ESTROGEN RECEPTOR-BETA GENOTYPE, SEVEN IMMUNOHISTOCHEMICAL MARKERS, AND HUMAN LUNG CANCER

by

Ji Young Song

B.S., University of Maryland at College Park, 2002

M.S., Medical University of South Carolina, 2004

Submitted to the Graduate Faculty of

Department of Epidemiology

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This dissertation was presented

by

Ji Young Song

It was defended on

July 1, 2010

and approved by

Dissertation Advisor:
Joel L. Weissfeld, MD, MPH
Associate Professor of Epidemiology
Assistant Professor of Medicine
Graduate School of Public Health
School of Medicine
University of Pittsburgh

Committee Members:
Jill M. Siegfried, Ph.D.
Professor of Pharmacology and Chemical Biology
School of Medicine
University of Pittsburgh

Stephanie Ruth Land, Ph.D.
Research Associate Professor of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Brenda Diergaarde, Ph.D.
Assistant Professor of Epidemiology
Graduate School of Public Health
University of Pittsburgh

Robert P. Bowser, Ph.D.
Associate Professor of Pathology
School of Medicine
University of Pittsburgh

Joel L. Weissfeld, MD, MPH

ESTROGEN RECEPTOR-BETA GENOTYPE, SEVEN IMMUNOHISTOCHEMICAL MARKERS, AND HUMAN LUNG CANCER

Ji Young Song, PhD

University of Pittsburgh, 2010

Lung cancer is the leading cause of cancer death in the United States. However, few new and effective treatments are available for lung cancer. A comprehensive understanding of the multiple signaling pathways that lead to tumor growth is a prerequisite for more effective and targeted cancer treatments. The purpose of this research was to investigate the relationship between proteins [hepatocypte growth factor (HGF), c-Met, and estrogen receptor-beta (ER-beta)] and gene expression of ER-beta (*ESR2*) and lung cancer survival along with identifying meaningful expression patterns of seven biomarkers [HGF, c-Met, ER-alpha, ER-beta, progesterone receptor (PR), aromatase, and epidermal growth factor receptor (EGFR)].

We used immunohistochemistry to quantify the expression of seven proteins in primary lung tumor tissues from the Lung Cancer Specialized Program of Research Excellence substudy (N=204). The generalized linear mixed model approach, which controlled for sample type (tissue microarray vs. whole-section), showed high HGF expression associated negatively with advanced cancer stage (P_{global}=0.05) and positively with smoking (P_{global}=0.14). After accounting for stage and other factors, neither HGF nor c-Met expression predicted survival.

Using a cluster algorithm, two groups were identified: (Cluster 1: high expression of ER-alpha ER-beta, cytoplasmic PR, EGFR, and aromatase; Cluster 2: high expression of HGF, c-Met, and nuclear PR). Two lung cancer subgroups exhibiting dissimilar 7-protein IHC

expression patterns were similar in terms of host and tumor characteristics and in terms of overall survival (log rank test: p=0.69).

Among 22 htSNPs of *ESR2* gene, we have identified that rare allele of rs1256061 is associated with the maximum ER-beta expression among patients with adenocarcinoma, but not with squamous cell carcinoma.

The results of this research enhanced the knowledge of the role of HGF and c-Met on lung cancer survival and also suggested that the relationship between genetic variation of *ESR2* gene and protein expression may differ by lung cancer histology. Understanding the roles, the expression patterns, and the genetic of steroid hormones, growth factors and their receptors in lung cancer is of great public health significance because it may enable biologically directed and individually tailored treatment and their possible use as biomarkers for early detection and prevention.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to take this opportunity to acknowledge the many individuals who have contributed both to this work and to my development as a cancer epidemiologist. I am extremely fortunate to have been surrounded by people who supported my education and who gave me the opportunity to pursue this line of research.

In the first place I would like to record my gratitude to my advisor, Dr. Joel Weissfeld, for his supervision, advice, and guidance from the very early stage of this research as well as giving me extraordinary experiences throughout the work. His questions and critiques have improved this research and have made me a better epidemiologist. Above all and the most needed, he provided me unflinching encouragement and support in various ways. I am indebted to him more than he knows. I cannot possibly imagine completing this work successfully without his endless help and guidance. He is a model of scientific excellence and a wisdom mentor that I will always look up to.

I am also appreciative of the funding support from the National Cancer Institute's Predoctoral Cancer Epidemiology Training Grant. I am grateful for the friendship with co-workers and fellow trainees: YeuhYing (Monica), Ashley, Kwame, Cher, Scott, Jesse, Harris, Stacy, Tracy, and Mary Yagjian.

I am heartily thankful to my dissertation committee members, Dr. Jill Siegfried, Dr. Brenda Diergaarde, Dr. Stephanie Land, and Dr. Robert Bowser for all of their supportive

suggestions and insightful commentary on my work. They have been an indispensable resource throughout the dissertation process. My immense gratitude to Dr. Siegfried for all the work she did initiating and executing the Lung Cancer Specialized Program of Research Excellence (SPORE), and for providing me with all the materials I needed for my work. I also thank Dr. Autumn Davis for her work coordinating the project. Dr. Laura Stabile guided me to acquire a comprehensive understanding of the original project and promptly provided me with all the information I needed for my work, for which I am most grateful. Dr. Brenda Diergaarde has fostered my interest in genetic and molecular epidemiology through her direct guidance to execute my *ESR2* genetic study of lung cancer.

I would like to thank the investigators and collaborators in Lung Cancer SPORE project for their efforts in establishing this unique study. I am especially indebted to Drs. Sanja Dacic, Rajiv Dhir, and Marjorie Romkes for their support of my work. I also benefited by outstanding works from the Health Sciences Tissue Bank, the Clinical Genomics Facility of Department of Pathology, and the Genomics and Proteomics Core laboratories.

I thank you my husband, Hyung Bo for being always there for me. His unfailing belief in my abilities has helped me to persevere at times when I was ready to quit. I could not have accomplished this goal without Hyung Bo's love and support.

Finally, I thank my family for their patience, dedication and support through all these years. I could not be who I am now without your endless prayers. Thank you, MOMs, DADs, my baby sister-Ji Hae, and only sister in law-Hyo Won. I only hope that I have made them proud. Mom and Dad, I dedicate this work to you.

Last but utmost, I praise my God Almighty for His Mercies and abundant grace on my journey of life.

1.0 INTRODUCTION

Lung cancer is the leading cause of cancer death in men and women the United States. 160,390 deaths are estimated in the United Stated in 2007 from lung cancer. However, epidemiology of male and female lung cancer differs. More women have died each year from lung cancer than from breast cancer since 1987.[1] Also, even though male lung cancer mortality has declined significantly by about 1.9% per year during 1991-2003, female death rates are approaching a plateau after continuously increasing for several decades.[1,2] It is well known that smoking is the major cause of lung cancer. 85% to 90% of all lung cancer patients have smoked cigarettes at some time in their lives.[3] However, lung cancer also occurs in non-smokers and not all smokers get lung cancer. Interestingly, nonsmokers diagnosed with lung cancer are predominately women.[4]

Gender differences in lung cancer susceptibility and prognosis are shown in many epidemiological studies. According to Surveillance Epidemiology and End Results (SEER) Statistics Review 1975-2005 from National Cancer Institute, women have a higher 5-year relative survival rate than men during recent 25 years (18.2 vs. 13.5).[1] Prognostic factors uniquely associated with female lung cancer may create new therapeutic opportunities. Identifying prognostic markers is critical because the 5-year survival rate for patients with all stages of lung cancer combined is only 15.7% and the 5 year survival rate for both men and women increased only 3% since 1975.[1]

Growth factors and their receptors are attractive targets for cancer therapy due to their involvement in cell division and cell survival which may contribute the imbalance in malignant cells through the signaling pathways.[5] Among many growth factors and their receptors, the hepatocypte growth factor (HGF) and its only known receptor, c-Met, are known to be promising targets for cancer therapy by its multiple biological functions such as cell proliferation, motility, angiogenesis (blood vessel formation), and morphogenesis.[6] HGF is a multifunctional cytokine and mainly detected in epithelial cells.[7] HGF is the ligand for the c-Met protein, a tyrosine kinase receptor which constitutively activated by mutations and expressed by both epithelial and endothelial cells.[5] Also HGF and/or c-Met is overexpressed in many human cancers such as breast, prostate, and lung.[8-10] Previous studies showed that overexpression of HGF and/or c-Met is associated with the poor prognosis of non-small-cell lung cancer (NSCLC) patients.[11-13] Also, Chen and colleagues found that overexpression of HGF has significant correlation with cigarette smoking and tumor stages.[14] Since women and younger lung cancer patients who have weaker association with smoking exposure develop adenocarcinoma (subtype of NSCLC) more often, HGF/c-Met in the lung tumor tissue may be a clue to the prognostic difference in gender and histological subtypes.[15] Despite the need and potential use of HGF and its receptor as the prognostic biomarker for lung cancer, only a few epidemiological studies were conducted.

These reasons, the purpose of the present research is as follows: 1) to identify factors associated with HGF and c-Met immunohiotochemistry (IHC) expression in lung tumor tissue, 2) To examine association between HGF and c-Met IHC expression and lung cancer survival, 3) to identify meaningful expression patterns of seven biomarkers [HGF, c-Met, ERα, ERβ, PR, aromatase, and EGFR], and 4) to examine the association between subjects *ESR2* genotype and

ER beta tumor IHC expression in patients with lung cancer. The following literature review presents an overview of lung cancer epidemiology and known risk factors for lung cancer. A more detailed background on HGF/c-Met biology and human genetic of *ESR2* as they relate to lung cancer is also provided.

1.1 SPECIFIC AIMS

In this dissertational research, I aimed to determine the relationship between protein (HGF, c-Met, and ER-beta) and gene marker expression (*ESR2*) and lung cancer survival and to identify meaningful expression patterns of seven biomarkers [HGF, c-Met, ERα, ERβ, PR, aromatase, and EGFR]. To accomplish this goal, I proposed three discrete, but related projects (below):

1.1.1 Project 1: HGF AND c-Met: Immunohistochemical expression and lung cancer survival

1.1.1.1 Specific Aim 1

Using data from the Lung and Thoracic Malignancies Program (LTMP) Tissue and Blood Bank [subjects consented from Genetic Markers of Lung Malignancy (a.k.a., the Carinal Biopsy Study)] and tissue microarray and whole section experiment using IHC detection method, Project #1investigated if HGF/c-Met can be a strong and independent predictor of survival in lung cancer. As a primary specific aim in Project #1, I examined the relationship of HGF/c-Met expression in tumor lung issue with the clinical parameters (smoking, gender, histology, and disease stage) of subjects with lung cancer and other lung cancer risk factors. I explicitly and

statistically tested the alternative hypothesis (H_A) of difference in the prevalence of high HGF/c-Met expression in lung tumor tissue between “histological types” of lung cancer, between “smoker and non-smoker”, and between “men and women” against the null hypothesis (H_0) of no difference in the prevalence of high HGF/c-Met expression in lung tumor tissue between “histological types” of lung cancer, between “smoker and non-smoker”, and between “men and women”.

1.1.1.2 Specific Aim 2

As a second specific aim in project #1, I evaluated the association between HGF/c-Met expression in tumor lung tissue and lung cancer survival rate and impact HGF/c-Met expression level by gender on lung cancer prognosis. Under the retrospective cohort study design, Project #1 explicitly and statistically tested the alternative hypothesis (H_A) of difference in survival rate of lung cancer patients between HGF/c-Met expression levels in tumor lung tissue against the null hypothesis (H_0) of no difference in the survival rate of lung cancer patients between high and low HGF/c-Met expression in tumor lung tissue. Additionally, the stratified test was performed to test the alternative hypothesis (H_A) of difference in the hazard ratio of lung cancer patients between subjects with and without high HGF/c-Met expression by tumor lung tissue between men and women against the null hypothesis (H_0) of no difference in the hazard ratio of lung cancer patients between subjects with and without high HGF/c-Met expression by tumor lung tissue between men and women.

1.1.2 Project 2: Validation study of Immunohistochemical expression patterns involving seven lung tumor markers

1.1.2.1 Specific Aim 1

Using data from the Lung and Thoracic Malignancies Program (LTMP) Tissue and Blood Bank [subjects consented from Genetic Markers of Lung Malignancy (a.k.a., the Carinal Biopsy Study)] and immunohistochemical expression, Project #2 investigated the inter-correlation among seven biomarkers: estrogen receptor alpha (ERα), estrogen receptor beta (ERβ), epidermal growth factor receptor (EGFR), hepatocyte growth factor (HGF), c-Met, aromatase, and progesterone receptor (PR). As a primary specific aim for this project, I evaluated the strength and direction of the relationship *(correlation)* of immunohistochemical expression in lung tumor tissue of seven markers. Also, I identified meaningful expression patterns involving these seven interesting and relevant proteins by using a cluster algorithm. I compared the identified clusters according to personal host characteristics, tumor stage and histology, and survival.

1.1.3 Project 3: *ESR2* polymorphisms and estrogen receptor beta expression in lung tumors

Using data from the Lung and Thoracic Malignancies Program (LTMP) Tissue and Blood Bank [subjects consented from Genetic Markers of Lung Malignancy (a.k.a., the Carinal Biopsy Study)] and the IHC expression of ERβ in lung tumors and the genotyping results of the estrogen receptor beta gene *(ESR2*), Project #3 determined if there an association between polymorphisms in *ESR2* and the protein expression of ERβ in terms of lung cancer survival.

1.1.3.1 Specific Aim 1

As a primary specific aim in project #3, I examined the relationship of both cytoplasmic and nuclear ERβ protein expression in lung tumors with the clinical parameters (smoking, gender, histology, and disease stage) of lung cancer patients and other lung cancer risk factors in order to evaluate if protein expression of ERβ can be a strong and independent predictor of lung cancer survival. I explicitly and statistically tested the alternative hypothesis (H_A) of difference in the median Allred scores of ERβ expression in lung tumor tissue between "histological types" of lung cancer, between "smoker and non-smoker", and between "men and women" against the null hypothesis (H_0) of no difference in the median Allred scores of ERβ expression in lung tumor tissue between "histological types" of lung cancer, between "smoker and non-smoker", and between "men and women".

1.1.3.2 Specific Aim 2

I assessed the association between the polymorphisms in *ESR2* gene and the ERβ expression status in lung tumor tissue. I explicitly and statistically tested the alternative hypothesis (H_A) of difference in prevalence of polymorphisms in *ESR2* gene between high and low ERβ expression against the null hypothesis (H_0) of no difference in prevalence of polymorphisms in *ESR2* gene between high and low ERβ expression status. Also, I examined whether there are histological types differences in the relation of the polymorphisms in *ESR* gene with ERβ expression in lung tumor tissue among study groups. Therefore, I additionally tested the alternative hypothesis (H_A) of difference in the distribution of the ER β expression among the polymorphisms in *ESR2* gene against the null hypothesis (H_0) of no difference in the distribution of the ER β expression among the polymorphisms in *ESR2* gene stratified by two major histological types (adenocarcinoma and squamous cell carcinoma) of lung cancer.

2.0 LITERATURE REVIEW

2.1 EPIDEMIOLOGY OF LUNG CANCER

It is predicted that 213,380 American will have been diagnosed with lung and bronchus cancer in 2007 alone.[1] Lung cancer is the second most commonly diagnosed cancer among men and women in the United States and accounts for 15% of all cancers in both men and women (excluding non-melanoma skin cancers and in situ cancers).[1] Among U.S. lung cancer ranks first in terms of cancer mortality in both men and women with 160,390 lung cancer deaths predicted for 2007.[1] Lung cancer deaths account for 29% of the burden of cancer mortality in the U.S. (31% for men and 26 for women).[1] Since 1990, the age-adjusted lung cancer death rate in men has been decreasing. However, mortality rate in women from lung cancer has increased more than two times in recent 25 years[1,2] According to Surveillance Epidemiology and End Results (SEER) Statistics Review 1975-2005 from National Cancer Institute, the percentage of women surviving at least five years after diagnosis has been higher than that of men during recent 25 years (18.2 vs. 13.5).[1] The percentage of men and women surviving at least five years after diagnosis is only 15.7% and has increased only 3% since 1975.[1] 49.5% of men and women with 5-year survival have localized disease.[1]

Lung cancer has two major histological types: small-cell lung cancer, and nonsmall-cell lung cancer. Non-small-cell lung cancer (NSCLC) is the major form of lung cancer which

accounts for 84.7% of invasive lung cancer in 2001-2005 and classified into three histologic types: adenocarcinoma, squamous cell carcinoma, and large cell carginoma.[1] The proportional occurrence of these histological subtypes differs significantly between men and women.[1] Adenocarcinoma is currently the most common histological subtype in both men (33.3%) and women(40.4%), and women have proportionally more adenocarcinoma and less squamous cell carcinoma compared to men.[1]

Cigarette smoking is the most important risk factor for lung cancer. 85% to 90% of all lung cancer patients have smoked cigarettes at some time in their lives.[3] However, lung cancer also occurs in non-smokers and not all smokers get lung cancer. Interestingly, nonsmokers diagnosed with lung cancer are predominately women (3 female: 1 male ratio in never smoker lung cancer patients) and the highest proportion of non-smoker with lung cancer developed adenocarcinoma.[4] Other risk factors for lung cancer are secondhand smoke, radon, asbestos, radiation, and a history of tuberculosis.[1] Genetic factors along with environmental factors play a role in lung cancer development at a younger age.[1]

2.2 HGF AND C-MET

2.2.1 Biology of HGF and c-met

Hepatocyte growth factor (HGF) was first discovered in the late 1980s.[16,17] HGF is a mainly paracrine growth factor that is secreted by fibroblasts in the lung and acts upon the c-Met receptor expressed by both epithelial and endothelial cells.[5,18] HGF is the ligand for the c-Met protein, a tyrosine kinase receptor and this ligand-receptor pair initiate signaling pathways

promoting proliferation, survival, angiogenesis, and invasion.[5] Since HGF has multiple biological functions, it is known to be a promising target for cancer therapy.[6] HGF is found many organs including the mammary gland, lung, kidney, and liver and HGF and/or c-Met is overexpressed in many human cancers such as breast, prostate, and lung.[8-10,18]

2.2.2 HGF and c-Met protein expression and human cancer

Many human cancer exhibit overexpression of HGF and/or c-Met.[8-10,18] Previous studies showed that overexpression of HGF and/or c-Met is associated with the poor prognosis of NSCLC patients.[11-13] Also, Chen and colleagues found that overexpression of HGF has significant correlation with cigarette smoking and tumor stages.[14] In vitro, nicotine upregulated HGF expression in lung cancer tissue and authors suggest that cigarette smoking may play a key role in promoting tumor progression via activation of HGF expression in tumor cells in patients with NSCLC.[14] Since women and younger lung cancer patients who have weaker association with smoking exposure develop adenocarcinoma (subtype of NSCLC) more often, HGF/c-Met in the lung tumor tissue may be a clue to the prognostic difference in gender and histological subtypes.[15]

Table 2-1 Studies for protein expression level of HGF and c-Met and lung cancer survival

Authors	Biomarker	Laboratory Assay	Participants	Mean Follow-up time	Results	Conclusion
Siegfried, Weissfeld et al. 1997[13]	HGF	Western blot[1]	Total N=56 NSCLC [ADC=47, ASC=3, BAC=3, SCC=3]	• 29 months for censored	• No significant association between ir-HGF and other clinical parameters [age (p=0.33), stage (p=0.20), smoking history (p=0.56), gender (p=0.43), histological groups (p=0.76)]. • Low ir-HGF showed significantly better overall survival compared with elevated ir-HGF survival (p=0.03, log-rank test). • In Cox model, risk continuously increases as ir-HGF (as a continuous variable) increases (RR=4.11 for ir-HGF level 70 vs. level 5)	Elevated ir-HGF is a negative prognostic indicator in NSCLC.
Siegfried, Weissfeld et al. 1998[18]	HGF	Western blot[1]	Total N=56 NSCLC [ADC=47, ASC=3, BAC=3, SCC=3]	• 29 months for censored • 12.2 months for deceased	• Elevated HGF is associated with poor disease-free and overall survival (p=0.01, log-rank test and Wilcoxon test for disease-free survival). • Elevated HGF with stage I had a worse survival than low HGF with high stage. • In a multivariate Cox analysis, RR=10 for HGF greater than 100 units vs. HGF level of 1 unit.	HGF is a negative prognostic indicator in lung cancer.
Siegfried, Luketich et al. 2004[15]	HGF	Western blot[1]	Total N=59 NSCLC [ADC=48, ASC=6, BAC=5]	• 61 months for censored	• No significant association between HGF levels and other variables (age, gender nodal status, stage, smoking history). • In multivariate Cox analysis, high-HGF group is statistically significant associated with poor survival (RR=2.2 for all-cause survival, 3.0 for lung cancer survival, and 3.3 for disease-free survival).	HGF is a negative prognostic indicator at all stages of disease for adenocarcinoma.

Table 2-1 (continued)

Ichimura, Maeshima et al. 1996[11]	• c-Met • HGF (only with 11 cell lines)	• Western blot[2] • IHC used only to confirm c-Met[2][3]	Total N=104 NSCLC [ADC=47, SCC=52, Others=5]	• No mean follow-up reported • survival curve with 4 years follow-up after surgery	• Adenocarcinoma with high c-Met protein expression showed worse outcome than those without c-Met expression (p<0.01). [KM method used for survival based on western blot analysis alone] • c-Met is more frequently expressed in ADC than in SCC. • Strong intensity of c-Met is more frequently expressed in the higher stage. • IHC results were identical with western blot in most cases, but 17 tumors (16.3%) showed a discrepancy.	c-Met is closely related to progression of adenocarcinomas of lung.
Takanami, Tanana et al. 1996[12]	• c-Met • HGF	IHC[6][7]	Total N=120 ADC	• No mean follow-up reported • survival curve with 5 year follow-up	• The prognosis was significantly worse in the HGF-positive or c-Met-positive patients than in the negative patients. • c-Met had a significant effect on the prognosis, whereas HGF did not. (based on multivariate analysis) • No significant relationship between clinicopathology and HGF expression. • Significant relationship between stage and c-Met expression (p>0.05).	• c-Met expression is an independent poor prognostic marker in ADC. • HGF tumor expression in ADC was a poor prognostic marker, but only in univariate analysis

Table 2-1 (continued)

Masuya, Huang et al. 2004[6]	• c-Met • HGF	IHC[2][4]	Total N=88 NSCLC [ADC=46, SCC=29, LCC=13]	• 49.8 ±36.1 months for all patients	• Frequency of intratumoral c-Met-negative tumors was significantly higher for less advanced stages (p=0.0169). • Intratumoral c-Met status is a significant factor for predicting the prognosis of NSCLC patients (RR=2.642, p=0.0029) • None of the carcinomas were stromal c-Met-positive. • Survival rate of patients with intratumoral c-Met positive tumours were significantly lower than for patients with c-Met negative tumours (p=0.0095) • No significant difference in survival among patients in relation to intratumoral and stromal HGF status.	• c-Met expression is a negative prognostic factor for NSCLC patients.
Nakamura, Niki et al. 2007[19]	• c-Met • HGF	IHC[5]	Total N=130 ADC [papillary=62, acinar=5, solid=21, BAC=15, mixed=27]	• No mean follow-up reported • survival curve with 80 months follow-up	• High levels of c-Met expression correlated with higher pathological stage(≥IIIA) (p=0.006). • No siginificant differences in survival among cases grouped according to their expression of HGF, c-Met and phosphor-c-Met. [Only Kaplan-Meir method used]	• Neither HGF nor c-Met expression are associated with survival of lung adenocarcinoma patients.

Abbreviations: NSCLC=non-small cell lung cancer, ADC=adenocarcinomas, ASC=adenosquamous carcinomas, BAC=bronchiole-alveolar carcinomas, SCC=squamous cell carcinomas, LCC=large-cell carcinomas, RR=relative risk, IHC=Immunohistochemistry, ir=immunoreactive

[1] a goat polyclonal anti-HGF antibody (R&D systems, Minneapolis, MN)

[2] a rabbit polyclonal anti-human c-met anti-body (SC-10, Santa Cruz Biotechnology, Inc., Delaware, CA)

[3] anti-c-met antibody (#18321, IBL Laboratories)

[4] a rabbit polyclonal antibody against HGF (SC-7949, Santa Cruz Biotechnology INC., Santa Cruz, CA)

[5] a rabbit polyclonal anti-HGF-α and rabbit polyclonal anti-c-Met antibodies (IBL, Gunma, Japan)

[6] a rabbit antihuman HGFα polyclonal antibody (#18131, Immune Biotechnology Lab., Fujioka, Gumma, Japan)

[7] a rabbit antihuman Met polyclonal antibody (SC-28, Santa Cruz Biotechnology INC., Santa Cruz, CA)

2.3 ESTROGEN RECEPTOR-BETA

2.3.1 Human genetics of *ESR2*

The official (HUGO) name of estrogen receptor beta is estrogen receptor 2 (ER-beta). The official symbol is *ESR2*. The aliases of this gene are Erb, ESRB, ESTRB, NR3A2, ER-BETA, and ESR-BETA. In the human genome, the *ESR2* gene is located on chromosome 14, band q23.2. The size of the entire coding sequence (introns and exons) of *ESR2* gene is approximately 61.2 kilobases. There are 8 exons in the human *ESR2* gene. Also, there are 2 additional untranslated exons, 0N and 0K, in the 5' region and an exon at the 3' end. It measures 468 bases at the 5' untranslated region (UTR), and 108 bases at the 3' UTR.[20,21] The total number of amino acids in *ESR2* gene (residue/ translational length) is 530.[22]

Since *ESR2* is a member of the nuclear receptor superfamily, it has common structural characteristics of this family including five distinguishable domanins named the A/B, C, D, E, and F, respectively.[23,24] The A/B domain is the N-terminal domain which is the most variable region and normally contains a transactivation domain that can interact directly with factors of the transcriptional machinery.[25] The C domain is the DNA binding domain which involved in specific DNA binding and the transactivation capacity of the receptor.[24,25] The D domain is referred as the hinge domain since it works as a flexible hinge between ligand binding domain and the DNA binding domain. The E domain is the ligand binding domain since it contains different sets of amino acids that bind to different ligands. Even though ERα and ERβ are the subtypes of estrogen receptors, these receptor subtypes only shares only 55% of the amino acids

sequence for the ligand binding domain. This may results in different affinities of ligand binding between ERα and ERβ.[25] The functions of F domain remain undefined.[24]

There are five full-length transcripts due to alternative splicing in the human *ESR2* gene.[24] ERβ1 is a full-length isoform of human ERβ protein with 530 amino acids and a molecular weight of 59.2 KDa translated from 8 exons. Other full-length ERβ2-5 are translated from same sequences with ERβ1 from exon 1 to exon 7 but a unique C-terminus, where the amino acids corresponding to exon 8.[26,27]

The *ESR2* gene has been related to those diseases such as Alzheimer's disease in women, breast cancer, bone mineral density, ovarian cancer, coronary artery disease, and prostate cancer (Table 2-2). However, there are no mutations known to cause any specific phenotypes or disease at this point. *ESR2* gene product is expressed in human tissues or cells from vascular endothelium and regions of the brain, retina, thyroid, lung, bladder, ovary (granulose cells), breast, colon, bone marrow, prostate (epithelium), and white blood cell.[28] The general function of the *ESR2* includes estrogen receptor activity, lipid binding, metal ion binding, protein binding, receptor antagonist activity, sequence-specific DNA binding, steroid binding, transcription coactivator activity, transcription factor activity, and zinc ion binding. It is involved in the estrogen receptor signaling pathway.[29]

2.3.2 *ESR2* expression in adult lung tissue

It is important to know whether or not normal adult lung tissue expresses *ESR2*. I hypothesize estrogen mediated sex-related differences in lung cancer risk and lung cancer related outcomes. Since women make more estrogen than men, estrogen may explain male-female differences in lung cancer. Moreover, I believe that the *ESR2* gene product (ERβ) is the biological factor

primarily for mediating these effects. It is hard to accept these notions, unless it can be shown that normal lung tissue and/or lung cancer tissue expresses *ESR2*.

Soon after cloning *ESR2*, the discoverers of *ESR2* examined *ESR2* expression in normal human tissues "obtained after surgery performed for different reasons".[28] Using four different *ESR2* oligonucleotide probes, *in situ* hybridization detected *ESR2* mRNA in lung parenchyma and pulmonary blood vessels. Taylor *et al*. (2000) studied "normal human tissue samples obtained from adult human cadavers *post mortem* or from patients at the time of surgery for various pathological conditions".[30] Immunohistochemistry with two polyclonal rabbit anti-rat ERβ antibodies, one against N-terminal and a second against C-terminal sequences, showed nuclear ERβ expression in bronchiolar columnar epithelial, intermediate, basal, and smooth muscle cells. Omoto *et al*. (2001) studied histologically normal lung tissue obtained at surgery from 35 lung cancer patients. In every case, staining with an anti-ERβ chicken IgY polyclonal antibody showed nuclear ERβ expression in more than a quarter of normal bronchial epithelial cells.[31]

Fasco *et al*. (2002) studied normal (uninvolved) and malignant (involved) lung tissues obtained surgically from 26 patients with stage I or II lung cancer.[32] Reverse transcription-polymerase chain reaction (RT-PCR) detected *ESR2* transcripts in uninvolved lung tissue from nine (35%) of the 26 lung cancer patients. Mollerup *et al*. (2002) studied normal (adjacent-to-tumor) lung tissue obtained at surgery from 46 non-small cell lung cancer patients.[33] In every instance, quantitative RT-PCR detected *ESR2* mRNA (mean optical densitometry units relative to GAPDH ± standard deviation, 1.06±0.81 in women and 1.16±0.77 in men).

At the University of Pittsburgh Cancer Institute, Stabile *et al*. (2002) studied six normal lung fibroblast cell lines and three primary bronchial epithelial cell cultures produced from upper

airway biopsies obtained from lung cancer patients at the time of surgery.[34] In every instance, RT-PCR detected *ESR2* mRNA. In addition, Western analysis with a rabbit polyclonal anti-ERβ antibody (PanVer, Madison, WI) directed against the C-terminus (amino acids 512-530) of ERβ detected the full length (59kDa) *ESR2* protein product. Finally, Schwartz *et al*. (2005) used the MCA-1974S antibody (Serotec, Oxford, UK), directed against the C-terminus of ERβ1, to stain ten normal lung tissue samples obtained at autopsy of patients without history of cancer.[35] Using a conservative threshold (at least weak (1+) staining of at least 10% of cells), these investigators observed "lung bronchus tissue" ERβ expression in 2 (20%) of ten samples.

Four of the seven studies cited used normal appearing lung tissue harvested from lung cancer patients. Nevertheless, the published literature permits the assertion that *ESR2* expression has been observed in normal lung tissue from adult humans. If *ESR2* is expressed in normal lung tissue, we can speculate that specific genotype (polymorphism) of *ESR2* may produce higher ERβ protein in normal lung tissue. In that case, subjects with ERβ overexpression in normal lung may be predisposed to lung cancer and ERβ overexpression or underexpression in lung tumor tissue may exert antitumoral effects. However, if the normal tissues are from the lung cancer patients, we cannot make any assumption on the role of the ERβ expression for the tumor development in normal lung tissue. Therefore, in this project, the tumor tissues from lung cancer patients were used to measure ERβ protein expression status and genotype variants of *ESR2* gene.

2.3.3 ER-beta (ERβ) protein expression and lung cancer

ERβ, a second isoform of ER, was discovered in 1996.[20] Until the discovery of the ERβ, the estrogen receptor studies could not distinguish between ERα and ERβ. Nuclear ERβ positivity was presented in 61% of lung tumor tissue and 20% of normal lung tissue sample by using

immunohistochemistry.[35] A study demonstrated the survival differences between gender: women with ERβ expression in tumor tissue had a increase in mortality, whereas men with ERβ expression had a significant reduction (55%, p=0.04) in mortality compared with those with ERβ negative tumors.[35] Overexpression of ERβ was significantly more frequent in tumors occurring in lung cancer patients without smoking history (53.5%) than in those with smoking history.[36] It is found that ERβ overexpression is statistically significant favorable prognostic indicator for lung cancer patients.[36] Kawai and the colleagues found that absence of ERβ expression is correlated with poorer overall survival and can be an independent factor predictive of poor disease outcome of non-small cell lung cancer patients (hazard ration, 1.9; 95% confidence interval, 1.1-3.4; p=0.0264).[37] These studies investigating the expression of ERβ in lung cancer were conducted by using immunohistochemical staining method. Based on the study findings, ERβ protein expression status can be a potential biomarker identifying patients at high risk.

2.3.4 *ESR2* gene variants and disease association studies

Three frequently studied *ESR2* genetic variants are (1) rs1256049 [RsaI]: a silent G1082A SNP in exon 6 (ligand binding domain), (2) rs4986938 [AluI]: A1730G SNP in the 3'-untranslated region of exon 8, and (3) CA dinucleotide repeat polymorphism in intron 5 (Table C-1). The inheritance of one or another of these three specific *ESR2* genetic variants has been studied in relation to cancers of the colon or rectum [38], endometrium [39], ovary [40], testis [41], prostate [42-44], and breast.[45-53]

From an OVID Medline literature search (see page140), sixteen articles investigated the association between *ESR2* genetic variants and human cancer are identified and evaluated (Table 2-2). Only one out of four studies showed the association between *ESR2* SNP variants and

prostate cancer risk: rs29877983 located in the promoter region was significantly associated with prostate cancer risk (OR=1.22, 95% CI=1.02–1.46) and with localized carcinomas (OR=1.33, 95% CI=1.08–1.64).[54] The one available study on *ESR2* gene and colon and rectal cancer showed that G allele of rs1256049 is associated with increased risk of rectal cancer among the total population if diagnosed before 60 years of age (OR, 1.68; 95% CI, 1.02-2.79).[38] Seven out of nine breast cancer studies found statistically significant association between breast cancer and either single variants or haplotypes or CA repeat of *ESR2* (Table 2-2). However, only two of them showed the association with single variants of *ESR2*: (1) rs8018687 (*5772G) and rs4986938 (*38A) are associated with breast cancer risk in women with benign breast disease[51], and (2) C(14206)T and rs1256054 are associated with breast in postmenopausal women.[46]

Many studies investigated the relationships between single nucleotide polymorphisms (SNPs) in the human *ESR2* gene and non-cancer disease. *ESR2* polymorphisms are significantly associated with bone mass in both men and women.[55] Caucasian women with the TC genotype for *ESR2* rs1256030 had lower LS-BMD than did those with the CC genotype (P=0.02).[56] Wang firstly detected significant association of *ESR2* with hip fractures (rs960070: P=0.0070, OR=1.43, 95%CI: 1.10-1.86) and this findings are also supported by haplotype analyses.[57]

A nested case-control study with Spanish population showed that rs1271572 SNP T variant of *ESR2* was significantly more common in patients who developed myocardial infarction (P < 0.001). Assuming a dominant model of inheritance, the association remained statistically significant in men [odds ratio (OR) 1.65, 95% CI 1.18-2.30; P = 0.003) but not in women (P = 0.754).[58] The rs1256030 and rs1256065 SNP of *ESR2* were associated with HDL cholesterol concentrations in Chinese women (P=0.05).[59] Single variants in the *ESR2* gene is associated with an increased risk of Alzheimer's disease in women (OR=1.87, 95% CI=1.21-

2.90), whereas it does not contribute to the disease susceptibility in men.[60] It is interesting to observe that the relationship between variations in *ESR2* gene and diseases may differ by gender.

