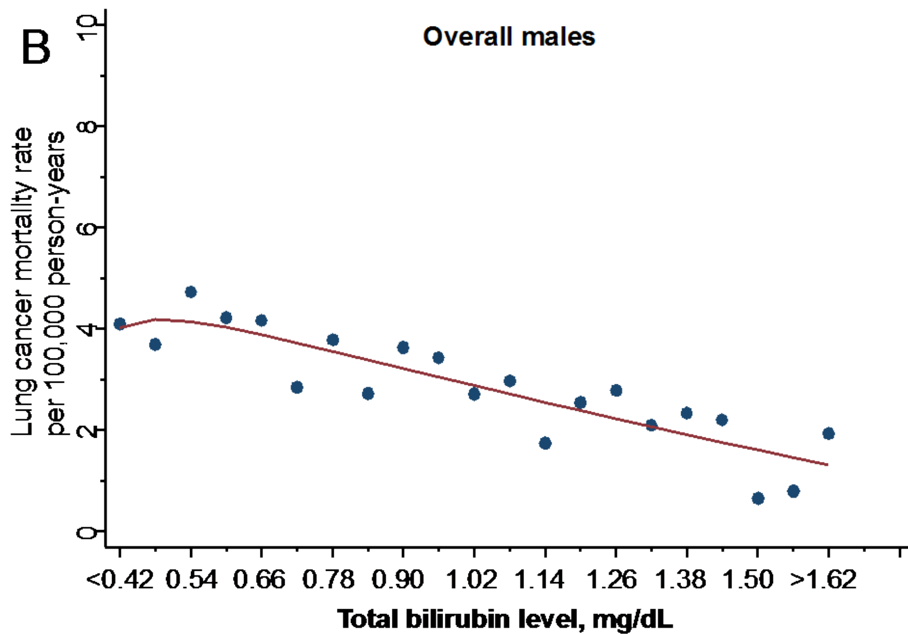
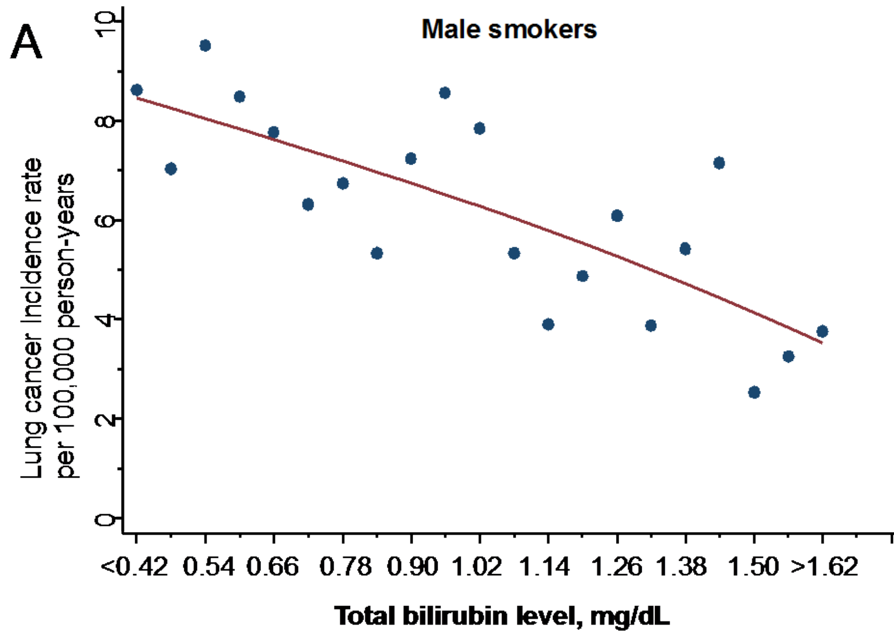


Figure 7. Serum total bilirubin levels and lung cancer incidence rates (A) and mortality rates (B) in overall males of the prospective cohort study.

3.1.5. The ability of bilirubin in identifying smokers with higher risk of lung cancer

We then assessed the association between bilirubin levels and lung cancer incidence or mortality rate stratified by smoking status. Among females, neither non-smokers or smokers showed significant association, as only 17,123 (8.3%) participants were smokers, and there were only 37 lung cancer cases among them. Among males, the association was only present in smokers. Compared to smokers with bilirubin levels in the highest tertile, smokers with bilirubin levels in the middle and lowest tertiles had significantly increased lung cancer risk (HRs, 1.29 and 1.55) and mortality (HRs, 1.37 and 1.66) (Table 10). The risk appeared to be stronger in light smokers: the HRs for the lowest tertile of bilirubin compared to the highest tertile were 1.77 for incidence and 2.56 for mortality in smokers of <30 pack years and 1.31 for incidence and 1.32 for mortality in smokers of ≥30 pack years, respectively (Table 10). We also plotted the lung cancer incidence and mortality rates against subgroups of bilirubin levels in smokers and introduced a best-fit model. Subjects with bilirubin levels <0.42 mg/dL showed more than two folds increase in both lung cancer incidence rate (8.62 vs 3.76 per 100,000 person-years, Figure 7A) and mortality rate (6.27 vs 3.05 per 100,000 person-years, Figure 7B) compared to the subgroup with bilirubin levels >1.62 mg/dL. The associations between bilirubin level as a continuous variable and both lung cancer incidence and lung cancer mortality were statistically significant in males. The logistic regression model showed a 5% (95% CI, 3%-8%, $P < 0.001$) increase in lung cancer incidence and 6% (95% CI, 3%-9%, $P < 0.001$) increase in lung cancer mortality per 0.1 mg/dL decrease in bilirubin, after adjusting for age,

BMI and educational level. The interaction between bilirubin levels and smoking status was not significant (log-likelihood test, $P>0.05$).



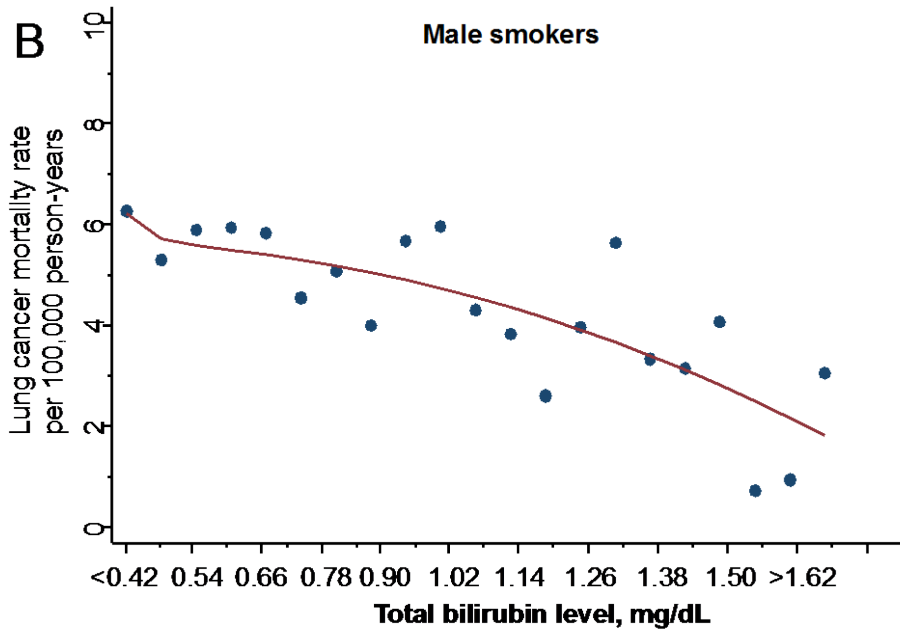


Figure 8. Serum total bilirubin levels and lung cancer incidence rates (A) and mortality rates (B) in male smokers of the prospective cohort study.

3.2. Discussion

The purpose of this study is to identify biomarkers among serum metabolites for lung cancer early diagnosis and to assist in identifying high-risk individuals for lung cancer development. To our knowledge, this is the first study that identified serum bilirubin as a lung cancer biomarker through unbiased global metabolomic profiling. Furthermore, lower bilirubin was validated to confer an increased risk for both lung cancer incidence and lung cancer specific mortality in a large prospective cohort. When stratifying by smoking status, the association was significant only in male ever-smokers. Through this multi-stage study, we have identified and validated serum bilirubin as a risk predictor for lung cancer incidence as well as mortality in male smokers. While smoking is a strong risk factor for lung cancer and shows a dose-response relationship, the smoking-related risk is particularly high among individuals with low levels of serum bilirubin, a 55% increase among those with bilirubin <0.75 mg/dL. Smokers with ≥30 pack years had a 4-fold increase in lung cancer risk, and within this group, those with bilirubin of <0.75 mg/dL had a 31% higher risk compared to those of >1 mg/dL. The potential of using serum bilirubin to identify smokers at particularly high-risk for lung cancer over and above the risk associated with heavy smoking is an important observation. The inverse relationship between bilirubin levels and lung cancer can be translated into a 5% increase in lung cancer risk and a 6% increase in lung cancer mortality for each 0.1 mg/dL decrease in bilirubin levels. In most clinical settings, emphasis is placed on elevated bilirubin for diagnosis of liver diseases; hence low values of bilirubin are

generally ignored. Making use of low serum bilirubin values to counsel heavy smokers who are at particularly high risk for lung cancer about smoking cessation can be carried out easily in many clinic settings.

Our findings may also have implications for the LDCT screening for lung cancer. It has been reported that LDCT screening prevented the most deaths from lung cancer among participants with the highest risk for lung cancer deaths − 60% of participants at the highest risk accounted for 88% of prevented lung-cancer deaths (7). Based on our results, male smokers with bilirubin level <0.75 mg/dL have a 66% increased risk for lung cancer mortality compared to those with bilirubin level >1 mg/dL, and for heavy smokers of ≥30 pack-years, the hazard ratio is smaller, but still significant (HR = 1.32, $P < 0.001$). Consideration of bilirubin levels might improve identifying participants with the highest risk for lung cancer mortality who would benefit the most from the screening, and help improve the specificity of LDCT screening. Furthermore, bilirubin results could be used to target and motivate both light and heavy smokers for smoking cessation. Indeed, the ability of low bilirubin in predicting high risk of lung cancer was not limited to smokers with ≥30 pack-years in our study. The relationship was seen for all smokers, regardless of pack years of smoking.

Elevated levels of serum bilirubin have been associated with a lower risk of respiratory diseases and lung cancer (23, 24). The mechanism of this association was credited to the antioxidant and anti-inflammatory properties of bilirubin. One study examined the association between serum bilirubin levels and all-cause, cardiovascular, and cancer mortality in 10,000 Belgian individuals who had been

followed up for 10 years (40). They found that higher bilirubin within the normal range was associated with lower cancer mortality only in men. Another recent large UK cohort study showed higher bilirubin levels within the normal-range were associated with lower risks of respiratory diseases including chronic obstructive pulmonary disease (COPD), lung cancer and all-cause mortality (41). While that study examined risks of respiratory diseases and all-cause mortality, our study focused on lung cancer incidence and lung cancer specific mortality. Interestingly, the UK study found significant association between bilirubin and lung cancer incidence in both men and women, while in our study only men exhibited significant association. Similarly, the Belgian study also found significant inverse association between bilirubin and cancer mortality only in men. It should be pointed out that percentage (58%) of female ever-smokers in the UK study was much higher than those of our study (8.3%) and the Belgian study (~25%). It is possible that the associations found in both men and women in the UK study reflect largely the associations between bilirubin levels and lung cancer incidence in male and female smokers.

Neither the UK nor the Belgian study performed stratified analysis of bilirubin and lung cancer risk by smoking status. Smoking significantly reduces serum bilirubin level and simply adjusting for smoking status during the analyses may not be sufficient to account for the effect of smoking on the association between bilirubin and lung cancer incidence and mortality, as adjusting for smoking in our study did not remove the association between bilirubin levels and lung cancer incidence and mortality in men. Only by examining the association in

smokers and non-smokers separately that we were able to detect the bilirubin-lung cancer associations in the subgroup of male ever-smokers. Notably, the association between bilirubin levels and lung cancer incidence and mortality rate in male ever-smokers remained strong in a series of sensitivity analyses in our study, including restricting bilirubin levels within the normal range, excluding participants who had abnormal liver enzyme assays or blood counts, or adjusting for more variables (drinking status, physical activity and systolic blood pressure) (data not shown). As bilirubin is a commonly ordered laboratory test, uncovering this potentially protective relationship is intriguing. This study, while in line with the reported conclusion, is the first to study the role of bilirubin as a risk factor for lung cancer mortality, to focus on the analysis in smokers in detail, and to quantify the hazards of low bilirubin.

We conducted a series of sensitivity analyses to strengthen our conclusion. We excluded participants with lung cancer diagnosed within 3 years of cohort enrollment. We restricted bilirubin levels within normal range, excluding participants with abnormal liver enzymes or blood counts. Additional variables (drinking status, physical activity and systolic blood pressure) were adjusted in the multivariable models. Results essentially remained unchanged after all of the above sensitivity analyses were carried out.

Recently, several research groups had applied metabolomic profiling of serum/plasma to unveil metabolic alterations associated with lung cancer. *Hori* and colleagues performed metabolomic profiling of serum samples in a Japanese case-control study using gas chromatography/mass spectrometry (GC/MS),

which detected a total of only 58 metabolites in serum (45). Among these, the levels of 23 metabolites were altered in lung cancer patients compared to healthy controls (45). In another study conducted in Japan, *Maeda et al.* studied 21 plasma amino acids in NSCLC patients by liquid chromatography/mass spectrometry (LC/MS) and showed that differences in the amino acid profiles could be used for screening NSCLC (35). *Jordan* and colleagues used nuclear magnetic resonance (NMR) to measure 21 metabolites and showed the potential of serum metabolomics to differentiate between lung cancer subtypes and between patients and controls (34). However, these studies were limited by the small number of metabolites detected. Our global unbiased metabolomic profiling approach identified 403 known metabolites from different stages of lung cancer, yielding a comprehensive picture of the metabolic profile changes associated with cancer progression. Validated with two additional study sets, bilirubin was found and confirmed as the most significant biomarker for lung cancer, which was further validated prospectively in a large cohort.

A few potential limitations should be considered in the interpretation of our findings. First, while we observed significant inverse associations between serum bilirubin levels and lung cancer in male smokers, the associations were not statistically significant in female smokers, which was most likely due to the lack of power resulting from a small number of female smokers (8.3% of total females) and very few number of lung cancer cases (n=37) among them. Second, although we observed an inverse relationship between bilirubin levels and lung cancer risk, the causality of the association remains unclear. Low bilirubin level could be a

consequence of cancer rather than a predisposing factor. It is noteworthy that the significant risk remained after we excluded lung cancer occurring within three years of the bilirubin tests. Third, only the bilirubin data at the time of enrollment were analyzed. In a subset of subjects that had two bilirubin tests performed longitudinally, we found highly correlative data, implying the stability of total bilirubin results over time.

In summary, low levels of serum bilirubin are associated with higher risk for lung cancer incidence and mortality in male smokers and can be used to identify higher risk smokers for lung cancer development and mortality. Future prospective studies that incorporate this variable into NLST selection criteria to fully assess its potential use for LDCT screening are warranted.

3.3. Clinical Laboratory Tests and Lung Cancer Clinical Outcome

3.3.1. Patient characteristics

This study included 2,675 NSCLC patients, among whom 623 (23.3%) were early stage and 2,052 (77.6%) were advanced stage patients (Table 11). Among early stage NSCLC patients, 313 (50.2%) were men and 310 (49.8%) were women, with a mean age of 66.8 years. The median survival time (MST) was 62.7 months, with a median follow-up time (MFT) of 36.8 months. Over 90% of the patients had a history of smoking, with 316 (50.7%) being former smokers and 246 (39.5%) being current smokers or having quit within one year prior to diagnosis. Three Hundred and ninety-six (63.6%) of them had stage I NSCLC, while the remaining 227 (36.4%) had stage II disease. Among them, over half (51.8%) had adenocarcinoma and 33.4% had squamous cell carcinoma. Among 2,052 advanced stage patients, 1,125 (54.8%) were men and 927 (45.2%) were women, with a mean age of 61.7 years. The MST and MFT were 13.8 and 13.0 months respectively. Eight hundred and nineteen (39.9%) of them were former smokers and 880 (42.9%) were current smokers or had quit within one year prior to diagnosis. There were 477 (23.2%) stage IIIA, 277 (13.5%) stage IIIB and 1,298 (63.3%) stage IV patients. Patients with adenocarcinoma or squamous cancer carcinoma represented 55.8% (n=1,146) or 22.1% (n=454) of all advanced stage patients, respectively. Among them, 1,192 (58.1%) received definitive chemotherapy, 978 of which were platinum-based. There was no statistically significant association between the routine pre-treatment laboratory tests investigated and recurrence or survival of early stage NSCLC patients (data not

shown), therefore, only results of advanced stage NSCLC patients are presented in details.

Table 11. Host characteristics of overall, early stage and late stage NSCLC patients

Variables	All Stage	Early Stage	Late Stage
	N (%)		
Mean age, y (SD)	62.9 (11.0)	66.8 (10.0)	61.7 (11.0)
Sex			
Male	1438 (53.8)	313 (50.2)	1125 (54.8)
Female	1237 (46.2)	310 (49.8)	927 (45.2)
Race			
White	2201 (82.3)	535 (85.9)	1666 (81.2)
Hispanic	107 (4.0)	19 (3.0)	88 (4.3)
Black	367 (13.7)	69 (11.1)	298 (14.5)
Smoking status			
Never	413 (15.4)	61 (9.8)	352 (17.2)
Former	1135 (42.4)	316 (50.7)	819 (39.9)
Current	1126 (42.1)	246 (39.5)	880 (42.9)
Clinical stage			
Stage I	396 (14.8)	396 (63.6)	
Stage II	227 (8.5)	227 (36.4)	
Stage IIIA	477 (17.8)		477 (23.2)
Stage IIIB	277 (10.4)		277 (13.5)
Stage IV	1298 (48.5)		1298 (63.3)
Histology			
Adenocarcinoma	1469 (54.9)	323 (51.8)	1146 (55.8)
Squamous	662 (24.7)	208 (33.4)	454 (22.1)
Others	544 (20.3)	92 (14.8)	452 (22.0)
Performance Status (ECOG score)			
0	539 (20.1)	153 (24.6)	386 (18.8)
1	1271 (47.5)	251 (40.3)	1020 (49.7)
2-4	421 (15.7)	63 (10.1)	358 (17.4)
Unknown	444 (16.6)	156 (25.0)	288 (14.0)
Histology grade			
Well differentiated	130 (4.9)	64 (10.3)	66 (3.2)
Moderately differentiated	474 (17.7)	180 (28.9)	294 (14.3)
Poorly differentiated and undifferentiated	1128 (42.2)	258 (41.4)	870 (42.4)
Unknown	943 (35.3)	121 (19.4)	822 (40.1)

No. of metastasis			
0	1537 (57.5)	622 (99.8)	915 (44.6)
1	742 (27.7)	1 (0.2)	741 (36.1)
2	262 (9.8)	0	262 (12.8)
≥3	134 (5.0)	0	134 (6.5)
Surgery			
No	2080 (77.8)	200 (32.1)	1880 (91.6)
Yes	595 (22.2)	423 (67.9)	172 (8.4)
Radiation			
No	1941 (72.6)	472 (75.8)	1469 (71.6)
Yes	734 (27.4)	151 (24.2)	583 (28.4)
Chemotherapy			
No	1297 (48.5)	437 (70.1)	860 (41.9)
Yes	1378 (51.5)	186 (29.9)	1192 (58.1)
Chemoradiation			
No	2174 (81.3)	577 (92.6)	1597 (77.8)
Yes	501 (18.7)	46 (7.4)	455 (22.2)
MST, mo	18.1	62.7	13.8
MFT, mo	16	36.8	13

Abbreviations: MST, median survival time; MFT, median follow-up time

3.3.2. Association of pre-treatment laboratory test levels with survival of advanced stage NSCLC patients

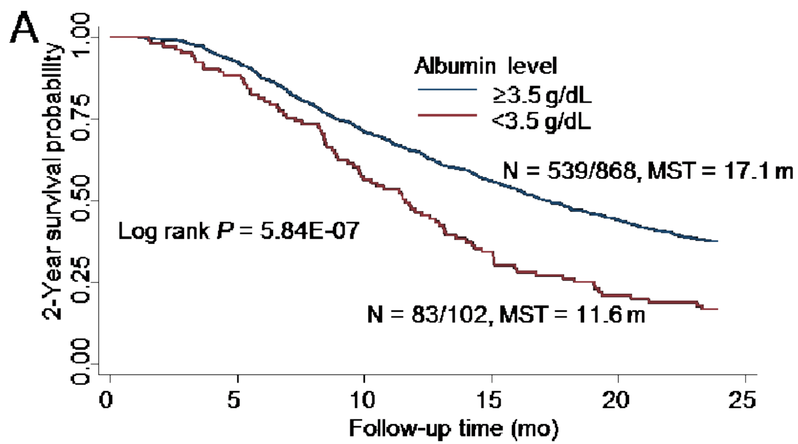
Among the seven routine clinical pre-treatment serum laboratory tests analyzed, low albumin, high ALP and high LDH were found statistically significantly associated with both 2-year and 5-year survival of advanced stage NSCLC patients subsequently treated with platinum-based chemotherapy (Table 12), after adjusting for other factors, i.e. age, gender, ethnicity, smoking status, clinical stage and treatment regimen. Lower than normal (<3.5 g/dL) serum pre-treatment albumin levels were associated with 47% and 39% increase risk of death at 2-year and 5-year survival of advanced stage NSCLC patients treated with platinum-based chemotherapy. This corresponded to a MST of 11.6 months among those with low serum albumin levels and a MST of 17.1 months among those with normal serum albumin levels (log-rank test $P < 0.01$) (Figure 8A and 9A). Similarly, advanced stage NSCLC patients with elevated and higher than normal serum ALP levels (>126 IU/L) or LDH levels (>618 IU/L) were at 31% or 60% increased risk of death at two years than those with normal serum ALP or LDH level, respectively. Similar increases were also found for 5-year survival among those patients (Table 12). MST for patients with normal serum ALP levels was 17.4 months compared to 12.6 months for those with higher than normal ALP levels (log-rank test $P < 0.0001$) (Figure 8B and 9B). Similarly, MST for patients with normal serum LDH levels was 18.0 months compared to 12.3 months for those with higher than normal LDH levels (log-rank test $P < 0.0001$) (Figure 8C and 9C).