Table 2-2 *ESR2* genetic variants and cancer studies

Author	Cancer	Associated	Variants studied	Results
Thellenberg-Karlsson, Lindstrom et al. 2006[54]	Prostate	Yes	28 single nucleotide polymorphisms (SNP) spanning the entire ERbeta gene from the promoter to the 3'-untranslated region	only one polymorphism (rs29877983) located in the promoter region was significantly associated with PC risk (OR=1.22, 95% CI=1.02–1.46) and with localized carcinomas (OR=1.33, 95% CI=1.08–1.64).
Chen, Kraft et al. 2007[43]	Prostate	Yes	Four haplotype tags (rs3020450, rs1256031, rs1256049(RsaI), rs4986938(AluI))	No association between the four tag SNPs in *ESR2* and PC risk. However, we observed that men carrying two copies of one of the variant haplotypes (TACC) had a 1.46-fold increased risk of prostate cancer (99% confidence interval, 1.06-2.01) compared with men carrying zero copies of this variant haplotype.
McIntyre, Kantoff et al. 2007[44]	Prostate	No	CA repeat polymorphism in intron 5 of *ESR2*	Unassociated with prostate cancer
Nicolaiew, Cancel-Tassin et al. 2009[61]	Prostate	No	Eleven polymorphisms: four in the coding region (rs1256049(RsaI), two in introns(rs944050), and five in the 3'UTR (rs4986938(AluI), rs928554 and rs28440970)	No association between those polymorphisms and PC risk
Leigh Pearce, Near et al. 2008[40]	Ovary	Yes	Five htSNPs (rs4986938(AluI), rs944046, rs1256049(RsaI), rs1256031, rs3020450)	No significant association between the five htSNPs and ovarian cancer. However, Haplotype D (CACAC) increased risk of invasive clear cell ovarian cancer (odds ratio, 3.88; 95% confidence interval, 1.28-11.73; P = 0.016). Haplotype D possibly associated with ovary cancer
Setiawan, Hankinson et al. 2004[39]	Endometrium	No	rs1256049(RsaI), rs1271572, CA repeat	Unassociated with endometrial cancer: rs1256049 (OR = 1.2; 95%CI: 0.7-2.3), rs1271572 (OR = 0.8; 95%CI: 0.5-1.1) and CA repeat (22 repeat allele versus > or = 22 repeat allele, OR = 1.1; 95%CI: 0.7-1.7)

Table 2-2 (continued)

Slattery, Sweeney et al. 2005[38]	Colon	Yes	rs1256049 (RsaI) and CA repeat	No association with risk of colon and rectal cancer. However, G allele of rs1256049 associated with increased risk of rectal cancer among the total population if diagnosed before 60 years of age (OR, 1.68; 95% CI, 1.02-2.79). Having two 25 or more CA repeats in ERß was associated with an increased relative risk of colon cancer in women [odds ratio (OR), 2.13; 95% confidence interval (95% CI), 1.24-3.64] but not in men
Maguire, Margolin et al. 2005[48]	Breast	Yes	rs1256049(RsaI), rs4986938(AluI), and rs928554 (Cx+56 A-->G)	No overall association for any of the SNPs studied. However, One haplotype possibly associated with sporadic breast cancer(OR = 3.0, p = 0.03)
Tsezou, Tzetis et al. 2008[52]	Breast	Yes	Repeat polymorphisms c. 1092+3607(CA)(10-26)	Associated with breast cancer
Breast and Prostate Cancer Cohort Consortium, Cox et al. 2008[53]	Breast	Yes	Four htSNPs: rs4986938(AluI), rs1256049(RsaI), rs1256031, rs3020450	None of the SNPs were independently associated with breast cancer risk; one haplotype possibly associated (OR 1.17, 95% CI 1.07-1.28, p = 0.0007)
Gallicchio, Berndt et al. 2006[51]	Breast	Yes	rs4986938(AluI,G1730A), rs8018687, rs928554, A5696G (no rs number)	*ESR2* rs8018687 (*5772G), rs4986938 (*38A) associated with breast cancer risk in women with benign breast disease
Gold, Kalush et al. 2004[47]	Breast	Yes	8 SNPs (rs1152579, rs1255998, rs1256030, rs1256049 (RsaI, G1082A), rs1271572, rs4986938 (AluI, G1730A),rs928554, E2Ex4CorT)	Several *ESR2* haplotypes associated with breast cancer risk

Table 2-2 (continued)

Zheng, Zheng et al. 2003[46]	Breast	Yes	eight sequence variants rs1271572, G(-11943)A, T(-11891)C, C(14206)T, rs1256049(RsaI), rs1256054, A(50766)G, G(50995)A	C(14206)T and rs1256054 associated with breast in postmenopausal women
Iobagiu, Lambert et al. 2006[49]	Breast	No	CA repeat	Unassociated with breast as single variant, possibly associated in combination with other ESR1 or AR repeat polymorphisms
Georgopoulos, Adonakis et al. 2006[50]	Breast	No	rs1256049(RsaI) and rs4986938(AluI) polymorphisms	Unrelated endometrial pathology in Tamoxifen treatement women or the stage of breast cancer
Forsti, Zhao et al. 2003[45]	Breast	No	six studied polymorphisms: rs1256049 (RsaI, G1082A), rs4986938 (AluI, G1730A), (Nt805(del21), G864A, A1505-4G, CA repeat in intron 5	No association

2.4 SIGNIFICANCE

It is estimated that 160,390 people died due to the lung cancer in 2007 in U.S alone.[1] However, the percentage of surviving at least five years after lung cancer diagnosis has increased only 3% since 1975.[1] Also, men and women have different proportion of histological subtype of lung cancer: women have proportionally more adenocarcinoma and less squamous cell carcinoma compared to men.[1]

There is only few effective treatment options are available for lung cancer patients. Therefore, understanding of gender difference in cancer development and susceptibility may lead to innovative therapeutic approaches. It is also important to understand the action of as growth factors and hormone receptors because not only their biological function which may be a clue to the prognostic difference in gender and histological subtypes but also their potential clinical implication as an indicator for the selection of appropriate treatment.

It is also important to investigate the inter-correlations among biomarkers because it may provide enhanced insight and understanding of the complexity of molecular mechanisms. Although strong experimental evidence suggests that *ESR2* play a role in carcinogenesis, the results of epidemiologic investigations are less persuasive. For example, a few polymorphic variants of the *ESR2* gene have been associated with an increased risk of common cancers like prostate, colorectal, and breast cancers in some studies (Table 2-2). In addition, no study we are aware of has yet examined the association between *ESR2* gene polymorphisms, ERβ protein expression status and lung cancer risk and survival.

3.0 HGF AND C-MET: IMMUNOHISTOCHEMICAL EXPRESSION AND LUNG CANCER SURVIVAL

To be submitted for publication

Ji Young Song,[1] Jill Siegfried,[3] Brenda Diergaarde,[1] Stephanie Land,[2] Robert Bowser,[4] Laura Stabile,[3] Sanja Dacic,[4] Rajiv Dhir,[4] Joel L. Weissfeld[1]

[1] Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

[2] Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

[3] Department of Pharmacology and Chemical Biology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania

[4] Department of Pathology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania

This work was supported in part by National Institutes of Health grant P50 CA090440 SPORE in Lung Cancer and R25 CA057703 Education Programs in Cancer Prevention.

Abbreviations Used: HGF, hepatocyte growth factor; IHC, immunohistochemistry; OS, overall survival; TMAs, tissue microarrays

3.1 ABSTRACT

Background: Previous studies have shown association between the lung tumor expression of hepatocyte growth factor (HGF) and factors (smoking and advanced stage) related to lung cancer outcome. These observations motivated a direct study of lung cancer survival association with the lung tumor expression of HGF and its receptor (c-Met).

Methods: We used immunohistochemistry (IHC) to measure HGF and c-Met expression in primary lung tumor tissue semi-quantitatively from n=180 patients, including n=115 represented as multiple cores on tissue microarrays (TMA) and n=65 represented as single whole tissue sections. We used sample type-specific (TMA vs. whole section) Allred score median cutpoints to distinguish high expression from low expression. To identify baseline factors related to HGF and c-Met expression, we used a generalized linear mixed model (GLIMMIX) approach, which controlled for sample type (TMA vs. whole section) and accounted for the correlated nature of the TMA core-level data. We used Cox (proportional hazards) regression to evaluate survival associations with expression. All models included factors for age, smoking, stage, sex, race, and histology.

Results: 43.8% and 44.1% of lung tumors showed high HGF and c-Met expression, respectively. GLIMMIX showed borderline significant associations between high HGF expression and stage [Odds Ratios (OR) relative to stage IA: stage IB 0.66 (95% confidence interval 0.28-1.58), stage II 1.26 (95% CI 0.47-3.39), and stage III/IV 0.43 (95% CI 0.19-0.98), P_{global}(Type III)=0.05] and between high HGF expression and smoking [(OR relative to never smoker: active smoker 2.35

(95% CI 0.74-7.43) and ex-smoker 3.08 (95% CI 0.99-9.57), P_{global}(Type III)=0.14]. Neither HGF [Hazard Ratio (HR) 0.87, (95% CI 0.59-1.29)] nor c-Met expression [HR 1.06, (95% CI 0.71-1.58)] predicted survival. Associations between high expression and survival were statistically similar in men and women.

Conclusion: HGF immunochemical expression in lung tumor correlates positively with smoking (an observation consistent with previous reports) and negatively with advanced cancer stage (an observation contrary to previous reports). After accounting for stage and other factors, neither HGF nor c-Met expression predicted survival.

3.2 INTRODUCTION

Lung cancer accounts for 29% of U.S. cancer deaths (31% for men and 26% for women). Female lung cancer mortality has increased more than two-fold in the last 25 years [1, 2].

Hepatocyte growth factor (HGF), first discovered in the late 1980s [3, 4] is a paracrine growth factor secreted by lung fibroblasts. Lung epithelial and endothelial cells express c-Met, a only known tyrosine kinase receptor of HGF [5, 6]. HGF c-Met ligand receptor binding initiates signaling pathways that promote cell proliferation, migration, survival, angiogenesis, and invasion [6]. These biological functions make the HGF c-Met pathway a promising cancer therapy target [7].

Some studies show association between HGF and/or c-Met expression and poor prognosis after diagnosis of non-small cell lung cancer (NSCLC) [8-10]. NSCLC accounted for 84.7% of 2001-2005 U.S. lung cancer cases[1]. Adenocarcinoma, the most common histological

NSCLC, is proportionally more common in women than men [1] and in never smokers than ever cigarette smokers [11].

Chen and colleagues found that overexpression of HGF has significant correlation with cigarette smoking and tumor stages [12]. In vitro, nicotine upregulated HGF expression in lung cancer tissue and authors suggest that cigarette smoking may play a key role in promoting tumor progression via activation of HGF expression in tumor cells in patients with NSCLC [12]. Since women and younger lung cancer patients who have weaker association with smoking exposure develop adenocarcinoma (subtype of NSCLC) more often, HGF/c-Met in the lung tumor tissue may be a clue to the prognostic difference in gender and histological subtypes [13]. The purpose of this study is to investigate the association between the expression of HGF/c-Met and risk factors, including sex, smoking status, and histology group, and to elucidate the prognostic significance of HGF/c-Met immunochemical expression in the tumor lung.

3.3 METHODS

3.3.1 Study Population

The study population included n=203 subjects age 21 years and older who received surgery at a University of Pittsburgh Medical Center hospital for the staging or treatment of pathologically confirmed primary lung cancer. We assembled risk factor, tumor, and follow-up information from several sources, including outpatient paper charts, inpatient and outpatient electronic medical records, hospital-based cancer registries, and Social Security Death Index database searches. The research used formalin-fixed and paraffin-embedded tissue specimens, processed

as tissue microarray (TMA) cores (n=126 subjects) or as whole tissue sections (n=77 subjects). Data analyses included 180 subjects (115 of 126 on TMA and 65 of 77 on whole sections) with non-missing lung tumor expression data and with known survival outcome. The University of Pittsburgh Institutional Review Board approved subject recruitment and tissue use protocols.

3.3.2 Laboratory Methods

TMA construction included three 0.6mm diameter lung tumor cores per subject with examination of hematoxylin- and eosin-stained sections to verify malignant content. Preparations for immunohistochemisty (IHC) included deparaffinization and hydration with xylene and ethanol, heat induced antigen retrieval with 10mM citrate buffer at pH 6, quenching endogenous peroxidase with 3% hydrogen peroxide for 5 min at room temperature, and blocking with non-immune normal serum for 5-20 min at room temperature. HGF staining used anti-HGF (AB-294-NA, R&D Systems) at 1:200 dilution in PBS and EnVision™ reagents (DAKO Corp., Carpinteria, CA). c-Met staining used anti-c-Met (SC-10, Santa Cruz) at 1:150 dilution in PBS and the MACH 4 Universal HRP-Polymer Kit with DAB (Biocare Medical, LLC. Concord, CA). Final steps consisted of incubation with diaminobenzidine (DAB) chromogenic substrate at room temperature for 5-10 min and counterstaining with hematoxylin for 2-2.5 min. Breast cancer tissue, with and without the application of primary antibodies, were used as positive and negative IHC controls.

The study lung pathologist (S.D.) assessed each TMA core and whole section for percentage of tumor cells stained and for intensity of staining. Scoring for the percentage of tumor cells stained used a six-level ordinal scale (0 to 5, respectively, for no cells stained, 0-1%

cells stained, 2-10% cells stained, 11-33% cells stained, 34-66% cells stained, and 67-100% cells stained). Scoring for intensity of staining used a four-level ordinal scale (0 to 3, respectively, for no, weak, moderate, and strong staining). Data analyses used a semi-quantitative measure of IHC expression in terms of the Allred score (range 0 to 8), the sum of the percentage and intensity scores [14].

3.3.3 Statistical Analysis

Variables used in data analyses included age at tissue collection (continuous and categorical), race (White, African-American), sex (women, men), smoking status (never, former, active), smoking dose duration among ever smokers (<50 pack-years, 50+ pack-years), pathologic stage group (IA, IB, IIA/IIB, III/IV), and histology group (squamous cell carcinoma, non-squamous non-small cell lung cancer, undifferentiated carcinoma, and small cell carcinoma). The non-squamous non-small cell lung cancer group included adenocarcinoma, adenosquamous carcinoma, bronchioloalveolar carcinoma, and malignant carcinoid. The undifferentiated carcinoma group included large cell carcinoma and undifferentiated non-small cell lung cancer. Ever cigarette smokers with unknown quit status were grouped with active smokers. For subjects without pathologic stage information, clinical stage information was used instead.

We used Wilcoxon rank sum and Fisher's exact tests to evaluate the statistical significance of differences according to sample type (TMA *vs.* whole section). Single variable analyses of factors related to HGF or c-Met expression used subject-specific Allred scores averaged across TMA cores and applied sample type-specific median cutpoints to distinguish between high and low expression. Multi-variable analyses of factors related to HGF and c-Met

expression preserved core-level Allred data and used generalized linear mixed models (SAS PROC GLIMMIX) to control for sample type and account for the correlated nature of the TMA core-level data. We specified a first-order autoregressive covariance structure for random core-level effects within subject.[i] All models included variables for age at tissue collection, sex, smoking status, and stage.

Survival analyses modeled times between dates of primary surgery and dates of death, with survivors censored on dates last contacted alive. We used the Kaplan-Meier product limit estimator and the log-rank statistic to estimate survival and to evaluate the statistical significance of differences according to IHC expression level (subject-specific averaged Allred values above or below sample type-specific median cutpoints). We used Cox proportional hazards regression to adjust survival for differences in age at tissue collection, sex, smoking status, and stage. Finally, we used standardized score process plots and Kolmogorov-type supremum tests to evaluate proportional hazards assumptions.

All analyses used SAS version 9.2 (SAS Institute, Inc., Cary, North Carolina) and two-sided *p*-values.

3.4 RESULTS

Fifty one percent of the subject group were women, 9.2% African-American, and 90.8% white (Table 3-1). Mean age was 66.4 years. Few were never smokers (5.8%). Fifty-eight percent had adenocarcinoma, bronchioloalveolar carcinoma, or adenosquamous carcinoma. Median Allred

[i] Statistical Analysis with the GLIMMIX Procedure Course Notes, Copyright©2009 SAS Institute Inc. Cary, NC, USA.: Page 2-87

scores were 7 for HGF and 7 for c-Met. With Allred score greater than 7 representing high expression, high HGF expression was observed in 58.8% of tumors from the TMA set and in 29.1% of tumors from the whole section set (p=0.0003). High c-Met expression was observed in 42.9% from the TMA set and in 53.5% from the whole section set (p=0.2532). A Table 1-2 footnote shows sample type-specific cutpoints used in subsequent analyses to distinguish between high and low expression. Representative immunostaining of each marker is shown in Figure 3-1.

Table 3-2 shows subject characteristics by HGF and c-Met expression status. High HGF expression was less frequent and high c-Met expression more frequent in tumors from African-American subjects (HGF: 28.6% African-American vs. 45.6% white; c-Met: 64.3% African-American vs. 42.7% white). High HGF expression was more frequent in tumors from women (47.7%) than in tumors from men (39.8%). Ex-smokers and active smokers more often had high HGF and c-Met expression than never smokers (HGF: 60.0% of ex-smokers, 46.3% of active smokers, versus 20.0% of never smokers; c-Met: 44.4% of ex-smokers, 45.1% of active smokers, versus 22.2% of never smokers). Interestingly, only 12.5% of stage IV tumors showed high HGF expression (Table 3-2).

In generalized linear mixed models (GLIMMIX) that adjusted for age, smoking, stage, and sex, men relative to women had 37% and 9% lower odds of high HGF and c-Met lung tumor expression, respectively (HGF: OR=0.63, 95% CI 0.36-1.11; c-Met: OR=0.91, 95% CI 0.52-1.59). However, these sex-related differences were not statistically significant (Table 3-3). High HGF expression was observed more often in ex-smokers (OR=3.08, 95% CI 0.99-9.57, p=0.05) and in active smokers (OR=2.35, 95% CI 0.74-7.43, p=0.15) than in never smokers and less often in stage III/IV than in stage IA (OR=0.43, 95% CI 0.19-0.98, p=0.05). High c-Met

expression was much more frequent in lung tumors from African-American subjects than in lung tumors from white subjects (OR=2.66, 95% CI 1.07-6.59, p=0.03). Lung tumor c-Met expression did not differ statistically according to age, smoking status, stage, or histology group.

Mean and median follow-up times were 5.5 years (standard deviation 0.46 years) and 3.3 years, respectively. Cumulative survival at one and three years was 78.1% (139 of 178) and 48.2% (80 of 166), respectively. Univariate (Kaplan-Meier) analyses showed similar survival in high and low HGF expression (p=0.33 log-rank test; Figure 3-2) and in high and low c-Met expression categories (p=0.82 log-rank test; and Figure 3-3).

In Cox proportional hazards regression models that adjusted for age, sex, stage, and smoking status, mortality was significantly higher in men than in women (HR=1.51, 95% CI 1.03-2.22), higher in active smokers (HR=2.60, 95% CI 1.15-5.86) and in ex-smokers (HR=1.27, 95% CI 0.56-2.88) than in never smokers, and higher in more advanced stage than in less advanced stage (Table 3-4). However, adjusted analyses still showed no statistically significant survival differences according to HGF or c-Met expression (HGF: HR=0.87, 95% CI 0.59-1.29; c-Met: HR=1.06, 95% CI 0.71-1.58). Statistically significant survival differences were not observed in subgroups restricted to male or female sex (Table 3-5) or in a subgroup restricted to non-squamous non-small cell histology (data not shown).

3.5 DISCUSSION

In this prospective cohort study of lung cancer patients, we found that HGF immunochemical expression in lung cancer was associated positively with smoking and negatively with more

advanced cancer stage. A lung cancer patient who is an active smoker, for example, had 135% higher odds of high HGF expression compared to an otherwise similar patient without a smoking history.

Out results are consistent with a previous finding of a positive association between high expression of HGF and cigarette smoking status, though our results did not achieve statistical significance [12]. Several studies have investigated the associations between the expression of HGF [7, 9, 10, 12, 15, 16] or c-Met [7-9, 15] in lung cancer and clinicopathological parameters. Chen *et al.*, for instance, observed the correlation between the high expression of HGF in lung tumor and higher stage group [12]; however, others did not show any significant association between HGF expression in lung cancer patients and other clinical parameters such as age, stage, smoking history, gender, histological groups [7, 9, 10, 15, 16]. While previous studies reported that high level of c-Met expression was correlated with higher pathological stage, our study did not show any statistical associations between c-Met expression and clinical parameters, except race [7-9, 15].

Few studies in humans have investigated the association between HGF or c-Met protein expression in relation to lung cancer survival. Detection of HGF or c-Met protein expression has been most commonly performed using Western blots [5, 8, 10, 16]; another method used is the immunohistochemical staining method [7, 9, 15]. Ichimura *et al.* also used IHC; however, just to confirm the results of western blot analysis for c-Met and showed 16.3% discrepancy between the results obtained by the two methods [8]. Depending on the method used, previous studies reported inconsistent findings on the prognostic significance of HGF protein expression in the tumor lung. All of our previous studies were based on western blot analysis and showed that elevated HGF expression in tumor tissue is associated with poor survival in non-small cell lung

cancer patients, specifically adenocarcinoma [5, 10, 16]. All of previous studies, which used the IHC method, did not show the HGF expression as a significant independent prognostic marker for lung cancer patients [7, 9, 15]. Only one IHC method based study reported that high HGF expression was associated with poor clinical outcome of lung cancer patients in the univariate analysis, but not in a multivariate context [9].

The contrary finding between our previous studies [5, 10, 16], which observed poorer survival in patients with non-small cell lung tumors expressing more HGF, and this study, which has a similar study population, might due to the difference in the methods of measuring the protein expression in lung tumors. Previous studies using western blot quantified the protein expression level in a uniformed manner; however, this study with IHC method only could measure the protein expression semi-quantitatively through assessing the percentage and intensity of tumor cells stained by a pathologist. This semi-quantitative method of measuring IHC expression of marker was used in this study because there is no standardized method to measure IHC expression of markers objectively.

Regardless of the methods used for protein detection, c-Met expression was a negative prognostic factor for lung cancer patients [7-9], except only one recent study of Nakamura and his colleagues [15]. Nakamura *et al*. performed immunohistochemical analyses on 130 patients with adenocarcinomas of lung, the largest sample size ever reported, and showed no significant differences in survival between patients in relation to expression of c-Met [15]. In our study, neither HGF nor c-Met expression was statistically significant biomarkers that predict the overall survival of lung cancer patients. Considering the type of method utilized for protein detection, our study supports the previous findings on the HGF expression in relation to lung cancer survival.

One of our goals was to investigate if HGF and c-Met may be the potential factor which explains the gender differences in prognosis shown in many epidemiological studies. While in our study, men were less likely to have high expression of HGF and c-Met compared to women, the risk of death for men who did show a high expression of HGF was decreased by 31% (p=0.25) while that of women decreased by only 8% (p=0.73); the association between high expression and survival; however, was statistically similar in men and women. No previous study evaluated the prognostic significance of HGF or c-Met protein expression in the tumor lung stratified by the gender.

A significant limitation to our study is the low statistical power due to the small sample size. However, our study has the largest sample size among the previously reported studies on the HGF or c-Met expression in lung cancer patients. The small sample size was problematic especially for investigating the association between high expression of biomarkers and risk factors such as smoking status, race, and histology which have disproportional distributions across the categories of factors. Only 10 subjects were never smoker in this study and there are only 3 subjects with small cell carcinomas of lung. Therefore, we had large 95% confidence intervals for the association between the risk factors and HGF and c-Met status. A majority of the study population was Whites, thus these results may not apply to lung cancer patients of other races or ethnicities. Interestingly, although only a small percentage of participants were African-American, African-American were statistically significantly associated with high c-Met expression in lung tumor. Since our study only used tissues cored from tumor epithelium, not stroma, the stromal production of HGF could not be evaluated within lung tumors.

The laboratory assay procedures were performed in blinded fashion to outcome-related information. In this study, a multilevel generalized linear mixed model was used to control for

sample type and to comply with repeated measures from TMA with discrete response. This model accounts the correlations among repeated IHC readings from TMA data on the same subject, and also for some possible heterogeneous variances among observations obtained on the same subject. Through modeling the correlation among repeated measures from TMA, we could obtain the best linear unbiased predictions. This statistical method used in this study is unique and strengthen our findings because there are no standardized method in quantifying the biomarker expression which can explain the results from different laboratory assays (western blot vs. IHC).

Our results showed both consistent and contrary findings to previous reports in association between HGF and c-Met expression and the risk factors for lung cancer outcome. In this study, we were not able to replicate our previous observations showing association between HGF expression and poor lung cancer survival [5, 10, 16], even though this study has a similar study population with previous studies, using immunohistochemistry to measure the protein expression. Future studies should investigate the potential factors which may result discrepancies in the observed relations of HGF and c-Met with the prognosis of lung cancer patients. In order to develop the well-defined biomarker of lung cancer prognosis, it is necessary to have a comprehensive understanding of various signaling pathways and the effects of genetic variation along with its interaction with environmental exposures. For example, the HGF/c-Met pathway shares important signal intermediates such as p44/p42 MAPK, PI3K/AKT, and STAT with the epidermal growth factor receptor (EGFR) pathway[17, 18], which are already in clinical use by tyrosine kinase inhibitor (TKI) drugs such as gefitinib and erlotinib. However, the clinical response rate to EGFR-TKI treatment is different between lung cancers with *EGFR* mutations (70%) and without mutations (10%) [19]. Therefore,

future studies should investigate the relationships between HGF/c-Met expression and *EGFR* mutations and how such relationships might impact overall survival of lung cancer patients.

3.6 REFERENCES

1. American Cancer Society, *Cancer Facts & Figures 2007*. 2007, American Cancer Society, Inc.: Atlanta.
2. Jemal, A., et al., *Cancer Statistics, 2007.* CA Cancer J Clin, 2007. **57**(1): p. 43-66.
3. Nakamura, T., et al., *Molecular cloning and expression of human hepatocyte growth factor.* Nature, 1989. **342**(6248): p. 440-3.
4. Gherardi, E. and M. Stoker, *Hepatocytes and scatter factor.* Nature, 1990. **346**(6281): p. 228.
5. Siegfried, J.M., et al., *The clinical significance of hepatocyte growth factor for non-small cell lung cancer.* Ann Thorac Surg, 1998. **66**(6): p. 1915-8.
6. Siegfried, J.M., et al., *Signaling pathways involved in cyclooxygenase-2 induction by hepatocyte growth factor in non small-cell lung cancer.* Mol Pharmacol, 2007. **72**(3): p. 769-79.
7. Masuya, D., et al., *The tumour-stromal interaction between intratumoral c-Met and stromal hepatocyte growth factor associated with tumour growth and prognosis in non-small-cell lung cancer patients.* Br J Cancer, 2004. **90**(8): p. 1555-62.
8. Ichimura, E., et al., *Expression of c-met/HGF receptor in human non-small cell lung carcinomas in vitro and in vivo and its prognostic significance.* Jpn J Cancer Res, 1996. **87**(10): p. 1063-9.
9. Takanami, I., et al., *Hepatocyte growth factor and c-Met/hepatocyte growth factor receptor in pulmonary adenocarcinomas: an evaluation of their expression as prognostic markers.* Oncology, 1996. **53**(5): p. 392-7.
10. Siegfried, J.M., et al., *Association of immunoreactive hepatocyte growth factor with poor survival in resectable non-small cell lung cancer.* Cancer Res, 1997. **57**(3): p. 433-9.
11. Belani, C.P., et al., *Women and lung cancer: epidemiology, tumor biology, and emerging trends in clinical research.* Lung Cancer, 2007. **55**(1): p. 15-23.
12. Chen, J.T., et al., *Cigarette smoking induces overexpression of hepatocyte growth factor in type II pneumocytes and lung cancer cells.* Am J Respir Cell Mol Biol, 2006. **34**(3): p. 264-73.
13. Stabile, L.P. and J.M. Siegfried, *Estrogen receptor pathways in lung cancer.* Curr Oncol Rep, 2004. **6**(4): p. 259-67.
14. Allred, D.C., et al., *Prognostic and predictive factors in breast cancer by immunohistochemical analysis.* Mod Pathol, 1998. **11**(2): p. 155-68.
15. Nakamura, Y., et al., *c-Met activation in lung adenocarcinoma tissues: an immunohistochemical analysis.* Cancer Science, 2007. **98**(7): p. 1006-13.
16. Siegfried, J.M., et al., *Elevated hepatocyte growth factor level correlates with poor outcome in early-stage and late-stage adenocarcinoma of the lung.* Chest, 2004. **125**(5 Suppl): p. 116S-9S.

17. Egloff, A.M. and J.R. Grandis, *Targeting Epidermal Growth Factor Receptor and Src Pathways in Head and Neck Cancer.* Seminars in Oncology, 2008. **35**(3): p. 286-297.
18. Knowles, L.M., et al., *HGF and c-Met participate in paracrine tumorigenic pathways in head and neck squamous cell cancer.* Clin Cancer Res, 2009. **15**(11): p. 3740-50.
19. Mitsudomi, T. and Y. Yatabe, *Mutations of the epidermal growth factor receptor gene and related genes as determinants of epidermal growth factor receptor tyrosine kinase inhibitors sensitivity in lung cancer.* Cancer Sci, 2007. **98**(12): p. 1817-24.

3.7 TABLES AND FIGURES

Table 3-1 Subject characteristics: TMA *vs*. Whole section

Variable	Measure	All n=180	Tissue source: Whole section n=65	Tissue source: TMA n=115	*p*-value[1]
Survival status	Dead, %	68.3	69.2	67.8	0.87
Sex	Women, %	51.1	50.8	51.3	1.00
Race	African-American, %	9.2	10.2	8.7	0.79
Age	30-59 years, %	22.2	30.8	17.4	0.13
	60-69 years, %	34.4	30.8	36.5	
	70+ years, %	43.3	38.5	46.1	
Smoking status	never smoker, %	5.8	6.3	5.5	0.09
	ex-smoker, %	43.4	32.8	49.5	
	active smoker, %	50.9	60.9	45.0	
Smoking dose-duration (among ever smokers)	<50 pack-years, %	56.3	54.4	57.4	0.74
	50+pack-years, %	43.7	45.6	42.6	
Stage	IA	17.9	15.6	19.1	0.42
	IB	25.7	31.3	22.6	
	IIA/B	19.6	20.3	19.1	
	III	27.4	28.1	27.0	
	IV	9.5	4.7	12.2	
Histology	squamous cell carcinoma	33.9	33.9	33.9	0.18
	non-squamous non-small cell	57.8	63.1	54.8	
	undifferentiated	6.7	1.5	9.6	
	small cell carcinoma	1.7	1.5	1.7	
HGF[3]	High expression[4], %	49.1	29.1	58.8	0.0003
	Allred, Median	7.0	6.0	7.5	<.0001[2]
c-Met[3]	High expression[4], %	50.0	42.9	53.5	0.25
	Allred, Median	7.1	7.0	7.3	0.87[2]

[1]Fisher exact test, except where indicated otherwise
[2] Wilcoxon rank sum test
[3]Using subject-specific Allred values averaged across TMA cores
[4]Allred >7
Whole section: 6 missing race, 1 missing smoking status, 3 missing smoking dose-duration (among ever smokers), 1 missing stage
TMA: 6 missing smoking status, 2 missing smoking dose-duration (among ever smokers)
Smoking dose duration Total N: All=163, Whole section=60 , TMA=103

Table 3-2 Frequency of high HGF and high c-Met IHC expression according to subject category

	HGF Total N=169		c-Met Total N=170	
	N	High (%)	N	High (%)
STATUS AT LAST CONTACT				
alive	52	53.8	53	35.8
dead	117	47.0	117	47.9
SEX				
women	86	47.7	88	45.5
men	83	39.8	82	42.7
RACE				
African-American	14	28.6	14	64.3
White	149	45.6	150	42.7
AGE (years)				
30-59	35	42.9	35	40.0
60-69	57	47.4	58	48.3
70+	77	41.6	77	42.9
SMOKING STATUS				
never smoker	10	20.0	9	22.2
active smoker	82	46.3	82	45.1
ex-smoker	70	60.0	72	44.4
SMOKING DOSE-DURATION (among ever smokers)				
<50 pack-years	83	45.8	84	47.6
50+ pack-years	64	48.4	65	41.5
STAGE				
IA	30	56.7	30	60.0
IB	44	47.7	44	34.1
IIA/B	34	55.9	34	38.2
III	45	53.3	46	47.8
IV	16	12.5	16	43.8
HISTOLOGY GROUP				
squamous cell carcinoma	57	49.1	58	55.2
non-squamous non-small cell	97	48.5	97	39.2
undifferentiated	12	58.3	12	33.3
small cell carcinoma	3	33.3	3	33.3

High HGF and c-Met expression defined by subject-specific averaged Allred values above IHC source-specific Allred median cutpoints (HGF cutpoints: 7.5 for TMA and 6.0 for whole section; c-Met cutpoints: 7.25 for TMA and 7.0 for whole section)

Total N=Number of subjects with non-missing IHC data

HGF: 6 missing race (33.3% high expression), 7 missing smoking status (14.3% high expression), 22 missing smoking dose-duration (22.7% high expression)

c-Met: 6 missing race (33.3% high expression), 7 missing smoking status (57.1% high expression), 21 missing smoking dose-duration (38.1% high expression)

Table 3-3 Results from generalized linear mixed models (SAS PROC GLIMMIX): Adjusted odds ratios (OR) and 95% confidence intervals (CI) for associations between personal characteristics and high HGF and high c-Met IHC expression

	HGF				c-Met			
	OR	**95% CI**		***p*-value***	**OR**	**95% CI**		***p*-value***
SEX								
Women	1.00				1.00			
Men	0.63	0.36	1.11	0.11	0.91	0.52	1.59	0.74
RACE								
White	1.00				1.00			
African-American	0.95	0.34	2.66	0.91	2.66	1.07	6.59	0.03
AGE (per year of age)	1.00	0.97	1.03	0.98	1.00	0.97	1.03	0.83
AGE (years)				0.67				0.40
30-59	1.00				1.00			
60-69	1.43	0.63	3.24	0.39	1.69	0.76	3.74	0.20
70+	1.37	0.62	3.02	0.43	1.24	0.60	2.57	0.56
SMOKING STATUS				0.14				0.91
never smoker	1.00				1.00			
active smoker	2.35	0.74	7.43	0.15	1.25	0.32	4.82	0.74
ex-smoker	3.08	0.99	9.57	0.05	1.33	0.35	5.13	0.67
STAGE				0.05				0.39
IA	1.00				1.00			
IB	0.66	0.28	1.58	0.35	0.57	0.24	1.33	0.19
IIA/B	1.26	0.47	3.39	0.64	0.48	0.19	1.22	0.12
III/IV	0.43	0.19	0.98	0.05	0.75	0.36	1.55	0.43
HISTOLOGY GROUP				0.88				0.49
non-squamous non-small cell	1.00				1.00			
undifferentiated	0.77	0.32	1.86	0.56	0.73	0.22	2.48	0.62
squamous cell carcinoma	0.85	0.46	1.57	0.60	1.39	0.76	2.54	0.29
small cell carcinoma	1.31	0.25	6.95	0.75	0.34	0.02	4.89	0.42

High HGF and c-Met expression defined by subject-specific averaged Allred values above IHC source-specific Allred median cutpoints (HGF cutpoints: 7.5 for TMA and 6.0 for whole section; c-Met cutpoints: 7.25 for TMA and 7.0 for whole section)

Every model includes terms for age (continuous), sex , smoking status, stage

*Wald chi-square test

Table 3-4 Results from Cox proportional hazards regression survival models: Adjusted hazards ratios (HR) and 95% confidence intervals (CI)

	HR (95% CI)	*p*-value*
SEX (men *vs*. women)	1.51 (1.03, 2.22)	0.03
RACE (African-American)	1.42 (0.72, 2.83)	0.31
AGE (per year of age)	1.03 (1.01, 1.05)	0.001
AGE (years)		0.01
30-59	1.00	
60-69	1.74 (1.00, 3.04)	0.05
70+	2.24 (1.34, 3.76)	0.002
SMOKING STATUS		0.001
never smoker	1.00	
active smoker	2.60 (1.15, 5.86)	0.02
ex-smoker	1.27 (0.56, 2.88)	0.56
STAGE		<.0001
IA	1.00	
IB	1.59 (0.83, 3.06)	0.17
IIA/B	4.39 (2.21, 8.72)	<.0001
III/IV	4.00 (2.17, 7.36)	<.0001
HISTOLOGY GROUP		0.29
non-squamous non-small cell	1.00	
undifferentiated	1.73 (0.83, 3.63)	0.15
squamous cell carcinoma	1.25 (0.82, 1.92)	0.30
small cell carcinoma	2.38 (0.70, 8.16)	0.17

Every model includes terms for age (continuous), sex, race, smoking status, stage, and histology group.
* Wald chi-square test.