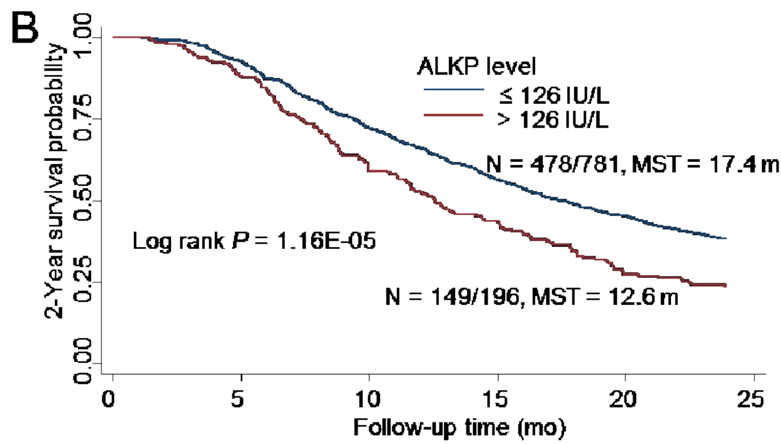
Table 12. Laboratory test levels associated with 2-year and 5-year survival in advanced NSCLC patients treated with platinum-based chemotherapy

Lab Test	2-year survival						5-year survival					
	Alive, N (%)	Dead, N (%)	Adjusted HR[a] (95% CI)	*P*	MST	Log-rank *P*	Alive, N (%)	Dead, N (%)	Adjusted HR[a] (95% CI)	*P*	MST	Log-rank *P*
Albumin (g/dL)												
≥3.5	328 (37.88)	538 (62.12)	1 (reference)		17.1		170 (19.63)	696 (80.37)	1 (reference)		17.1	
<3.5	19 (18.63)	83 (81.37)	1.47 (1.15-1.87)	0.002	11.6	5.75E-07	9 (8.82)	93 (91.18)	1.39 (1.11-1.75)	0.004	11.6	1.31E-06
ALP (IU/L)												
38-126	302 (38.77)	477 (61.23)	1 (reference)		17.4		154 (19.77)	625 (80.23)	1 (reference)		17.4	
>126	47 (23.98)	149 (76.02)	1.31 (1.08-1.58)	0.006	12.6	1.16E-05	25 (12.76)	171 (87.24)	1.24 (1.04-1.48)	0.01	12.6	6.06E-05
LDH (IU/L)												
≤618	299 (39.45)	459 (60.55)	1 (reference)		18.0		155 (20.45)	603 (79.55)	1 (reference)		18.0	
>618	47 (21.96)	167 (78.04)	1.60 (1.34-1.92)	3.51E-07	12.3	5.29E-09	24 (11.21)	190 (88.79)	1.56 (1.32-1.85)	1.74E-07	12.3	2.58E-08

Abbreviations: ALP, ; LDH,

[a] Adjusted for age, gender, ethnicity, smoking status, clinical stage, performance status and treatment with radiation.





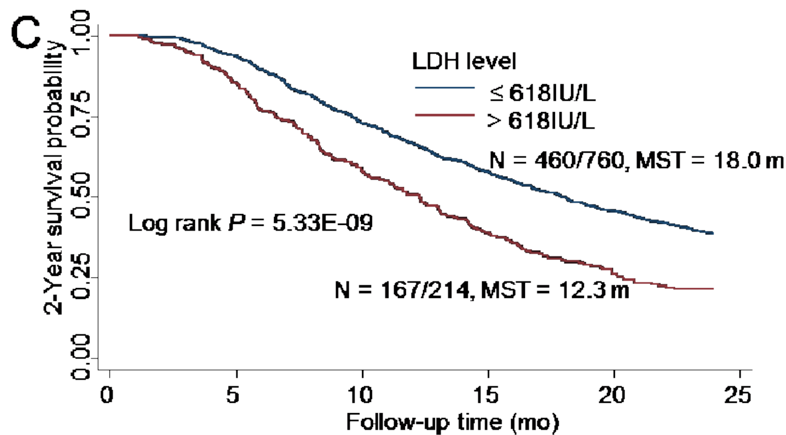
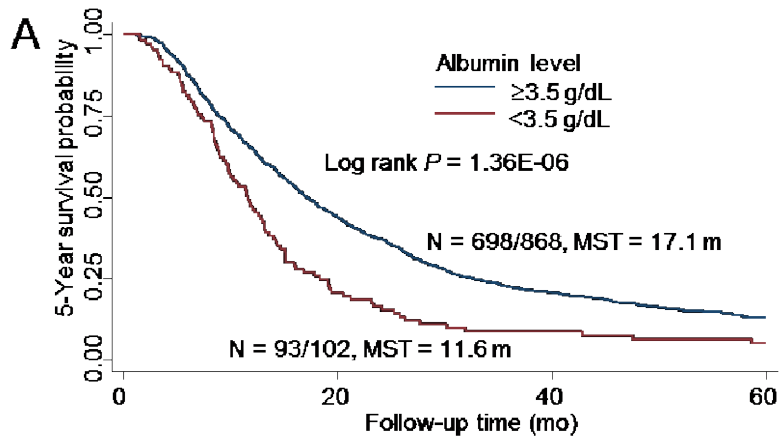
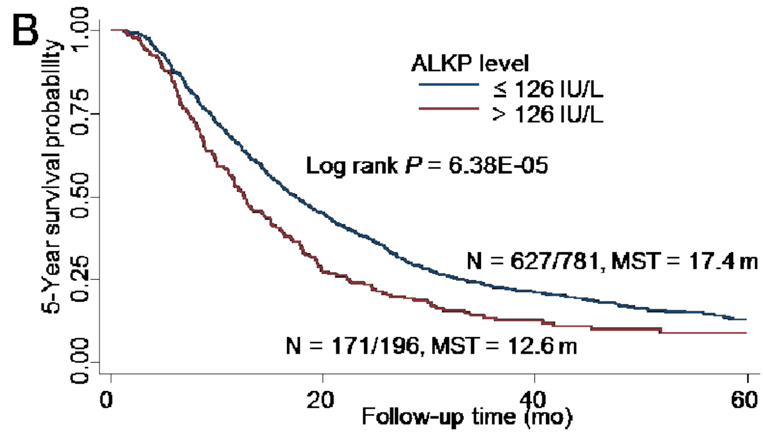


Figure 9. Kaplan-Meier 2-year survival curves for patients with advanced NSCLC treated with platinum-based chemotherapy grouped by individual lab test level: (A) Albumin, (B) ALP, (C) LDH.





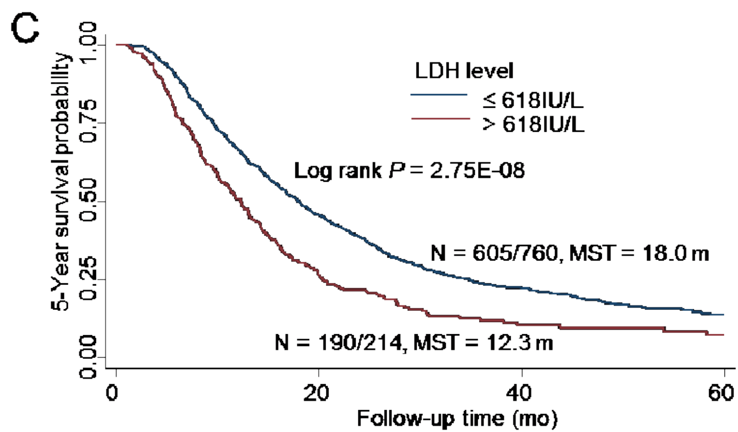


Figure 10. Kaplan-Meier 5-year survival curves for patients with advanced NSCLC treated with platinum-based chemotherapy grouped by individual lab test level: (A) Albumin, (B) ALP, (C) LDH.

3.3.3. Cumulative effects of unfavorable laboratory test levels

We conducted joint analysis to investigate whether advanced stage NSCLC patients with more unfavorable pre-treatment laboratory test levels were more likely to suffer from death, and the results were statistically significant (Table 13). Compared to those with none of the three adverse laboratory test levels for albumin, ALP and LDH, advanced stage NSCLC patients who had one adverse laboratory test level had a 34% increase in risk of death at 2 years (HR = 1.34, 95% CI: 1.12-1.60, $P < 0.001$). Those with more than one adverse laboratory test levels were two folds more likely to die at 2 years (HR = 2.00, 95% CI: 1.57-2.56, $P < 0.0001$). Similar results were also found for 5-year survival of these patients. The corresponding MSTs were 19.6, 14.0 and 10.0 months for advanced stage NSCLC patients with 0, 1 and ≥2 unfavorable pre-treatment laboratory test levels (log-rank test $P < 0.0001$) (Figure 10A and 11A).

Table 13. Cumulative effect of unfavorable laboratory test levels associated with 2-year and 5-year survival in advanced NSCLC patients treated with platinum-based chemotherapy

No. of Adverse Lab Test	2-year survival						5-year survival					
	Alive, N (%)	Dead, N (%)	Adjusted HR[a] (95% CI)	*P*	MST	Log-rank *P*	Alive, N (%)	Dead, N (%)	Adjusted HR[a] (95% CI)	*P*	MST	Log-rank *P*
0	247 (43.33)	323 (56.67)	1 (reference)		19.6		130 (22.81)	440 (77.19)	1 (reference)		19.6	
1	82 (28.28)	208 (71.72)	1.34 (1.12-1.60)	1.29E-03	14.0		39 (13.45)	251 (86.55)	1.32 (1.12-1.55)	6.97E-04	14.0	
≥2	14 (13.73)	88 (86.27)	2.00 (1.57-2.56)	2.68E-08	10.0	1.90E-14	9 (8.82)	93 (91.18)	1.80 (1.43-2.27)	7.34E-07	10.0	1.17E-12

[a] Adjusted for age, gender, ethnicity, smoking status, clinical stage, performance status and treatment with radiation.

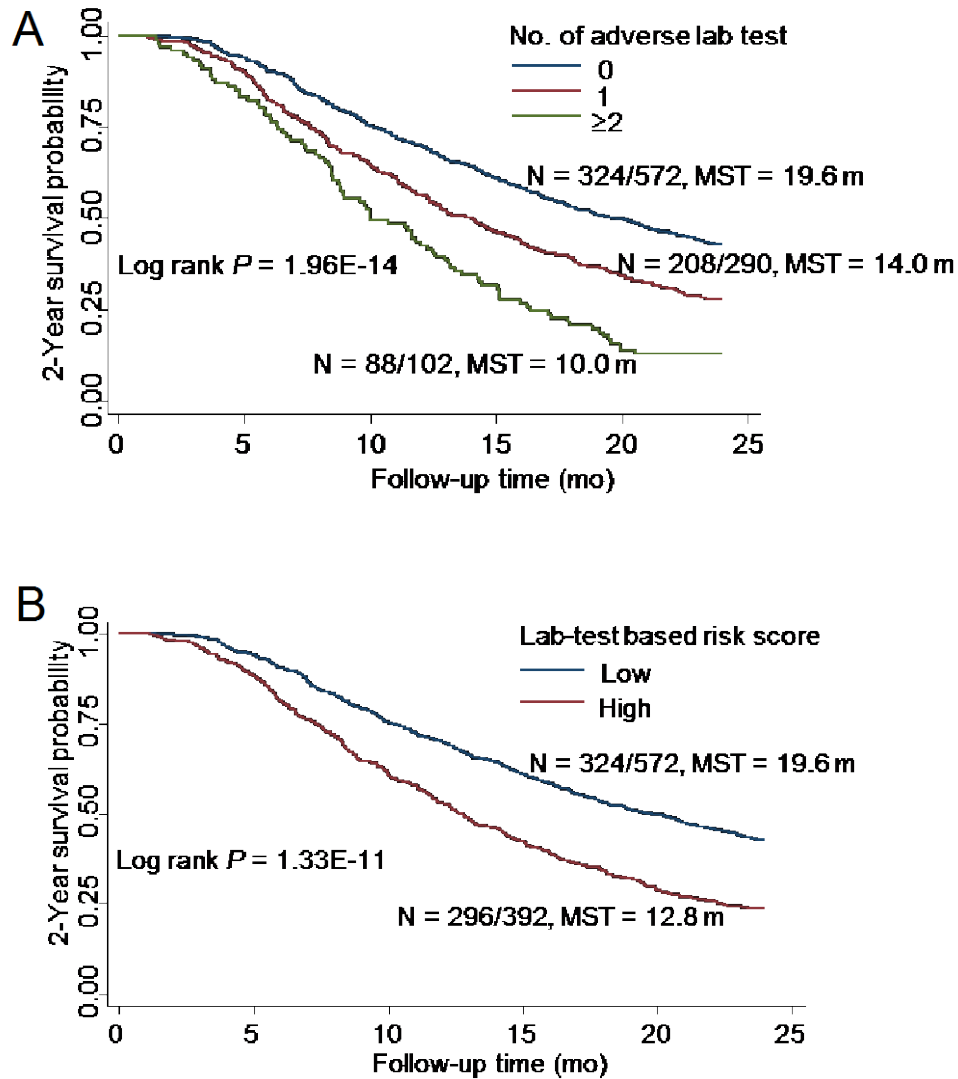


Figure 11. Kaplan-Meier 2-year survival curves for patients with advanced NSCLC treated with platinum-based chemotherapy grouped by cumulative effect of adverse lab test (A) and lab-test based risk score (B).

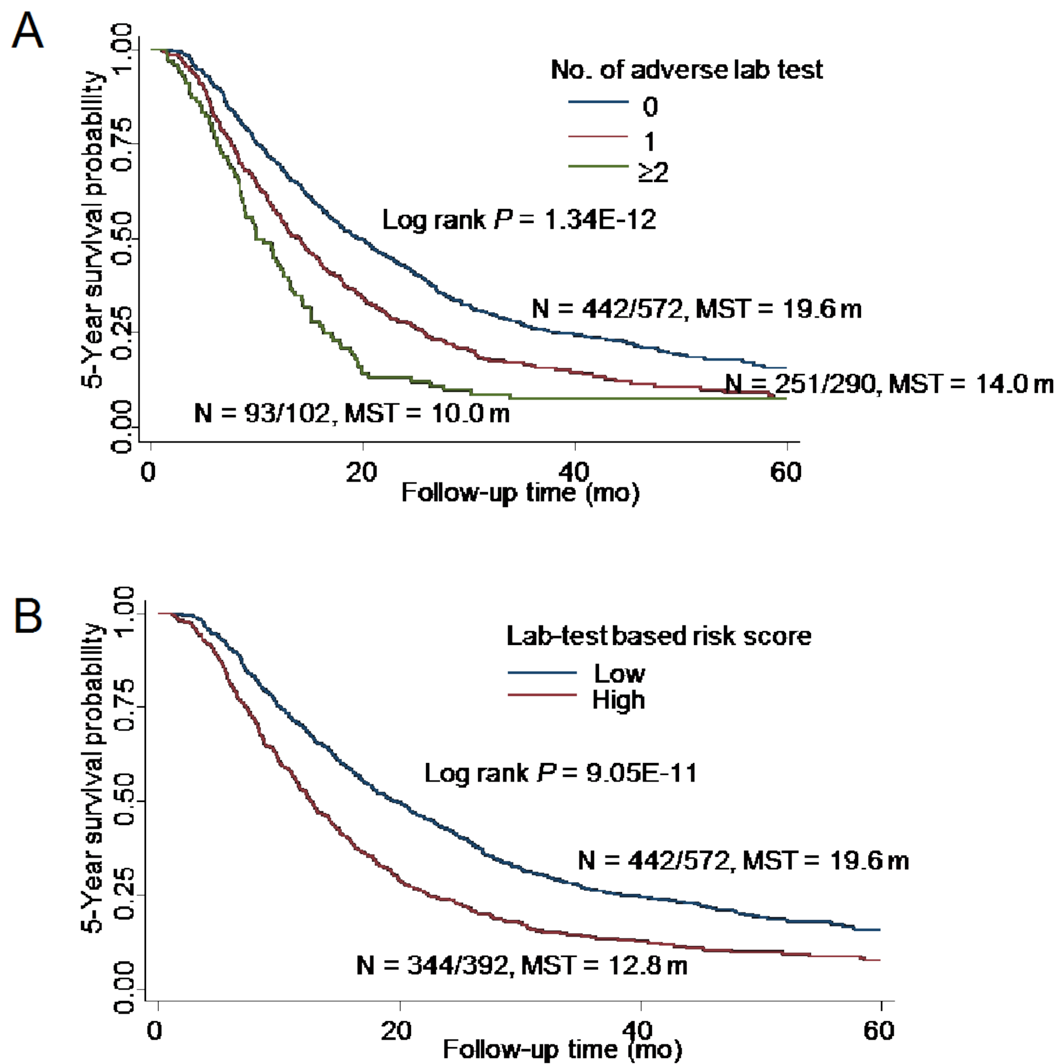


Figure 12. Kaplan-Meier 5-year survival curves for patients with advanced NSCLC treated with platinum-based chemotherapy grouped by cumulative effect of adverse lab test (A) and lab-test based risk score (B).

3.3.4. Lab-test based risk score associated with survival of advanced NSCLC patients

We generated a risk score based on the three pre-treatment serum laboratory test levels associated with survival of advanced stage NSCLC patients treated with platinum-based chemotherapy. Compared to those with a low-risk score, patients with a high-risk score had 49% increased risk of death at two years (HR = 1.49, 95% CI: 1.26-1.75, $P < 0.0001$). Similar results were found for 5-year survival of those patients (Table 14). Kaplan-Meier survival analysis showed that the corresponding MST was 19.6 months for those with a low-risk score and 12.8 months for those with a high-risk score (log-rank test $P < 0.0001$) (Figure 10B and 11B).

Table 14. Lab-test based risk score associated with 2-year and 5-year survival in advanced NSCLC patients treated with platinum-based chemotherapy

	2-year survival						5-year survival					
Risk Score	Alive, N (%)	Dead, N (%)	Adjusted HR[a] (95% CI)	*P*	MST	Log-rank *P*	Alive, N (%)	Dead, N (%)	Adjusted HR[a] (95% CI)	*P*	MST	Log-rank *P*
Low	247 (43.33)	323 (56.67)	1 (reference)		19.6		130 (22.81)	440 (77.19)	1 (reference)		19.6	
High	96 (24.49)	296 (75.51)	1.49 (1.26-1.75)	1.71E-06	12.8	1.30E-11	48 (12.24)	344 (87.76)	1.42 (1.23-1.64)	2.28E-06	12.8	7.92E-11

[a] Adjusted for age, gender, ethnicity, smoking status, clinical stage, performance status and treatment with radiation.

3.4. Discussion

In this study, we evaluated seven serum pre-treatment clinical laboratory tests indicative of overall metabolic status and investigated their prognostic values in 622 early stage and 2,053 advanced stage NSCLC patients. While we did not observe any significant associations between any of these laboratory test levels and recurrence or survival of early stage NSCLC patients, we identified and confirmed three serum pre-treatment clinical laboratory test levels (lower than normal levels of albumin, higher than normal levels of ALP and LDH) as prognostic factors for advanced stage NSCLC patients treated with platinum-based chemotherapy. Not only was each of the three clinical laboratory tests significantly associated with overall survival among patients with advanced stage NSCLC treated with platinum-based chemotherapy, there was also a significant cumulative effect – those with two or more than two adverse laboratory test levels were two-fold more likely to die at two years compared to those with none of the adverse laboratory test levels, corresponding to a MST of 10.0 months and 19.6 months, respectively.

Serum albumin level has been used as a biomarker for long-term nutritional status (46, 47). Serum albumin is produced in the liver and is the most abundant protein in human plasma/serum. Its major functions include maintaining the intravascular oncotic pressure of the blood compartment, transporting molecules of low water solubility and serving as a free radical scavenger (48). In adults, normal range of serum albumin is 3.5 to 5.0 g/dL. Low serum albumin level (<3.5 g/dL in adults), also known as hypoalbuminemia, may be caused by liver

diseases, malabsorption, pregnancy, genetic variations, as well as cancers. Albumin level is often used as an indicator for liver disease or acute infection (49, 50). It's also closely correlated with malnutrition, which commonly occurs in cancer patients resulting from the host response to the tumor as well as to therapeutic treatments. Studies have investigated the association of malnutrition with several clinical outcomes among cancer patients, including quality of life (QOL), response to treatment as well as overall survival. Malnutrition has been found to be associated with worse overall clinical outcomes, which includes worse QOL and shorter overall survival among cancer patients. In addition, hypoalbuminemia has also been shown to be associated with worse clinical outcomes in cancer patients. Studies have shown that low serum albumin level is a negative prognostic factor in several kinds of malignancies, including colorectal cancer (51, 52), gastric cancer (53), pancreatic cancer and breast cancer (54).

Alkaline phosphatase (ALP) is a ubiquitous enzyme present in all human body tissues, catalyzing hydrolysis of phosphate esters and removing phosphate groups (also called "dephosphorylation") from different types of molecules in an alkaline environment (55). There are several different ALP isoenzymes, depending on where it is produced in the body, e.g., hepatic ALP, placenta ALP or bone ALP isoenzyme, etc. Among these, ALP levels are highest in liver and bone tissues (56). Serum total ALP is often measured as part of routine liver panel tests or when presented with symptoms of liver or bone disorders to screen for a variety of liver or bone disorders, including extrahepatic bile obstruction, intrahepatic cholestasis, hepatitis, and bone disorders, e.g. Paget's disease, etc (56). Serum

total ALP levels have been shown to be elevated in several metastatic malignancies (57, 58). Of note, bone-specific alkaline phosphatase (bALP), which is produced by osteoblasts, is an indicator of osteoblastic activity and a marker of bone formation. bALP levels have been shown to be significantly elevated in patients with solid malignancies experiencing metastases to the bone (57-60). Several studies have examined bALP as a potential marker to detect occult bone metastases in patients with lung cancer (61-63). In NSCLC, patients with skeletal metastases had significantly higher levels of bALP than those without skeletal metastases (64). Moreover, Brown et. al showed that among 238 patients with NSCLC or other solid tumors, high baseline serum bALP levels were significantly associated with an increased risk of adverse outcomes, including skeletal-related events (SREs), disease progression, and overall survival. Patients with high bALP levels had a 1.5-fold increased risk of death (RR = 1.53, 95% CI: 1.15-2.03, $P = 0.003$) (57). In their study, serum bALP levels were assessed using a chemical inhibition and differential inactivation assay (57, 65). While the results are consistent with our finding and have a similar magnitude of relative risk, their analysis was conducted on pooled NSCLC and other solid tumors together, presumably because of the small sample size of NSCLC patients (N = 115). With 978 advanced stage NSCLC patients treated with platinum-based chemotherapy in our study, our results may reflect a more accurate risk estimate and are more statistically significant (HR = 1.38, $P < 0.001$). However, in our study, we didn’t measure the levels of different ALP isoenzymes, and rather only assessed overall serum ALP levels.

Lactate dehydrogenase (LDH) is a glycolytic enzyme involved in the conversion of pyruvate, which is the final product of the glycolysis pathway, and lactate. LDH is ubiquitously expressed in human body cells and serves as a general marker for tissue injuries since it is commonly released from the cells to blood after tissue damage (36). Elevated serum LDH levels may be caused by many pathologic conditions, including hemolytic anemia, myocardial infarction, liver disease, pulmonary disease, kidney disease and many types of malignancies, as cancer cells have a higher turnover rate (36). In addition, cancer cells have been long known to reprogram their energy metabolism towards glycolysis even under aerobic conditions, which was first observed and reported by Otto Warburg in the 1930s (66, 67). Elevated serum LDH levels have been shown to be associated with tumor mass, tumor aggressiveness and unfavorable prognosis in several malignancies (68-70). Moreover, several studies have identified serum LDH levels as a significant variable associated with survival in patients with NSCLC as well (71, 72).