Table 3-5 Results from Cox proportional hazards regression: Adjusted hazards ratios (HR) and 95% confidence intervals (CI) expressing associations between HGF and c-Met IHC expression and survival

		All Subjects		Women		Men	
		HR (95% CI)	**p-value***	**HR (95% CI)**	**p-value***	**HR (95% CI)**	**p-value***
HGF	**High vs. low expression**	0.87 (0.59, 1.29)	0.49	0.91 (0.53, 1.56)	0.73	0.69 (0.37, 1.30)	0.25
	Per Allred unit	1.02 (0.91, 1.13)	0.79	1.02 (0.87, 1.20)	0.79	1.02 (0.88, 1.18)	0.77
c-Met	**High vs. low expression**	1.06 (0.71, 1.58)	0.79	0.91 (0.53, 1.58)	0.74	1.26 (0.70, 2.29)	0.44
	Per Allred unit	1.08 (0.95, 1.24)	0.25	1.07 (0.90, 1.28)	0.46	1.10 (0.90, 1.33)	0.35

Every model includes terms for age, sex (where appropriate), smoking status, and stage.
High HGF and c-Met expression defined by subject-specific averaged Allred values above IHC source-specific Allred median cutpoints (HGF cutpoints: 7.5 for TMA and 6.0 for whole section; c-Met cutpoints: 7.25 for TMA and 7.0 for whole section)

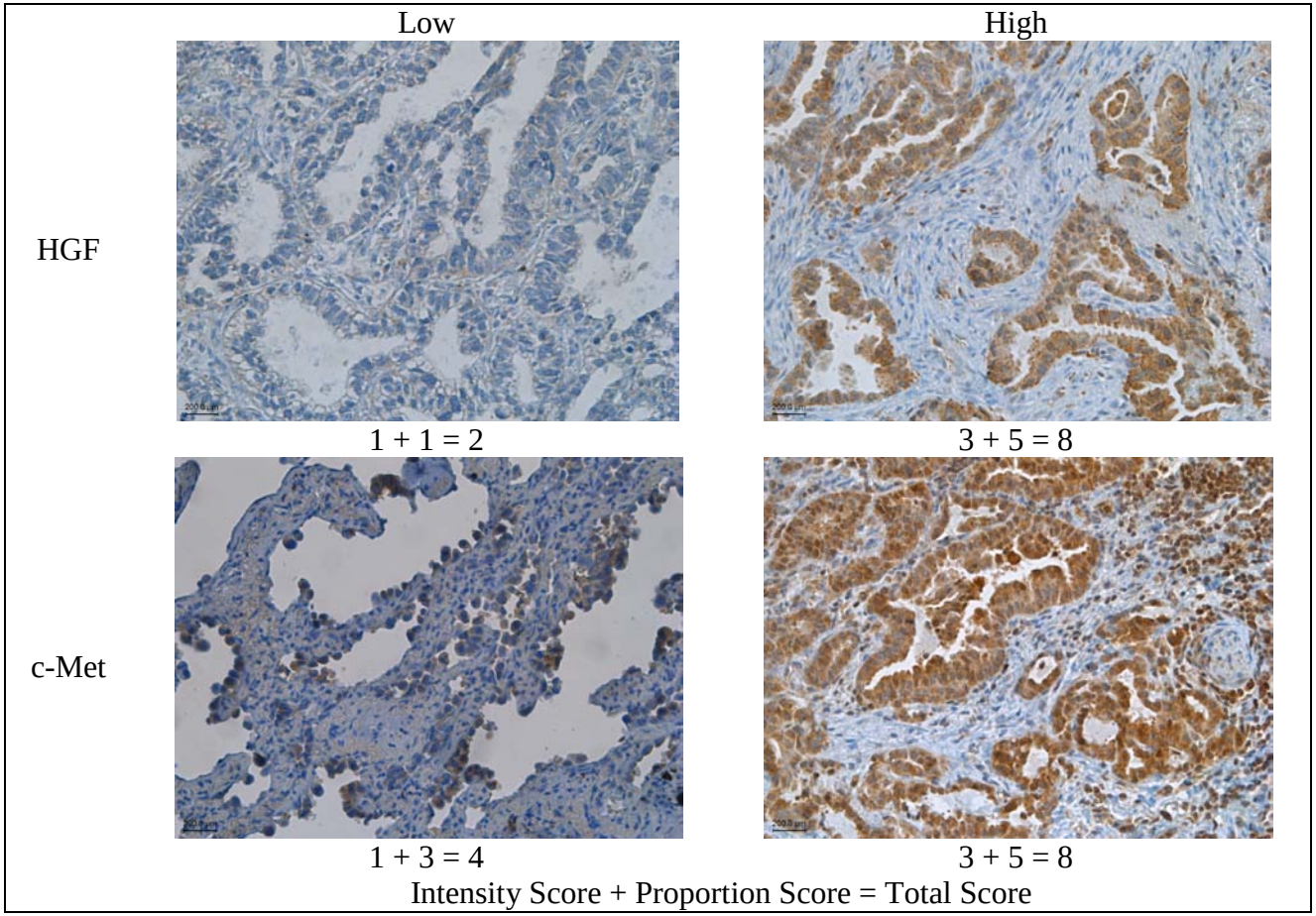


Figure 3-1 Representative immunohistochemical staining pattern of HGF and c-Met from whole section of lung tumor tissues. An intensity score and proportion scores were added to obtain a total immunohistochemical score. Each score is shown under the photos.

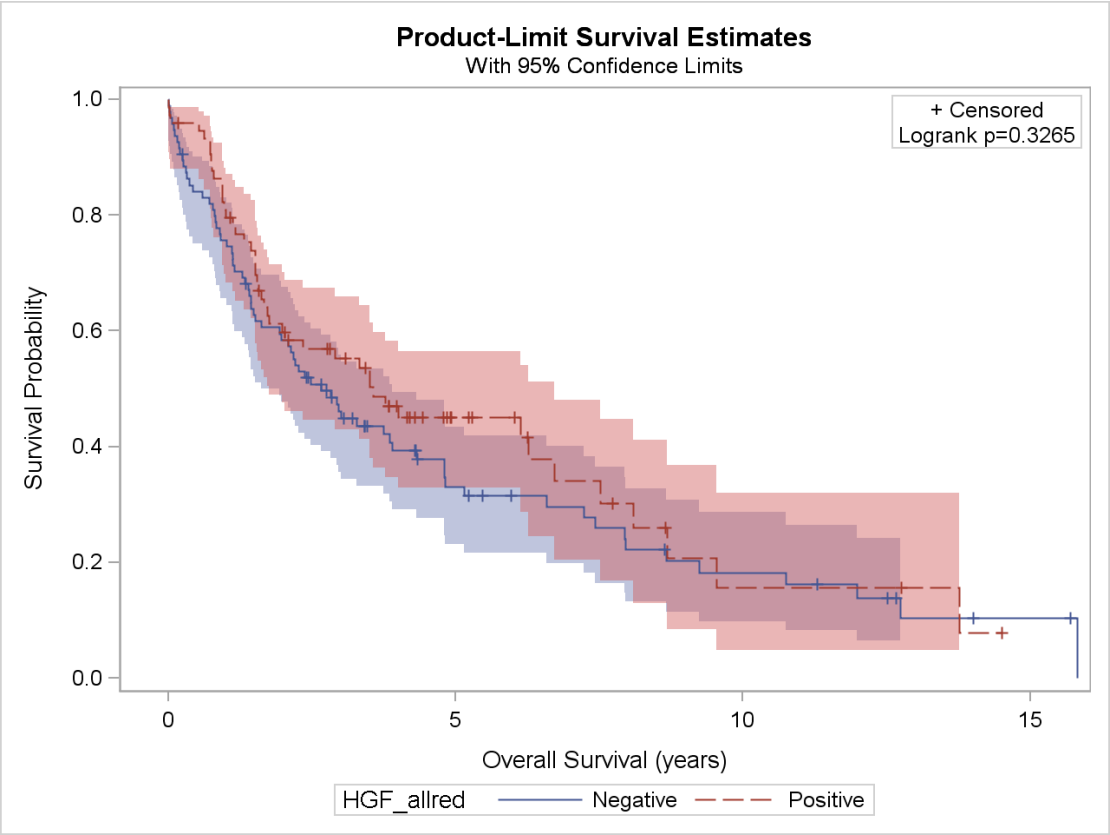


Figure 3-2 Kaplan Meier survival curves for high (Positive) and low (Negative) HGF expression level categories

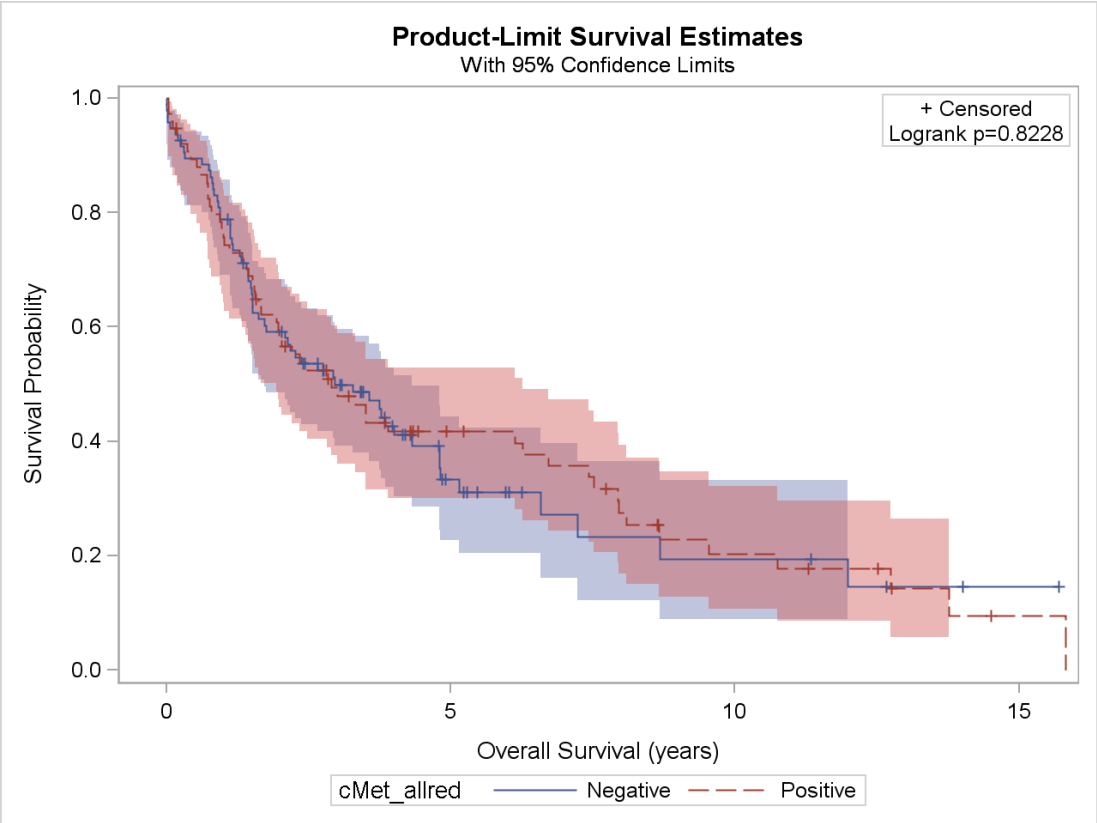


Figure 3-3 Kaplan Meier survival curves for high (Positive) and low (Negative) c-Met expression level categories

4.0 VALIDATION STUDY OF IMMUNOHISTOCHEMICAL EXPRESSION PATTERNS INVOLVING SEVEN LUNG TUMOR MARKERS

To be submitted for publication

Ji Young Song,[1] Jill Siegfried,[3] Brenda Diergaarde,[1] Stephanie Land,[2] Robert Bowser,[4] Laura Stabile,[3] Sanja Dacic,[4] Rajiv Dhir,[4] Joel L. Weissfeld[1]

[1] Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

[2] Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

[3] Department of Pharmacology and Chemical Biology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania

[4] Department of Pathology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania

This work was supported in part by National Institutes of Health grant P50 CA090440 SPORE in Lung Cancer and R25 CA057703 Education Programs in Cancer Prevention.

Abbreviations Used:; EGFR, epidermal growth factor receptor; ERα, estrogen receptor alpha; ERβ, estrogen receptor beta; HGF, hepatocyte growth factor; IHC, immunohistochemistry; PR, progesterone receptor; TMA, tissue microarray

4.1 ABSTRACT

Background: Treatment options for lung cancer are few. Steroid hormones, growth factors, and their receptors are attractive therapeutic targets. We attempted to identify meaningful expression patterns in lung tumors. These expression patterns may enable biologically directed and patient tailored treatments for lung cancer.

Method: We analyzed primary lung tumors for immunohistochemical (IHC) expression of the seven proteins: (1) hepatocyte growth factor (HGF), (2) c-Met, (3) estrogen receptor alpha (ERα), (4) estrogen receptor beta (ERβ), (5) progesterone receptor (PR), (6) aromatase, and (7) epidermal growth factor receptor (EGFR). We used a cluster algorithm (implemented in Cluster 2.11, http://rana.lbl.gov/EisenSoftware.htm) to sort 175 lung tumors into two more inter-homogenous grouped IHC expression clusters. We used the standard statistical techniques to compare clusters according to personal host characteristics, tumor stage and histology, and survival.

Results: ERα, ERβ, cytoplasmic PR, EGFR, and aromatase expression characterized the 77 tumors grouped into one cluster (cluster 1) and HGF, c-Met, and nuclear PR expression characterized the 98 tumors grouped into the second cluster (cluster 2). Clinicopathologic features, including age, race, gender, smoking history, histopathology, and stage were statistically similar in the two clusters. There were no significant survival differences between the two clusters (log rank test: p=0.6909).

Conclusion: Two lung cancer subgroups exhibiting dissimilar 7-protein IHC expression patterns were similar in terms of host and tumor characteristics and in terms of overall survival.

4.2 INTRODUCTION

Five year lung cancer survival, all stages, is only 15.7%. Five-year survival has improved only 3% since 1975 [1]. Few new and effective treatments are available for lung cancer.

Steroid hormones, growth factors and their receptors are attractive targets for cancer therapy because these molecules control many biological processes, including cell proliferation, apoptosis, motility, angiogenesis, and morphogenesis [2], [3]. For example, small-molecule tyrosine kinase inhibitors (TKIs) of epidermal growth factor receptor (EGFR) have entered clinical use for treating lung cancer. Clinical response to EGFR tyrosine kinase inhibition differs for lung cancer with (70%) and without *EGFR* mutations (10%) [4]. The HGF/c-Met pathway shares signal intermediates with the EGFR pathway [5, 6]. Many human cancers, including breast, prostate, and lung cancer, over-express HGF or c-Met [7-9]. In some studies, HGF or c-Met over-expression predicts poor non-small-cell lung cancer (NSCLC) prognosis [10-12]. Immunohistochemistry (IHC) detects nuclear expression of ERβ in 61% of lung tumor and 20% of normal lung samples [13]. In vitro studies show cross-talk between EGFR and estrogen receptor pathways [14, 15]. Recently, Dr. Nose and colleagues showed correlation between ERβ expression and *EGFR* mutation in lung adenocarcinoma [16, 17]. Immunohistochemistry also detects aromatase in NSCLC and aromatase inhibition prevents the tumor growth in vivo [18].

A comprehensive understanding of the multiple signaling pathways that lead to tumor growth is a prerequisite for more effective and targeted cancer treatments. This work examines the correlations between immunohistochemical (IHC) expression of seven protein markers, hepatocyte growth factor (HGF), c-Met, estrogen receptor alpha (ERα), estrogen receptor beta (ERβ), progesterone receptor (PR), aromatase, and epidermal growth factor receptor (EGFR), in tumor tissue from lung cancer patients. We presume that protein expression patterns transmit fundamental information about underlying tumor biology. Therefore, we aim to identify meaningful expression patterns involving these seven interesting and relevant proteins. Lung cancer patient clusters based on the expression patterns of multiple markers may distinguish subgroups with better or worse survival. These expression patterns may enable biologically directed and individually tailored treatment.

4.3 METHODS

4.3.1 Study Population

The study population included n=188 persons aged 21 year-old and older who received surgery at a University of Pittsburgh Medical Center hospital for the staging or treatment of pathologically confirmed primary lung cancer. We assembled risk factor, tumor, and follow-up information from several sources, including outpatient paper charts, inpatient and outpatient electronic medical records, hospital-based cancer registries, and Social Security Death Index database searches. The research used formalin-fixed and paraffin-embedded tissue specimens,

processed as tissue microarray (TMA) cores or as whole tissue sections. The University of Pittsburgh Institutional Review Board approved subject recruitment and tissue use protocols.

4.3.2 Laboratory Methods

TMA construction included three 0.6 mm diameter lung tumor cores per subject with examination of hematoxylin- and eosin-stained sections to verify malignant content. Preparations for immunohistochemisty (IHC) included deparaffinization and hydration with xylene and ethanol, heat-induced antigen retrieval with 10 mM citrate buffer at pH 6, quenching endogenous peroxidase with 3% hydrogen peroxide for 5 min at room temperature, and blocking with non-immune normal serum for 5-20 min at room temperature. ERα, ERβ, PR, aromatase, EGFR, HGF, and c-Met staining used anti-ERα (HC-20, Santa Cruz), anti-ERβ (MCA1974ST, Serotec), anti-PR (MAB429, Chemicon International), anti-cytochrome P450 aromatase (MCA2077, Serotec), anti-EGFR (E3138, Sigma Diagnostics), anti-HGF (AB-294-NA, R&D Systems), and anti-c-Met (SC-10, Santa Cruz). Antibodies were diluted in PBS as follows: ERα: 1:200 dilution for 30 min at room temperature, ERβ: 1:20 dilution overnight at 4°C, PR: 1:80 dilution for 1 hour at room temperature, aromatase: 1:50 dilution overnight at 4°C, EGFR: 1:7,500 dilution for 30 min at room temperature, HGF: 1:200 dilution for 1 hour at room temperature, and c-Met: 1:150 dilution for 30 minutes at room temperature.

ERα, ERβ, EGFR, and HGF staining used the EnVision method (DAKO Corp., Carpinteria, CA), PR and aromatase staining the Vector ABC method (Vector Labs, Burlingame, CA), c-Met staining the MACH 4 Universal HRP-Polymer Kit with DAB (Biocare Medical, LLC., Concord, CA). Final steps consisted of incubation with diaminobenzidine (DAB)

chromogenic substrate at room temperature for 5-10 min and counterstaining with hematoxylin for 2-2.5 min. IHC runs used breast cancer as positive control for ERa, ERβ, PR, HGF, and c-Met, placenta as positive control for aromatase, and laryngeal squamous cell carcinoma as positive control for EGFR. Assessments for background staining eliminated the primary antibody.

For each IHC assay except EGFR, the study lung pathologist (S.D.) assessed each TMA core and whole section for percentage of tumor cells stained and for intensity of staining. Scoring for the percentage of tumor cells stained used a six-level ordinal scale (0 to 5, respectively, for no cells stained, 0-1% cells stained, 2-10% cells stained, 11-33% cells stained, 34-66% cells stained, and 67-100% cells stained). Scoring for intensity of staining used a four-level ordinal scale (0 to 3, respectively, for no, weak, moderate, and strong staining). Data analyses expressed IHC expression in terms of the Allred score (range 0 to 8), the sum of the percentage and intensity scores.[1] IHC scoring for EGFR expression used a simple four-level ordinal scale (0 (staining in less than 10% cells), 1 (light staining in more than 10% cells), 2 (moderate staining in more than 10% cells), 3 (strong staining in more than 10% cells)). Scores were averaged for the multiple cores from each patient. For each patient, there is a nuclear score and a cytoplasmic score for ERα, ERβ, PR, and aromatase. Through the IHC assay evaluation, we obtained the eleven IHC measures (HGF, c-Met, ERα cytoplasmic, ERα nuclear, ERβ cytoplasmic, ERβ nuclear, PR cytoplasmic, PR nuclear, aromatase cytoplasmic, aromatase nuclear, and EGFR) derived from the initial seven proteins in lung tumors.

4.3.3 Statistical Analysis

For each IHC measures, lung tumors were divided into two (low and medium-high) or three (low, median, and high) ordered and integer scored categories based on IHC expression scores (Table 4-1). Cytoplasmic and nuclear integer scores for ERα, ERβ, and aromatase, were used to form single combined scores for ERα, ERβ, and aromatase because of the moderate to high correlations observed between nuclear and cytoplasmic expression values. We used the following equation to form combined scores for ERα and ERβ, combined score = 0.25*(nuclear integer score + cytoplasmic integer score) (Figure 4-1). The combined aromatase score used the following transformation, 0.00 to indicate no cytoplasmic or nuclear expression, 0.33 to indicate medium cytoplasmic and no nuclear expression, 0.67 to indicate high cytoplasmic and no nuclear expression, and 1.00 to indicate any nuclear expression (Figure 4-1). Therefore, the eight constructed markers, instead of the eleven IHC measures, were used in the statistical analyses. Pearson correlation coefficient (rho) was used to measure the correlations among protein expression levels for lung cancer patients with complete expression data for all eight constructed markers.

Clustering of lung cancer patients was performed with Cluster software, version 2.11, after mean centering and normalizing IHC scores[ii]. In this clustering procedure, only 175 lung tumors with non-missing expression information for at least six of the eight constructed markers were included. We used standard Pearson correlations to express distances between observations and the average linkage clustering algorithm to calculate distances that involve clusters with

[ii] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patters. PNAS 95(25):14863-14868, 1998; http://rana.lbl.gov/EisenSoftware.htm, last accessed March 4, 2010.

more than one observation. Heat map was generated with TreeView program in Cluster software to visualize the clusters and IHC expression of the eight constructed markers for lung cancer patients.

We used Fisher's exact tests to evaluate the statistical significance of differences between the last two clusters (cluster 1, n=77 *vs.* 2, n=98) of subjects combined by the average linkage clustering algorithm based on IHC expression of the eight constructed markers. Variables used in analyses comparing two clusters included sample source (TMA, whole section), status at last contact (alive, dead), age at tissue collection (30-59, 60-69, 70+ years), race (White, African-American), sex (women, men), smoking status (active, ex, smoker-NOS, never), smoking dose duration among ever smokers (1-25, 26-50, 51-75, 75+ pack-years), pathologic stage group (IA, IB, IIA/IIB, III, IV), and histology group (squamous cell carcinoma, non-squamous non-small cell lung cancer, undifferentiated carcinoma, and small cell carcinoma). The non-squamous non-small cell lung cancer group included adenocarcinoma, adenosquamous carcinoma, bronchioloalveolar carcinoma, and malignant carcinoid. The undifferentiated carcinoma group included large cell carcinoma and undifferentiated non-small cell lung cancer. Smokers without current status were categorized into active smoker group for smoking status variable. For subjects without pathologic stage information, clinical stage information was used instead.

We used the Kaplan-Meier product limit estimator and the log-rank statistic to estimate survival and to evaluate the statistical significance of differences between cluster groups. Pearson correlations, comparison between the clusters, and survival analysis used SAS version 9.2 (SAS Institute, Inc., Cary, North Carolina) and two-sided *p*-values.

4.4 RESULTS

The results for the eleven IHC measures in 188 lung tumors were classified into three categories based on the expression distribution of each IHC measures and the classification summary is described in Table 4-1. Three expression levels have different but relatively similar Allred score ranges for each marker. The distributions of the eleven IHC measures are equally proportional for each levels, except nuclear aromatase (low=88.8%, medium=11.2%, and no high expression) and EGFR (low=69.0%, medium=13.8%, and high=17.2).

Expression of several constructed markers was correlated (Table 4-2). Constructed markers that had a strong positive correlation ($p<0.0001$; rho $\geq 0.3$15) include the following: ERα/ERβ, ERα/PR cyto, ERα/EGFR, ERα/aromatase, ERβ/PR cyto, ERβ/EGFR, ERβ/c-Met, PR cyto/EGFR, PR cyto/aromatase, EGFR/aromatase. PR nuclear had weak to moderate negative correlations with other constructed markers (rho range: -0.22 to -0.04), except with HGF (rho=0.27). All others had a weak to moderate correlation.

Data analyses included 175 subjects (119 from TMA and 56 from whole sections) with non-missing expression information for at least six of the eight constructed markers. Heat map shows the tree view of lung cancer patient clusters based on their expression of the eight constructed markers which were color-coded as: green=negative expression, black=zero, red=positive expression, gray=missing (Figure 4-2). First level clustering in the heat map grouped patients into relatively homogeneous two clusters based on expression of the eight constructed markers. Cluster 1 has positive expression of ERα, ERβ, PR cytoplasm, EGFR, and aromatase and negative expression of HGF, c-Met, and PR nuclear. Cluster 2 has the opposite expression of the eight constructed markers.

Subject characteristics of 175 participants according to major IHC expression categories (cluster 1 and 2) are shown in Table 4-3. 77 patients were grouped as cluster 1 (positive expression of ERα, ERβ, PR cytoplasm, EGFR, and aromatase) while 98 patients in cluster 2 (positive expression of HGF, c-Met, and PR nuclear). There was no difference in sample source (TMA *vs.* whole section) between cluster 1 (TMA: 66.2%) and cluster 2 (TMA: 69.4%) (p=0.6571). There were no associations between cluster groups and clinicopathologic features, such as age, race, gender, smoking history, pathologic type and clinical stage (Table 4-3).

Survival analysis using the Kaplan-Meier method was carried out to assess the prognostic significance of the clusters which are defined as sub-groups of lung cancer patients with relatively homogenous expressions of eight markers. There were no significant differences in survival among lung cancer patients between two clusters (log rank test: p=0.6909) (Figure 4-3).

4.5 DISCUSSION

In this study of lung cancer patients, we evaluated the intercorrelation of IHC expression of eight biomarkers. We also have identified two distinct clusters of patients based on the protein expression level of the eight constructed markers. However, two clusters did not show statistically significant difference in patient survival. Therefore, we fail to identify a cluster of the patients who are characterized by the IHC expression of the eight constructed markers and who have a distinct survival pattern.

Two clusters identified in this study were characterized as positive expression of ERα, ERβ, PR cytoplasm, EGFR, and aromatase group (cluster 1) and as positive expression of HGF,

c-Met, and PR nuclear group (cluster 2). Identified homogenous expression of the eight constructed markers within each cluster may be explained by their biological functions, interactions with other markers, and impact on survival.

Estrogen receptors (α, β) mediate cellular response to estrogen. It is known that ERβ overexpression is a favorable prognostic indicator for lung cancer patients [13, 17, 19-21], while ERα expression is associated with a poorer prognosis [19, 22]. Aromatase is a key enzyme for estrogen synthesis and was detected in non-small cell lung tumor specimens [14, 18, 23], suggesting the autocrine ligand-receptor mechanism of estrogen and its receptors in the lung tumors. Therefore, we were not surprised to observe that both estrogen receptors and aromatase were grouped together as one cluster. Mah *et al.* has demonstrated that aromatase expression is a negative prognostic factor for early-stage NSCLC [16]. Also, recently Abe *et al.* shows that ERβ and aromatase are frequently expressed together in NSCLC [24]. Positive IHC expression of ERα showed statistically significant association with positive expression of ERβ and PR [25]. Also, in addition to the presence of the cross-talk between EGFR and estrogen receptors (ERs), recent studies demonstrated the correlation between ER-beta expression and *EGFR* mutations [17, 22]. These findings may be helpful in making medical decisions for individual cancer therapy since the clinical response rate to currently available EGFR- tyrosine kinase inhibitors treatment is different between lung cancers with *EGFR* mutations (70%) and without mutations (10%)[4]. In our study, patients in cluster 1 also had positive cytoplasmic PR expression.

HGF is the ligand for the c-Met protein, a tyrosine kinase receptor constitutively activated by mutations and expressed by both epithelial and endothelial cells [2]. HGF has multiple biological functions such as cell proliferation, motility, angiogenesis (blood vessel formation), and morphogenesis [3]. Previous studies showed that c-Met expression was a

negative prognostic factor for lung cancer patients [3, 10, 11], except one recent study by Nakamura *et al*. [26]. While some studies used western blot analysis reported that elevated HGF expression in tumor tissue is associated with poor survival in non-small cell lung cancer patients [12, 27, 28], studies with the IHC method did not show the HGF expression as a significant independent prognostic marker for lung cancer patients [3, 11, 26]. It is expected that HGF and c-Met are expressed homogeneously among lung cancer patients due to their autocrine mechanism in tumor cells regardless of their impact on survival. However, we were surprised to have positive expression of nuclear PR in cluster 2 groups along with HGF and c-Met since previous study with IHC method has reported that nuclear PR expression is a favorable prognostic factor for NSCLC [29]. However, Raso and colleagues did not show any association between the expression of estrogen receptor and nuclear PR expression and overall survival [22]. Also, no correlations were observed in nuclear PR expression and EGFR mutation status [22]. These inconsistent results of nuclear PR expression with a direction of positive association with survival may attenuate the negative impact of HGF and c-Met and may explain no survival difference between the two clusters identified in our study.

Our hypothesis for this study was that a useful prognostic marker or a set of markers for lung cancer patients can be developed through identifying meaningful expression patterns involving the seven interesting and relevant proteins. This hypothesis was an alternative approach that aims to validate the results of Dr. Stabile's recent study [30]. Dr. Stabile's study utilized the survival information of lung cancer patients through the Cox proportional hazards model in order to identify the significant prognostic proteins. The results showed that patients with high expression of cytoplasmic ERβ, aromatase, and EGFR with low PR total expression had a higher risk of death than the patients with the opposite protein characteristics [30]. We

identified a cluster of subjects (cluster 1) with similar protein expression characteristics with Dr. Stabile's results, except expression of PR. However, in terms of finding a helpful set of prognostic markers, our analytic approach did not to replicate Dr. Stabile's results.

The results of this study demonstrate that expressions of several markers were positively correlated, except PR nuclear expression. Two major clusters identified in this study were based on IHC expression information derived from the seven proteins. Our method, which ignores outcome information when grouping tumors according to IHC expression, did not identify two major subgroups with differing host and tumor characteristics or clinical outcomes. However, two major clusters identified in this study are interesting due to the biological functions of the proteins composed in each cluster. For example, the patients in the second cluster have relatively high expression of HGF, c-Met, and PR nuclear when compared with the others. This identified cluster supports the idea that autocrine HGF-c-Met signaling plays significant roles in the progression of lung tumors.

Due to multiple interactions between hormones and growth factors, it is difficult to predict the patient survival based on a single marker expression. This may explain the null finding of our study which investigated the association between the expression of HGF/c-Met and lung cancer survival without accounting for the role of other related hormones and growth factors. Therefore, future studies investigating the prognostic significance of a single protein marker in the tumor lung should consider the impact of multiple interactions between other relevant hormones and growth factors on overall survival of lung cancer patients through identifying more specific and meaningful expression patterns.

4.6 REFERENCES

1. Allred, D.C., et al., *Prognostic and predictive factors in breast cancer by immunohistochemical analysis.* Mod Pathol, 1998. **11**(2): p. 155-68.
2. Siegfried, J.M., et al., *Signaling pathways involved in cyclooxygenase-2 induction by hepatocyte growth factor in non small-cell lung cancer.* Mol Pharmacol, 2007. **72**(3): p. 769-79.
3. Masuya, D., et al., *The tumour-stromal interaction between intratumoral c-Met and stromal hepatocyte growth factor associated with tumour growth and prognosis in non-small-cell lung cancer patients.* Br J Cancer, 2004. **90**(8): p. 1555-62.
4. Mitsudomi, T. and Y. Yatabe, *Mutations of the epidermal growth factor receptor gene and related genes as determinants of epidermal growth factor receptor tyrosine kinase inhibitors sensitivity in lung cancer.* Cancer Sci, 2007. **98**(12): p. 1817-24.
5. Egloff, A.M. and J.R. Grandis, *Targeting Epidermal Growth Factor Receptor and Src Pathways in Head and Neck Cancer.* Seminars in Oncology, 2008. **35**(3): p. 286-297.
6. Knowles, L.M., et al., *HGF and c-Met participate in paracrine tumorigenic pathways in head and neck squamous cell cancer.* Clin Cancer Res, 2009. **15**(11): p. 3740-50.
7. Olivero, M., et al., *Overexpression and activation of hepatocyte growth factor/scatter factor in human non-small-cell lung carcinomas.* Br J Cancer, 1996. **74**(12): p. 1862-8.
8. Edakuni, G., et al., *Expression of the hepatocyte growth factor/c-Met pathway is increased at the cancer front in breast carcinoma.* Pathol Int, 2001. **51**(3): p. 172-8.
9. Kurimoto, S., et al., *Co-expression of hepatocyte growth factor and its receptor in human prostate cancer.* Histochem J, 1998. **30**(1): p. 27-32.
10. Ichimura, E., et al., *Expression of c-met/HGF receptor in human non-small cell lung carcinomas in vitro and in vivo and its prognostic significance.* Jpn J Cancer Res, 1996. **87**(10): p. 1063-9.
11. Takanami, I., et al., *Hepatocyte growth factor and c-Met/hepatocyte growth factor receptor in pulmonary adenocarcinomas: an evaluation of their expression as prognostic markers.* Oncology, 1996. **53**(5): p. 392-7.
12. Siegfried, J.M., et al., *Association of immunoreactive hepatocyte growth factor with poor survival in resectable non-small cell lung cancer.* Cancer Res, 1997. **57**(3): p. 433-9.
13. Schwartz, A.G., et al., *Nuclear estrogen receptor beta in lung cancer: expression and survival differences by sex.* Clin Cancer Res, 2005. **11**(20): p. 7280-7.
14. Pietras, R.J., et al., *Estrogen and growth factor receptor interactions in human breast and non-small cell lung cancer cells.* Steroids, 2005. **70**(5-7): p. 372-81.
15. Stabile, L.P., et al., *Combined targeting of the estrogen receptor and the epidermal growth factor receptor in non-small cell lung cancer shows enhanced antiproliferative effects.* Cancer Res, 2005. **65**(4): p. 1459-70.
16. Mah, V., et al., *Aromatase expression predicts survival in women with early-stage non small cell lung cancer.* Cancer Res, 2007. **67**(21): p. 10484-90.
17. Nose, N., et al., *Association between estrogen receptor-beta expression and epidermal growth factor receptor mutation in the postoperative prognosis of adenocarcinoma of the lung.* J Clin Oncol, 2009. **27**(3): p. 411-7.

18. Weinberg, O.K., et al., *Aromatase inhibitors in human lung cancer therapy.* Cancer Res, 2005. **65**(24): p. 11287-91.
19. Kawai, H., et al., *Estrogen receptor alpha and beta are prognostic factors in non-small cell lung cancer.* Clin Cancer Res, 2005. **11**(14): p. 5084-9.
20. Skov, B.G., B.M. Fischer, and H. Pappot, *Oestrogen receptor beta over expression in males with non-small cell lung cancer is associated with better survival.* Lung Cancer, 2008. **59**(1): p. 88-94.
21. Wu, C.T., et al., *The significance of estrogen receptor beta in 301 surgically treated non-small cell lung cancers.* J Thorac Cardiovasc Surg, 2005. **130**(4): p. 979-86.
22. Raso, M.G., et al., *Immunohistochemical expression of estrogen and progesterone receptors identifies a subset of NSCLCs and correlates with EGFR mutation.* Clin Cancer Res, 2009. **15**(17): p. 5359-68.
23. Pezzi, V., et al., *Profiling transcript levels for steroidogenic enzymes in fetal tissues.* J Steroid Biochem Mol Biol, 2003. **87**(2-3): p. 181-9.
24. Abe, K., et al., *Highly concordant coexpression of aromatase and estrogen receptor beta in non-small cell lung cancer.* Hum Pathol. **41**(2): p. 190-8.
25. Toh, C.K., et al., *Correlation between epidermal growth factor receptor mutations and expression of female hormone receptors in East-Asian lung adenocarcinomas.* J Thorac Oncol. **5**(1): p. 17-22.
26. Nakamura, Y., et al., *c-Met activation in lung adenocarcinoma tissues: an immunohistochemical analysis.* Cancer Science, 2007. **98**(7): p. 1006-13.
27. Siegfried, J.M., et al., *The clinical significance of hepatocyte growth factor for non-small cell lung cancer.* Ann Thorac Surg, 1998. **66**(6): p. 1915-8.
28. Siegfried, J.M., et al., *Elevated hepatocyte growth factor level correlates with poor outcome in early-stage and late-stage adenocarcinoma of the lung.* Chest, 2004. **125**(5 Suppl): p. 116S-9S.
29. Ishibashi, H., et al., *Progesterone receptor in non-small cell lung cancer--a potent prognostic factor and possible target for endocrine therapy.* Cancer Res, 2005. **65**(14): p. 6450-8.
30. Stabile, L.P., et al., *Combined analysis ofestrogen receptor β and progesterone receptor expression identifies lung cancer patients with poor outcome.* Clinical Cancer Research, 2010. **Submitted, not published yet**.

4.7 TABLES AND FIGURES

Table 4-1 Immunohistochemisty results, 11 IHC measures in n=188 lung tumors

		Expression level, Allred range			Expression level, %		
Marker	n	Low	Medium	High	Low	Medium	High
HGF	169	0.0-6.2	6.3-7.9	8.0	33.7	31.4	34.9
c-Met	170	0.0-6.5	6.6-7.7	7.8-8.0	34.1	34.7	31.2
ERα cytoplasmic	177	0.0	0.1-6.3	6.4-8.0	27.7	36.2	36.2
ERα nuclear	177	0.0	0.1-6.4	6.5-8.0	46.9	26.6	26.6
ERβ cytoplasmic	176	0.0	0.1-7.0	7.1-8.0	21.0	40.9	38.1
ERβ nuclear	176	0.0-6.5	6.6-7.9	8.0	22.7	21.0	56.3
PR cytoplasmic	175	0.0	0.1-5.9	6.0-8.0	49.1	25.1	25.7
PR nuclear	178	0.0	0.1-6.9	7.0-8.0	17.4	39.9	42.7
Aromatase cytoplasmic	178	0.0	0.1-4.2	4.3-8.0	38.8	29.8	31.5
Aromatase nuclear	178	0.0	0.1-8.0		88.8	11.2	
EGFR	174	0.0	0.1-0.6	0.7-3.0	69.0	13.8	17.2

Table 4-2 Pearson correlations matrix for eight constructed markers, n=148 lung tumors with complete data

	ERα	ERβ	PR cyto	EGFR	Aromatase	c-Met	HGF
ERβ	0.63						
PR cyto	0.55	0.44					
EGFR	0.47	0.41	0.36				
Aromatase	0.32	0.21	0.36	0.38			
c-Met	0.14	0.34	0.08	0.18	0.09		
HGF	0.20	0.12	0.24	0.16	0.17	0.15	
PR nuclear	-0.04	-0.22	-0.06	-0.15	-0.08	-0.15	0.27

$|r| \geq 0.162$, $p<0.05$, $|r| \geq 0.211$, $p<0.01$, $|r| \geq 0.267$, $p<0.001$, $|r| \geq 0.315$, $p<0.0001$

Table 4-3 Characteristics of lung tumor according to major immunohistochemical expression category.

	Cluster 1 Total N=77		Cluster 2 Total N=98		p-value
	N	%	N	%	
SAMPLE SOURCE					0.6571
TMA	51	66.2	68	69.4	
Whole section	26	33.8	30	30.6	
STATUS AT LAST CONTACT					0.2151
alive	20	26.0	34	34.7	
dead	57	74.0	64	65.3	
SEX					0.7238
women	38	49.4	51	52.0	
men	39	50.6	47	48.0	
RACE					0.4350
African-American	8	10.8	7	7.4	
White	66	89.2	88	92.6	
AGE (years)					0.6477
30-59	14	18.2	22	22.5	
60-69	29	37.7	31	31.6	
70+	34	44.2	45	45.9	
SMOKING STATUS					0.2803
active smoker	22	30.1	38	40.0	
ex-smoker	33	45.2	41	43.2	
smoker, Not Otherwise Specified	10	13.7	12	12.6	
never smoker	8	11.0	4	4.2	
SMOKING DOSE-DURATION (among ever smokers)					0.1512
1-25 pack-years	14	22.2	9	10.1	
26-50 pack-years	26	41.3	35	39.3	
51-75 pack-years	11	17.5	23	25.8	
>76 pack-years	12	19.0	22	24.7	
STAGE					0.5616
IA	13	16.9	17	17.4	
IB	19	24.7	23	23.5	
IIA/B	19	24.7	16	16.3	
III	17	22.1	31	31.6	
IV	9	11.7	11	11.2	
HISTOLOGY GROUP					
squamous cell carcinoma	28	36.4	30	30.6	
non-squamous non-small cell	43	55.8	58	59.2	
undifferentiated	5	6.5	8	8.2	
small cell carcinoma	1	1.3	2	2.0	
HISTOLOGY GROUP					0.4863
squamous cell carcinoma	28	39.4	30	34.1	
non-squamous non-small cell	43	60.6	58	65.9	

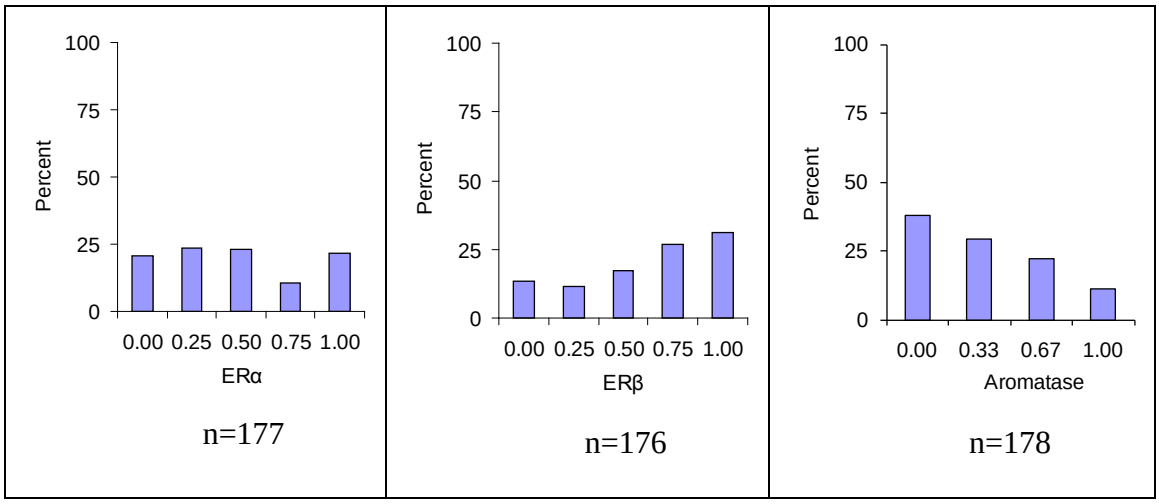


Figure 4-1 Three derived immunohistochemical measures, distribution of results for n=188 lung tumors

1. ERα = 0.25*(ERα cytoplasmic score + ERα nuclear score) and ERβ = 0.25*(ERβ cytoplasmic score + ERβ nuclear score).
2. Aromatase = 0.00 indicates no cytoplasmic or nuclear expression, Aromatase = 0.33 indicates medium cytoplasmic and no nuclear expression, Aromatase=0.67 indicates high cytoplasmic and no nuclear expression, and Aromatase=1.00 indicates any nuclear expression.

Figure 4-2 Heat map for n=175 lung tumors with non-missing expression information for at least six of eight immunohistochemical markers. Tumor clustering uses Cluster version 2.11, after mean centering and normalizing immunohistochemical scores (Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patters. PNAS 95(25):14863-14868, 1998; http://rana.lbl.gov/EisenSoftware.htm, last accessed March 4, 2010). We used standard Pearson correlations to express distances between observations and the average linkage clustering algorithm to calculate distances that involve clusters with more than one observation.

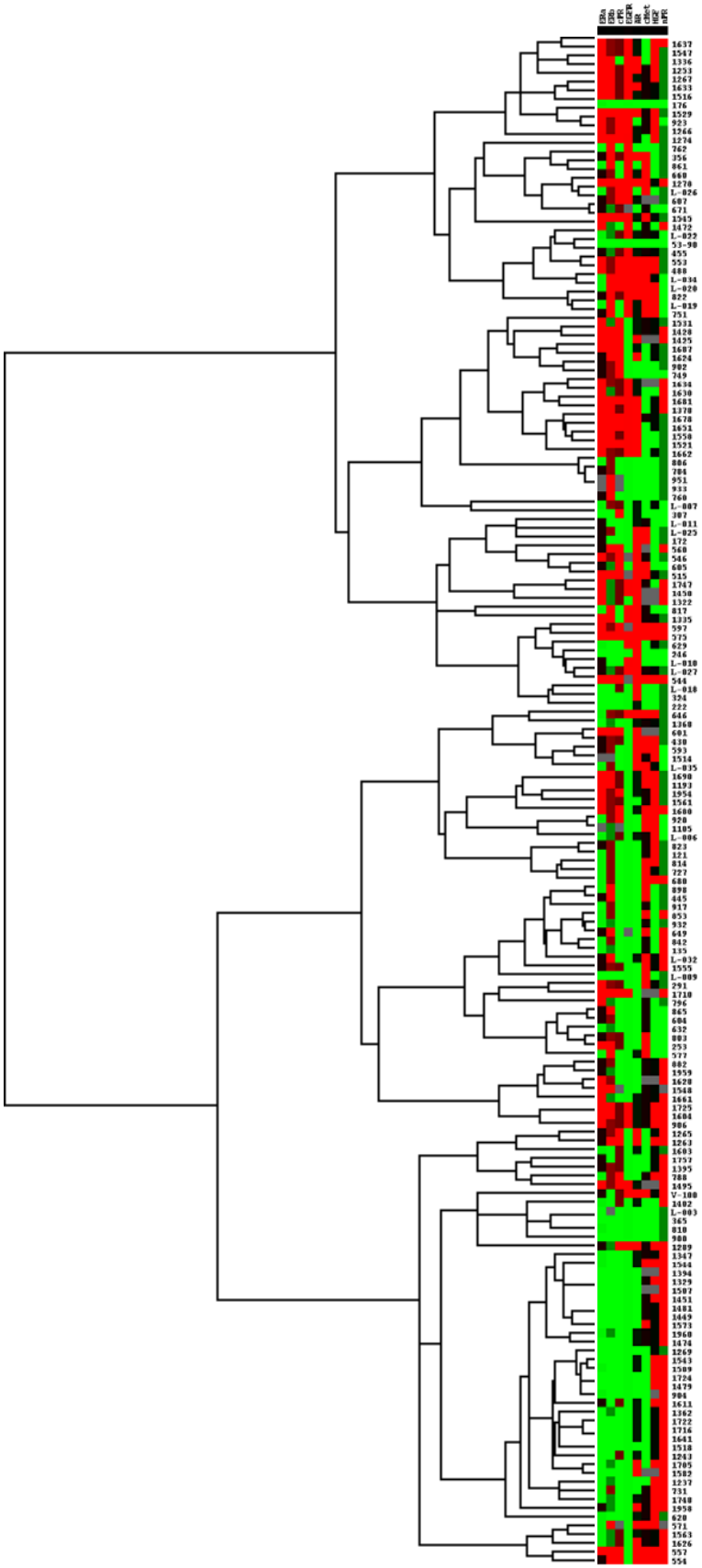


Figure 4-2 Heat map for n=175 lung tumors with non-missing expression information for at least six of eight immunohistochemical markers.

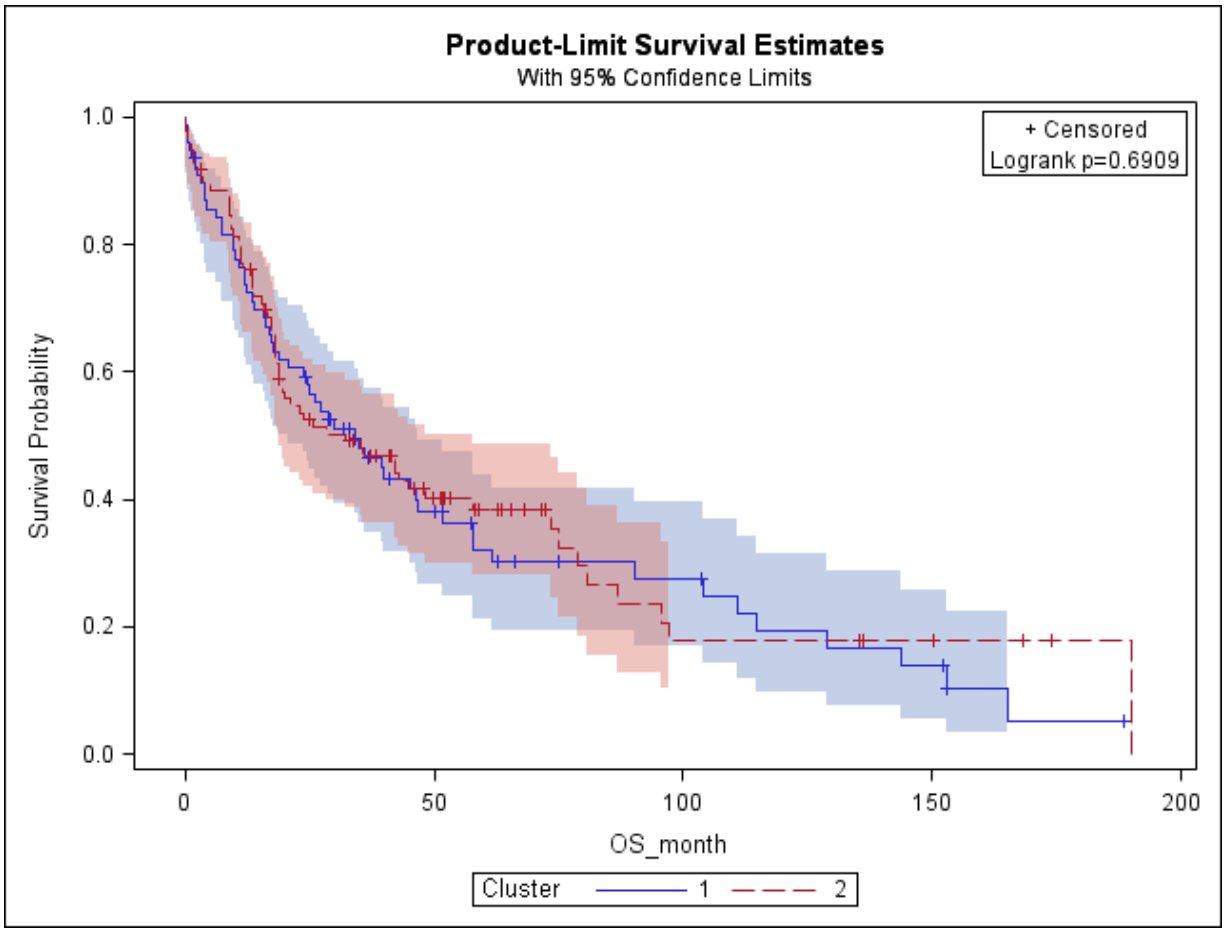


Figure 4-3 Kaplan-Meier survival curves according to major immunohistochemical expression category, with 77 (57 deaths) and 97 (64 deaths) in clusters 1 and 2, respectively

5.0 *ESR2* POLYMORPHISMS AND ESTROGEN RECEPTOR BETA EXPRESSION IN LUNG TUMORS

To be submitted for publication

Ji Young Song,[1] Jill Siegfried,[3] Brenda Diergaarde,[1] Stephanie Land,[2] Robert Bowser,[4] Laura Stabile,[3] Sanja Dacic,[4] Rajiv Dhir,[4] Marjorie Romkes,[5] Joel L. Weissfeld,[1]

[1] Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

[2] Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

[3] Department of Pharmacology and Chemical Biology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania

[4] Department of Pathology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania

[5] Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania

This work was supported in part by National Institutes of Health grant P50 CA090440 SPORE in Lung Cancer and R25 CA057703 Education Programs in Cancer Prevention.

Abbreviations Used: ERβ, estrogen receptor beta; *ESR2*, estrogen receptor beta gene; htSNPs, haplotype tagging single nucleotide polymorphisms; IHC, immunohistochemistry; TMA, tissue microarray

5.1 ABSTRACT

Objective: To investigate the association between the genetic variations in the ERβ gene (*ESR2*) and ERβ protein expression in lung tumors.

Methods: We used genetic results of 135 lung cancer patients with nuclear and cytoplasmic expression of ERβquantified by immunohistochemistry (IHC) on tissue microarrays (TMA) or on single whole tissue sections. A total of 22 single nucleotide polymorphisms (SNPs) were selected using literature search, NCBI Entrez SNP, the Cancer Genome Anatomy Project (CGAP) SNP500Cancer Database, HapMap Project, and FastSNP. Genotyping was done using Squenom iPlexGold. The Jonckheere-Terpstra test was used to test the null hypothesis that the distribution of the ERβ IHC expression does not differ among genotypes of 22 htSNPs. Unconditional logistic regression model was fitted to assess the association between genotype of three htSNPs (rs8021944, rs1256061, and rs10146204) and cytoplasmic and nuclear ER-beta expression score in lung tumors for all subjects.

Results: Three *ESR2* htSNPs (rs8021944, rs1256061, and rs10146204) were associated with nuclear ERβ expression. Subgroup analysis based on histological types of lung cancer suggests that the rs1256061 association with ERβ expression may be specific to adenocarcinoma of lung. Maximum ERβ expression (Allred score=8) was observed more often in tumors from patients with the CA or AA genotype [nuclear: Odds Ratios (OR) relative to Allred ≤ 6: 3.54 (95% confidence interval 1.22-10.3) and cytoplasmic: OR relative to Allred=0: 5.08 (95% CI 1.47-17.6)] than the CC genotype at rs1256061. Maximum ERβ expression was observed more often in patients with the GA or AA genotype [nuclear: OR relative to Allred ≤ 6: 3.71 (95% CI 1.31-

10.6) and cytoplasmic: OR relative to Allred=0: 4.00, 95% CI 1.26-12.7)] than the GG genotype at rs10146204.

Conclusion: We found that individuals with at least one rare allele of two htSNPs (rs1256061 and rs10146204) are associated with maximum expression of both cytoplasmic and nuclear ERβ expression in the dominant inheritance model, compared to non-carriers.

5.2 INTRODUCTION

ERβ, a second estrogen receptor (ER) isoform was discovered in 1996 [1]. Until the discovery of the ERβ, the estrogen receptor studies could not distinguish between ERα and ERβ. Nuclear ER-beta positivity was present in 61% of lung tumor tissue and 20% of normal lung tissue sample by using immunohistochemistry [2]. A study demonstrated the survival differences between genders: women with ER-beta expression in tumor tissue had a increase in mortality, whereas men with ER-beta expression had a significant reduction (55%, p=0.04) in mortality compared with those with ER-beta negative tumors [2]. Overexpression of ER beta was significantly more frequent in tumors occurring in lung cancer patients without smoking history (53.5%) than in those with smoking history (36.6%, P = .004) [3].

ESR2 is the estrogen receptor 2 gene. In the human genome, the ESR2 gene is located on chromosome 14, band q23.2. The size of the entire coding sequence (introns and exons) of ESR2 gene is approximately 61.2 kilobases. There are 8 exons in the human ESR2 gene. Also, there are 2 additional untranslated exons, 0N and 0K, in the 5' region and an exon at the 3' end.

It measures 468 bases at the 5' untranslated region (UTR), and 108 bases at the 3' UTR [1, 4]. The total number of amino acids in ESR2 gene (residue/ translational length) is 530 [5] .

Although strong experimental evidence suggests that *ESR2* plays a role in carcinogenesis, the results of epidemiologic investigations are less persuasive. For example, a few polymorphic variants of the *ESR2* gene have been associated with an increased risk of common cancers like prostate [6, 7], colorectal [8], and breast cancers [9-14] in some studies. Only one [6] out of four studies showed the association between *ESR2* SNP variant and prostate cancer risk: rs29877983 located in the promoter region was significantly associated with prostate cancer risk and with localized carcinomas [6, 7, 15, 16]. The one available study on *ESR2* gene and colon and rectal cancer showed that G allele of rs1256049 is associated with increased risk of rectal cancer among the total population if diagnosed before 60 years of age [8]. Six out of nine breast cancer studies found statistically significant association between breast cancer and either single nucleotide variants [10, 14] or haplotypes [9, 11, 12] or CA repeat [13] of *ESR2* [9-14, 17-19]. However, only two of them showed the association with single nucleotide variants of *ESR2*: (1) rs8018687 (*5772G) and rs4986938 (*38A) are associated with breast cancer risk in women with benign breast disease[10], and (2) C(14206)T and rs1256054 are associated with breast in postmenopausal women [14]. In addition, no study has yet examined the association between *ESR2* gene polymorphisms and ERβ protein expression in tumor lung.

We conducted a study of *ESR2* gene polymorphisms in relation to immunohistochemical expression of ERβ in lung tumors. Our hypothesis was that genetic variation in the ER-beta gene might alter the protein expression level of the gene in lung tumor. In addition to 22 selected single nucleotide polymorphisms, we also investigated associations between *ESR2* haplotypes and both cytoplasmic and nuclear expression.

5.3 METHODS

5.3.1 Study Population

The study population included n=204 21 year-old and older persons who received surgery at a University of Pittsburgh Medical Center hospital for the staging or treatment of pathologically confirmed primary lung cancer. We assembled risk factor, tumor, and follow-up information from several sources, including outpatient paper charts, inpatient and outpatient electronic medical records, hospital-based cancer registries, and Social Security Death Index database searches.

As described in Figure 5-1, a DNA sample was obtained for 185 (90.7%) of 204 study subjects. Due to lack of sufficient amounts and purity of DNA sample for genotyping, 13 subjects were dropped and genotyping attempted for only 172 subjects. Among 172 subjects genotyped, only 146 subjects had good genetic data (high call rates). Subjects with less than four missing SNPs out of 18 SNPs in the first plex or with less than two missing out of 4 SNPs in the second plex were considered as having the high call rates. In this study, statistical analyses were performed only with 135 subjects with non-missing lung tumor expression data and with good *ESR2* genotyping data. The research used formalin-fixed and paraffin-embedded tissue specimens, processed as tissue microarray (TMA) cores or as whole tissue sections in order to obtain the immunohistochemical expression of ERβ protein in lung tumors. The University of Pittsburgh Institutional Review Board approved subject recruitment and tissue use protocols.

5.3.2 Single nucleotide polymorphisms (SNPs) selection methodology

We conducted searches for known *ESR2* SNPs in the human from five data sources: (1) OVID Medline®, (2) NCBI Entrez SNP [3], (3) the Cancer Genome Anatomy Project (CGAP) SNP500Cancer Database[4] [20], (4) the International HapMap Project[5], and (5) FastSNP[6] [21].

Three frequently studied *ESR2* variants in relation to cancers were identified through OVID Medline literature search: (1) rs1256049 (RsaI): a silent G1082A SNP in exon 6 (ligand binding domain), (2) rs4986938 (AluI): A1730G SNP in the 3'-untranslated region of exon 8, and (3) D14S1026: a CA dinucleotide repeat polymorphism in intron 5 [22].

A HapMap Data Rel 24/phase II Nov 08 database (NCBI build 36) query restricted to the CEU population (N=90 Utah residents with ancestry from northern and western Europe) identified 169 SNPs in chromosome 14 (position 63743506 to 63895021), a 151.5 kb genomic region spanning 20 kb upstream and 20 kb downstream of the estrogen receptor beta isoform 2 (NM_001040276).

SNP500Cancer, Entrez SNP, FastSNP, and CEU HapMap database searches identified a total of 1,149 SNPs according to dbSNP identifier ("rs number"), including 154 SNPs common to CEU HapMap and non-HapMap sources (SNP500Cancer, Entrez SNP, and FastSNP). SNPs were considered as high priority SNPs if they were included in SNP500Cancer SNPs, coding SNPs in Entrez SNP or FastSNP, and promoter-regulator SNPs in FastSNP. SNP500Cancer,

[3] http://www.ncbi.nlm.nih.gov/sites/entrez
[4] http://snp500cancer.nci.nih.gov/home_1.cfm
[5] http://www.hapmap.org/
[6] http://fastsnp.ibms.sinica.edu.tw/pages/input_CandidateGeneSearch.jsp

Entrez SNP, and FastSNP database searches identified 29 high priority SNPs, including 11 CEU HapMap SNPs.

5.3.3 Haplotype tag-SNP (htSNP) selection procedure

As noted above, a HapMap search initially identified 169 CEU *ESR2* Phase II SNPs. However, 49 *ESR2* SNPs had a zero minor allele frequency (MAF) in the CEU population. The de Bakker pairwise Tagger algorithm [23] at an $R^2 = 0.80$ threshold, as implemented in Haploview 4.1 [24], was used to select TagSNPs and the AluI SNP (rs4986938) and the RsaI SNP (rs1256049), identified from the literature search, and four eligible high priority SNPs (rs8006145, rs1256031, rs1256030, and rs3020450) were forced in the selection . Tagger selected 34 htSNPs, including 28 SNPs within the *ESR2* gene, capturing all 120 SNPs with mean $R^2 = 0.967$. Nine of the 34 htSNPs captured only low-frequency-low-priority SNPs (MAF < 0.05). The SNP500Cancer SNPs rs1256031 captured the six SNPs tagged by the adjacent SNP500Cancer SNP rs1256030.

Therefore, in total, 25 htSNPs remained after excluding rs1256030 and the low-frequency-low-priority SNPs. Replacing two low priority SNPs with linked alternatives, a set of 25 htSNPs could be genotyped. These 25 htSNPs captured 104 (87%) of the 120 CEU HapMap SNPs within 20 kb of *ESR2* at $R^2 \geq 0.80$ with mean $R^2 = 0.961$. Due to unusual amount of white powder in one of the primers for rs1256031 (high priority SNP), a set of 24 htSNPs were genotyped on two Sequenom multi-plex panels.

5.3.4 Laboratory Assay

5.3.4.1 TMA construction, immunohistochemical staining, and evaluation

TMA construction included three 0.6mm diameter lung tumor cores per subject with examination of hematoxylin- and eosin-stained sections to verify malignant content. Preparations for immunohistochemisty (IHC) included deparaffinization and hydration with xylene and ethanol, heat induced antigen retrieval with 10mM citrate buffer at pH 6, quenching endogenous peroxidase with 3% hydrogen peroxide for 5 min at room temperature, and blocking with non-immune normal serum for 5-20 min at room temperature. ERβ staining used anti-ERβ (MCA1974ST, Serotec) at 1:20 dilution in PBS overnight at 4°C and EnVision™ reagents (DAKO Corp., Carpinteria, CA). Final steps consisted of incubation with diaminobenzidine (DAB) chromogenic substrate at room temperature for 5-10 min and counterstaining with hematoxylin for 2-2.5 min. Breast cancer tissue, with and without the application of primary antibodies, were used as positive and negative IHC controls.

The study lung pathologist (S.D.) assessed each TMA core and whole section for percentage of tumor cells stained and for intensity of staining. Scoring for the percentage of tumor cells stained used a six-level ordinal scale (0 to 5, respectively, for no cells stained, 0-1% cells stained, 2-10% cells stained, 11-33% cells stained, 34-66% cells stained, and 67-100% cells stained). Scoring for intensity of staining used a four-level ordinal scale (0 to 3, respectively, for no, weak, moderate, and strong staining). Data analyses expressed IHC expression in terms of the Allred score (range 0 to 8), the sum of the percentage and intensity scores [25]. The Allred scores were averaged for the multiple cores from each patient. For each patient, there is a nuclear score, a cytoplasmic score, and a total score (sum of nuclear and cytoplasmic scores).

5.3.4.2 DNA preparation

DNA was extracted by three different methods depending on the source of DNA available. For either whole blood or tissue samples, DNA (50 subjects) was isolated by using Gentra Systems Inc. (Minneapolis, MN) DNA isolation kits (M.R.). The EASY-DNA Kit[7] from Invitrogen Corporation (Carlsbad, CA) was used to extract DNA from frozen lung tissues of 89 subjects (J.S. and J.Y.S.). Genomic DNA was also isolated from formalin-fixed and paraffin-embedded (FFPE) tissue specimens (33 subjects) by standard methods using the DNeasy Kit[8] from Qiagen Inc. (Valencia, CA). DNA extraction from FFPE was performed by Clinical Genomics Facility of Department of Pathology at University of Pittsburgh Medical School.

DNA quantity and quality was assessed using the Thermo Scientific Nanodrop[9] 1000 full-spectrum UV/Vis spectrophotometer.

5.3.4.3 Genotypeing Method

Individual genotyping were performed at the University of Pittsburgh Genomics and Proteomics Core laboratories (GPCL) using MassARRAY® iPLEX Gold (Sequenom, Inc., San Diego, CA). All SNP specific and mass extend oligonucleotides were designed using Sequenom RealSNP (www.realsnp.com); assays were designed using MassARRAY Assay Design version 3.1 (Sequenom, Inc., San Diego, CA). To monitor genotyping quality, a control DNA sample and a DNA-free ('negative') control were included, in duplicate, on every plate.

---

[7] Protocol #3 from Invitrogen's instruction manual for Easy-DNA Kit For genomic DNA isolation (Catalog no. K1800-01). Version F July 21, 2003 25-0056. (http://www.invitrogen.com/site/us/en/home.html)
[8] The extraction protocol is based on the March 2004 revision of the DNeasy Tissue Handbook supplied by Qiagen and modified for PET with support from Qiagen technical support. (http://www.qiagen.com/)
[9] Thermo Scientific Nanodrop (http://www.nanodrop.com)

***Primer Design*:** Three primers are designed for each locus of interest using MassARRAY Assay Design 3.1. The two amplification primers flank the polymorphic site to provide for sample amplification, while the single MassExtend primer lies immediately adjacent to allow for allelic discrimination via single base extension. Assay Design software determines how primer sets can be pooled to optimize multiplex reactions. Mass modifications are incorporated in the design of the MassExtend primers to maximize the mass differential between primers of different loci within a given multiplex pool. Multiplex pools can be designed for up to 36 loci, depending on primer compatibility for the specific loci being assayed.

Sample Amplification: Target loci are amplified within the samples by multiplex PCR in 1X PCR buffer (Qiagen) containing 3.5 mM MgCl2, 25 mM dNTPs, 500 nM each forward and reverse amplification primer within the multiplex pool and 2.5 U HotStar Taq (Qiagen). PCR conditions are: 95^{o}C for 15 minutes for Taq activation followed by 45 cycles of 94^{o}C for 20 seconds, 56^{o}C for 20 seconds and 72^{o}C for 1 minute. A single extension for 1 minute at 72^{o}C completes the PCR reaction. dNTPs and primers are removed by incubation with 0.5 U shrimp alkaline phosphotase (SAP) at 37 ^{o}C for 40 minutes. SAP is inactivated by incubation at 87 ^{o}C for 5 minutes.

MassExtend: Excess MassExtend primers corresponding to the loci represented by the amplification primers used are pooled. Higher mass primers are added at a higher concentration to adjust for signal drop off during spectra acquisition. Single base extension is carried out in 0.2X iPLEX buffer plus, 1X termination mix (containing mass modified termination nucleotides), 1X iPLEX enzyme and primers at 0.84 ºM, 1.04 ºM and 1.25 ºM as appropriate to the relative mass of the primer. A double cycle amplification program performs 40 cycles of denaturation at 94 oC for 5 seconds followed by 5 cycles of 52oC for 5 seconds, 80 oC for 5

seconds, back to 94 oC for a total of 200 cycles. A final extension at 72 oC for 3 minutes completes the amplification. Clean resin and water is added to the MassExtend reaction products. Samples are incubated in clean resin at room temperature with mixing for 5 minutes and centrifuged at 3200 x g for 5 minutes.

NanoDispense, Spectra acquisition and analysis: Samples are dispensed to a SpectraChip using the MassArray Nanodispenser according to manufacturer's instructions. Spectra chips are loaded into the MassArray analyzer and spectra acquired for each sample. MassArray Typer software uses the known mass of the MassExtend primers to identify each locus, and the increase caused by each distinct nucleotide to identify the alleles present in the sample.

We observed 100% concordance rates in replicated samples. Centre d'Etude du Polymorphisme Humain (CEPH#7038) positive controls and water negative controls were included in two 192 well plates as part of quality control measures. Since two htSNPs (both rs1273196 and rs8018687 had 0% call rate) failed the genotyping, genetic information of only 22 htSNPs were used in the analysis. For 21 of 22 SNP assays we were able to obtain genotyping results for over 98% of study subjects with good genetic data (N=132 in plex1 and N=144 in plex2). One SNP assay (rs1152589) produced a genotype result in 94.7%.

5.3.5 Statistical Analysis

The total number of subjects included for analysis in this study is 135 subjects. A test for deviation from Hardy-Weinberg Equilibrium genotype frequencies was done for each htSNPs among all study subjects and Whites. Statistical analyses used Kruskal-Wallis Test or Wilcoxon

rank-sum tests to evaluate the significance of differences in the clinicopathologic features for IHC expression scores. The Jonckheere-Terpstra test was used to test the null hypothesis that the distribution of the ERβ IHC expression does not differ among genotypes of 22 htSNPs. We assumed a dominant model of inheritance to evaluate the magnitude of association [Odds ratios (OR) and 95% confidence interval (CI)] between *ESR2* genotype and ER-beta protein expression. For those htSNPs which showed statistically different distributions of the ERβ IHC expression among genotypes based on the Jonckheere-Terpstra test, unconditional logistic regression model was fitted to assess the association between genotype of htSNPs and cytoplasmic and nuclear ER-beta expression score in lung tumors for all subjects.

Haplotype-based analyses used the Expectation-Maximization (EM) algorithm implemented in SAS Genetics (PROC HAPLOTYPE) to estimate group-level haplotype frequencies and to generate subject-level haplotype probability weights. EM algorithm refers to a statistical method commonly used to estimate haplotype frequencies from genotype data where genetic phase is ambiguous for individuals who are heterozygous at more than one loci.[10] I then used logistic regression (implemented in SAS PROC LOGISTIC) to estimate independent associations between the haplotype probability weights and ERβ expression category. A standard LD-plot was produced by Haploview for Whites. All analyses used SAS 9.2 (SAS Institute, Inc., Cary, North Carolina) and two-sided *p*-values.