Our study adds significant value to the current literature body. First, this is one of the largest studies investigating the prognostic values of pre-treatment clinical laboratory test levels on NSCLC patients' outcomes, thus statistically powered to reveal significant associations. Moreover, of the three laboratory tests identified (albumin, ALP and LDH), we found that not only did each of their pre-treatment level affect survival of advanced stage NSCLC patients treated with platinum-based chemotherapy, cumulatively patients with more adverse tests also had worse overall survival. Furthermore, compared to other large studies

investigating the roles of clinical laboratory test levels in NSCLC, our study has the advantage that all patients were recruited from the same institution, thus offering a more homogenous population in terms of quality of diagnosis, treatment, laboratory test techniques and reference values used. These findings also have practical importance in the clinical setting. Since these laboratory tests are inexpensive, minimally invasive for the patients and routinely performed in the clinical setting, their values can be easily assessed and evaluated. There are also limitations with our study. Although all laboratory test levels were measured before treatment, this study was designed retrospectively, therefore is susceptible to the limitations of possible selection bias and reverse causation. Future prospective studies investigating the prognostic values of clinical laboratory tests on NSCLC outcomes are warranted.

In conclusion, in this study, we have utilized the large NSCLC patient population at the University of Texas MD Anderson Cancer Center, and found lower than normal albumin levels, higher than normal levels of ALP and LDH are all significant adverse prognostic factors for advanced stage NSCLC patients treated with platinum-based chemotherapy, both individually and cumulatively. Our results suggested that pre-treatment clinical laboratory test levels could be utilized to enhance predictions of survival among advanced stage NSCLC patients treated with platinum-based chemotherapy.

Chapter 4. Conclusions

In conclusion, our studies have identified clinically relevant novel serum biomarkers for lung cancer early diagnosis and clinical outcome. Through our multi-phase study comprised of global metabolomic profiling and prospective validation in a large cohort, we have identified and validated bilirubin as a risk predictor for lung cancer incidence as well as mortality in smokers. Our results suggested that low serum bilirubin levels can be used to identify higher risk smokers for lung cancer. Addition of this variable into National Lung Screening Trial (NLST) selection criteria for low-dose computed tomography (LDCT) screening might help identify higher risk smokers who would benefit more from LDCT screening and reduce false positives. We have also found that three pre-treatment clinical laboratory test levels indicative of metabolic status were both individually and jointly associated with overall survival among advanced stage NSCLC patients treated with platinum-based chemotherapy. These tests could further be utilized to enhance predictions of survival among advanced stage NSCLC patients treated with platinum-based chemotherapy.

Taken together, our results suggested that metabolic alterations associated with lung cancer could serve as novel serum biomarkers with clinical significance for both lung cancer early detection and clinical outcome. With future prospective validation studies, these biomarkers could be applied to the clinics to screen for lung cancer at an early stage and to predict clinical outcome.

Chapter 5. Future Directions

In the current studies, we investigated global metabolomic profiles of lung cancer patients and matched controls and identified bilirubin as a biomarker for lung cancer early diagnosis, which can be used to identify higher risk smokers for lung cancer. For the future studies, our group is planning to determine the sensitivity, specificity, positive predictive value and negative predictive value of bilirubin as a lung cancer biomarker, and to incorporate bilirubin with other known lung cancer risk factors into a risk prediction model for lung cancer risk prediction. It would be also interesting to further conduct other "omic" profilings in these samples, e.g. genomic, transcriptomic and proteomic profiling to identify biomarkers at DNA, RNA and protein levels. Moreover, it would be of great interest and value to correlate our metabolomics profiling data with these other "omics" data to investigate the correlations among DNA, RNA, protein and metabolite levels to identify and elucidate novel molecular pathways underlying lung cancer development. In the future, we will also determine the metabolomic profiles among lung cancer patients with different prognosis to identify novel prognostic biomarkers for lung cancer clinical outcomes.

Our studies have also identified three clinical laboratory test levels as biomarkers for advanced stage non-small cell lung cancer patients treated with platinum-based chemotherapy. While this is one of the largest studies so far investigating the association between clinical laboratory test levels and NSCLC clinical outcomes, we will seek collaboration in other independent patient populations with similarly detailed information on epidemiological and clinical follow-up data as well as data on clinical laboratory test levels, to validate our

findings. In addition, we will do the meta-analysis to further validate our results. In the meanwhile, prospective studies, in which clinical laboratory test levels of newly diagnosed NSCLC patients are recorded and the clinical outcomes of these patients are followed up prospectively would significantly provide further support to the research findings. Lastly, we will incorporate the clinical laboratory test levels and other NSCLC prognostic factors to build a prediction model for NSCLC patients' survival prediction.

Bibliography

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA: a cancer journal for clinicians. 2011;61:69-90.

2. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA: a cancer journal for clinicians. 2014;64:9-29.

3. Ramalingam SS, Owonikoko TK, Khuri FR. Lung cancer: New biological insights and recent therapeutic advances. CA: a cancer journal for clinicians. 2011;61:91-112.

4. Thun MJ, Henley SJ, Burns D, Jemal A, Shanks TG, Calle EE. Lung cancer death rates in lifelong nonsmokers. Journal of the National Cancer Institute. 2006;98:691-9.

5. Hammond EC, Selikoff IJ, Seidman H. Asbestos exposure, cigarette smoking and death rates. Annals of the New York Academy of Sciences. 1979;330:473-90.

6. Siegel R, DeSantis C, Virgo K, Stein K, Mariotto A, Smith T, Cooper D, Gansler T, Lerro C, Fedewa S, Lin C, Leach C, Cannady RS, Cho H, Scoppa S, Hachey M, Kirch R, Jemal A, Ward E. Cancer treatment and survivorship statistics, 2012. CA: a cancer journal for clinicians. 2012;62:220-41.

7. Coate LE, John T, Tsao MS, Shepherd FA. Molecular predictive and prognostic markers in non-small-cell lung cancer. The lancet oncology. 2009;10:1001-10.

8. Andrews J, Yeh P, Pao W, Horn L. Molecular predictors of response to chemotherapy in non-small cell lung cancer. Cancer journal. 2011;17:104-13.

9. Hoang T, Xu R, Schiller JH, Bonomi P, Johnson DH. Clinical model to predict survival in chemonaive patients with advanced non-small-cell lung cancer treated with third-generation chemotherapy regimens based on eastern cooperative oncology group data. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2005;23:175-83.

10. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365:395-409.

11. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, Byers T, Colditz GA, Gould MK, Jett JR, Sabichi AL, Smith-Bindman R, Wood DE, Qaseem A, Detterbeck FC. Benefits and harms of CT screening for lung cancer: a systematic review. JAMA. 2012;307:2418-29.

12. Wood DE, Eapen GA, Ettinger DS, Hou L, Jackman D, Kazerooni E, Klippenstein D, Lackner RP, Leard L, Leung AN, Massion PP, Meyers BF, Munden RF, Otterson GA, Peairs K, Pipavath S, Pratt-Pozo C, Reddy C, Reid ME, Rotter AJ, Schabath MB, Sequist LV, Tong BC, Travis WD, Unger M, Yang SC. Lung cancer screening. Journal of the National Comprehensive Cancer Network : JNCCN. 2012;10:240-65.

13. Jaklitsch MT, Jacobson FL, Austin JH, Field JK, Jett JR, Keshavjee S, MacMahon H, Mulshine JL, Munden RF, Salgia R, Strauss GM, Swanson SJ, Travis WD, Sugarbaker DJ. The American Association for Thoracic Surgery guidelines for lung cancer screening using low-dose computed tomography scans

for lung cancer survivors and other high-risk groups. The Journal of thoracic and cardiovascular surgery. 2012;144:33-8.

14. Wender R, Fontham ET, Barrera E, Jr., Colditz GA, Church TR, Ettinger DS, Etzioni R, Flowers CR, Gazelle GS, Kelsey DK, LaMonte SJ, Michaelson JS, Oeffinger KC, Shih YC, Sullivan DC, Travis W, Walter L, Wolf AM, Brawley OW, Smith RA. American Cancer Society lung cancer screening guidelines. CA: a cancer journal for clinicians. 2013;63:107-17.

15. Kovalchik SA, Tammemagi M, Berg CD, Caporaso NE, Riley TL, Korch M, Silvestri GA, Chaturvedi AK, Katki HA. Targeting of low-dose CT screening according to the risk of lung-cancer death. N Engl J Med. 2013;369:245-54.

16. Kulasingam V, Pavlou MP, Diamandis EP. Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer. Nature reviews Cancer. 2010;10:371-8.

17. Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, Fleisher M, Lilja H, Brogi E, Boyd J, Sanchez-Carbayo M, Holland EC, Cordon-Cardo C, Scher HI, Tempst P. Differential exoprotease activities confer tumor-specific serum peptidome patterns. The Journal of clinical investigation. 2006;116:271-84.

18. Baron JA. Screening for cancer with molecular markers: progress comes with potential problems. Nature reviews Cancer. 2012;12:368-71.

19. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Molecular & cellular proteomics : MCP. 2002;1:845-67.

20. Deberardinis RJ, Sayed N, Ditsworth D, Thompson CB. Brick by brick: metabolism and tumor cell growth. Curr Opin Genet Dev. 2008;18:54-61.
21. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646-74.
22. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science. 2009;324:1029-33.
23. 2020 visions. Nature. 2010;463:26-32.
24. Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, Li Y, Nyati MK, Ahsan A, Kalyana-Sundaram S, Han B, Cao X, Byun J, Omenn GS, Ghosh D, Pennathur S, Alexander DC, Berger A, Shuster JR, Wei JT, Varambally S, Beecher C, Chinnaiyan AM. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature. 2009;457:910-4.
25. Gu H, Pan Z, Xi B, Asiago V, Musselman B, Raftery D. Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: application to the detection of breast cancer. Analytica chimica acta. 2011;686:57-63.
26. Ritchie SA, Ahiahonu PW, Jayasinghe D, Heath D, Liu J, Lu Y, Jin W, Kavianpour A, Yamazaki Y, Khan AM, Hossain M, Su-Myat KK, Wood PL, Krenitsky K, Takemasa I, Miyake M, Sekimoto M, Monden M, Matsubara H, Nomura F, Goodenowe DB. Reduced levels of hydroxylated, polyunsaturated

ultra long-chain fatty acids in the serum of colorectal cancer patients: implications for early screening and detection. BMC medicine. 2010;8:13.

27. Zhang J, Liu L, Wei S, Nagana Gowda GA, Hammoud Z, Kesler KA, Raftery D. Metabolomics study of esophageal adenocarcinoma. The Journal of thoracic and cardiovascular surgery. 2011;141:469-75, 75 e1-4.

28. Ikeda A, Nishiumi S, Shinohara M, Yoshie T, Hatano N, Okuno T, Bamba T, Fukusaki E, Takenawa T, Azuma T, Yoshida M. Serum metabolomics as a novel diagnostic approach for gastrointestinal cancer. Biomedical chromatography : BMC. 2012;26:548-58.

29. Xue R, Lin Z, Deng C, Dong L, Liu T, Wang J, Shen X. A serum metabolomic investigation on hepatocellular carcinoma patients by chemical derivatization followed by gas chromatography/mass spectrometry. Rapid communications in mass spectrometry : RCM. 2008;22:3061-8.

30. Gao H, Dong B, Liu X, Xuan H, Huang Y, Lin D. Metabonomic profiling of renal cell carcinoma: high-resolution proton nuclear magnetic resonance spectroscopy of human serum with multivariate data analysis. Analytica chimica acta. 2008;624:269-77.

31. Zhou J, Xu B, Huang J, Jia X, Xue J, Shi X, Xiao L, Li W. 1H NMR-based metabonomic and pattern recognition analysis for detection of oral squamous cell carcinoma. Clinica chimica acta; international journal of clinical chemistry. 2009;401:8-13.

32. Bathe OF, Shaykhutdinov R, Kopciuk K, Weljie AM, McKay A, Sutherland FR, Dixon E, Dunse N, Sotiropoulos D, Vogel HJ. Feasibility of identifying pancreatic

cancer based on serum metabolomics. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2011;20:140-7.

33. Hori S, Nishiumi S, Kobayashi K, Shinohara M, Hatakeyama Y, Kotani Y, Hatano N, Maniwa Y, Nishio W, Bamba T, Fukusaki E, Azuma T, Takenawa T, Nishimura Y, Yoshida M. A metabolomic approach to lung cancer. Lung Cancer. 2011;74:284-92.

34. Jordan KW, Adkins CB, Su L, Halpern EF, Mark EJ, Christiani DC, Cheng LL. Comparison of squamous cell carcinoma and adenocarcinoma of the lung by metabolomic analysis of tissue-serum pairs. Lung Cancer. 2010;68:44-50.

35. Maeda J, Higashiyama M, Imaizumi A, Nakayama T, Yamamoto H, Daimon T, Yamakado M, Imamura F, Kodama K. Possibility of multivariate function composed of plasma amino acid profiles as a novel screening index for non-small cell lung cancer: a case control study. BMC Cancer. 2010;10:690.

36. Sagman U, Feld R, Evans WK, Warr D, Shepherd FA, Payne D, Pringle J, Yeoh J, DeBoer G, Malkin A, et al. The prognostic significance of pretreatment serum lactate dehydrogenase in patients with small-cell lung cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1991;9:954-61.

37. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, Shete S, Etzel CJ. A risk model for prediction of lung cancer. Journal of the National Cancer Institute. 2007;99:715-26.

38. Lawton KA, Berger A, Mitchell M, Milgram KE, Evans AM, Guo L, Hanson RW, Kalhan SC, Ryals JA, Milburn MV. Analysis of the adult human plasma metabolome. Pharmacogenomics. 2008;9:383-97.

39. Dehaven CD, Evans AM, Dai H, Lawton KA. Organization of GC/MS and LC/MS metabolomics data into chemical libraries. J Cheminform. 2010;2:9.

40. Temme EH, Zhang J, Schouten EG, Kesteloot H. Serum bilirubin and 10-year mortality risk in a Belgian population. Cancer Causes Control. 2001;12:887-94.

41. Horsfall LJ, Rait G, Walters K, Swallow DM, Pereira SP, Nazareth I, Petersen I. Serum bilirubin and risk of respiratory disease and death. JAMA : the journal of the American Medical Association. 2011;305:691-7.

42. Ryter SW, Morse D, Choi AM. Carbon monoxide and bilirubin: potential therapies for pulmonary/vascular injury and disease. Am J Respir Cell Mol Biol. 2007;36:175-82.

43. Lin JP, O'Donnell CJ, Schwaiger JP, Cupples LA, Lingenhel A, Hunt SC, Yang S, Kronenberg F. Association between the UGT1A1*28 allele, bilirubin levels, and coronary heart disease in the Framingham Heart Study. Circulation. 2006;114:1476-81.

44. Novotny L, Vitek L. Inverse relationship between serum bilirubin and atherosclerosis in men: a meta-analysis of published studies. Exp Biol Med (Maywood). 2003;228:568-71.

45. Hori S, Nishiumi S, Kobayashi K, Shinohara M, Hatakeyama Y, Kotani Y, Hatano N, Maniwa Y, Nishio W, Bamba T, Fukusaki E, Azuma T, Takenawa T,

Nishimura Y, Yoshida M. A metabolomic approach to lung cancer. Lung Cancer. 2011.

46. Laky B, Janda M, Bauer J, Vavra C, Cleghorn G, Obermair A. Malnutrition among gynaecological cancer patients. European journal of clinical nutrition. 2007;61:642-6.

47. Fuhrman MP, Charney P, Mueller CM. Hepatic proteins and nutrition assessment. Journal of the American Dietetic Association. 2004;104:1258-64.

48. Jin Y, Zhao L, Peng F. Prognostic impact of serum albumin levels on the recurrence of stage I non-small cell lung cancer. Clinics. 2013;68:686-93.

49. Gupta D, Lis CG. Pretreatment serum albumin as a predictor of cancer survival: a systematic review of the epidemiological literature. Nutrition journal. 2010;9:69.

50. Franch-Arcas G. The meaning of hypoalbuminaemia in clinical practice. Clinical nutrition. 2001;20:265-9.

51. Cengiz O, Kocer B, Surmeli S, Santicky MJ, Soran A. Are pretreatment serum albumin and cholesterol levels prognostic tools in patients with colorectal carcinoma? Medical science monitor : international medical journal of experimental and clinical research. 2006;12:CR240-7.

52. Boonpipattanapong T, Chewatanakornkul S. Preoperative carcinoembryonic antigen and albumin in predicting survival in patients with colon and rectal carcinomas. Journal of clinical gastroenterology. 2006;40:592-5.

53. Onate-Ocana LF, Aiello-Crocifoglio V, Gallardo-Rincon D, Herrera-Goepfert R, Brom-Valladares R, Carrillo JF, Cervera E, Mohar-Betancourt A. Serum albumin

as a significant prognostic factor for patients with gastric carcinoma. Annals of surgical oncology. 2007;14:381-9.

54. Lis CG, Grutsch JF, Vashi PG, Lammersfeld CA. Is serum albumin an independent predictor of survival in patients with breast cancer? JPEN Journal of parenteral and enteral nutrition. 2003;27:10-5.

55. Reichling JJ, Kaplan MM. Clinical use of serum enzymes in liver disease. Digestive diseases and sciences. 1988;33:1601-14.

56. Saif MW, Alexander D, Wicox CM. Serum Alkaline Phosphatase Level as a Prognostic Tool in Colorectal Cancer: A Study of 105 patients. The journal of applied research. 2005;5:88-95.

57. Brown JE, Cook RJ, Major P, Lipton A, Saad F, Smith M, Lee KA, Zheng M, Hei YJ, Coleman RE. Bone turnover markers as predictors of skeletal complications in prostate cancer, lung cancer, and other solid tumors. Journal of the National Cancer Institute. 2005;97:59-69.

58. Klompje J, Petrelli NJ, Herrera L, Mittelman A. The prognostic value of preoperative alkaline phosphatase for resection of solitary liver metastasis from colorectal carcinoma. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 1987;13:345-7.

59. Jonsson PE, Bengtsson G, Carlsson G, Jonson G, Tryding N. Value of serum-5-nucleotidase, alkaline phosphatase and gamma-glutamyl transferase for prediction of liver metastases preoperatively in colorectal cancer. Acta chirurgica Scandinavica. 1984;150:419-23.

60. Demers LM, Costa L, Lipton A. Biochemical markers and skeletal metastases. Cancer. 2000;88:2919-26.

61. Alatas F, Alatas O, Metintas M, Colak O, Erginel S, Harmanci E. Usefulness of bone markers for detection of bone metastases in lung cancer patients. Clinical biochemistry. 2002;35:293-6.

62. Coleman R, Brown J, Terpos E, Lipton A, Smith MR, Cook R, Major P. Bone markers and their prognostic value in metastatic bone disease: clinical evidence and future directions. Cancer treatment reviews. 2008;34:629-39.

63. Dane F, Turk HM, Sevinc A, Buyukberber S, Camci C, Tarakcioglu M. Markers of bone turnover in patients with lung cancer. Journal of the National Medical Association. 2008;100:425-8.

64. Mountzios G, Ramfidis V, Terpos E, Syrigos KN. Prognostic significance of bone markers in patients with lung cancer metastatic to the skeleton: a review of published data. Clinical lung cancer. 2011;12:341-9.

65. Moss DW. Alkaline phosphatase isoenzymes. Clinical chemistry. 1982;28:2007-16.

66. Warburg O. On the origin of cancer cells. Science. 1956;123:309-14.

67. Warburg O. On respiratory impairment in cancer cells. Science. 1956;124:269-70.

68. Osterlind K, Andersen PK. Prognostic factors in small cell lung cancer: multivariate model based on 778 patients treated with chemotherapy with or without irradiation. Cancer research. 1986;46:4189-94.

69. Albain KS, Crowley JJ, LeBlanc M, Livingston RB. Determinants of improved outcome in small-cell lung cancer: an analysis of the 2,580-patient Southwest Oncology Group data base. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1990;8:1563-74.
70. Cerny T, Blair V, Anderson H, Bramwell V, Thatcher N. Pretreatment prognostic factors and scoring system in 407 small-cell lung cancer patients. International journal of cancer Journal international du cancer. 1987;39:146-9.
71. O'Connell JP, Kris MG, Gralla RJ, Groshen S, Trust A, Fiore JJ, Kelsen DP, Heelan RT, Golbey RB. Frequency and prognostic importance of pretreatment clinical characteristics in patients with advanced non-small-cell lung cancer treated with combination chemotherapy. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1986;4:1604-14.
72. Albain KS, Crowley JJ, LeBlanc M, Livingston RB. Survival determinants in extensive-stage non-small-cell lung cancer: the Southwest Oncology Group experience. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1991;9:1618-26.

Appendix A: Publications

1. Wen CP*, **Zhang F*,** Liang D, Wen C, Gu J, Skinner H, Chow WH, Ye Y, Pu X, Hildebrandt MA, Huang M, Chen CH, Hsiung CA, Tsai MK, Tsao CK, Wu X. The ability of bilirubin in identifying smokers with higher risk of lung cancer: a large cohort study in conjunction with metabolomic profiling. Clinical Cancer Research. In revision. *Equal Contribution.

2. **Zhang F,** Meng QH, Chen M, Ye Y, Hildebrandt MA, Huang M, Wu X. Pre-treatment clinical laboratory tests indicative of metabolic status are associated with survival in patients with non-small cell lung cancer." In submission.

3. Gully CP, Velazquez-Torres G, Shin JH, Fuentes-Mattei E, Wang E, Carlock C, Chen J, Rothenberg D, Adams HP, Choi HH, Guma S, Phan L, Chou PC, Su CH, **Zhang F,** Chen JS, Yang TY, Yeung SC, Lee MH. Aurora B kinase phosphorylates and instigates degradation of p53. Proc Natl Acad Sci U S A. 2012;109:E1513-22.

4. Matthias N, Lockworth CR, **Zhang F,** Lee MH, Yeung SC, Tsai KY, Hamir AN. Multiple cystic sweat gland tumors in transgenic mice. Comp Med. 2012;62:27-30.

5. Zhao R, Yeung SC, Chen J, Iwakuma T, Su CH, Chen B, Qu C, **Zhang F,** Chen YT, Lin YL, Lee DF, Jin F, Zhu R, Shaikenov T, Sarbassov D, Sahin A, Wang H, Lai CC, Tsai FJ, Lozano G, Lee MH. Subunit 6 of the COP9 signalosome promotes tumorigenesis in mice through stabilization of MDM2 and is upregulated in human cancers. J Clin Invest. 2011;121:851-65.

6. Su CH, Zhao R, **Zhang F,** Qu C, Chen B, Feng YH, Phan L, Chen J, Wang H, Yeung SC, Lee MH. 14-3-3sigma exerts tumor-suppressor activity mediated by regulation of COP1 stability. Cancer Res. 2011;71:884-94.