Variables used in data analyses included age at tissue collection (continuous and categorical), race (White, African-American), sex (women, men), smoking status (never, former, active), smoking dose duration among ever smokers (1-25, 26-50, 51-75, >76 pack-years), stage

[10] Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Molecular Biology & Evolution. 1995 Sep;12(5):921-927.

group (I, II, III, IV, recurrent), and histology group (adenocarcinoma, bronchioloalveolar carcinoma(BAC), adenosquamous carcinoma, squamous cell carcinoma, large cell carcinoma, undifferentiated carcinoma, malignant carcinoid, small cell carcinoma). In some analyses, ever cigarette smokers with unknown quit status were grouped with active smokers. For subjects without pathologic stage information, clinical stage information was used instead.

Sample size calculation was performed with significance level of alpha=0.05 (two-sided), 80% power (beta=0.20), and various minor allele frequencies of *ESR2* SNPs. The sample size calculation was performed for both the recessive and dominant models by treating ER-beta protein expression as categorical variable. The power analysis software, Power Analysis and Sample Size (PASS)[11], were used to perform the sample size calculation. This may provide less power for other hypotheses testing including stratifications by gender, histological types of lung cancer, and smoking history.

5.4 RESULTS

Selected subject characteristics are shown in Table 5-1. A total of 135 lung cancer patients who were satisfactorily genotyped and had ER-beta lung tumor expression data were included in our analyses. Fifty four percent were women, 88.9% Whites, and 5.2% African-Americans (Table 5-1). Few were never smokers (9.6%). Forty-eight percent had adenocarcinoma while 36.3% had squamous cell carcinoma of lung. Median Allred scores were 7.0 for cytoplasmic ERβ and 8.0 for nuclear ERβ expression.

[11] Pass 2000 (January 21, 2005): Hintze J. (2004). NCSS and Pass. Number Cruncher Statistical Systems. Kaysville, Utah. www.ncss.com

Table 5-2 shows subject characteristics by total, cytoplasmic, and nuclear ERβ expression scores. African Americans have higher cytoplasmic ERβ expression than Whites (p=0.05). Median nuclear ERβ expression score was lower in tumors from younger (30-59 years) lung cancer patients than older (greater than 60 years) (p=0.01). Interestingly, lung cancer patients who are dead at the last contact date had statistically significantly higher cytoplasmic ERβ expression score than those who are alive (p=0.03).

Twenty two selected htSNPs were genotyped satisfactorily. All htSNPs were in Hardy-Weinberg equilibrium among all study subjects (p>0.05). One htSNP (rs1256120) departed from HWE among White lung cancer patients (p=0.031). Except two htSNPs (rs1273196 and rs8018687) failure, no other genotyping error was detected among the duplicates, corresponding to an estimated error rate of 0.0%.

The distributions of nuclear ERβ IHC expression scores for all study subjects differed significantly among the genotypes of three htSNPs (rs8021944, rs1256061, and rs10146204) (Table 5-3). The genotype of TG or GG in first SNP (rs8021944) had higher overall nuclear ERβ expression distribution than the wild-type TT (p=0.029). Subjects with rare variant allele (CA+AA) in rs1256061 also showed significantly higher distribution of nuclear ERβ expression than the subjects with the wild-type CC (p=0.022). Only one SNP (rs10146204) showed significant differences of distribution in both cytoplasmic and nuclear ERβ expression scores among genotypes [cytoplasmic (3 level comparison of GG, GA, and AA): p=0.032 and nuclear (2 levels of GG and GA+AA): p=0.025]. For the other 19 htSNPs, no significant difference was found in distribution of cytoplasmic or nuclear ERβ expression scores among genotypes.

Subgroup analysis based on histological types of lung cancer suggests that the rs1256061 association with ERβ expression may be specific to adenocarcinoma of lung (Table 5-4).

Patients diagnosed with adenocarcinoma of lung showed statistically significant associations between that rare variant allele (CA+AA) and high scores of both cytoplasmic and nuclear ERβ expression (p=0.023 and p=0.027, respectively). But, no association was observed among squamous cell carcinoma patients.

To perform logistic regression analyses for three htSNPs which showed significant results from the Jonckheere-Terpstra test, ERβ expression was classified into three groups based on the distribution of expression results in either cytoplasm or nuclear. Cytoplasmic expression had three ordered categories of Allred=0 as a reference group, Allred > 0 and <8, and Allred = 8 while nuclear expression had Allred ≤ 6 as a reference group, Allred > 6 and <8, and Allred =8. Increasing doses of the variant allele (A) at rs1256061 and at rs10146204 were associated with increased risk of having maximum nuclear ERβ expression (Allred = 8) (Table 5-5). Maximum ERβ expression was observed more often in tumors from patients with the CA or AA genotype [nuclear: Odds Ratios (OR) relative to Allred ≤ 6: 3.54 (95% confidence interval 1.22-10.3) and cytoplasmic: OR relative to Allred=0: 5.08 (95% CI 1.47-17.6)] than the CC genotype at rs1256061 (Table 5-5). Maximum ERβ expression was observed more often in patients with the GA or AA genotype [nuclear: OR relative to Allred ≤ 6: 3.71 (95% CI 1.31-10.6) and cytoplasmic: OR relative to Allred=0: 4.00, 95% CI 1.26-12.7)] than the GG genotype at rs10146204 (Table 5-5).

Only five of the eight possible haplotypes had a frequency of ≥ 1% in White study subjects, based on the three selected htSNPs. Two haplotypes (G-C-A and G-C-G) had zero frequency and one haplotype (G-A-G) had very low estimated frequency (0.0006). Maximum nuclear ERβ expression score (Allred score=8 vs. ≤ 6) was observed more often in patients with the haplotype T-A-A (OR=11.43, 95% CI 1.06-123) than the haplotype T-C-G (Table 5-6). We

found no additional evidence of association between ERβ expression in lung tumor and any of the haplotypes.

5.5 DISCUSSION

We investigated the association of *ESR2* gene polymorphisms and the ERβ expression in lung tumors. To our knowledge, this is the first study evaluating genetic variation of *ESR2* gene and its relationship with immunihistochemical expression of nuclear and cytoplasmic ERβ in lung tumor. In this study, we identified statistically significant ERβ expression associations with three htSNPs (rs8021944, rs1256061, and rs10146204).

All of three identified htSNPs were not our high priority SNPs since they were included in SNP500Cancer SNPs, coding SNPs in Entrez SNP or FastSNP, and promoter-regulator SNPs in FastSNP. One of three selected htSNPs (rs8021944) is a spectrin repeat containing nuclear envelope 2 (*SYNE2*) gene while another htSNPs (rs10146204) is not a part of *ESR2* transcription region. These two htSNPs could be included in our study since we used a 151.5 kb genomic region spanning 20 kb upstream and 20 kb downstream of the *ESR2* gene to select tagger SNPs.

In this study, two frequently studied *ESR2* SNPs (rs1256049 [RsaI] and rs4986938 [AluI]) did not show an association with ERβ expression in lung tumors. While the inheritance of one or another of these two specific *ESR2* SNPs has been studied in relation to cancers of the colon or rectum [8], endometrium [26], ovary [27], prostate [7, 15, 28], and breast [9-12, 14, 18, 19], only few studies demonstrated significant association between the SNPs and cancer risk: rs1256049 (RsaI) with increased risk of rectal cancer [8] and rs4986938 (AluI) with breast cancer risk [10].

Three identified htSNPs were in week linkage disequilibrium (LD) with one another (highest r^2=0.24), Figure 5-2. According to LD-plot (Figure 5-2), rs1256061 had moderate linkage equilibrium with rs4986938 [AluI] (r^2=0.54) while rs4986938 [AluI] is also in moderate LD with rs8006145, the high priority (r^2=0.65). SNP (rs10146204) and rs3020450 (high priority) were also in moderate linkage disequilibrium (r^2=0.60). The exact functions of two htSNPs (rs1256061 and rs10146204) that showed significant association with ERβ expression in this study was not known; however, their moderate LD with frequently studied SNP and high priority SNPs supports the thought that they might be potentially functional.

Logistic regression analyses revealed that increasing doses of the variant allele (A) at rs1256061 were associated with increased risk of having maximum nuclear ERβ expression (P_{trend}=0.0147). Also, htSNP (rs1256061) is associated with ERβ expression among patients with adenocarcinoma, but not with squamous cell carcinoma. Its association specific to one histological subtype of lung cancer and the dose-response relationship with ERβ expression supports the causality assumption based our hypothesis: the genetic variation in the ER-beta gene might alter the protein expression level of the gene in lung tumor.

Previous studies reported on nuclear ERβ expression as a favorable prognostic factor for lung cancer [2, 3, 29-31]. Thus, the negative association of ERβ expression with advanced stages of lung cancer was expected. However, we did not observe significant association between stage and nuclear ERβ expression in lung tumor (p=0.16) while patients with stage IV had lower median scores than patients with lower stages (I-III).

In this study, lung cancer patients who died during the follow-up had significantly higher cytoplasmic ERβ expression score than those who survived. This result agrees with our previous report of cytoplasmic ERβ as a negative prognostic factor for lung cancer patients (not published)

[32]. We also evaluated the role of variant alleles of 22 htSNPs in lung cancer survival but no significant association was found (no data shown).

This study had several limitations. The small sample size was problematic especially for conducting haplotype analyses and subgroup analyses with specific histological groups and race. Therefore, we had large 95% confidence intervals for the association between *ESR2* gene polymorphisms and immunohistochemical expression of ERβ in lung tumors. In addition to the fact that all of our study subjects were lung cancer patients, who may have different genotype distribution from general population, the deviation of HWE observed in our study may have been influenced by the small size. Also, majority of study population was Whites, thus these results may not apply to lung cancer patients of other races or ethnicities. Interestingly, although only a small percentage of participants were African-American, African Americans had statistically significantly higher cytoplasmic ERβ expression than Whites. DNA samples were extracted by three different methods from the different types of sources such as whole blood, FFPE, and frozen tissue. The call rates of genotyping results were significantly different among the extraction methods. This may result the selection bias. Since some of DNA samples were extracted from lung tumors, genetic variants observed in the research may not be inherited but acquired and influenced by environmental factors. Also, our study may have the misclassification bias due to the two different methods (TMA and whole section) used for the IHC expression assay. Since the IHC expression results from TMA were averaged value of the multiple cores, we cannot account possible heterogeneous variances among observations obtained on the same subject.

In spite of these limitations, our study had strengths. Our study did not have the observer bias since the genotyping procedures were performed in blinded fashion to IHC expression

results. Also, in our knowledge, this is the first study investigated the relationship of genetic variants of *ESR2* with both cytoplasmic and nuclear expressions of ERβ in lung tumors, which were detected by IHC method.

In conclusion, we found that individuals with individuals with at least one rare allele of two htSNPs (rs1256061 and rs10146204) had statistically significant association with maximum expression of both cytoplasmic and nuclear ERβ expression in the dominant inheritance model, compared to non-carriers. Our finding that one SNP (rs1256061) is associated with ERβ expression among patients with adenocarcinoma, but not with squamous cell carcinoma, suggests the need to perform subgroup analysis with various histological groups of lung cancer patients. The genetic variants examined in this study should be investigated with a larger cohort of lung cancer patients to replicate our findings.

5.6 REFERENCE

1. Kuiper, G.G., et al., *Cloning of a novel receptor expressed in rat prostate and ovary.* Proc Natl Acad Sci U S A, 1996. **93**(12): p. 5925-30.
2. Schwartz, A.G., et al., *Nuclear estrogen receptor beta in lung cancer: expression and survival differences by sex.* Clin Cancer Res, 2005. **11**(20): p. 7280-7.
3. Wu, C.T., et al., *The significance of estrogen receptor beta in 301 surgically treated non-small cell lung cancers.* J Thorac Cardiovasc Surg, 2005. **130**(4): p. 979-86.
4. Kuiper, G.G. and J.A. Gustafsson, *The novel estrogen receptor-beta subtype: potential role in the cell- and promoter-specific actions of estrogens and anti-estrogens.* FEBS Lett, 1997. **410**(1): p. 87-90.
5. Ogawa, S., et al., *The complete primary structure of human estrogen receptor beta (hER beta) and its heterodimerization with ER alpha in vivo and in vitro.* Biochem Biophys Res Commun, 1998. **243**(1): p. 122-6.
6. Thellenberg-Karlsson, C., et al., *Estrogen receptor beta polymorphism is associated with prostate cancer risk.* Clinical Cancer Research, 2006. **12**(6): p. 1936-41.
7. Chen, Y.-C., et al., *Sequence variants of estrogen receptor beta and risk of prostate cancer in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium.* Cancer Epidemiology, Biomarkers & Prevention, 2007. **16**(10): p. 1973-81.
8. Slattery, M.L., et al., *Associations between ERalpha, ERbeta, and AR genotypes and colon and rectal cancer.* Cancer Epidemiology, Biomarkers & Prevention, 2005. **14**(12): p. 2936-42.
9. Breast and Prostate Cancer Cohort Consortium, et al., *Haplotypes of the estrogen receptor beta gene and breast cancer risk.* International Journal of Cancer, 2008. **122**(2): p. 387-92.
10. Gallicchio, L., et al., *Polymorphisms in estrogen-metabolizing and estrogen receptor genes and the risk of developing breast cancer among a cohort of women with benign breast disease.* BMC Cancer, 2006. **6**: p. 173.
11. Gold, B., et al., *Estrogen receptor genotypes and haplotypes associated with breast cancer risk.* Cancer Research, 2004. **64**(24): p. 8891-900.
12. Maguire, P., et al., *Estrogen receptor beta (ESR2) polymorphisms in familial and sporadic breast cancer.* Breast Cancer Research & Treatment, 2005. **94**(2): p. 145-52.
13. Tsezou, A., et al., *Association of repeat polymorphisms in the estrogen receptors alpha, beta (ESR1, ESR2) and androgen receptor (AR) genes with the occurrence of breast cancer.* Breast, 2008. **17**(2): p. 159-66.
14. Zheng, S.L., et al., *Joint effect of estrogen receptor beta sequence variants and endogenous estrogen exposure on breast cancer risk in Chinese women.* Cancer Research, 2003. **63**(22): p. 7624-9.
15. McIntyre, M.H., et al., *Prostate cancer risk and ESR1 TA, ESR2 CA repeat polymorphisms.* Cancer Epidemiology, Biomarkers & Prevention, 2007. **16**(11): p. 2233-6.
16. Nicolaiew, N., et al., *Association between estrogen and androgen receptor genes and prostate cancer risk.* European Journal of Endocrinology, 2009. **160**(1): p. 101-6.

17. Iobagiu, C., et al., *Microsatellite profile in hormonal receptor genes associated with breast cancer.* Breast Cancer Research & Treatment, 2006. **95**(2): p. 153-9.
18. Georgopoulos, N.A., et al., *Estrogen receptor polymorphisms in tamoxifen-treated women with breast cancer.* Gynecological Endocrinology, 2006. **22**(4): p. 185-9.
19. Forsti, A., et al., *Polymorphisms in the estrogen receptor beta gene and risk of breast cancer: no association.* Breast Cancer Research & Treatment, 2003. **79**(3): p. 409-13.
20. Packer, B., et al., *SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes.* Nucleic Acids Research, 2006. **34 Database issue**: p. D617-D621.
21. Yuan, H.-Y., et al., *FASTSNP: An always up-to-date and extendable service for SNP function analysis and prioritization.* Nucleic Acids Research, 2006. **34**: p. W635-W641.
22. Gennari, L., et al., *Estrogen receptor gene polymorphisms and the genetics of osteoporosis: a HuGE review.* American Journal of Epidemiology, 2005. **161**(4): p. 307-20.
23. de Bakker, P., et al., *Efficiency and power in genetic association studies.* Nature Genetics, 2005. **37**: p. 1217-23.
24. Barrett, J., et al., *Haploview: Analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. **21**: p. 263–5.
25. Allred, D.C., et al., *Prognostic and predictive factors in breast cancer by immunohistochemical analysis.* Mod Pathol, 1998. **11**(2): p. 155-68.
26. Setiawan, V.W., et al., *Estrogen receptor beta (ESR2) polymorphisms and endometrial cancer (United States).* Cancer Causes & Control, 2004. **15**(6): p. 627-33.
27. Leigh Pearce, C., et al., *Comprehensive evaluation of ESR2 variation and ovarian cancer risk.* Cancer Epidemiology, Biomarkers & Prevention, 2008. **17**(2): p. 393-6.
28. Sun, Y.-h., et al., *[Association between single-nucleotide polymorphisms in estrogen receptor beta gene and risk of prostate cancer]*. Chung-Hua Wai Ko Tsa Chih [Chinese Journal of Surgery], 2005. **43**(14): p. 948-51.
29. Kawai, H., et al., *Estrogen receptor alpha and beta are prognostic factors in non-small cell lung cancer.* Clin Cancer Res, 2005. **11**(14): p. 5084-9.
30. Skov, B.G., B.M. Fischer, and H. Pappot, *Oestrogen receptor beta over expression in males with non-small cell lung cancer is associated with better survival.* Lung Cancer, 2008. **59**(1): p. 88-94.
31. Nose, N., et al., *Association between estrogen receptor-beta expression and epidermal growth factor receptor mutation in the postoperative prognosis of adenocarcinoma of the lung.* J Clin Oncol, 2009. **27**(3): p. 411-7.
32. Stabile, L.P., et al., *Combined analysis ofestrogen receptor β and progesterone receptor expression identifies lung cancer patients with poor outcome.* Clinical Cancer Research, 2010. **Submitted, not published yet**.

5.7 TABLES AND FIGURES

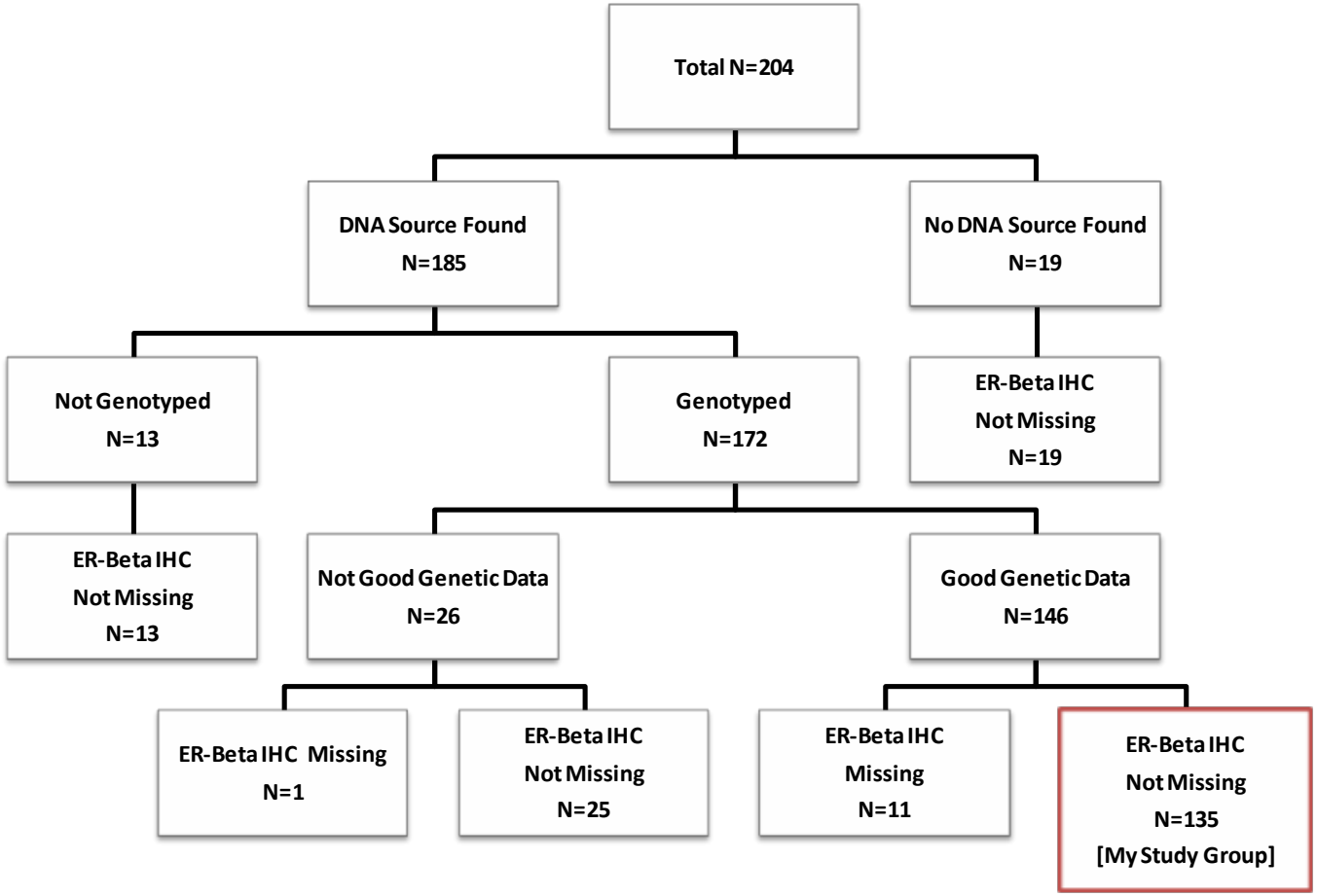


Figure 5-1 Study subject selection flow chart

*[Good Genetic Data] is defined as [subjects with less than 4 missing SNPs out of 18 SNPs in Plex#1] or [subjects with less than 2 missing SNPs out of 4 SNPs in Plex#2]

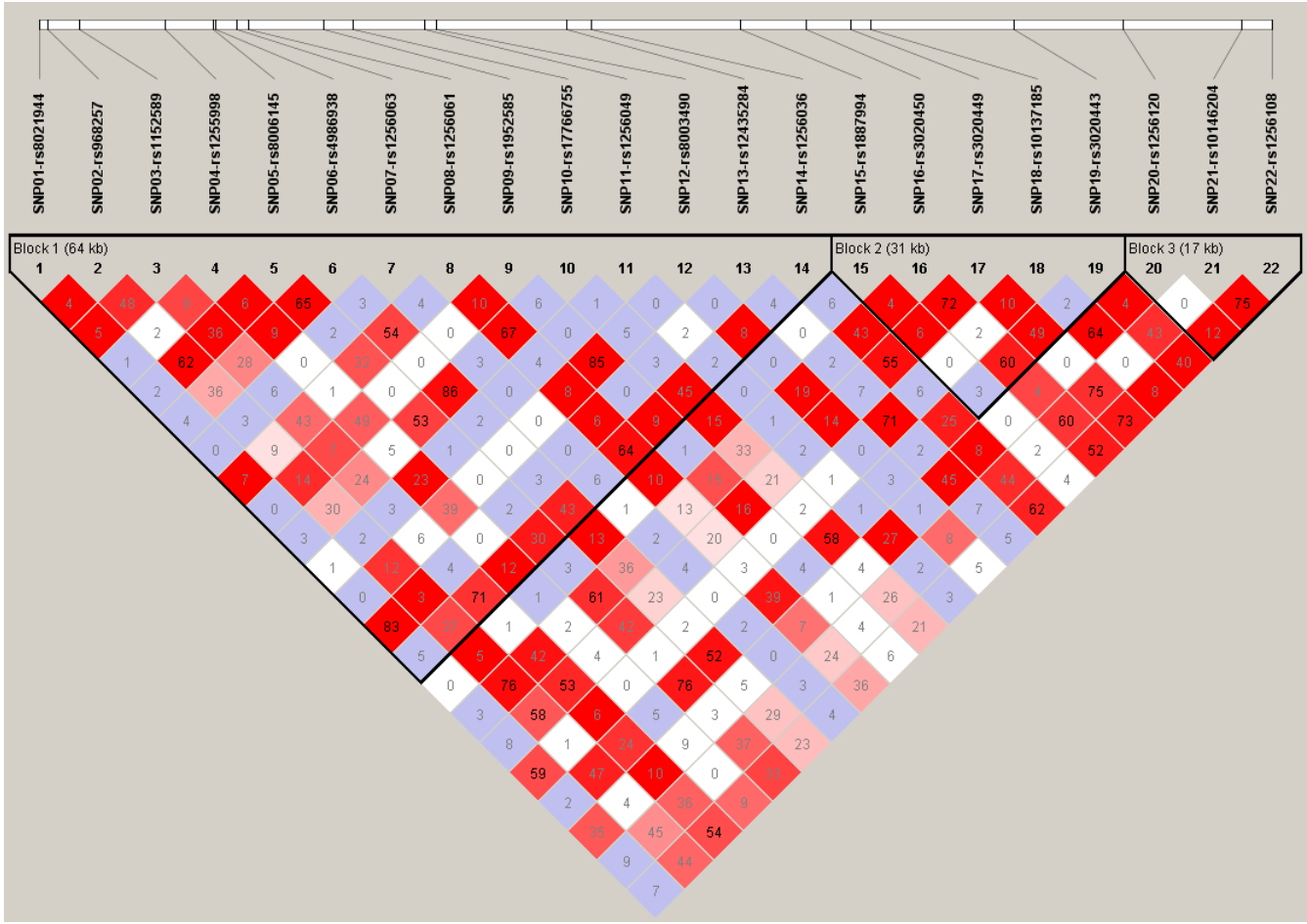


Figure 5-2 Linkage disequilibrium (LD) plot, drawn by Haploview using Solid Spine of LD method, is for White only. The 22 SNPs were in three haplotype blocks (as highlighted). The pattern of LD among Whites was indicated by a different color scheme: Bright red: D-prime=1 and LOD>=2, Shades of pink and red: D-prime<1 and LOD>=2, Blue: D-prime=1 and LOD<2, and White: D-prime<1 and LOD<2.

Table 5-1 Patient Characteristics

		All (n=135)	
Characteristic	**Measure**	**No.**	**Percent**
Survival status	Dead	87	64.4
	Alive	40	29.6
	Unknown	8	5.9
Sex	Men	62	45.9
	Women	73	54.1
Race	White	120	88.9
	Afircan-American	7	5.2
	Unknown	8	5.9
Age	30-59 years	31	23.0
	60-69 years	42	31.1
	70+ years	62	45.9
Smoking status	never smoker	13	9.6
	ex-smoker	54	40.0
	active smoker	62	45.9
	Unknown	6	4.4
Smoking dose-duration	1-25 pack-years	20	17.2
(among ever smokers=116)	26-50 pack-years	45	38.8
	51-75 pack-years	22	19.0
	>76 pack-years	25	21.6
	Unknown	4	3.4
Stage	I	53	39.3
	II	24	17.8
	III	40	29.6
	IV	5	3.7
	recurrent	7	5.2
	Unknown	6	4.4
Histology	Adenocarcinoma	65	48.1
	BAC	1	0.7
	Adenosquamous	4	3.0
	Squamous cell	49	36.3
	Large cell	6	4.4
	Undifferentiated	3	2.2
	Malignant carcinoid	1	0.7
	Small cell	2	1.5
	Unknown	4	3.0
Histology class	Adenocarcinoma	65	48.1
	Squamous cell	49	36.3
	Other/unknown	21	15.6
ERβ expression score	nuclear	135[a]	7.14 (8.0)[b]
	cytoplasmic	135[a]	5.38 (7.0)[b]
	total	135[a]	12.52 (14.75)[b]

[a] Number of subjects with non-missing IHC data
[b] Mean and median of Allred score, medians in parentheses.

Table 5-2 Associations between median ER beta Allred scores and personal characteristics

	Total ERβ Total N=135			Cytoplasmic ERβ Total N=135			Nuclear ERβ Total N=135		
	N[a]	Median	*p*-value*	N[a]	Median	*p*-value*	N[a]	Median	*p*-value*
STATUS AT LAST CONTACT			0.04			0.03			0.28
Dead	87	15		87	7		87	8	
Alive	40	14		40	6		40	8	
SEX			0.51			0.57			0.45
Male	62	15		62	7		62	8	
Female	73	14.6		73	7		73	8	
RACE			0.04			0.05			0.27
African-American	7	16		7	8		7	8	
White	120	14.75		120	7		120	8	
AGE (years)			0.45			0.75			0.01
30-59	31	14		31	6.5		31	7.5	
60-69	42	14.75		42	7		42	8	
70+	62	15		62	7		62	8	
SMOKING STATUS			0.84			0.70			0.85
active smoker	43	15		43	7		43	8	
ex-smoker	54	14.2		54	6.775		54	8	
smoker, NOS	19	14		19	7		19	8	
never smoker	13	15		13	7		13	8	
SMOKING DOSE-DURATION (among ever smokers)			0.58			0.66			0.23
1-25	20	15		20	7		20	8	
26-50	45	15		45	7		45	8	
51-75	22	13.875		22	6.375		22	8	
>76	25	14		25	6.8		25	7.6	

Table 5-2 (continued)

	Total ERβ Total N=135			Cytoplasmic ERβ Total N=135			Nuclear ERβ Total N=135		
	N[a]	Median	*p*-value*	N[a]	Median	*p*-value*	N[a]	Median	*p*-value*
STAGE			0.70			0.55			0.16
I	53	15		53	7		53	8	
II	24	14.875		24	7		24	8	
III	40	15		40	7		40	8	
IV	5	14.6		5	7.2		5	7.4	
recurrent	7	15.25		7	7.25		7	7.8	
HISTOLOGY			0.73			0.75			0.47
Adenocarcinoma	65	14.75		65	7		65	8	
BAC	1	15		1	7		1	8	
Adenosquamous	4	11.5		4	3.5		4	8	
Squamous cell	49	15		49	7		49	8	
Large cell	6	12		6	5.5		6	7.5	
Undifferentiated	3	10		3	3.5		3	6.5	
Malignant carcinoid	1	16		1	8		1	8	
Small cell	2	9.625		2	3.625		2	6	
HISTOLOGY CLASS			0.19			0.14			0.17
Adenocarcinoma	65	14.8		65	7		65	8	
Squamous cell	49	15		49	7		49	8	
Other/missing	21	10		21	5		21	7.5	

Total N=Number of subjects with non-missing IHC data

[a] Number of subjects with non-missing IHC data

*Wilcoxon rank sum test (Wilcoxon two-sample test) with a continuity correction of 0.5 for comparing two independent groups (e.g. sex and race)

* Kruskal-Wallis Test for comparing more than two non-parametric independent groups

Table 5-3 Association between *ESR2* SNPs and ER-beta IHC expression for all study subjects (N=135)

		Cytoplasmic ERβ					Nuclear ERβ				
SNP	Genotype	N	P25	Med	P75	p-value*	N	P25	Med	P75	p-value*
rs8021944	TT	118	3.00	7.00	8.00	0.083	118	7.00	8.00	8.00	0.028
	TG	14	6.00	7.83	8.00		14	8.00	8.00	8.00	
	GG	1	7.33	7.33	7.33		1	8.00	8.00	8.00	
	TG+GG	15	6.00	7.75	8.00	0.081	15	8.00	8.00	8.00	0.029
rs968257	AA	44	0.75	7.00	8.00	0.826	44	6.00	8.00	8.00	0.312
	AG	55	5.00	7.00	8.00		55	7.00	8.00	8.00	
	GG	22	0.00	6.00	7.70		22	7.00	8.00	8.00	
	AG+GG	77	4.00	7.00	8.00	0.576	77	7.00	8.00	8.00	0.439
rs1152589	AA	31	4.00	6.80	8.00	0.586	31	7.75	8.00	8.00	0.189
	AT	59	4.80	7.00	8.00		59	7.00	8.00	8.00	
	TT	26	0.00	6.25	7.75		26	6.00	8.00	8.00	
	AT+TT	85	3.20	7.00	8.00	0.864	85	6.50	8.00	8.00	0.064
rs1255998	CC	100	3.35	7.00	8.00	0.240	100	7.00	8.00	8.00	0.164
	CG	32	4.00	7.00	7.33		32	7.20	7.78	8.00	
	GG	1	3.50	3.50	3.50		1	6.50	6.50	6.50	
	CG+GG	33	4.00	7.00	7.25	0.259	33	7.00	7.75	8.00	0.185
rs8006145	CC	61	3.00	7.00	7.90	0.600	61	6.50	8.00	8.00	0.730
(Priority)	CA	49	4.00	7.00	8.00		49	7.00	8.00	8.00	
	AA	11	0.00	6.75	8.00		11	8.00	8.00	8.00	
	CA+AA	60	3.75	7.00	8.00	0.570	60	7.00	8.00	8.00	0.068
rs4986938	GG	44	1.50	7.00	7.63	0.397	44	6.50	8.00	8.00	0.086
(AluI)	GA	61	4.00	7.00	8.00		61	7.00	8.00	8.00	
	AA	16	4.75	7.00	8.00		16	7.50	8.00	8.00	
	GA+AA	77	4.00	7.00	8.00	0.462	77	7.00	8.00	8.00	0.137
rs1256063	CC	108	3.50	7.00	8.00	0.756	108	7.00	8.00	8.00	0.271
	CT	13	2.50	7.00	7.25		13	7.00	7.75	8.00	
	CT+TT	13	2.50	7.00	7.25	0.756	13	7.00	7.75	8.00	0.271
rs1256061	CC	33	0.00	6.50	7.25	0.551	33	6.00	7.67	8.00	0.632
	CA	67	4.00	7.00	8.00		67	7.00	8.00	8.00	
	AA	21	4.00	7.00	8.00		21	7.75	8.00	8.00	
	CA+AA	88	4.00	7.00	8.00	0.054	88	7.25	8.00	8.00	0.022
rs1952585	TT	96	3.35	7.00	8.00	0.133	96	7.00	8.00	8.00	0.190
	TC	24	3.25	5.80	7.23		24	6.65	7.68	8.00	
	CC	1	6.75	6.75	6.75		1	8.00	8.00	8.00	
	TC+CC	25	3.50	6.00	7.20	0.130	25	6.80	7.75	8.00	0.173
rs17766755	GG	46	0.00	7.00	7.50	0.322	46	6.50	8.00	8.00	0.097
	GA	62	3.50	7.00	8.00		62	7.00	8.00	8.00	
	AA	12	4.75	7.00	8.00		12	7.50	8.00	8.00	
	GA+AA	74	3.50	7.00	8.00	0.375	74	7.00	8.00	8.00	0.140
rs1256049	GG	112	3.35	7.00	8.00	0.421	112	7.00	8.00	8.00	0.584
(RsaI)	GA	8	2.00	6.00	7.20		8	6.75	7.80	8.00	
	GA+AA	8	2.00	6.00	7.20	0.421	8	6.75	7.80	8.00	0.584

Table 5-3 (continued)

		Cytoplasmic ERβ					Nuclear ERβ				
SNP	Genotype	N	P25	Med	P75	p-value*	N	P25	Med	P75	p-value*
rs8003490	GG	110	4.00	7.00	8.00	0.072	110	7.00	8.00	8.00	0.062
	GA	22	0.00	5.55	7.20		22	6.50	7.50	8.00	
	AA	1	6.75	6.75	6.75		1	8.00	8.00	8.00	
	GA+AA	23	0.00	5.60	7.20	0.054	23	6.50	7.50	8.00	0.119
rs12435284	CC	109	3.00	7.00	8.00	0.073	109	7.00	8.00	8.00	0.087
	CT	12	6.50	7.95	8.00		12	8.00	8.00	8.00	
	CT+TT	12	6.50	7.95	8.00	0.073	12	8.00	8.00	8.00	0.087
rs1256036	AA	33	3.50	6.00	8.00	0.541	33	7.00	8.00	8.00	0.220
	AG	67	5.00	7.00	8.00		67	7.00	8.00	8.00	
	GG	21	0.00	6.00	7.50		21	6.00	8.00	8.00	
	AG+GG	88	3.10	7.00	8.00	0.774	88	6.90	8.00	8.00	0.434
rs1887994	GG	102	3.00	7.00	8.00	0.584	102	7.00	8.00	8.00	0.981
	GT	19	4.80	7.00	8.00		19	7.00	8.00	8.00	
	GT+TT	19	4.80	7.00	8.00	0.584	19	7.00	8.00	8.00	0.981
rs3020450	GG	52	2.25	7.00	8.00	0.582	52	6.75	8.00	8.00	0.354
(Priority)	GA	53	5.00	7.00	8.00		53	7.00	8.00	8.00	
	AA	16	1.50	6.38	8.00		16	7.50	8.00	8.00	
	GA+AA	69	3.50	7.00	8.00	0.886	69	7.00	8.00	8.00	0.727
rs3020449	TT	38	0.00	6.45	7.75	0.394	38	6.00	8.00	8.00	0.092
	TC	61	5.50	7.00	8.00		61	7.40	8.00	8.00	
	CC	21	3.00	6.75	8.00		21	7.60	8.00	8.00	
	TC+CC	82	4.00	7.00	8.00	0.156	82	7.40	8.00	8.00	0.114
rs10137185	CC	106	3.00	7.00	8.00	0.087	106	6.80	8.00	8.00	0.155
	CT	14	6.80	7.13	8.00		14	7.90	8.00	8.00	
	TT	1	8.00	8.00	8.00		1	8.00	8.00	8.00	
	CT+TT	15	6.80	7.25	8.00	0.080	15	7.90	8.00	8.00	0.149
rs3020443	AA	66	3.00	7.00	7.90	0.432	66	6.50	8.00	8.00	0.140
	AC	45	4.00	7.00	8.00		45	7.00	8.00	8.00	
	CC	9	0.00	7.00	8.00		9	8.00	8.00	8.00	
	AC+CC	54	3.50	7.00	8.00	0.433	54	7.00	8.00	8.00	0.109
rs1256120	TT	100	3.00	7.00	8.00	0.805	100	6.90	8.00	8.00	0.400
	TC	16	5.75	7.00	7.58		16	7.30	7.95	8.00	
	CC	3	4.00	8.00	8.00		3	8.00	8.00	8.00	
	TC+CC	19	5.50	7.00	8.00	0.567	19	7.60	8.00	8.00	0.843
rs10146204	GG	42	0.00	6.63	7.33	0.032	42	6.00	7.75	8.00	0.258
	GA	57	5.75	7.00	8.00		57	7.50	8.00	8.00	
	AA	22	3.00	5.50	8.00		22	7.00	8.00	8.00	
	GA+AA	79	4.00	7.00	8.00	0.051	79	7.40	8.00	8.00	0.025
rs1256108	TT	30	0.00	6.20	7.75	0.494	30	5.75	8.00	8.00	0.255
	TC	67	5.60	7.00	8.00		67	7.40	8.00	8.00	
	CC	34	3.50	6.78	8.00		34	7.50	8.00	8.00	
	TC+CC	101	4.00	7.00	8.00	0.119	101	7.50	8.00	8.00	0.211

*Jonckheere-Terpstra Test

Table 5-4 Association between rs1256061 genotype variants and ER-beta IHC expression among lung cancer patients with adenocarcinoma or squamous cell carcinoma

		Cytoplasmic ERβ					Nuclear ERβ				
Histology	Genotype	N	P25	Med	P75	p-value*	N	P25	Med	P75	p-value*
Adenocarcinoma	CC	15	0.00	5.50	7.00	0.496	15	5.50	7.67	8.00	0.651
	CA	36	4.90	7.00	8.00		36	7.30	8.00	8.00	
	AA	9	6.00	7.00	8.00		9	8.00	8.00	8.00	
	CA+AA	45	5.50	7.00	8.00	0.023	45	7.75	8.00	8.00	0.027
squamous cell carcinoma	CC	12	5.00	7.23	7.88	0.990	12	6.95	8.00	8.00	0.925
	CA	20	6.20	7.00	7.88		20	7.55	8.00	8.00	
	AA	8	1.75	6.88	8.00		8	7.25	8.00	8.00	
	CA+AA	28	4.75	7.00	8.00	0.952	28	7.55	8.00	8.00	0.785

*Jonckheere-Terpstra Test

Table 5-5 Crude Odds Ratios for the association between three SNPs and cytoplasmic and nuclear ER-Beta IHC expression scores among all study subjects (N=135)

	ERβ cytoplasmic expression							ERβ nuclear expression						
	Allred = 0	Allred > 0 AND Allred < 8			Allred = 8			Allred ≤ 6	Allred > 6 AND Allred < 8			Allred = 8		
Genotype	n	n	OR	95% CI	n	OR	95% CI	n	n	OR	95% CI	n	OR	95% CI
rs8021944														
TT	25	60	Ref		33	Ref		19	31	Ref		68	Ref	
TG	1	7	2.92	0.34-25.0	6	4.55	0.51-40.2	1	1	0.61	0.04-10.4	12	3.35	0.41-27.4
GG	0	1			0			0	0			1		
TG+GG	1	8	3.33	0.40-28.0	6	4.55	0.51-40.2	1	1	0.61	0.04-10.4	13	3.63	0.45-29.6
rs1256061														
CC	11	17	Ref		5	Ref		9	9	Ref		15	Ref	
CA	9	37	2.66	0.93-7.61	21	5.13	1.38-19.1	9	14	1.56	0.45-5.41	44	2.93	0.98-8.76
AA	4	8	1.29	0.31-5.35	9	4.95	1.02-24.1	1	5	5.00	0.48-51.8	15	9.00	1.01-80.1
CA+AA	13	45	2.24	0.84-5.96	30	5.08	1.47-17.6	10	19	1.90	0.57-6.31	59	3.54	1.22-10.3
rs10146204														
GG	12	23	Ref		7	Ref		11	11	Ref		20	Ref	
GA	7	29	2.16	0.73-6.37	21	5.14	1.45-18.2	6	12	2.00	0.55-7.25	39	3.58	1.15-11.1
AA	5	10	1.04	0.29-3.76	7	2.40	0.55-10.5	2	5	2.50	0.40-15.7	15	4.13	0.79-21.5
GA+AA	12	39	1.70	0.66-4.39	28	4.00	1.26-12.7	8	17	2.13	0.65-6.95	54	3.71	1.31-10.6

Table 5-6 *ESR2* haplotypes and ER-beta IHC expression among only white subjects

Haplotype weight*	Freq	ERβ cytoplasmic expression: Allred > 0 AND Allred < 8 vs. Allred = 0 OR	95% CI	ERβ cytoplasmic expression: Allred = 8 vs. Allred = 0 OR	95% CI	ERβ nuclear expression: Allred > 6 AND Allred < 8 vs. Allred ≤ 6 OR	95% CI	ERβ nuclear expression: Allred = 8 vs. Allred ≤ 6 OR	95% CI
T-C-G	0.44	Ref		Ref		Ref		Ref	
T-A-A	0.25	0.668	0.12-3.83	4.07	0.54-31.0	4.15	0.32-54.0	11.43	1.06-123
T-A-G	0.15	1.46	0.12-18.5	0.92	0.04-21.6	10.19	0.45-233	1.56	0.09-27.2
T-C-A	0.10	0.992	0.08-11.9	0.11	0.00-4.48	26.87	0.61-	1.39	0.06-31.9
G-A-A	0.06	4.857	0.05-454	40.98	0.37-	1.066	0.00-479	28.42	0.34-

*Haplotype is composed of alleles in the order of rs8021944, rs1256061, and rs10146204.

6.0 DISCUSSION

In all three projects, the immunohistochemical assay was used to detect protein expression level. Also, the Tissue microarrays (TMAs) were constructed using randomly selected formalin-fixed, paraffin-embedded lung tumor tissue blocks from each patient specimen and the protein expression status of biomarkers. The laboratory assay procedures were performed in blinded fashion to outcome-related information.

In the first project, a multilevel generalized linear mixed model was used to control for sample type and to comply with repeated measures from TMA with discrete response. This model accounts the correlations among repeated IHC readings from TMA data on the same subject, and also for some possible heterogeneous variances among observations obtained on the same subject. Through modeling the correlation among repeated measures from TMA, we could obtain the best linear unbiased predictions.

In the second project, we presumed that protein expression patterns transmit fundamental information about underlying tumor biology and attempted to identify meaningful expression patterns involving these seven interesting and relevant proteins. Even though our study did not identify two major subgroups with differing host and tumor characteristics or clinical outcomes, our finding is important due to the biological functions of the proteins composed in each cluster which supports the idea of autocrine HGF-c-Met signaling plays significant roles in the progression of lung tumors.

At the last project, we found that individuals with at least one rare allele of two htSNPs (rs1256061 and rs10146204) are associated with maximum expression of both cytoplasmic and nuclear ERβ expression in the dominant inheritance model, compared to non-carriers. The last project produced results from first study of the relationship between *ESR2* gene variation and ERβ lung tumor expression.

The main limitation of these projects is the limited study population diversity: approximately 90% of study population is Caucasian. However, our study has the largest sample size among the previously reported studies on the HGF or c-Met expression in lung cancer patients. The small sample size may provide less power for other hypotheses testing including stratifications by gender, histological types of lung cancer, and smoking history. These projects have the retrospective cohort study design. Since the analysis of the study depends on preexisting records, I have limited control over the incompleted datasets. Therefore, unmeasured confounders, measurement error, and missing datasets could influence the study results.

More research is needed to fully understand the association between the immunohistochemical expression of protein markers in lung tumors and the lung cancer survival. It would be useful to replicate our findings regarding the ESR2 genetic variation and ERβ expression in lung tumors with a large cohort study where various host, tumor, and outcome information collection procedures were taken as part of the study protocol. Large cohort study will provide more power to perform subgroup analysis with various histological groups of lung cancer. This may eliminate the selection bias and measurement errors.

APPENDIX A

SUPPLEMENTAL TABLES AND FIGURES FOR PROJECT#1

A.1 DESCRIPTION FOR HGF AND C-MET DATABASE

Table A-1 Subjects Elimination Steps for Cleaned Database of HGF and c-Met

	Description	Subject ID	Total Number of Subjects
Start	Received Laboratory Data: TMA=126 & Whole section=77		203
	6 Duplicated observations: select only TMA data	430, 448, 593, 604, 671, 920	197
	Case status is not Lung cancer and Lung cancer histology="N/A"	V-101, V-102, 1520,1542, 1701, 1744	191
	IHC from Whole section which used "normal lung tissue"	948, 999	189
	Younger than 21 years old (age_at_tissue_collection)	V-101, V-102, 682	188
	Overall survival time is zero, died on the same day (surgery date), age = 7	682	188
Final	Survival Time is missing due to no death status information	V-101, V-102,1520,1542, 1701, 1744, 301, 317, 683, L-012, L-024, L-031, L-033, L-037,	180

NOTE: Whole-section: n=65 and TMA: n=115

Table A-2 Distribution of the difference between Age at diagnosis and Age at tissue collection

Age at tissue collection − Age at diagnosis	N
0	150
1	15
2	6
4	3
5	3
6	1
7	1
11	1

Table A-3 Percent missing between the TMA study and the Whole-section study

	Whole section n=65	TMA n=115	p-value*
Race	9.2	0.0	0.0019
Smoking status	1.5	5.2	0.4245
Smoking level	6.2	7.0	1.0000
Stage	1.5	0.0	0.3611

*Fisher exact test

Table A-4 HGF and c-Met expression and non-expression frequencies among total study subject (N=180)

		Total (N=180)		
		c-Met		
		Missing	Non-Missing	Total
HGF	Missing	9	2	11
	Non-Missing	1	168	169
	Total	10	170	180

		Whole section (N=65)		
		c-Met		
		Missing	Non-Missing	Total
HGF	Missing	8	2	10
	Non-Missing	1	54	55
	Total	9	56	65

		TMA (N=115)		
		c-Met		
		Missing	Non-Missing	Total
HGF	Missing	1	0	1
	Non-Missing	0	114	114
	Total	1	114	115

Table A-5 Subject characteristics: TMA *vs.* Whole section

Variable	Measure	All n=180	Tissue source: Whole section n=65	Tissue source: TMA n=115	*p*-value[1]
Survival status	Dead, %	68.3	69.2	67.8	0.87
Sex	Women, %	51.1	50.8	51.3	1.00
Race	African-American, %	9.2	10.2	8.7	0.79
Age	30-59 years, %	22.2	30.8	17.4	0.13
	60-69 years, %	34.4	30.8	36.5	
	70+ years, %	43.3	38.5	46.1	
Smoking status	never smoker, %	5.8	6.3	5.5	0.09
	ex-smoker, %	43.4	32.8	49.5	
	current smoker, %	50.9	60.9	45.0	
Smoking dose-duration (among ever smokers)	<50 pack-years, %	56.3	54.4	57.4	0.74
	50+pack-years, %	43.7	45.6	42.6	
Stage	IA	17.9	15.6	19.1	0.42
	IB	25.7	31.3	22.6	
	IIA/B	19.6	20.3	19.1	
	III	27.4	28.1	27.0	
	IV	9.5	4.7	12.2	
Histology	squamous cell carcinoma	33.9	33.9	33.9	0.18
	non-squamous non-small cell	57.8	63.1	54.8	
	undifferentiated	6.7	1.5	9.6	
	small cell carcinoma	1.7	1.5	1.7	
HGF[3]	High expression[4], %	49.1	29.1	58.8	0.0003
	Allred, Median	7.0	6.0	7.5	<.0001[2]
c-Met[3]	High expression[4], %	50.0	42.9	53.5	0.25
	Allred, Median	7.1	7.0	7.3	0.87[2]

[1]Fisher exact test, except where indicated otherwise
[2] Wilcoxon rank sum test
[3]Using subject-specific Allred values averaged across TMA cores
[4]Allred >7
Whole section: 6 missing race, 1 missing smoking status, 3 missing smoking dose-duration (among ever smokers), 1 missing stage
TMA: 6 missing smoking status, 2 missing smoking dose-duration (among ever smokers)
Smoking dose duration Total N: All=163, Whole section=60 , TMA=103

A.2 ASSOCIATIONS BETWEEN HGF AND C-MET AND SUBJECTS CHARACTERISTICS: DATASET WITH AVERAGED ALLRED SCORE

Table A-6 Frequency of high HGF and high c-Met IHC expression according to subject category

	HGF (N=169)			**c-Met** (N=170)		
	N	High (%)	*p*-value*	N	High (%)	*p*-value*
STATUS AT LAST CONTACT			0.09			0.18
Alive	52	53.8		53	35.8	
Dead	117	47.0		117	47.9	
SEX			0.35			0.76
women	86	47.7		88	45.5	
men	83	39.8		82	42.7	
RACE			0.27			0.16
African-American	14	28.6		14	64.3	
White	149	45.6		150	42.7	
AGE			0.81			0.73
30-59	35	42.9		35	40.0	
60-69	57	47.4		58	48.3	
70+	77	41.6		77	42.9	
SMOKING STATUS			0.04			0.46
never smoker	10	20.0		9	22.2	
active smoker	82	46.3		82	45.1	
ex-smoker	70	60.0		72	44.4	
SMOKING DOSE-DURATIONS (among ever smokers)			0.87			0.51
<50 pack-years	83	45.8		84	47.6	
50+ pack-years	64	48.4		65	41.5	
PATHOLOGIC STAGE			0.02			0.24
IA	30	56.7		30	60.0	
IB	44	47.7		44	34.1	
IIA/B	34	55.9		34	38.2	
III	45	53.3		46	47.8	
IV	16	12.5		16	43.8	
HISTOLOGY GROUP			0.70			0.19
SCCA	57	49.1		58	55.2	
Adeno, SQUAM, BAC, Carcinoid	97	48.5		97	39.2	
NSCLS, large cell carcinoma	12	58.3		12	33.3	
small cell carcinoma	3	33.3		3	33.3	

High HGF and c-Met expression defined by subject-specific averaged Allred values above IHC source-specific Allred median cutpoints (HGF cutpoints: 7.5 for TMA and 6.0 for whole section; c-Met cutpoints: 7.25 for TMA and 7.0 for whole section)

Total N=Number of subjects with non-missing IHC data

HGF: 6 missing race (33.3% high expression), 7 missing smoking status (14.3% high expression), 22 missing smoking dose-duration (22.7% high expression)

c-Met: 6 missing race (33.3% high expression), 7 missing smoking status (57.1% high expression), 21 missing smoking dose-duration (38.1% high expression)

*Fisher exact test

Table A-7 Bivariate associations between HGF and c-Met Allred scores and personal characteristics

	N	Median HGF	*p*-value*	N	Median c-Met	*p*-value*
STATUS AT LAST CONTACT			0.3167			0.0901
Alive	52	7.6		53	7.0	
Dead	117	7.0		117	7.3	
Sex			0.4138			0.8519
women	86	7.3		88	6.1	
men	83	7.0		82	6.0	
RACE			0.5015			0.1574
African-American	14	6.8		14	7.6	
White	149	7.3		150	7.0	
AGE			0.5403			0.4196
30-59	35	7.0		35	7.0	
60-69	57	7.3		58	7.3	
70+	77	7.0		77	7.0	
SMOKING STATUS			0.0768			0.2127
never smoker	10	6.8		9	6.0	
active smoker	82	7.0		82	7.3	
ex-smoker	70	7.5		72	7.0	
Smoking dose-duration (among ever smokers)			0.8147			0.6568
<50 pack-years	83	7.3		84	7.3	
50+ pack-years	64	7.4		65	7.0	
PATHOLOGIC STAGE			0.0595			0.6585
IA	30	7.3		30	7.6	
IB	44	7.0		44	7.0	
IIA/B	34	7.5		34	7.0	
III	45	7.3		46	7.3	
IV	16	6.0		16	7.2	
HISTOLOGY GROUP			0.9771			0.2766
SCCA	57	7.0		58	7.5	
Adeno, SQUAM, BAC, Carcinoid	97	7.0		97	7.0	
NSCLS, large cell carcinoma	12	7.3		12	6.1	
small cell carcinoma	3	7.0		3	5.8	

*Wilcoxon rank sum test (Wilcoxon two-sample test) with a continuity correction of 0.5 for comparing two independent groups (e.g. sex and race)

* Kruskal-Wallis Test for comparing more than 2 non-parametric independent groups

Table A-8 Subject Characteristics and Median HGF and c-Met Allred Score: Whole section vs. TMA

	HGF						c-Met					
	Whole section			TMA			Whole section			TMA		
	N	Median	*p**	N	Median	*p**	N	Median	*p**	N	Median	*p**
STATUS AT LAST CONTACT			0.90			0.08			0.46			0.10
Alive	16	6.0		36	7.8		17	7.0		36	6.8	
Dead	39	6.0		78	7.3		39	7.0		78	7.3	
SEX			0.51			0.72			0.54			0.77
women	28	6.0		58	7.6		30	7.0		58	7.3	
men	27	5.0		56	7.5		26	7.0		56	7.2	
RACE			0.59			0.09			0.38			0.21
African-American	4	6.5		10	6.8		4	7.5		10	7.6	
White	45	6.0		104	7.6		46	7.0		104	7.2	
AGE			0.09			0.92			0.84			0.50
30-59	15	6.0		20	7.3		15	7.0		20	7.2	
60-69	16	6.5		41	7.5		17	7.0		41	7.3	
70+	24	5.0		53	7.5		24	7.0		53	7.0	
SMOKING STATUS			0.11			0.44			0.15			0.76
never smoker	4	3.0		6	7.0		3	6.0		6	6.4	
active smoker	34	6.0		48	7.4		34	7.5		48	7.3	
ex-smoker	16	5.5		54	7.7		18	6.5		54	7.1	
SMOKING DOSE-DURATION (among ever smokers)			0.89			0.43			0.47			0.22
<50 pack-years	26	6.0		57	7.5		27	7.0		57	7.3	
50+ pack-years	21	6.0		43	7.5		22	7.0		43	7.0	
PATHOLOGIC STAGE			0.83			0.00			0.36			0.70
IA	8	6.0		22	7.8		8	8.0		22	7.3	
IB	18	5.5		26	7.5		18	7.0		26	6.9	
IIA/B	12	6.0		22	7.9		12	7.0		22	7.0	
III	15	6.0		30	7.6		16	7.0		30	7.4	
IV	2	3.0		14	6.0		2	3.5		14	7.3	
HISTOLOGY GROUP			0.10			0.83			0.41			0.06
SCCA	18	5.0		39	7.5		19	7.0		39	7.5	
Adeno, SQUAM, BAC, Carcinoid	35	6.0		62	7.6		35	7.0		62	7.2	
NSCLS, large cell carcinoma	1	4.0		11	7.3		1	5.0		11	6.3	
small cell carcinoma	1	5.0		2	7.5		1	8.0		2	5.6	

*Wilcoxon rank sum test (Wilcoxon two-sample test) with a continuity correction of 0.5 for comparing two independent groups (e.g. sex and race).
* Kruskal-Wallis Test for comparing more than 2 non-parametric independent groups.

A.3 RESULTS FROM GENERALIZED LINEAR MIXED MODELS (SAS PROC GLIMMIX): ASSOCIATIONS BETWEEN HGF AND C-MET AND SUBJECTS CHARACTERISTICS: CORRELATED DATASETS [TMA ALLRED SCORE CLUSTERED BY SUBJECTS]

Table A-9 Crude odds ratios (OR) and 95% confidence intervals (CI) for associations between personal characteristics and high HGF and high c-Met IHC expression

	High HGF Expression				High c-Met Expression			
	OR	**95% CI**		***p*-value***	**OR**	**95% CI**		***p*-value***
SEX								
Women	1.00				1.00			
Men	0.64	0.37	1.10	0.11	0.97	0.57	1.65	0.90
RACE								
White	1.00				1.00			
African-American	0.56	0.22	1.41	0.22	2.29	1.02	5.14	0.04
AGE (years)	1.01	0.98	1.04	0.56	0.99	0.96	1.01	0.39
AGE				0.32				0.27
30-59	1.00				1.00			
60-69	1.81	0.82	3.99	0.14	1.56	0.74	3.25	0.24
70+	1.56	0.75	3.23	0.23	0.97	0.49	1.95	0.94
SMOKING STATUS				0.20				0.85
never smoker	1.00				1.00			
active smoker	1.66	0.63	4.37	0.30	1.44	0.40	5.17	0.57
ex-smoker	2.30	0.87	6.09	0.09	1.44	0.40	5.22	0.57
Smoking dose-duration (among ever smokers)				0.44				0.21
<50 pack-years	1.00				1.00			
50+ pack-years	1.26	0.70	2.27	0.44	0.69	0.39	1.23	0.21
PATHOLOGIC STAGE				0.06				0.48
IA	1.00				1.00			
IB	0.64	0.28	1.47	0.28	0.57	0.25	1.32	0.19
IIA/B	0.91	0.36	2.28	0.83	0.56	0.23	1.34	0.19
III/IV	0.40	0.18	0.87	0.02	0.77	0.37	1.58	0.47
HISTOLOGY GROUP				0.78				0.39
Adeno, SQUAM, BAC, Carcinoid	1.00				1.00			
NSCLS, large cell carcinoma	0.81	0.31	2.10	0.66	0.77	0.23	2.55	0.66
SCCA	0.75	0.42	1.34	0.32	1.47	0.83	2.58	0.18
small cell carcinoma	1.09	0.19	6.22	0.92	0.38	0.03	4.40	0.44

HGF and c-Met expression (high vs. low) is defined by median Allred scores. High expression means Allred score>median. Low expression means Allred score ≤ median. [median for HGF from TMA=7.5 and whole section=6.0] and [median for c-Met from TMA=7.25 and whole section=7.0]

Odds ratios comparing individuals with high HGF/c-Met expression to those with low expression unless otherwise specified

*Wald Method for Testing Global Null Hypothesis: beta=0 and Wald's Chi-Square Test (p-value) for each stratified level based on analysis of maximum likelihood estimates

Table A-10 Adjusted odds ratios (OR) and 95% confidence intervals (CI) for associations between personal characteristics and high HGF and high c-Met IHC expression

	High HGF Expression				High c-Met Expression			
	OR	**95% CI**		***p*-value***	**OR**	**95% CI**		***p*-value***
SEX								
Women	1.00				1.00			
Men	0.63	0.36	1.11	0.11	0.91	0.52	1.59	0.74
RACE								
White	1.00				1.00			
African-American	0.95	0.34	2.66	0.91	2.66	1.07	6.59	0.03
AGE (years)	1.00	0.97	1.03	0.98	1.00	0.97	1.03	0.83
AGE				0.67				0.40
30-59	1.00				1.00			
60-69	1.43	0.63	3.24	0.39	1.69	0.76	3.74	0.20
70+	1.37	0.62	3.02	0.43	1.24	0.60	2.57	0.56
SMOKING STATUS				0.14				0.91
never smoker	1.00				1.00			
active smoker	2.35	0.74	7.43	0.15	1.25	0.32	4.82	0.74
ex-smoker	3.08	0.99	9.57	0.05	1.33	0.35	5.13	0.67
PATHOLOGIC STAGE				0.05				0.39
IA	1.00				1.00			
IB	0.66	0.28	1.58	0.35	0.57	0.24	1.33	0.19
IIA/B	1.26	0.47	3.39	0.64	0.48	0.19	1.22	0.12
III/IV	0.43	0.19	0.98	0.05	0.75	0.36	1.55	0.43
HISTOLOGY GROUP				0.88				0.49
Adeno, SQUAM, BAC, Carcinoid	1.00				1.00			
NSCLS, large cell carcinoma	0.77	0.32	1.86	0.56	0.73	0.22	2.48	0.62
SCCA	0.85	0.46	1.57	0.60	1.39	0.76	2.54	0.29
small cell carcinoma	1.31	0.25	6.95	0.75	0.34	0.02	4.89	0.42

High HGF and c-Met expression defined by subject-specific averaged Allred values above IHC source-specific Allred median cutpoints (HGF cutpoints: 7.5 for TMA and 6.0 for whole section; c-Met cutpoints: 7.25 for TMA and 7.0 for whole section)

Odds ratios comparing individuals with high HGF/c-Met expression to those with low expression unless otherwise specified.

*Wald Chi-Square Test (Type 3 Analysis of Effects) for overall and each stratified level based on analysis of maximum likelihood estimates.

Adjusted for age (continuous) , smoking, stage, and sex

Table A-11 Fully odds ratios (OR) and 95% confidence intervals (CI) for associations between personal characteristics and high HGF and high c-Met IHC expression

	High HGF Expression				High c-Met Expression			
	OR	95% CI		*p*-value*	OR	95% CI		*p*-value*
SEX								
Women	1.00				1.00			
Men	0.62	0.35	1.11	0.11	0.82	0.45	1.47	0.49
RACE								
White	1.00				1.00			
African-American	0.90	0.32	2.55	0.84	2.70	1.07	6.77	0.04
AGE (years)	1.00	0.96	1.04	0.98	0.99	0.96	1.02	0.36
SMOKING STATUS				0.14				0.82
never smoker	1.00				1.00			
active smoker	2.65	0.76	9.24	0.13	1.16	0.31	4.37	0.83
ex-smoker	3.48	1.00	12.13	0.05	1.36	0.36	5.08	0.65
PATHOLOGIC STAGE				0.05				0.49
IA	1.00				1.00			
IB	0.75	0.314	1.768	0.50	0.6	0.246	1.441	0.25
IIA/B	1.52	0.551	4.172	0.42	0.49	0.187	1.295	0.15
III/IV	0.46	0.2	1.062	0.07	0.74	0.357	1.537	0.42
HISTOLOGY GROUP				0.79				0.52
Adeno, SQUAM, BAC, Carcinoid	1.00				1.00			
NSCLS, large cell carcinoma	0.72	0.30	1.77	0.47	0.78	0.23	2.66	0.68
SCCA	0.79	0.41	1.49	0.46	1.42	0.76	2.65	0.27
small cell carcinoma	1.29	0.23	7.08	0.77	0.39	0.03	5.43	0.49

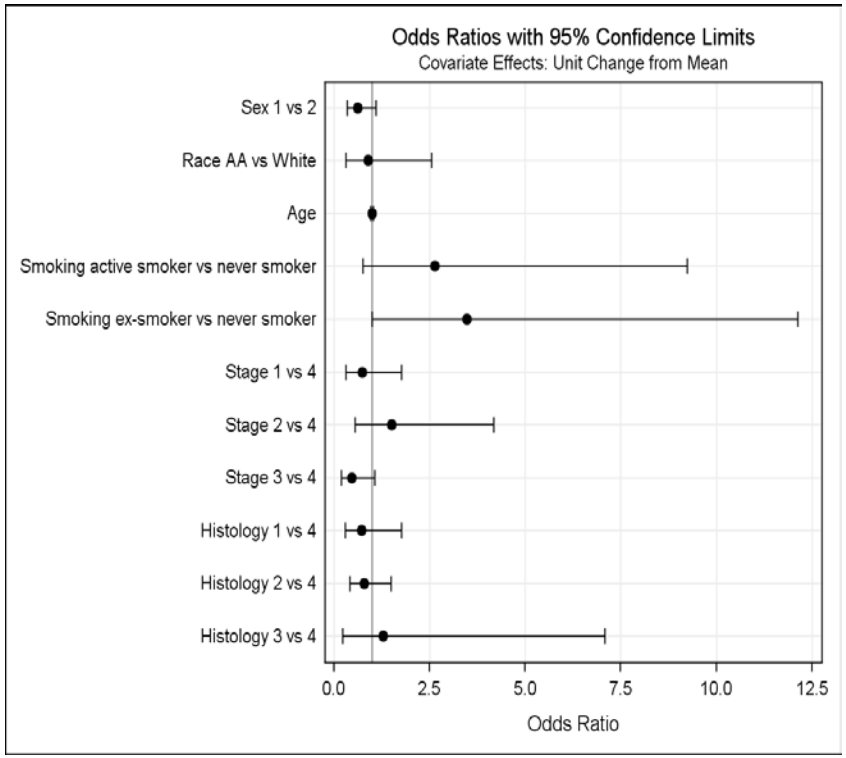
High HGF and c-Met expression defined by subject-specific averaged Allred values above IHC source-specific Allred median cutpoints (HGF cutpoints: 7.5 for TMA and 6.0 for whole section; c-Met cutpoints: 7.25 for TMA and 7.0 for whole section)

Odds ratios comparing individuals with high HGF/c-Met expression to those with low expression unless otherwise specified.

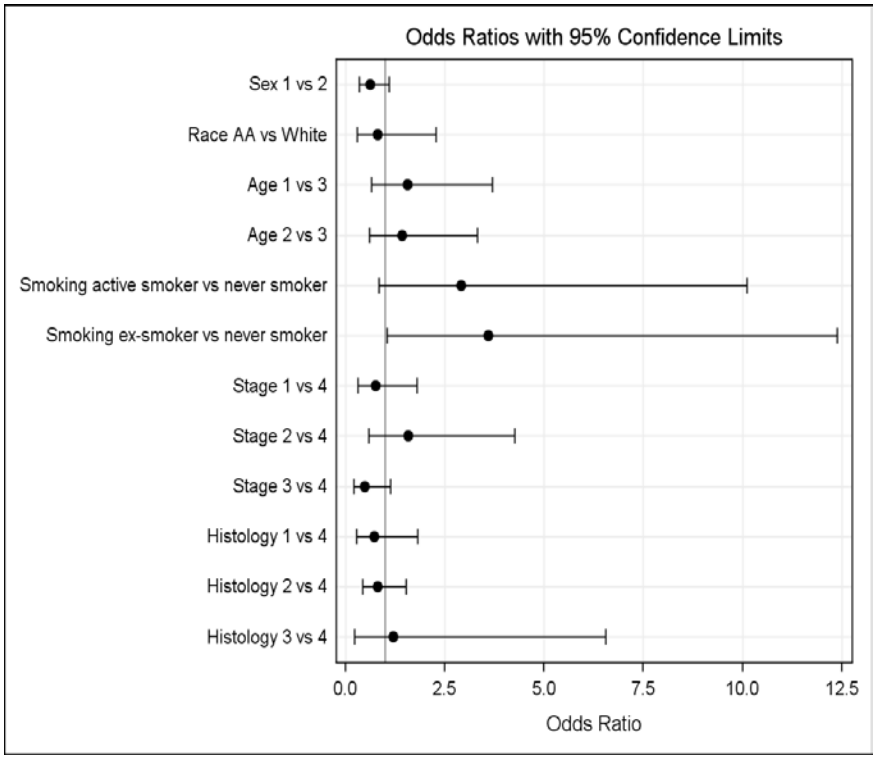
*Wald Chi-Square Test (Type 3 Analysis of Effects) for overall and each stratified level based on analysis of maximum likelihood estimates.

Adjusted for age (continuous) , smoking, stage, and sex, race, and histology

Age in Years



Age categories



Sex: 1=Men, 2=Women; Stage: 1=IB, 2=IIA/B, 3=III/IV, 4=IA; Histology: 1=NSCLS, large cell carcinoma, 2=SCCA, 3=small cell carcinoma, 4=Adeno, SQUAM, BAC, Carcinoid; Age: 1=30-59, 2=60-69, 3=70+.

HGF and c-Met expression (high vs. low) is defined by median Allred scores. High expression means Allred score>median. Low expression means Allred score≤ median. [Median for HGF from TMA=7.5 and whole section=6.0] and [median for c-Met from TMA=7.25 and whole section=7.0].

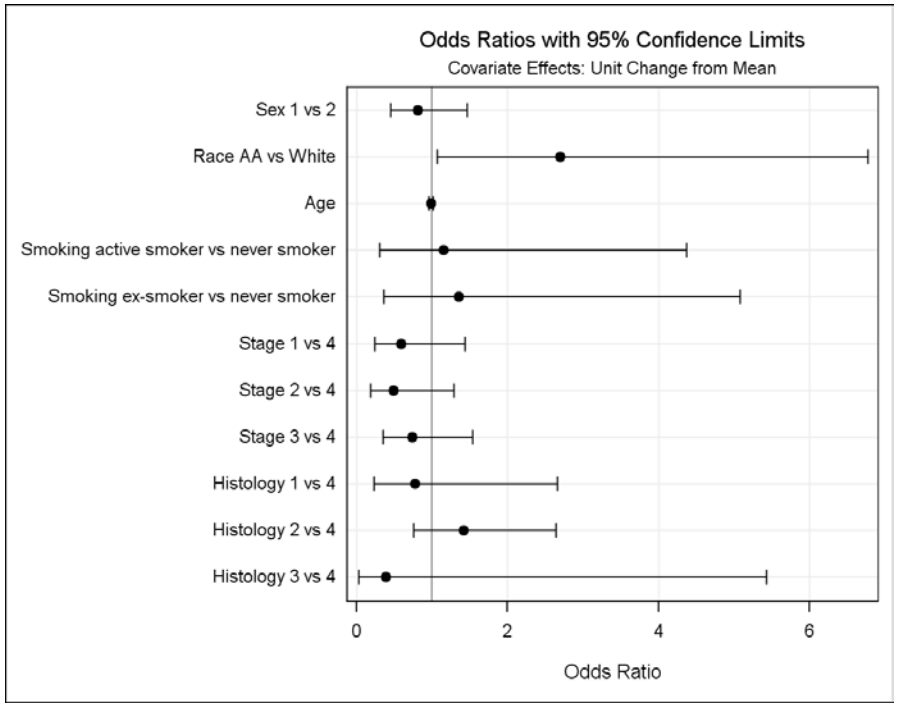
Odds ratios comparing individuals with high HGF/c-Met expression to those with low expression unless otherwise specified.