7. Gully CP, **Zhang F,** Chen J, Yeung JA, Velazquez-Torres G, Wang E, Yeung SC, Lee MH. Antineoplastic effects of an Aurora B kinase inhibitor in breast cancer. Mol Cancer. 2010;9:42.

Appendix B: Supplementary table

List of the 306 metabolites passing QC and their belonging pathways

Biochemical	Super-pathway	Sub-pathway
1,2-propanediol	Lipid	Ketone bodies
1,5-anhydroglucitol (1,5-AG)	Carbohydrate	Glycolysis, gluconeogenesis, pyruvate metabolism
1,6-anhydroglucose	Carbohydrate	Glycolysis, gluconeogenesis, pyruvate metabolism
1,7-dimethylurate	Xenobiotics	Xanthine metabolism
10-heptadecenoate (17:1n7)	Lipid	Long chain fatty acid
10-nonadecenoate (19:1n9)	Lipid	Long chain fatty acid
12-HETE	Lipid	Eicosanoid
13-HODE + 9-HODE	Lipid	Fatty acid, monohydroxy
13-methylmyristic acid	Lipid	Fatty acid, branched
17-methylstearate	Lipid	Fatty acid, branched
1-arachidonoyl-GPC* (20:4)*	Lipid	Lysolipid
1-arachidonoyl-GPE* (20:4)*	Lipid	Lysolipid

1-arachidonoyl-GPI* (20:4)*	Lipid	Lysolipid
1-docosahexaenoyl-GPC* (22:6)*	Lipid	Lysolipid
1-docosapentaenoyl-GPC* (22:5)*	Lipid	Lysolipid
1-eicosatrienoyl-GPC* (20:3)*	Lipid	Lysolipid
1-heptadecanoyl-GPC (17:0)	Lipid	Lysolipid
1-linoleoylglycerol (18:2)	Lipid	Monoacylglycerol
1-linoleoyl-GPC (18:2)	Lipid	Lysolipid
1-linoleoyl-GPE (18:2)*	Lipid	Lysolipid
1-methyladenosine	Nucleotide	Purine metabolism, adenine containing
1-myristoyl-GPC (14:0)	Lipid	Lysolipid
1-oleoylglycerol (18:1)	Lipid	Monoacylglycerol
1-oleoyl-GPC (18:1)	Lipid	Lysolipid
1-oleoyl-GPE (18:1)	Lipid	Lysolipid
1-palmitoleoyl-GPC* (16:1)*	Lipid	Lysolipid
1-palmitoylglycerol (16:0)	Lipid	Monoacylglycerol
1-palmitoyl-GPC (16:0)	Lipid	Lysolipid
1-palmitoyl-GPE (16:0)	Lipid	Lysolipid
1-palmitoyl-GPI* (16:0)*	Lipid	Lysolipid
1-palmitoylplasmenylethanolamine*	Lipid	Lysolipid

1-pentadecanoylglycerophosphocholine*	Lipid	Lysolipid
1-stearoylglycerol (18:0)	Lipid	Monoacylglycerol
1-stearoyl-GPC (18:0)	Lipid	Lysolipid
1-stearoyl-GPE (18:0)	Lipid	Lysolipid
1-stearoyl-GPI (18:0)	Lipid	Lysolipid
2-aminobutyrate	Amino acid	Butanoate metabolism
2-arachidonoyl-GPC* (20:4)*	Lipid	Lysolipid
2-arachidonoyl-GPE* (20:4)*	Lipid	Lysolipid
2-ethylhexanoic acid	Xenobiotics	Chemical
2-hydroxybutyrate (AHB)	Amino acid	Cysteine, methionine, SAM, taurine metabolism
2-hydroxyglutarate	Lipid	Fatty acid, dicarboxylate
2-hydroxyisobutyrate	Amino acid	Valine, leucine and isoleucine metabolism
2-hydroxypalmitate	Lipid	Fatty acid, monohydroxy
2-hydroxystearate	Lipid	Fatty acid, monohydroxy
2-linoleoyl-GPC* (18:2)*	Lipid	Lysolipid
2-linoleoyl-GPE* (18:2)*	Lipid	Lysolipid
2-methylbutyroylcarnitine (C5)	Amino acid	Valine, leucine and isoleucine metabolism
2-oleoyl-GPC* (18:1)*	Lipid	Lysolipid
2-oleoyl-GPE* (18:1)*	Lipid	Lysolipid

2-palmitoyl-GPC* (16:0)*	Lipid	Lysolipid
2-palmitoyl-GPE* (16:0)*	Lipid	Lysolipid
2-pyrrolidinone	Xenobiotics	Chemical
2-stearoyl-GPC* (18:0)*	Lipid	Lysolipid
3-(4-hydroxyphenyl)lactate (HPLA)	Amino acid	Phenylalanine & tyrosine metabolism
3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)	Lipid	Fatty acid, dicarboxylate
3-dehydrocarnitine*	Lipid	Carnitine metabolism
3-hydroxy-2-ethylpropionate	Amino acid	Valine, leucine and isoleucine metabolism
3-hydroxybutyrate (BHBA)	Lipid	Ketone bodies
3-hydroxyisobutyrate	Amino acid	Valine, leucine and isoleucine metabolism
3-hydroxylaurate	Lipid	Fatty acid, monohydroxy
3-indoxyl sulfate	Amino acid	Tryptophan metabolism
3-methoxytyrosine	Amino acid	Phenylalanine & tyrosine metabolism
3-methyl-2-oxobutyrate	Amino acid	Valine, leucine and isoleucine metabolism
3-methyl-2-oxovalerate	Amino acid	Valine, leucine and isoleucine metabolism
3-methylhistidine	Amino acid	Histidine metabolism
4-acetamidobutanoate	Amino acid	Guanidino and acetamido metabolism
4-androsten-3beta,17beta-diol disulfate 1*	Lipid	Sterol/Steroid

4-androsten-3beta,17beta-diol disulfate 2*	Lipid	Sterol/Steroid
4-hydroxyphenylacetate	Amino acid	Phenylalanine & tyrosine metabolism
4-methyl-2-oxopentanoate	Amino acid	Valine, leucine and isoleucine metabolism
5-dodecenoate (12:1n7)	Lipid	Medium chain fatty acid
5-oxoproline	Amino acid	Glutathione metabolism
7-HOCA	Lipid	Sterol/Steroid
acetylcarnitine (C2)	Lipid	Carnitine metabolism
acetylphosphate	Energy	Oxidative phosphorylation
adrenate (22:4n6)	Lipid	Long chain fatty acid
ADSGEGDFXAEGGGVR*	Peptide	Fibrinogen cleavage peptide
alanine	Amino acid	Alanine and aspartate metabolism
allantoin	Nucleotide	Purine metabolism, urate metabolism
alpha-hydroxyisovalerate	Amino acid	Valine, leucine and isoleucine metabolism
alpha-ketoglutarate	Energy	Krebs cycle
alpha-tocopherol	Cofactors and vitamins	Tocopherol metabolism
andro steroid monosulfate 2*	Lipid	Sterol/Steroid
androsterone sulfate	Lipid	Sterol/Steroid
arabinose	Carbohydrate	Nucleotide sugars, pentose metabolism
arabitol	Carbohydrate	Nucleotide sugars, pentose metabolism

arachidonate (20:4n6)	Lipid	Long chain fatty acid
arginine	Amino acid	Urea cycle; arginine-, proline-, metabolism
ascorbate (Vitamin C)	Cofactors and vitamins	Ascorbate and aldarate metabolism
asparagine	Amino acid	Alanine and aspartate metabolism
aspartate	Amino acid	Alanine and aspartate metabolism
aspartylphenylalanine	Peptide	Dipeptide
azelate (nonanedioate; C9)	Lipid	Fatty acid, dicarboxylate
benzoate	Xenobiotics	Benzoate metabolism
beta-alanine	Amino acid	Alanine and aspartate metabolism
beta-hydroxyisovalerate	Amino acid	Valine, leucine and isoleucine metabolism
betaine	Amino acid	Glycine, serine and threonine metabolism
bilirubin	Cofactors and vitamins	Hemoglobin and porphyrin metabolism
bilirubin (E,E)*	Cofactors and vitamins	Hemoglobin and porphyrin metabolism
bradykinin, des-arg(9)	Peptide	Polypeptide
butyrylcarnitine (C4)	Lipid	Fatty acid metabolism (also BCAA metabolism)
caffeine	Xenobiotics	Xanthine metabolism
campesterol	Lipid	Sterol/Steroid
caprate (10:0)	Lipid	Medium chain fatty acid
caproate (6:0)	Lipid	Medium chain fatty acid

caprylate (8:0)	Lipid	Medium chain fatty acid
carnitine	Lipid	Carnitine metabolism
catechol sulfate	Xenobiotics	Benzoate metabolism
C-glycosyltryptophan*	Amino acid	Tryptophan metabolism
cholesterol	Lipid	Sterol/Steroid
choline	Lipid	Glycerolipid metabolism
cis-4-decenoyl carnitine	Lipid	Carnitine metabolism
cis-aconitate	Energy	Krebs cycle
citrate	Energy	Krebs cycle
citrulline	Amino acid	Urea cycle; arginine-, proline-, metabolism
cortisol	Lipid	Sterol/Steroid
cortisone	Lipid	Sterol/Steroid
creatine	Amino acid	Creatine metabolism
creatinine	Amino acid	Creatine metabolism
cyclo(phe-phe)	Peptide	Dipeptide
cysteine	Amino acid	Cysteine, methionine, SAM, taurine metabolism
cystine	Amino acid	Cysteine, methionine, SAM, taurine metabolism
decanoylcarnitine (C10)	Lipid	Carnitine metabolism
dehydroisoandrosterone sulfate (DHEA-S)	Lipid	Sterol/Steroid

deoxycarnitine	Lipid	Carnitine metabolism
deoxycholate	Lipid	Bile acid metabolism
dihomolinoleate (20:2n6)	Lipid	Long chain fatty acid
dihomolinolenate (20:3n3 or 3n6)	Lipid	Essential fatty acid
dimethylarginine (ADMA + SDMA)	Amino acid	Urea cycle; arginine-, proline-, metabolism
dimethylglycine	Amino acid	Glycine, serine and threonine metabolism
docosadienoate (22:2n6)	Lipid	Long chain fatty acid
docosahexaenoate (DHA; 22:6n3)	Lipid	Essential fatty acid
docosapentaenoate (DPA; 22:5n3)	Lipid	Essential fatty acid
docosapentaenoate (n6 DPA; 22:5n6)	Lipid	Essential fatty acid
dodecanedioate (C12)	Lipid	Fatty acid, dicarboxylate
DSGEGDFXAEGGGVR*	Peptide	Fibrinogen cleavage peptide
eicosapentaenoate (EPA; 20:5n3)	Lipid	Essential fatty acid
eicosenoate (20:1n9 or 1n11)	Lipid	Long chain fatty acid
epiandrosterone sulfate	Lipid	Sterol/Steroid
erythritol	Xenobiotics	Sugar, sugar substitute, starch
erythronate*	Carbohydrate	Aminosugars metabolism
ethanolamine	Lipid	Glycerolipid metabolism
fructose	Carbohydrate	Fructose, mannose, galactose, starch, and