*Wald Chi-Square Test (Type 3 Analysis of Effects) for overall and each stratified level based on analysis of maximum likelihood estimates

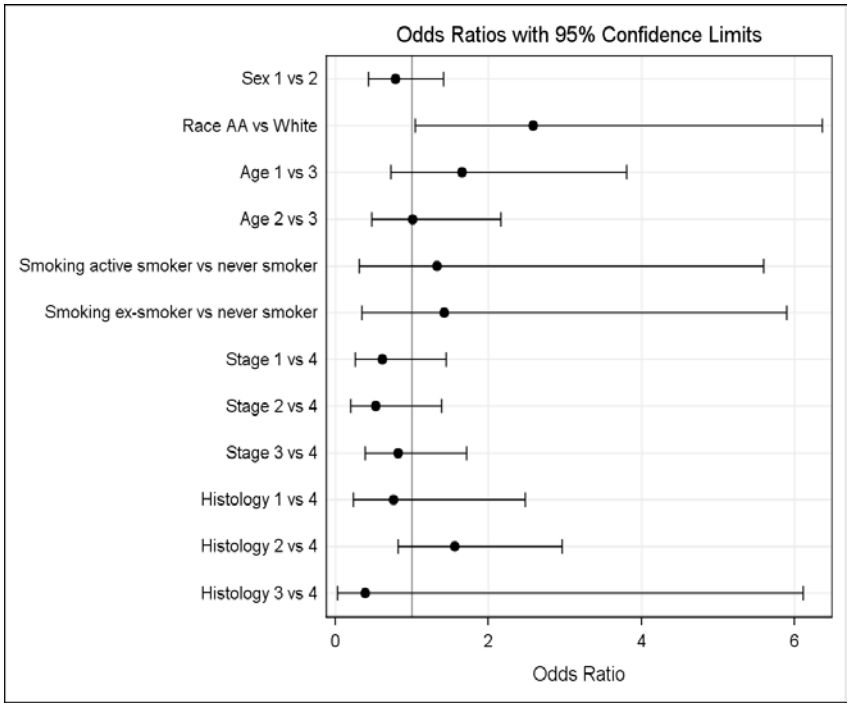
Adjusted for age, smoking, stage, and sex, race, and histology

Figure A-1 Fully odds ratios and 95% confidence intervals (CI) for the association between personal characteristics and HGF status [Age in years vs. Age categories]

Age in Years



Age categories



Sex: 1=Men, 2=Women; Stage: 1=IB, 2=IIA/B, 3=III/IV, 4=IA; Histology: 1=NSCLS, large cell carcinoma, 2=SCCA, 3=small cell carcinoma, 4=Adeno, SQUAM, BAC, Carcinoid; Age: 1=30-59, 2=60-69, 3=70+.

HGF and c-Met expression (high vs. low) is defined by median Allred scores. High expression means Allred score>median. Low expression means Allred score≤ median. [Median for HGF from TMA=7.5 and whole section=6.0] and [median for c-Met from TMA=7.25 and whole section=7.0].

Odds ratios comparing individuals with high HGF/c-Met expression to those with low expression unless otherwise specified.

*Wald Chi-Square Test (Type 3 Analysis of Effects) for overall and each stratified level based on analysis of maximum likelihood estimates

Adjusted for age, smoking, stage, and sex, race, and histology

Figure A-2 Fully odds ratios and 95% confidence intervals (CI) for the association between personal characteristics and c-Met status [Age in years vs. Age categories]

A.4 SURVIVAL ANALYSIS OF LUNG CANCER PATIENTS

Table A-12 Univariate Cox Regression for the Overall Survival

	Hazard Ratio (95%CI)	p-value*
HGF (High Expression)	0.830 (0.571, 1.205)	0.328
HGF Allred Score	1.018 (0.920, 1.127)	0.725
c-Met (High Expression)	0.959 (0.662, 1.388)	0.823
c-Met Allred Score	1.047 (0.918, 1.193)	0.497
Sex (Men)	1.287 (0.901, 1.837)	0.165
RACE (African-American)	1.427 (0.783, 2.602)	0.245
Age (continuous variable)	1.021 (1.002, 1.039)	0.030
AGE		0.041
30-59	1.00	
60-69	1.094 (0.664, 1.804)	0.724
70+	1.675 (1.049, 2.673)	0.031
SMOKING STATUS		0.208
never smoker	1.00	
active smoker	1.670 (0.763, 3.654)	0.200
ex-smoker	1.258 (0.568, 2.789)	0.571
Smoking dose-duration (among ever smokers)		
<50 pack-years	1.00	
50+ pack-years	1.193 (0.815, 1.748)	0.364
PATHOLOGIC STAGE		0.001
IA	1.00	
IB	1.467 (0.788, 2.732)	0.227
IIA/B	2.364 (1.274, 4.387)	0.006
III/IV	2.950 (1.664, 5.232)	0.0002
HISTOLOGY GROUP		0.292
Adeno, SQUAM, BAC, Carcinoid	1.00	
NSCLS, large cell carcinoma	1.312 (0.651, 2.644)	0.448
SCCA	1.396 (0.949, 2.053)	0.090
small cell carcinoma	1.895 (0.593, 6.052)	0.281

*Wald Test (Type 3 test) for Testing Global Null Hypothesis: beta=0 and Chi-Square Test for each stratified level based on analysis of maximum likelihood estimates.

HGF and c-Met expression (high vs. low) is defined by median Allred scores. High expression means Allred score>median. Low expression means Allred score ≤ median. [median for HGF from TMA=7.5 and whole section=6.0] and [median for c-Met from TMA=7.25 and whole section=7.0]

Table A-13 Univariate Cox Regression for the Progression Free Survival

	Hazard Ratio (95%CI)	p-value*
HGF (High Expression)	0.794 (0.541, 1.164)	0.237
HGF Allred Score	0.991 (0.898, 1.095)	0.861
c-Met (High Expression)	1.079 (0.742, 1.571)	0.691
c-Met Allred Score	1.068 (0.932, 1.224)	0.342
Sex (Men)	1.362 (0.948, 1.957)	0.095
RACE (African-American)	1.298 (0.714, 2.362)	0.393
Age (continuous variable)	1.009 (0.991, 1.027)	0.328
AGE		0.425
30-59	1.00	
60-69	1.044 (0.630, 1.729)	0.868
70+	1.304 (0.817, 2.079)	0.266
SMOKING STATUS		0.553
never smoker	1.00	
active smoker	1.217 (0.525, 2.819)	0.647
ex-smoker	0.989 (0.422, 2.318)	0.980
Smoking dose-duration (among ever smokers)		
<50 pack-years	1.00	
50+ pack-years	1.247 (0.846, 1.839)	0.264
PATHOLOGIC STAGE		0.001
IA	1.00	
IB	1.326 (0.710, 2.477)	0.376
IIA/B	2.451 (1.317, 4.562)	0.005
III/IV	2.661 (1.503, 4.713)	0.001
HISTOLOGY GROUP		0.714
Adeno, SQUAM, BAC, Carcinoid	1.00	
NSCLS, large cell carcinoma	1.374 (0.682, 2.766)	0.374
SCCA	1.139 (0.770, 1.687)	0.515
small cell carcinoma	1.524 (0.478, 4.865)	0.477

*Wald Test (Type 3 test) for Testing Global Null Hypothesis: beta=0 and Chi-Square Test for each stratified level based on analysis of maximum likelihood estimates

HGF and c-Met expression (high vs. low) is defined by median Allred scores. High expression means Allred score>median. Low expression means Allred score ≤ median. [median for HGF from TMA=7.5 and whole section=6.0] and [median for c-Met from TMA=7.25 and whole section=7.0]

Table A-14 Hazard ratios of HGF and c-Met for the overall and progression free survival among lung cancer patients

	Age Adjusted		Minimally Adjusted[a]		Additionally Adjusted[b]		Fully Adjusted[c]	
Overall Survival								
	HR (95% CI)	***p****	**HR (95% CI)**	***p****	**HR (95% CI)**	***p****	**HR (95% CI)**	***p****
HGF (High Expression)	0.819 (0.564, 1.190)	0.295	0.830 (0.565, 1.220)	0.344	0.869 (0.586, 1.291)	0.488	0.88 (0.59, 1.31)	0.518
HGF Allred Score	1.010 (0.912, 1.119)	0.849	1.011 (0.909, 1.126)	0.835	1.015 (0.912, 1.129)	0.790	1.02 (0.92, 1.14)	0.686
c-Met (High Expression)	0.942 (0.649, 1.366)	0.752	0.937 (0.640, 1.372)	0.737	1.056 (0.707, 1.578)	0.791	1.01 (0.67, 1.52)	0.963
c-Met Allred Score	1.046 (0.915, 1.195)	0.511	1.023 (0.897, 1.167)	0.736	1.081 (0.946, 1.236)	0.254	1.08 (0.95, 1.24)	0.244
Progression Free Survival								
	Age Adjusted		**Minimally Adjusted[a]**		**Additionally Adjusted[b]**		**Fully Adjusted[c]**	
	HR (95% CI)	***p****	**HR (95% CI)**	***p****	**HR (95% CI)**	***p****	**HR (95% CI)**	***p****
HGF (High Expression)	0.792 (0.540, 1.162)	0.233	0.785 (0.527, 1.169)	0.233	0.855 (0.568, 1.287)	0.453	0.87 (0.58, 1.32)	0.523
HGF Allred Score	0.987 (0.893, 1.091)	0.799	1.001 (0.901, 1.111)	0.990	1.003 (0.903, 1.114)	0.953	1.00 (0.90, 1.12)	0.934
c-Met (High Expression)	1.068 (0.733, 0.156)	0.731	1.031 (0.697, 1.526)	0.877	1.288 (0.855, 1.941)	0.226	1.30 (0.85, 1.99)	0.22
c-Met Allred Score	1.069 (0.931, 1.226)	0.345	1.059 (0.923, 1.215)	0.417	1.126 (0.974, 1.301)	0.109	1.14 (0.98, 1.33)	0.088

**p*-value from Wald Test (Type 3 test)

Abbreviations: HR, hazard ratio; CI, confidence interval.

[a] Adjusted for age (continuous) and smoking

[b] Adjusted for age (continuous) , smoking, stage, and sex (Variables selected from Modeling)

[c] Adjusted for age (continuous) , smoking, stage, sex, race, and histology

Table A-15 Multivariate Cox Regression for the overall Survival

	Age Adjusted		Minimally Adjusted[a]		Additionally Adjusted[b]		Fully Adjusted[c]	
	HR (95% CI)	***p****	**HR (95% CI)**	***p****	**HR (95% CI)**	***p****	**HR (95% CI)**	***p****
HGF (High Expression)	0.819 (0.564, 1.190)	0.295	0.830 (0.565, 1.220)	0.344	0.869 (0.586, 1.291)	0.488	0.88 (0.59, 1.31)	0.518
HGF Allred Score	1.010 (0.912, 1.119)	0.849	1.011 (0.909, 1.126)	0.835	1.015 (0.912, 1.129)	0.790	1.02 (0.92, 1.14)	0.686
c-Met (High Expression)	0.942 (0.649, 1.366)	0.752	0.937 (0.640, 1.372)	0.737	1.056 (0.707, 1.578)	0.791	1.01 (0.67, 1.52)	0.963
c-Met Allred Score	1.046 (0.915, 1.195)	0.511	1.023 (0.897, 1.167)	0.736	1.081 (0.946, 1.236)	0.254	1.08 (0.95, 1.24)	0.244
Sex (Men)	1.317 (0.922, 1.881)	0.131	1.435 (0.995, 2.070)	0.053	1.617 (1.115, 2.345)	0.113	1.51 (1.03, 2.22)*	0.034
RACE (African-American)	1.372 (0.752, 2.504)	0.303	1.112 (0.577, 2.145)	0.751	1.245 (0.639. 2.428)	0.520	1.42 (0.72, 2.83)	0.313
Age (continous variable)	**	**	1.026 (1.006, 1.046)	0.011	1.037 (1.017, 1.056)	0.0002	1.03 (1.01, 1.05)*	0.001
AGE				0.025		0.002		0.009
30-59	**	**	1.00		1.00		1.00	
60-69	**	**	1.122 (0.665, 1.892)	0.667	1.617 (0.937, 2.792)	0.085	1.74 (1.00, 3.04)	0.052
70+	**	**	1.775 (1.095, 2.879)	0.020	2.453 (1.479, 4.067)	0.001	2.24 (1.34, 3.76)*	0.002
SMOKING STATUS		0.079				0.005		0.001
never smoker	1.00		***	***	1.00		1.00	
active smoker	1.858 (0.847, 4.077)	0.122	***	***	2.544 (1.130, 5.727)	0.024	2.60 (1.15, 5.86)*	0.021
ex-smoker	1.269 (0.572, 2.812)	0.558	***	***	1.405 (0.627, 3.151)	0.409	1.27 (0.56, 2.88)	0.56
PATHOLOGIC STAGE		<.0001		<.0001		<.0001		<.0001
IA	1.00		1.00		1.00		1.00	
IB	1.472 (0.788, 2.749)	0.225	1.413 (0.751, 0.658)	0.284	1.709 (0.893, 3.269)	0.105	1.59 (0.83, 3.06)	0.166
IIA/B	2.737 (1.462, 5.123)	0.002	2.771 (1.460, 5.258)	0.002	3.914 (1.995, 7.679)	<.0001	4.39 (2.21, 8.72)*	<.0001
III/IV	3.333 (1.863, 5.964)	<.0001	3.291 (1.831, 5.917)	<.0001	3.997 (2178, 7.335)	<.0001	4.00 (2.17, 7.36)*	<.0001
HISTOLOGY GROUP		0.554		0.532		0.329		0.290
Adeno, SQUAM, BAC, Carcinoid	1.00		1.00		1.00		1.00	
NSCLS, large cell carcinoma	1.227 (0.607, 2.483)	0.569	1.315 (0.644, 2.686)	0.452	1.701 (0.819, 3.535)	0.155	1.73 (0.83, 3.63)	0.146
SCCA	1.283 (0.864 , 1.905)	0.216	1.259 (0.836, 1.896)	0.271	1.223 (0.802, 1.867)	0.350	1.25 (0.82, 1.92)	0.300
small cell carcinoma	1.690 (0.526, 5.429)	0.378	1.885 (0.580, 6.125)	0.292	2.230 (0.655, 7.586)	0.199	2.38 (0.70, 8.16)	0.167

*p-values from Wald Test (Type 3 test) for Testing Global Null Hypothesis: beta=0 and Chi-Square Test for each stratified level based on analysis of maximum likelihood estimates

Abbreviations: HR, hazard ratio; CI, confidence interval.

[a] Adjusted for age (continuous) and smoking; [b] Adjusted for age (continuous) , smoking, stage, and sex (Variables selected from Modeling); [c] Adjusted for age (continuous) , smoking, stage, sex, race, and histology

same as univariate hazard ratio; * same as age adjusted hazard ratio

Table A-16 Multivariate Cox Regression for the progression free survival

	Age Adjusted		Minimally Adjusted[a]		Additionally Adjusted[b]		Fully Adjusted[c]	
	HR (95% CI)	***p****	**HR (95% CI)**	***p****	**HR (95% CI)**	***p****	**HR (95% CI)**	***p****
HGF (High Expression)	0.792 (0.540, 1.162)	0.233	0.785 (0.527, 1.169)	0.233	0.855 (0.568, 1.287)	0.453	0.87 (0.58, 1.32)	0.523
HGF Allred Score	0.987 (0.893, 1.091)	0.799	1.001 (0.901, 1.111)	0.990	1.003 (0.903, 1.114)	0.953	1.00 (0.90, 1.12)	0.934
c-Met (High Expression)	1.068 (0.733, 0.1556)	0.731	1.031 (0.697, 1.526)	0.877	1.288 (0.855, 1.941)	0.226	1.30 (0.85, 1.99)	0.22
c-Met Allred Score	1.069 (0.931, 1.226)	0.345	1.059 (0.923, 1.215)	0.417	1.126 (0.974, 1.301)	0.109	1.14 (0.98, 1.33)	0.088
Sex (Men)	1.386 (0.963, 1.994)	0.079	1.483 (1.019, 2.157)	0.040	1.606 (1.098, 2.348)	0.015	1.59 (1.07, 2.36)*	0.021
RACE (African-American)	1.279 (0.702, 2.329)	0.421	1.101 (0.572, 2.119)	0.773	1.254 (0.644, 2.441)	0.506	1.38 (0.70, 2.73)	0.358
Age (continous variable)	**	**	1.011 (0.992, 1.030)	0.254	1.019 (1.001, 1.038)	0.040	1.02 (1.00, 1.04)	0.068
AGE				0.417		0.134		0.232
30-59	**	**	1.00		1.00		1.00	
60-69	**	**	1.045 (0.617, 1.772)	0.869	1.380 (0.799, 2.385)	0.249	1.37 (0.77, 2.43)	0.278
70+	**	**	1.316 (0.811, 2.135)	0.266	1.660 (1.009, 2.732)	0.046	1.56 (0.94, 2.59)	0.087
SMOKING STATUS		0.458				0.101		0.053
never smoker	1.00		***	***	1.00		1.00	
active smoker	1.275 (0.549, 2.964)	0.572	***	***	1.499 (0.619, 3.630)	0.370	1.45 (0.59, 3.54)	0.413
ex-smoker	1.005 (0.429, 2.355)	0.992	***	***	0.963 (0.403, 2.299)	0.932	0.87 (0.36, 2.10)	0.749
PATHOLOGIC STAGE		0.0003		0.0002		<.0001		<.0001
IA	1.00		1.00		1.00		1.00	
IB	1.332 (0.712, 2.491)	0.370	1.258 (0.667, 2.374)	0.479	1.395 (0.724, 2.687)	0.320	1.28 (0.66, 2.48)	0.463
IIA/B	2.594 (1.389, 4.846)	0.003	2.716 (1.435, 5.139)	0.002	3.393 (1.734, 6.639)	0.0004	3.71 (1.88, 7.35)*	0.0002
III/IV	2.880 (1.613, 5.145)	0.0004	2.918 (1.620, 5.258)	0.0004	3.072 (1.698, 5.558)	0.0002	3.05 (1.68, 5.53)*	0.0002
HISTOLOGY GROUP		0.797		0.717		0.294		0.280
Adeno, SQUAM, BAC, Carcinoid	1.00		1.00		1.00		1.00	
NSCLS, large cell carcinoma	1.331 (0.658, 2.692)	0.426	1.422 (0.696, 2.905)	0.335	1.903 (0.921, 3.934)	0.082	1.94 (0.93, 4.05)	0.077
SCCA	1.105 (0.742, 1.647)	0.623	1.070 (0.706, 1.622)	0.749	1.031 (0.670, 1.586)	0.890	1.06 (0.69, 1.64)	0.796
small cell carcinoma	1.458 (0.455, 4.674)	0.526	1.564 (0.483, 5.064)	0.456	1.793 (0.532, 6.043)	0.346	1.87 (0.55, 6.35)	0.313

*p-values from Wald Test (Type 3 test) for Testing Global Null Hypothesis: beta=0 and Chi-Square Test for each stratified level based on analysis of maximum likelihood estimates

Abbreviations: HR, hazard ratio; CI, confidence interval.

[a] Adjusted for age (continuous) and smoking; [b] Adjusted for age (continuous) , smoking, stage, and sex (Variables selected from Modeling); [c] Adjusted for age (continuous) , smoking, stage, sex, race, and histology

same as univariate hazard ratio; * same as age adjusted hazard ratio

Table A-17 Hazard ratios of HGF and c-Met by sex for the Overall Survival among Lung cancer patients

	Crude					
	All Subjects		Women		Men	
	HR (95% CI)	p-value*	HR (95% CI)	p-value*	HR (95% CI)	p-value*
HGF (High Expression)	0.830 (0.571, 1.205)	0.328	0.842 (0.500, 1.417)	0.518	0.821 (0.476, 1.413)	0.476
HGF Allred Score	1.018 (0.920, 1.127)	0.725	1.037 (0.902, 1.193)	0.609	0.987 (0.850, 1.146)	0.864
c-Met (High Expression)	0.959 (0.662, 1.388)	0.823	0.812 (0.483, 1.365)	0.432	1.193 (0.703, 2.026)	0.513
c-Met Allred Score	1.047 (0.918, 1.193)	0.497	1.036 (0.874, 1.230)	0.682	1.073 (0.866, 1.328)	0.520
	Age Adjusted					
	All Subjects		Women		Men	
	HR (95% CI)	p-value*	HR (95% CI)	p-value*	HR (95% CI)	p-value*
HGF (High Expression)	0.819 (0.564, 1.190)	0.295	0.820 (0.487, 1.380)	0.455	0.829 (0.481, 1.429)	0.499
HGF Allred Score	1.010 (0.912, 1.119)	0.849	1.025 (0.888, 1.182)	0.737	0.983 (0.848, 1.141)	0.825
c-Met (High Expression)	0.942 (0.649, 1.366)	0.752	0.800 (0.474, 1.349)	0.402	1.151 (0.676, 1.961)	0.604
c-Met Allred Score	1.046 (0.915, 1.195)	0.511	1.032 (0.866, 1.229)	0.727	1.074 (0.870, 1.326)	0.505
	Multivariable Adjusted[1]					
	All Subjects		Women		Men	
	HR (95% CI)	p-value*	HR (95% CI)	p-value*	HR (95% CI)	p-value*
HGF (High expression)	0.83 (0.56, 1.23)	0.36	0.91 (0.53, 1.56)	0.73	0.69 (0.37, 1.30)	0.25
HGF Allred Score	1.01 (0.91, 1.12)	0.86	1.02 (0.87, 1.20)	0.79	1.02 (0.88, 1.18)	0.77
c-Met (High expression)	1.02 (0.68, 1.51)	0.93	0.91 (0.53, 1.58)	0.74	1.26 (0.70, 2.29)	0.44
c-Met Allred Score	1.08 (0.95, 1.23)	0.26	1.07 (0.90, 1.28)	0.46	1.10 (0.90, 1.33)	0.35

[1] Cox proportional hazards models adjusted for age at tissue collection, smoking, and stage

High HGF and c-Met expression defined by subject-specific averaged Allred values above IHC source-specific Allred median cutpoints (HGF cutpoints: 7.5 for TMA and 6.0 for whole section; c-Met cutpoints: 7.25 for TMA and 7.0 for whole section)

Table A-18 Hazard ratios of HGF and c-Met by sex for the progression free survival among Lung cancer patients

	Crude					
	All Subjects		Women		Men	
	HR (95% CI)	p-value*	HR (95% CI)	p-value*	HR (95% CI)	p-value*
HGF (High Expression)	0.794 (0.541, 1.164)	0.237	0.843 (0.488, 1.458)	0.541	0.772 (0.448, 1.329)	0.350
HGF Allred Score	0.991 (0.898, 1.095)	0.861	1.017 (0.884, 1.169)	0.817	0.964 (0.834, 1.115)	0.625
c-Met (High Expression)	1.079 (0.742, 1.571)	0.691	0.969 (0.567, 1.655)	0.907	1.262 (0.744, 2.138)	0.388
c-Met Allred Score	1.068 (0.932, 1.224)	0.342	1.078 (0.895, 1.298)	0.427	1.070 (0.867, 1.320)	0.527
	Age Adjusted					
	All Subjects		Women		Men	
	HR (95% CI)	p-value*	HR (95% CI)	p-value*	HR (95% CI)	p-value*
HGF (High Expression)	0.792 (0.540, 1.162)	0.233	0.843 (0.487, 1.457)	0.540	0.772 (0.448, 1.331)	0.353
HGF Allred Score	0.987 (0.893, 1.091)	0.799	1.014 (0.880, 1.168)	0.845	0.962 (0.833, 1.111)	0.600
c-Met (High Expression)	1.068 (0.733, 0.1556)	0.731	0.964 (0.564, 1.649)	0.964	1.233 (0.726, 2.094)	0.437
c-Met Allred Score	1.069 (0.931, 1.226)	0.345	1.077 (0.894, 1.298)	0.436	1.075 (0.873, 1.323)	0.497
	Multivariable Adjusted[1]					
	All Subjects		Women		Men	
	HR (95% CI)	p-value*	HR (95% CI)	p-value*	HR (95% CI)	p-value*
HGF (High Expression)	0.834 (0.554, 1.257)	0.387	1.044 (0.567, 1.923)	0.889	0.687 (0.371, 1.271)	0.231
HGF Allred Score	1.002 (0.901, 1.113)	0.975	1.047 (0.891, 1.230)	0.577	1.027 (0.891, 1.184)	0.711
c-Met (High Expression)	1.218 (0.813, 1.825)	0.339	1.131 (0.632, 2.025)	0.678	1.599 (0.877, 2.915)	0.125
c-Met Allred Score	1.124 (0.975, 1.295)	0.106	1.141 (0.940, 1.386)	0.182	1.153 (0.935, 1.423)	0.183

[1] Cox proportional hazards models adjusted for age at tissue collection, smoking, and stage

High HGF and c-Met expression defined by subject-specific averaged Allred values above IHC source-specific Allred median cutpoints (HGF cutpoints: 7.5 for TMA and 6.0 for whole section; c-Met cutpoints: 7.25 for TMA and 7.0 for whole section)

Table A-19 Cox proportional hazards model for overall survival of lung cancer patients: Three models with HGF and c-Met treated as continuous variables

HGF and cMet (continous)

	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	HR	(95% CI)	
HGF Allred Score	1	-0.01268	0.06569	0.0372	0.847	0.987	0.868	1.123
cMet Allred Score	1	0.05205	0.08391	0.3848	0.535	1.053	0.894	1.242
Sex (Men)	1	0.53174	0.19756	7.2441	0.0071	1.702	1.155	2.507
Age (continous variable)	1	0.03792	0.00982	14.9035	0.0001	1.039	1.019	1.059
SMOKING STATUS (reference=never smoker)								
active smoker	1	0.93131	0.44152	4.4493	0.0349	2.538	1.068	6.029
ex-smoker	1	0.31123	0.44155	0.4968	0.4809	1.365	0.575	3.244
PATHOLOGIC STAGE (reference=IA)								
IB	1	0.41138	0.34073	1.4577	0.2273	1.509	0.774	2.942
IIA/B	1	1.24196	0.35635	12.147	0.0005	3.462	1.722	6.961
III/IV	1	1.31106	0.31764	17.0361	<.0001	3.71	1.991	6.915

Type 3 Tests

	DF	Wald Chi-Square
HGF_Allred	1	0.0372
cMet_Allred	1	0.3848
Sex	1	7.2441
Age_at_tissue_collection	1	14.9035
Smoking	2	10.598
stage_grp	3	24.6017

HGF (continous)

	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	HR	(95% CI)	
HGF Allred Score	1	0.0145	0.05454	0.0706	0.7904	1.015	0.912	1.129
Sex (Men)	1	0.51424	0.19502	6.9534	0.0084	1.672	1.141	2.451
Age (continous variable)	1	0.03872	0.00984	15.4844	<.0001	1.039	1.02	1.06
SMOKING STATUS (reference=never smoker)								
active smoker	1	0.99775	0.4158	5.758	0.0164	2.712	1.201	6.127
ex-smoker	1	0.37268	0.41679	0.7995	0.3712	1.452	0.641	3.286
PATHOLOGIC STAGE (reference=IA)								
IB	1	0.40235	0.33699	1.4255	0.2325	1.495	0.772	2.895
IIA/B	1	1.20307	0.34644	12.0597	0.0005	3.33	1.689	6.567
III/IV	1	1.31367	0.31598	17.2846	<.0001	3.72	2.002	6.91

Type 3 Tests

	DF	Wald Chi-Square
HGF_Allred	1	0.0706
Sex	1	6.9534
Age_at_tissue_collection	1	15.4844
Smoking	2	11.544
stage_grp	3	24.7532

cMet (continous)

	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	HR	(95% CI)	
cMet Allred Score	1	0.07787	0.06827	1.3011	0.254	1.081	0.946	1.236
Sex (Men)	1	0.54383	0.19586	7.7097	0.0055	1.723	1.173	2.529
Age (continous variable)	1	0.03706	0.00973	14.509	0.0001	1.038	1.018	1.058
SMOKING STATUS (reference=never smoker)								
active smoker	1	0.92859	0.44137	4.4263	0.0354	2.531	1.066	6.011
ex-smoker	1	0.30448	0.44101	0.4767	0.4899	1.356	0.571	3.218
PATHOLOGIC STAGE (reference=IA)								
IB	1	0.42077	0.33934	1.5375	0.215	1.523	0.783	2.962
IIA/B	1	1.27317	0.3472	13.447	0.0002	3.572	1.809	7.054
III/IV	1	1.29181	0.31686	16.6216	<.0001	3.639	1.956	6.772

Type 3 Tests

	DF	Wald Chi-Square
cMet_Allred	1	1.3011
Sex	1	7.7097
Age_at_tissue_collection	1	14.509
Smoking	2	10.8089
stage_grp	3	24.5159

Table A-20 Cox proportional hazards model for overall survival of lung cancer patients: Three models with HGF and c-Met treated as categorical variables

HGF and cMet (categorical)

	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	HR	(95% CI)	
HGF (High expression)	1	-0.15763	0.20485	0.5921	0.4416	0.85	0.57	1.276
cMet (High expression)	1	0.05551	0.21052	0.0695	0.792	1.06	0.7	1.597
Sex (Men)	1	0.51372	0.19833	6.7093	0.0096	1.67	1.13	2.466
Age (continous variable)	1	0.03756	0.00995	14.2631	0.0002	1.04	1.02	1.059
SMOKING STATUS (reference=never smoker)								
active smoker	1	0.96537	0.44448	4.7172	0.0299	2.63	1.1	6.275
ex-smoker	1	0.34516	0.44417	0.6039	0.4371	1.41	0.59	3.373
PATHOLOGIC STAGE (reference=IA)								
IB	1	0.38861	0.34489	1.2696	0.2598	1.48	0.75	2.9
IIA/B	1	1.22631	0.35676	11.8154	0.0006	3.41	1.69	6.859
III/IV	1	1.26074	0.32073	15.4519	<.0001	3.53	1.88	6.615

Type 3 Tests

	DF	Wald Chi-Square	Pr > ChiSq
HGF_Allred	1	0.5921	0.4416
cMet_Allred	1	0.0695	0.792
Sex	1	6.7093	0.0096
Age_at_tissue_collection	1	14.2631	0.0002
Smoking	2	10.783	0.0046
stage_grp	3	23.3121	<.0001

HGF (categorical)

	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	HR	(95% CI)	
HGF (High expression)	1	-0.13994	0.20164	0.4816	0.4877	0.87	0.59	1.291
Sex (Men)	1	0.5011	0.19578	6.551	0.0105	1.65	1.13	2.423
Age (continous variable)	1	0.03905	0.00978	15.9363	<.0001	1.04	1.02	1.06
SMOKING STATUS (reference=never smoker)								
active smoker	1	1.04533	0.42048	6.1804	0.0129	2.84	1.25	6.485
ex-smoker	1	0.42354	0.41977	1.0181	0.313	1.53	0.67	3.477
PATHOLOGIC STAGE (reference=IA)								
IB	1	0.39314	0.33765	1.3557	0.2443	1.48	0.76	2.872
IIA/B	1	1.21175	0.34599	12.2656	0.0005	3.36	1.71	6.619
III/IV	1	1.2781	0.31874	16.079	<.0001	3.59	1.92	6.705

Type 3 Tests

	DF	Wald Chi-Square	Pr > ChiSq
HGF_Allred	1	0.4816	0.4877
Sex	1	6.551	0.0105
Age_at_tissue_collection	1	15.9363	<.0001
Smoking	2	11.7748	0.0028
stage_grp	3	23.7339	<.0001

cMet (categorical)

	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	HR	(95% CI)	
cMet (High expression)	1	0.05437	0.20486	0.0704	0.7907	1.06	0.71	1.578
Sex (Men)	1	0.54605	0.19628	7.7392	0.0054	1.73	1.18	2.536
Age (continous variable)	1	0.03746	0.00993	14.2261	0.0002	1.04	1.02	1.059
SMOKING STATUS (reference=never smoker)								
active smoker	1	0.92778	0.44165	4.413	0.0357	2.53	1.06	6.01
ex-smoker	1	0.28777	0.44064	0.4265	0.5137	1.33	0.56	3.163
PATHOLOGIC STAGE (reference=IA)								
IB	1	0.39455	0.34423	1.3138	0.2517	1.48	0.76	2.913
IIA/B	1	1.24204	0.35341	12.3516	0.0004	3.46	1.73	6.922
III/IV	1	1.24453	0.31797	15.3191	<.0001	3.47	1.86	6.474

Type 3 Tests

	DF	Wald Chi-Square	Pr > ChiSq
cMet_Allred	1	0.0704	0.7907
Sex	1	7.7392	0.0054
Age_at_tissue_collection	1	14.2261	0.0002
Smoking	2	11.1886	0.0037
stage_grp	3	23.2162	<.0001

Table A-21 Cox proportional hazards model for progression free survival of lung cancer patients: Three models with HGF and c-Met treated as continuous variables

HGF and cMet (continous)

	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	HR	(95% CI)	
HGF Allred Score	1	-0.04543	0.06596	0.4744	0.491	0.956	0.84	1.087
cMet Allred Score	1	0.12007	0.08835	1.8469	0.1741	1.128	0.948	1.341
Sex (Men)	1	0.48951	0.20237	5.8512	0.0156	1.632	1.097	2.426
Age (continous variable)	1	0.02224	0.00945	5.5445	0.0185	1.022	1.004	1.042
SMOKING STATUS (reference=never smoker)								
active smoker	1	0.53617	0.49016	1.1965	0.274	1.709	0.654	4.468
ex-smoker	1	0.09668	0.48692	0.0394	0.8426	1.102	0.424	2.861
PATHOLOGIC STAGE (reference=IA)								
IB	1	0.33809	0.34713	0.9486	0.3301	1.402	0.71	2.769
IIA/B	1	1.21277	0.36349	11.1319	0.0008	3.363	1.649	6.857
III/IV	1	1.16492	0.31794	13.4243	0.0002	3.206	1.719	5.978

Type 3 Tests

	DF	Wald Chi-Square	Pr > ChiSq
HGF_Allred	1	0.4744	0.491
cMet_Allred	1	1.8469	0.1741
Sex	1	5.8512	0.0156
Age_at_tissue_collection	1	5.5445	0.0185
Smoking	2	4.5544	0.1026
stage_grp	3	20.9011	0.0001

HGF (continous)

	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	HR	(95% CI)	
HGF Allred Score	1	0.00315	0.05363	0.0035	0.9531	1.003	0.903	1.114
Sex (Men)	1	0.47872	0.20062	5.694	0.017	1.614	1.089	2.392
Age (continous variable)	1	0.02123	0.0095	4.9935	0.0254	1.021	1.003	1.041
SMOKING STATUS (reference=never smoker)								
active smoker	1	0.41399	0.45333	0.834	0.3611	1.513	0.622	3.679
ex-smoker	1	-0.0292	0.45057	0.0042	0.9483	0.971	0.402	2.349
PATHOLOGIC STAGE (reference=IA)								
IB	1	0.25761	0.34308	0.5638	0.4527	1.294	0.66	2.535
IIA/B	1	1.10337	0.34781	10.0634	0.0015	3.014	1.524	5.96
III/IV	1	1.08798	0.31322	12.0655	0.0005	2.968	1.607	5.484

Type 3 Tests

	DF	Wald Chi-Square	Pr > ChiSq
HGF_Allred	1	0.0035	0.9531
Sex	1	5.694	0.017
Age_at_tissue_collection	1	4.9935	0.0254
Smoking	2	4.3095	0.1159
stage_grp	3	19.9498	0.0002

cMet (continous)

	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	HR	(95% CI)	
cMet Allred Score	1	0.1183	0.07374	2.5734	0.1087	1.126	0.974	1.301
Sex (Men)	1	0.49086	0.20012	6.0166	0.0142	1.634	1.104	2.418
Age (continous variable)	1	0.02096	0.00938	4.9935	0.0254	1.021	1.003	1.04
SMOKING STATUS (reference=never smoker)								
active smoker	1	0.53639	0.48951	1.2007	0.2732	1.71	0.655	4.463
ex-smoker	1	0.07805	0.48505	0.0259	0.8722	1.081	0.418	2.797
PATHOLOGIC STAGE (reference=IA)								
IB	1	0.33682	0.34514	0.9524	0.3291	1.4	0.712	2.755
IIA/B	1	1.21559	0.35153	11.958	0.0005	3.372	1.693	6.717
III/IV	1	1.14027	0.31576	13.0406	0.0003	3.128	1.684	5.807

Type 3 Tests

	DF	Wald Chi-Square	Pr > ChiSq
cMet_Allred	1	2.5734	0.1087
Sex	1	6.0166	0.0142
Age_at_tissue_collection	1	4.9935	0.0254
Smoking	2	4.9824	0.0828
stage_grp	3	21.0006	0.0001

Table A-22 Cox proportional hazards model for progression free survival of lung cancer patients: Three models with HGF and c-Met treated as categorical variables

HGF and cMet (categorical)

	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	HR	(95% CI)	
HGF (High expression)	1	-0.21561	0.2146	1.0095	0.315	0.806	0.529	1.228
cMet (High expression)	1	0.274	0.21823	1.5764	0.2093	1.315	0.858	2.017
Sex (Men)	1	0.48644	0.20387	5.6932	0.017	1.627	1.091	2.425
Age (continous variable)	1	0.01972	0.00951	4.2969	0.0382	1.02	1.001	1.039
SMOKING STATUS (reference=never smoker)								
active smoker	1	0.59097	0.49905	1.4023	0.2363	1.806	0.679	4.802
ex-smoker	1	0.15818	0.49648	0.1015	0.75	1.171	0.443	3.1
PATHOLOGIC STAGE (reference=IA)								
IB	1	0.34521	0.34874	0.9798	0.3222	1.412	0.713	2.798
IIA/B	1	1.22485	0.36421	11.3098	0.0008	3.404	1.667	6.95
III/IV	1	1.113	0.31832	12.2257	0.0005	3.043	1.631	5.68

Type 3 Tests

	DF	Wald Chi-Square	Pr > ChiSq
HGF_Allred	1	1.0095	0.315
cMet_Allred	1	1.5764	0.2093
Sex	1	5.6932	0.017
Age_at_tissue_collection	1	4.2969	0.0382
Smoking	2	4.5924	0.1006
stage_grp	3	20.0557	0.0002

HGF (categorical)

	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	HR	(95% CI)	
HGF (High expression)	1	-0.15678	0.20869	0.5644	0.4525	0.855	0.568	1.287
Sex (Men)	1	0.47213	0.20121	5.5055	0.019	1.603	1.081	2.379
Age (continous variable)	1	0.02125	0.00944	5.07	0.0243	1.021	1.003	1.041
SMOKING STATUS (reference=never smoker)								
active smoker	1	0.49228	0.46366	1.1273	0.2884	1.636	0.659	4.059
ex-smoker	1	0.05851	0.46049	0.0161	0.8989	1.06	0.43	2.614
PATHOLOGIC STAGE (reference=IA)								
IB	1	0.25156	0.34349	0.5364	0.4639	1.286	0.656	2.521
IIA/B	1	1.1145	0.34798	10.2576	0.0014	3.048	1.541	6.029
III/IV	1	1.05612	0.31561	11.1976	0.0008	2.875	1.549	5.337

Type 3 Tests

	DF	Wald Chi-Square	Pr > ChiSq
HGF_Allred	1	0.5644	0.4525
Sex	1	5.5055	0.019
Age_at_tissue_collection	1	5.07	0.0243
Smoking	2	4.3423	0.114
stage_grp	3	19.369	0.0002

cMet (categorical)

	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	HR	(95% CI)	
cMet (High expression)	1	0.25303	0.20915	1.4636	0.2264	1.288	0.855	1.941
Sex (Men)	1	0.51625	0.20108	6.5916	0.0102	1.676	1.13	2.485
Age (continous variable)	1	0.01997	0.00952	4.3978	0.036	1.02	1.001	1.039
SMOKING STATUS (reference=never smoker)								
active smoker	1	0.51683	0.49023	1.1115	0.2918	1.677	0.641	4.383
ex-smoker	1	0.04623	0.48587	0.0091	0.9242	1.047	0.404	2.714
PATHOLOGIC STAGE (reference=IA)								
IB	1	0.34982	0.34776	1.0119	0.3144	1.419	0.718	2.805
IIA/B	1	1.23242	0.3584	11.8247	0.0006	3.43	1.699	6.923
III/IV	1	1.10628	0.31567	12.282	0.0005	3.023	1.628	5.612

Type 3 Tests

	DF	Wald Chi-Square	Pr > ChiSq
cMet_Allred	1	1.4636	0.2264
Sex	1	6.5916	0.0102
Age_at_tissue_collection	1	4.3978	0.036
Smoking	2	5.2188	0.0736
stage_grp	3	20.1245	0.0002

A.5 PROC GLIMMIX EXAMPLE IN SAS 9.2

To identify baseline factors related to HGF and c-Met expression, we used a generalized linear mixed model approach, which controlled for data source (TMA vs. whole section) and accounted for the correlated nature of the TMA core-level data.