		sucrose metabolism
fucose	Carbohydrate	Aminosugars metabolism
fumarate	Energy	Krebs cycle
gamma-CEHC	Cofactors and vitamins	Tocopherol metabolism
gamma-glutamylalanine	Peptide	gamma-glutamyl
gamma-glutamylglutamine	Peptide	gamma-glutamyl
gamma-glutamylleucine	Peptide	gamma-glutamyl
gamma-glutamylmethionine	Peptide	gamma-glutamyl
gamma-glutamylphenylalanine	Peptide	gamma-glutamyl
gamma-glutamylthreonine*	Peptide	gamma-glutamyl
gamma-glutamyltyrosine	Peptide	gamma-glutamyl
gamma-glutamylvaline	Peptide	gamma-glutamyl
gamma-tocopherol	Cofactors and vitamins	Tocopherol metabolism
gluconate	Carbohydrate	Nucleotide sugars, pentose metabolism
glucose	Carbohydrate	Glycolysis, gluconeogenesis, pyruvate metabolism
glucuronate	Carbohydrate	Glycolysis, gluconeogenesis, pyruvate metabolism
glutamate	Amino acid	Glutamate metabolism

glutamine	Amino acid	Glutamate metabolism
glutaroylcarnitine (C5)	Amino acid	Lysine metabolism
glycerate	Carbohydrate	Glycolysis, gluconeogenesis, pyruvate metabolism
glycerol	Lipid	Glycerolipid metabolism
glycerol 2-phosphate	Xenobiotics	Chemical
glycerol 3-phosphate (G3P)	Lipid	Glycerolipid metabolism
glycerophosphorylcholine (GPC)	Lipid	Glycerolipid metabolism
glycine	Amino acid	Glycine, serine and threonine metabolism
glycochenodeoxycholate	Lipid	Bile acid metabolism
glycocholate	Lipid	Bile acid metabolism
glycocholenate sulfate*	Lipid	Bile acid metabolism
glycodeoxycholate	Lipid	Bile acid metabolism
glycolate (hydroxyacetate)	Xenobiotics	Chemical
glycolithocholate sulfate*	Lipid	Bile acid metabolism
glycoursodeoxycholate	Lipid	Bile acid metabolism
glycylphenylalanine	Peptide	Dipeptide
glycylvaline	Peptide	Dipeptide
heme*	Cofactors and vitamins	Hemoglobin and porphyrin

heptanoate (7:0)	Lipid	Medium chain fatty acid
hexadecanedioate (C16)	Lipid	Fatty acid, dicarboxylate
hexanoylcarnitine (C6)	Lipid	Carnitine metabolism
hippurate	Xenobiotics	Benzoate metabolism
histidine	Amino acid	Histidine metabolism
homostachydrine*	Xenobiotics	Food component/Plant
HWESASXX*	Peptide	Polypeptide
hydroxyproline	Amino acid	Urea cycle; arginine-, proline-, metabolism
hypoxanthine	Nucleotide	Purine metabolism, (hypo)xanthine/inosine containing
indoleacetate	Amino acid	Tryptophan metabolism
indolelactate	Amino acid	Tryptophan metabolism
indolepropionate	Amino acid	Tryptophan metabolism
inosine	Nucleotide	Purine metabolism, (hypo)xanthine/inosine containing
inositol 1-phosphate (I1P)	Lipid	Inositol metabolism
isobutyrylcarnitine (C4)	Amino acid	Valine, leucine and isoleucine metabolism
isoleucine	Amino acid	Valine, leucine and isoleucine metabolism
isovalerylcarnitine (C5)	Amino acid	Valine, leucine and isoleucine metabolism

kynurenine	Amino acid	Tryptophan metabolism
lactate	Carbohydrate	Glycolysis, gluconeogenesis, pyruvate metabolism
lathosterol	Lipid	Sterol/Steroid
laurate (12:0)	Lipid	Medium chain fatty acid
leucine	Amino acid	Valine, leucine and isoleucine metabolism
leucylleucine	Peptide	Dipeptide
levulinate (4-oxovalerate)	Amino acid	Valine, leucine and isoleucine metabolism
linoleate (18:2n6)	Lipid	Essential fatty acid
linolenate (18:3n3 or 3n6)	Lipid	Essential fatty acid
lysine	Amino acid	Lysine metabolism
malate	Energy	Krebs cycle
mannitol	Carbohydrate	Fructose, mannose, galactose, starch, and sucrose metabolism
mannose	Carbohydrate	Fructose, mannose, galactose, starch, and sucrose metabolism
margarate (17:0)	Lipid	Long chain fatty acid
methionine	Amino acid	Cysteine, methionine, SAM, taurine metabolism
methyl palmitate (15 or 2)	Lipid	Fatty acid, branched

methyl palmitate (16:0)	Lipid	Fatty acid, methyl ester
methylphosphate	Nucleotide	Purine and pyrimidine metabolism
myo-inositol	Lipid	Inositol metabolism
myristate (14:0)	Lipid	Long chain fatty acid
myristoleate (14:1n5)	Lipid	Long chain fatty acid
N6-acetyllysine	Amino acid	Lysine metabolism
N-acetylalanine	Amino acid	Alanine and aspartate metabolism
N-acetyl-beta-alanine	Amino acid	Alanine and aspartate metabolism
N-acetylglycine	Amino acid	Glycine, serine and threonine metabolism
N-acetylmethionine	Amino acid	Cysteine, methionine, SAM, taurine metabolism
N-acetylneuraminate	Carbohydrate	Aminosugars metabolism
N-acetylornithine	Amino acid	Urea cycle; arginine-, proline-, metabolism
N-acetylserine	Amino acid	Glycine, serine and threonine metabolism
N-acetylthreonine	Amino acid	Glycine, serine and threonine metabolism
N-formylmethionine	Amino acid	Cysteine, methionine, SAM, taurine metabolism
nicotinamide	Cofactors and vitamins	Nicotinate and nicotinamide metabolism
nonadecanoate (19:0)	Lipid	Long chain fatty acid
octadecanedioate (C18)	Lipid	Fatty acid, dicarboxylate
octanoylcarnitine (C8)	Lipid	Carnitine metabolism

oleate (18:1n9)	Lipid	Long chain fatty acid
oleoylcarnitine (C18)	Lipid	Carnitine metabolism
ornithine	Amino acid	Urea cycle; arginine-, proline-, metabolism
oxalate (ethanedioate)	Carbohydrate	Glyoxylate and dicarboxylate metabolism
palmitate (16:0)	Lipid	Long chain fatty acid
palmitoleate (16:1n7)	Lipid	Long chain fatty acid
palmitoyl sphingomyelin	Lipid	Sphingolipid
palmitoylcarnitine (C16)	Lipid	Carnitine metabolism
pantothenate (Vitamin B5)	Cofactors and vitamins	Pantothenate and CoA metabolism
paraxanthine	Xenobiotics	Xanthine metabolism
p-cresol sulfate	Amino acid	Phenylalanine & tyrosine metabolism
pelargonate (9:0)	Lipid	Medium chain fatty acid
pentadecanoate (15:0)	Lipid	Long chain fatty acid
phenol sulfate	Amino acid	Phenylalanine & tyrosine metabolism
phenylacetylglutamine	Amino acid	Phenylalanine & tyrosine metabolism
phenylalanine	Amino acid	Phenylalanine & tyrosine metabolism
phenylalanylleucine*	Peptide	Dipeptide
phenylalanylphenylalanine	Peptide	Dipeptide
phenyllactate (PLA)	Amino acid	Phenylalanine & tyrosine metabolism

phosphate	Energy	Oxidative phosphorylation
pipecolate	Amino acid	Lysine metabolism
piperine	Xenobiotics	Food component/Plant
pregn steroid monosulfate*	Lipid	Sterol/Steroid
pregnen-diol disulfate*	Lipid	Sterol/Steroid
pregnenolone sulfate	Lipid	Sterol/Steroid
proline	Amino acid	Urea cycle; arginine-, proline-, metabolism
prolylhydroxyproline	Peptide	Dipeptide
propionylcarnitine (C3)	Lipid	Fatty acid metabolism (also BCAA metabolism)
pseudouridine	Nucleotide	Pyrimidine metabolism, uracil containing
pyridoxate	Cofactors and vitamins	Vitamin B6 metabolism
pyroglutamine*	Amino acid	Glutamate metabolism
pyroglutamylglycine	Peptide	Dipeptide
pyrophosphate (PPi)	Energy	Oxidative phosphorylation
pyruvate	Carbohydrate	Glycolysis, gluconeogenesis, pyruvate metabolism
quinate	Xenobiotics	Food component/Plant
ribitol	Carbohydrate	Nucleotide sugars, pentose metabolism
ribose	Carbohydrate	Nucleotide sugars, pentose metabolism

scyllo-inositol	Lipid	Inositol metabolism
sebacate (decanedioate)	Lipid	Fatty acid, dicarboxylate
serine	Amino acid	Glycine, serine and threonine metabolism
serotonin (5HT)	Amino acid	Tryptophan metabolism
S-methylcysteine	Amino acid	Cysteine, methionine, SAM, taurine metabolism
sphingomyelin	Lipid	Sphingolipid
stachydrine	Xenobiotics	Food component/Plant
stearamide	Lipid	Fatty acid, amide
stearate (18:0)	Lipid	Long chain fatty acid
stearidonate (18:4n3)	Lipid	Long chain fatty acid
stearoylcarnitine (C18)	Lipid	Carnitine metabolism
suberate (octanedioate)	Lipid	Fatty acid, dicarboxylate
succinate	Energy	Krebs cycle
succinylcarnitine (C4)	Energy	Krebs cycle
sucrose	Carbohydrate	Fructose, mannose, galactose, starch, and sucrose metabolism
taurochenodeoxycholate	Lipid	Bile acid metabolism
taurocholenate sulfate*	Lipid	Bile acid metabolism
taurodeoxycholate	Lipid	Bile acid metabolism

taurolithocholate 3-sulfate	Lipid	Bile acid metabolism
theobromine	Xenobiotics	Xanthine metabolism
theophylline	Xenobiotics	Xanthine metabolism
threitol	Carbohydrate	Nucleotide sugars, pentose metabolism
threonate	Cofactors and vitamins	Ascorbate and aldarate metabolism
threonine	Amino acid	Glycine, serine and threonine metabolism
triethyleneglycol	Xenobiotics	Chemical
trizma acetate	Xenobiotics	Chemical
tryptophan	Amino acid	Tryptophan metabolism
tryptophan betaine	Amino acid	Tryptophan metabolism
tyrosine	Amino acid	Phenylalanine & tyrosine metabolism
undecanedioate	Lipid	Fatty acid, dicarboxylate
undecanoate (11:0)	Lipid	Medium chain fatty acid
urate	Nucleotide	Purine metabolism, urate metabolism
urea	Amino acid	Urea cycle; arginine-, proline-, metabolism
uridine	Nucleotide	Pyrimidine metabolism, uracil containing
vaccenate (18:1n7)	Lipid	Long chain fatty acid
valine	Amino acid	Valine, leucine and isoleucine metabolism
xanthine	Nucleotide	Purine metabolism, (hypo)xanthine/inosine

		containing
xylonate	Carbohydrate	Nucleotide sugars, pentose metabolism
xylose	Carbohydrate	Nucleotide sugars, pentose metabolism

VITA

Fanmao Zhang, the second son of Wenping Zhang and Xianxiu Xu, was born in Yueqing, Zhejiang, People's Republic of China on July 1st, 1986. In September 2002, he entered College of Life Sciences at Wuhan University in Wuhan, China and earned his Bachelor of Science degree in 2006 with a major in Biological Sciences and Biotechnology. In August 2007, he enrolled in the Ph.D. program at Graduate School of Biomedical Sciences at The University of Texas Health Science Center at Houston. He joined Department of Epidemiology, The University of Texas MD Anderson Cancer Center in February 2011, where he did his dissertation research under the supervision of Dr. Xifeng Wu, M.D., Ph.D.