SAS editor example using Proc Glimmix

```
/*********************Fully adjusted odds ratios*****************/
/****Adjusted by age, smoking, stage,  sex, race and histology****/
/************************************************************/
%let data=all_corr_median;
%let insert=HGF_allred; /*Insert "HGF_allred" or "cMet_allred"*/

ods rtf file= 'I:\HGFcMet_Analysis\HGFcMet_SAS_Code\OR_corr_final.rtf'
style=journal;
Ods graphics on;

Proc glimmix data=&data plots=oddsratio method=RSPL empirical;
Class subjectID_final datasource sex stage smoking race histology;
Model  &insert(event="Positive")=  sex race age smoking  stage histology/
dist=binary link=logit solution oddsratio;
Random _residual_/subject=subjectID_final(datasource) type=AR(1) ;
format  histology histo_group. stage stg_grp. smoking smoking. HGF_allred
HGF. cMet_allred cMet. ;
run;

Ods graphics off;
ods rtf close;
```

SAS output from Proc Glimmix

The SAS System

The GLIMMIX Procedure

Model Information	
Data Set	WORK.ALL_CORR_MEDIAN
Response Variable	HGF_allred
Response Distribution	Binary
Link Function	Logit
Variance Function	Default
Variance Matrix Blocked By	SubjectID(DataSourc)
Estimation Technique	Residual PL
Degrees of Freedom Method	Between-Within
Fixed Effects SE Adjustment	Sandwich - Classical

Class Level Information

Class	Levels	Values
SubjectID_Final	156	1105 1193 1237 1243 1253 1263 1265 1266 1267 1269 1270 1274 1278 1289 1292 1329 1335 1336 1347 1362 1368 1378 1395 1402 1428 1449 1451 1472 1474 1479 1481 1509 1514 1516 1518 1521 1529 1531 1543 1544 1545 1547 1548 1555 1558 1561 1563 1573 1603 1604 1611 1624 1626 1630 1633 1637 1641 1651 1661 1662 1678 1680 1681 1687 1690 1705 1716 1722 1724 1725 1747 1748 1757 176 1954 1958 1959 1960 246 253 307 356 365 430 445 448 455 488 515 53-90 544 546 553 554 557 560 571 575 577 593 597 602 604 605 620 629 632 646 649 660 671 680 727 731 749 751 760 762 784 788 796 803 806 810 814 817 822 823 824 842 853 861 882 890 898 900 902 906 917 920 923 932 933 951 L-006 L-007 L-009 L-010 L-011 L-019 L-022 L-025 L-029 L-032 L-034 V-100
DataSource	2	0 1
Sex	2	1 2
Stage	4	1 2 3 4
Smoking	3	active smoker ex-smoker never smoker
Race	2	AA White
Histology	4	1 2 3 4

Number of Observations Read	619
Number of Observations Used	522

Response Profile

Ordered Value	HGF_allred	Total Frequency
1	Negative	218
2	Positive	304

The GLIMMIX procedure is modeling the probability that HGF_allred='Positive'.

The SAS System

The GLIMMIX Procedure

Dimensions	
R-side Cov. Parameters	2
Columns in X	17
Columns in Z per Subject	0
Subjects (Blocks in V)	162
Max Obs per Subject	10

Optimization Information	
Optimization Technique	Newton-Raphson with Ridging
Parameters in Optimization	1
Lower Boundaries	1
Upper Boundaries	1
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	0	2	2189.1685635	0.20898326	0.000051
1	0	2	2161.0997963	0.09401794	1.551E-7
2	0	1	2161.8366787	0.00104497	1.419E-6
3	0	1	2161.8365996	0.00001272	1.85E-10
4	0	0	2161.8365918	0.00000000	3.704E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	2161.84
Generalized Chi-Square	534.94
Gener. Chi-Square / DF	1.05

The SAS System

The GLIMMIX Procedure

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
AR(1)	SubjectID(DataSourc)	0.5391	0.04189
Residual		1.0489	0.08072

Solutions for Fixed Effects

Effect	Sex	Race	Stage	Smoking	Histology	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept						-0.1175	1.3919	150	-0.08	0.9329
Sex	1					-0.4810	0.2952	150	-1.63	0.1053
Sex	2					0	.	.	.	.
Race		AA				-0.1062	0.5281	150	-0.20	0.8410
Race		White				0	.	.	.	.
Age						-0.00057	0.01819	150	-0.03	0.9751
Smoking				active smoker		0.9730	0.6331	150	1.54	0.1264
Smoking				ex-smoker		1.2458	0.6326	150	1.97	0.0508
Smoking				never smoker		0	.	.	.	.
Stage			1			-0.2941	0.4371	150	-0.67	0.5021
Stage			2			0.4160	0.5124	150	0.81	0.4182
Stage			3			-0.7738	0.4221	150	-1.83	0.0687
Stage			4			0	.	.	.	.
Histology					1	-0.3241	0.4517	150	-0.72	0.4741
Histology					2	-0.2409	0.3239	150	-0.74	0.4581
Histology					3	0.2512	0.8636	150	0.29	0.7715
Histology					4	0	.	.	.	.

Odds Ratio Estimates

Sex	Stage	Smoking	Race	Histology	Age	Sex	Stage	_Smoking	Race	Histology	_Age	Estimate	DF
1					66.929	2					66.929	0.618	150
			AA		66.929				White		66.929	0.899	150
					67.929						66.929	0.999	150
		active smoker			66.929			never smoker			66.929	2.646	150
		ex-smoker			66.929			never smoker			66.929	3.476	150
	1				66.929		4				66.929	0.745	150
	2				66.929		4				66.929	1.516	150
	3				66.929		4				66.929	0.461	150

The SAS System

The GLIMMIX Procedure

Odds Ratio Estimates

Sex	Stage	Smoking	Race	Histology	Age	Sex	Stage	_Smoking	Race	Histology	_Age	Estimate	DF
				1	66.929					4	66.929	0.723	150
				2	66.929					4	66.929	0.786	150
				3	66.929					4	66.929	1.286	150

Effects of continuous variables are assessed as one unit offsets from the mean. The AT suboption modifies the reference value and the UNIT suboption modifies the offsets.

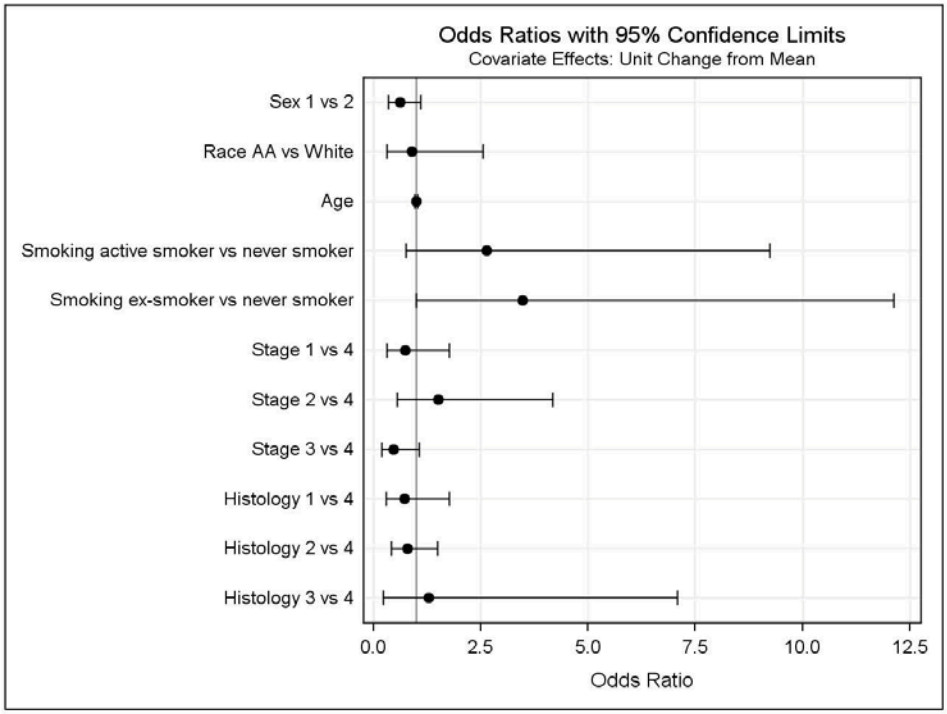
Odds Ratio Estimates

Sex	Stage	Smoking	Race	Histology	Age	Sex	Stage	_Smoking	Race	Histology	_Age	95% Confidence Limits	
1					66.929	2					66.929	0.345	1.108
			AA		66.929				White		66.929	0.317	2.553
					67.929						66.929	0.964	1.036
		active smoker			66.929			never smoker			66.929	0.757	9.244
		ex-smoker			66.929			never smoker			66.929	0.996	12.130
	1				66.929		4				66.929	0.314	1.768
	2				66.929		4				66.929	0.551	4.172
	3				66.929		4				66.929	0.200	1.062
				1	66.929					4	66.929	0.296	1.765
				2	66.929					4	66.929	0.414	1.490
				3	66.929					4	66.929	0.233	7.083

Effects of continuous variables are assessed as one unit offsets from the mean. The AT suboption modifies the reference value and the UNIT suboption modifies the offsets.

The SAS System

The GLIMMIX Procedure



Type III Tests of Fixed Effects

Effect	*Num DF*	*Den DF*	*F Value*	*Pr > F*
Sex	1	150	2.66	0.1053
Race	1	150	0.04	0.8410
Age	1	150	0.00	0.9751
Smoking	2	150	2.01	0.1369
Stage	3	150	2.60	0.0546
Histology	3	150	0.35	0.7876

APPENDIX B

SAMPLE SIZE AND POWER CALCULATION FOR PROJECT#3

The sample size calculation is based on the first hypothesis of specific aim #3 from Project #3: the prevalence of polymorphisms in *ESR2* gene is not different between lung cancer patients with and without ER beta protein expression in lung tumor tissue. Sample size calculation was performed with significance level of alpha=0.05 (two-sided), 80% power (beta=0.20), and various minor allele frequencies of *ESR2* SNPs selected in the project #3. The sample size calculation was performed for both the recessive and dominant models by treating ER-beta protein expression as categorical variable.

Based on previous analysis, we had 60% ER-beta protein expression among study samples. Therefore, assuming 120 samples with ER-beta protein expression and 80 samples without the expression, odds ratios of ER-beta protein expression associated with various SNPs for *ESR2* were obtained with continuity correction for both models (Table B-1). For recessive model of *ESR2* SNP with minor allele frequency of 0.35 in the Caucasian population, it is calculated that we will have 12.3% minor genotype prevalence among subjects with ER-beta protein expression and 29.4% among subjects without ER-beta protein expression to show an odds ratio of 2.97 for the lung cancer patients with a 80% power at 5% significance. For

dominant model of *ESR2* SNP with minor allele frequency of 0.35 in the Caucasian population, it is calculated that we will have 57.8% minor genotype prevalence among subjects with ER-beta protein expression and 77.7% among subjects without ER-beta protein expression to show an odds ratio of 2.54 for the lung cancer patients with a 80% power at 5% significance. In both model, odds ratios are calculated to be greater than 2.5 showing that subjects without ER-beta protein expression are much more likely to have the minor allele of *ESR2* SNP than those with ER-beta protein expression.

Assuming dominant model of inheritance, we can detect the odds ratios less than 3.5 with 80% power at 5% significant for the polymorphisms with extreme minor allele frequency as 0.05 (lower) and 0.5 (higher). However, in recessive model, only extreme odds ratios such as 35.86 and 12.64 can be detected with 80% power at 5% significance for polymorphisms with lower minor allele frequencies: 0.05 and 0.10, respectively. The power analysis software, Power Analysis and Sample Size (PASS)[1], were used to perform the sample size calculation. This may provide less power for other hypotheses testing including stratifications by gender, histological types of lung cancer, and smoking history.

[1] Pass 2000 (January 21, 2005): Hintze J. (2004). NCSS and Pass. Number Cruncher Statistical Systems. Kaysville, Utah. www.ncss.com

Table B-1 Sample Size and Power Calculation

	Genotype Frequency			Recessive model			Dominant model		
Allele Freq	**AA**	**AB**	**BB**	**P0**	**P1**	**OR**	**P0**	**P1**	**OR**
0.050	0.003	0.095	0.903	0.003	0.097	35.86	0.098	0.259	3.22
0.100	0.010	0.180	0.810	0.010	0.113	12.64	0.190	0.379	2.61
0.150	0.023	0.255	0.723	0.023	0.140	6.89	0.278	0.481	2.41
0.200	0.040	0.320	0.640	0.040	0.170	4.93	0.360	0.570	2.35
0.250	0.063	0.375	0.563	0.063	0.207	3.89	0.438	0.648	2.36
0.300	0.090	0.420	0.490	0.090	0.248	3.34	0.510	0.716	2.42
0.350	0.123	0.455	0.423	0.123	0.294	2.97	0.578	0.777	2.54
0.400	0.160	0.480	0.360	0.160	0.342	2.73	0.640	0.828	2.72
0.450	0.203	0.495	0.303	0.203	0.395	2.57	0.698	0.874	3.00
0.500	0.250	0.500	0.250	0.250	0.451	2.46	0.750	0.912	3.46

N0 (ER-beta positive)=120, N1 (ER-beta negative)=80, alpha=0.05 (two-sided, difference between two proportions with continuity correction), beta=0.20, allele A=minor allele, OR=[P1/(1-P1)]/[P0/(1-P0)]

APPENDIX C

SNP SELECTION METHDOLOGY FOR PROJECT#3

C.1 SNP SELECTION METHODOLOGY

Candidate *ESR2* single nucleotide polymorphisms (SNPs)

Five data sources provided information about genetic variation in the human *ESR2* gene, 1) OVID Medline®, 2) NCBI Entrez SNP[13], 3) the Cancer Genome Anatomy Project (CGAP) SNP500Cancer Database[14] [1], 4) the International HapMap Project[15], and 5) FastSNP[16] [2].

1. OVID Medline®

An OVID Medline literature search (conducted on 01/28/2009) for articles indexed under the keywords "Estrogen Receptor beta" and ("Polymorphism, Genetic" or "Polymorphism, Restriction Fragment Length" or "Polymorphism, Single Nucleotide" or "Polymorphism, Single-Stranded Conformational") produced 119 citations published between 1998 and 2009.

[13] http://www.ncbi.nlm.nih.gov/sites/entrez
[14] http://snp500cancer.nci.nih.gov/home_1.cfm
[15] http://www.hapmap.org/
[16] http://fastsnp.ibms.sinica.edu.tw/pages/input_CandidateGeneSearch.jsp

Three *ESR2* variants of scientific interest have included a silent G1082A SNP in exon 6 (ligand binding domain), A1730G SNP in the 3′-untranslated region of exon 8, and a CA dinucleotide repeat polymorphism in intron 5 [3] (Table C-1). The inheritance of one or another of these three specific *ESR2* genetic variants has been studied in relation to cancers of the colon or rectum [4], endometrium [5], ovary [6], testis [7], prostate [8-10], and breast [11-19].

Table C-1 Three frequently studied *ESR2* genetic variants.

Identifier	Restriction site	Description	MAF[1]
rs1256049	RsaI	Silent G1082A SNP in exon 6 (ligand binding domain)	0.025
rs4986938	AluI	A1730G SNP in the 3′-untranslated region of exon 8	0.398
D14S1026		CA dinucleotide repeat polymorphism in intron 5	

1. Minor allele frequency (MAF) in the CEU population (Utah residents with ancestry from northern and western Europe), HapMap Data Rel 24/phase II Nov 08 database.

Table C-2 lists the 15 *ESR2* SNPs included in haplotype or genome-wide association studies of cancer [9, 13, 20].

Table C-2 *ESR2* SNPs included in haplotype or genome-wide association studies of cancer.

Identifier	BPC3	CAPS	CGEMS
rs1256031	X		
rs1256049 (RsaI)	X		
rs3020450	X		X
rs4986938 (AluI)	X		
rs1256040		X	
rs1256062		X	X
rs1887994		X	
rs2987983		X	
rs1255998			
rs1256030			
rs1256065			
rs10137185			X
rs1256044			X
rs1269056			X
rs944045			X

Legend: BPC3 – Breast and Prostate Cancer Cohort Consortium, CAPS – Cancer Prostate in Sweden, CGEMS– Cancer Genetic Markers of Susceptibility

2. NCBI Entrez SNP

A 01/28/2009 query of the NCBI Entrez SNP database (*ESR2*[All Fields] AND ("homo sapiens"[Organism] AND "snp"[Snp_Class]) identified 571 SNPs. The search identified 10 coding SNPs, all synonymous (sense).

3. Cancer Genome Anatomy Project (CGAP) SNP500Cancer Database

A 01/28/2009 query of the Cancer Genome Anatomy Project (CGAP) SNP500Cancer Database identified 13 SNPs with variation in either the SNP500Cancer or Human Diversity Panel (HDP) populations.

4. International HapMap Project

A HapMap Data Rel 24/phase II Nov 08 database (NCBI build 36) query restricted to the CEU population (N=90 Utah residents with ancestry from northern and western Europe) identified 169 SNPs in chromosome 14 (position 63743506 to 63895021), a 151.5 kb genomic region spanning 20 kb upstream and 20 kb downstream of the estrogen receptor beta isoform 2 (NM_001040276).

5. FastSNP

A 01/29/2009 Excel spreadsheet FastSNP download of variants (coding type = ALL) in the *ESR2* ENST00000358599 transcript contained 754 SNPs. The search identified ten SNPs with possible functional significance, including one conservative missense and three sense SNPs in an *ESR2* coding region and six non-coding SNPs in an *ESR2* promoter or regulatory region.

List of candidate SNPs

SNP500Cancer, Entrez SNP, FastSNP, and CEU HapMap database searches identified a total of 1,149 SNPs according to dbSNP identifier (“rs number”), including 154 SNPs common to CEU

HapMap and non-HapMap sources (SNP500Cancer, Entrez SNP, and FastSNP), 980 SNPs unique to non-HapMap sources, and 15 SNPs unique to CEU HapMap (Table C-3). SNP500Cancer, Entrez SNP, and FastSNP database searches identified 29 high priority SNPs, including 11 CEU HapMap SNPs (Table C-3).

Table C-3 N=1,149 SNPs identified through SNP500Cancer, Entrez SNP, FastSNP, and HapMap database searches

	CEU HapMap SNP	
	No (N=980)	Yes (N=169)
In SNP500Cancer	5	8
In Entrez SNP	434	133
In FastSNP	623	128
In SNP500, EntrezSNP, or FastSNP	980	154
High Priority SNP[1]	18	11
Not in SNP500Cancer, EntrezSNP, or FastSNP		15

1. High priority SNPs include SNP500Cancer SNPs, coding SNPs in Entrez SNP or FastSNP, and promoter-regulator SNPs in FastSNP.

Haplotype tag-SNP (htSNP) selection procedure

As noted above, a HapMap search initially identified 169 CEU *ESR2* Phase II SNPs (Table C-3). However, 49 *ESR2* SNPs had a zero minor allele frequency (MAF) in the CEU population. Figure C-1 displays measures (D') of linkage disequilibrium (LD) for the remaining 120 *ESR2* SNPs with non-zero MAF in the CEU population. To select htSNPs for the SNPs shown in Figure C-1, I forced selection of the AluI SNP (rs4986938), the RsaI SNP (rs1256049), and four eligible high priority SNPs (rs8006145, rs1256031, rs1256030, and rs3020450) and used the de

Bakker pairwise Tagger algorithm [21] at an $R^2 = 0.80$ threshold, as implemented in Haploview 4.1 [22]. Tagger selected 34 htSNPs, including 28 SNPs within the *ESR2* gene (Table C-4), capturing all 120 SNPs with mean $R^2 = 0.967$. Nine of the 34 htSNPs captured only low-frequency-low-priority SNPs (MAF < 0.05). The SNP500Cancer SNPs rs1256031 captured the six SNPs tagged by the adjacent SNP500Cancer SNP rs1256030. Twenty-five htSNPs remained after excluding rs1256030 and the low-frequency-low-priority SNPs. Replacing two low priority SNPs with linked alternatives, a final set of 25 htSNPs could be genotyped on two Sequenom multi-plex panels (Table C-5). These 25 ht SNPs captured 104 (87%) of the 120 CEU HapMap SNPs within 20 kb of *ESR2* at $R^2 \geq 0.80$ with mean $R^2 = 0.961$. Table C-6 lists the HapMap SNPs not captured by the 25 htSNPs in Table C-5.

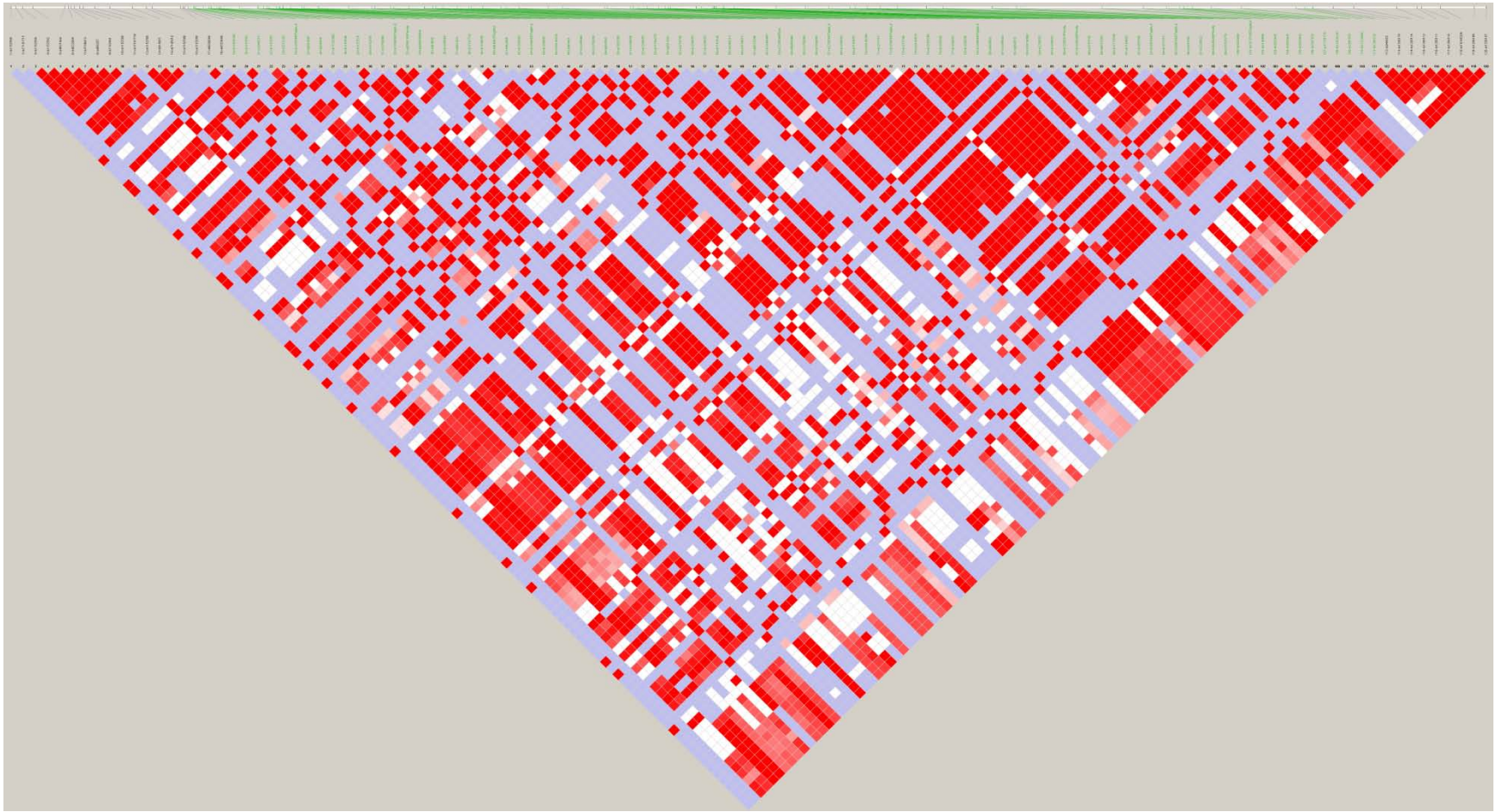


Figure C-1 Linkage disequilibrium (LD) display of N=120 HapMap Phase II SNPs within 20 kb of the *ESR2* with non-zero MAF in the CEU population. Color key – White= D′<1 and LOD<2, Blue: D′=1 and LOD<2, Shades of pink and red: D′<1 and LOD≥2, and Bright red: D′=1 and LOD≥2.

Table C-4 Haplotype tagging SNPs (htSNP) for *ESR2* HapMap Phase II SNPs.

			SNPs captured			Max MAF <
Position	htSNP	Forced [1]	N	Min MAF	Max MAF	0.05
63763624	rs1255998		2	0.086	0.092	
63763835	rs8018687		3	0.059	0.067	
63764502	rs17225885		1	0.033	0.033	X
63769203	rs8006145	Priority	3	0.308	0.317	
63769569	rs4986938	AluI	3	0.398	0.425	
63770894	rs17101732		1	0.026	0.026	X
63771970	rs1256063		1	0.083	0.083	
63773346	rs1256061		4	0.440	0.475	
63782108	rs1952585		5	0.114	0.176	
63785526	rs17766755		1	0.331	0.331	
63787406	rs1256052		1	0.033	0.033	X
63788870	rs7157428		2	0.085	0.092	
63790285	rs2738415		1	0.008	0.008	X
63793804	rs1256049	RsaI	10	0.017	0.026	
63809258	rs1273196		1	0.059	0.059	
63810273	rs12435284		2	0.125	0.125	
63813085	rs1256036		20	0.321	0.461	
63815932	rs1256031	Priority	1	0.415	0.415	
63816923	rs1256030	Priority	6	0.417	0.458	
63838055	rs3020450	Priority	4	0.325	0.342	
63843145	rs3020449		2	0.483	0.500	
63845529	rs10137185		2	0.158	0.158	
63854189	rs7146908		3	0.009	0.033	X
63862093	rs3020443		1	0.216	0.216	
63865264	rs11629158		5	0.025	0.034	X
63868313	rs2987976		2	0.083	0.083	
63873910	rs17226088		1	0.042	0.042	X
63874754	rs1256120		1	0.183	0.183	

1. Forced selection as htSNP, rs1256049 because of location in coding region and rs8006145, rs4986938, rs1256031, rs1256030, and rs3020450 because of membership in SNP500Cancer.

Table C-5 Proposed htSNPs.

	Tagged SNP		
htSNP identifier	Identifier	MAF	R^2 with htSNP
1 rs8021944	rs8021944	0.110	1.000
	rs8022694	0.108	1.000
	rs7145919	0.100	0.914
	rs12434245	0.100	0.914
2 rs968257	rs1152594	0.400	0.833
	rs1152592	0.408	0.867
	rs968257	0.392	1.000
	rs1152590	0.408	0.867
3 rs1152589	rs2738413	0.450	0.935
	rs1152591	0.450	0.935
	rs1152589	0.467	1.000
4 rs1255998	rs1152583	0.075	1.000
	rs1048315	0.070	1.000
	rs1255998	0.086	1.000
	rs1256064	0.092	0.901
5 rs8018687	rs8020646	0.051	1.000
	rs8018687	0.059	1.000
	rs1109056	0.067	1.000
	rs944045	0.067	1.000
6 rs8006145	rs2772163	0.316	0.843
	rs8006145	0.317	1.000
	rs867443	0.308	0.887
7 rs4986938	rs4986938	0.398	1.000
	rs3783736	0.425	0.865
	rs17179740	0.422	0.860
8 rs1256063	rs1256063	0.083	1.000
9 rs1256061	rs1256061	0.475	1.000
	rs4365213	0.440	0.901
	rs6573549	0.440	0.901
	rs12435857	0.449	0.903
10 rs1952585	rs1256062	0.176	0.805
	rs10144225	0.125	1.000
	rs8017441	0.125	1.000
	rs1952585	0.125	1.000
	rs2274705	0.114	1.000
11 rs17766755	rs17766755	0.331	1.000
12 rs1256049	rs1152596	0.025	1.000
	rs1152585	0.017	1.000
	rs1152580	0.017	1.000
	rs1256066	0.017	1.000
	rs944050	0.025	1.000
	rs944460	0.025	1.000
	rs944461	0.025	1.000
	rs1256060	0.025	1.000
	rs953592	0.026	1.000
	rs1256055	0.025	1.000
	rs1256053	0.025	1.000
	rs1256049	0.025	1.000

Table C-5 Continued

htSNP identifier	Tagged SNP Identifier	MAF	R^2 with htSNP
13 rs8003490	rs8003490	0.085	1.000
	rs7157428	0.092	1.000
14 rs1273196	rs1273196	0.059	1.000
15 rs12435284	rs12435284	0.125	1.000
	rs7159462	0.125	1.000
16 rs1256036	rs915057	0.383	0.863
	rs1152588	0.405	0.930
	rs1152582	0.397	0.894
	rs928554	0.365	0.840
	rs1152579	0.377	0.855
	rs1152578	0.386	0.891
	rs1256065	0.384	0.891
	rs1256059	0.397	0.930
	rs1256056	0.400	0.932
	rs1256048	0.383	1.000
	rs1256045	0.383	1.000
	rs1256044	0.383	1.000
	rs1256043	0.321	1.000
	rs10148269	0.383	1.000
	rs1271573	0.381	1.000
	rs1256038	0.373	1.000
	rs1256037	0.390	1.000
	rs1256036	0.383	1.000
	rs1269056	0.383	1.000
	rs960069	0.370	1.000
	rs1271572	0.414	0.931
	rs3020445	0.461	0.924
	rs2357479	0.433	0.813
17 rs1256031	rs1256040	0.420	1.000
	rs10143616	0.450	0.871
	rs960070	0.450	0.871
	rs1256033	0.417	1.000
	rs1256031	0.415	1.000
	rs1256030	0.417	1.000
	rs6573553	0.458	0.842
18 rs1887994	rs1887994	0.083	1.000
	rs2987976	0.083	1.000
19 rs3020450	rs7154455	0.342	0.963
	rs2987983	0.333	1.000
	rs3020450	0.333	1.000
	rs3020444	0.325	0.963
20 rs3020449	rs2978381	0.500	0.967
	rs3020449	0.483	1.000
21 rs10137185	rs1952586	0.158	1.000
	rs10137185	0.158	1.000
22 rs3020443	rs3020443	0.216	1.000

Table C-5 Continued

	Tagged SNP		
htSNP identifier	Identifier	MAF	R^2 with htSNP
23 rs1256120	rs1256120	0.183	1.000
	rs944052	0.192	0.947
	rs1256116	0.192	0.947
	rs1256114	0.192	0.947
24 rs10146204	rs10146204	0.465	1.000
25 rs1256108	rs1256112	0.408	0.872
	rs1256111	0.442	1.000
	rs1256110	0.490	0.925
	rs1256108	0.442	1.000
	rs1256107	0.442	1.000

Table C-6 SNPs not captured by htSNPs in Table C-5.

Identifier	Position	MAF
rs17101715	63744135	0.008
rs17101718	63752147	0.025
rs9323448	63761209	0.025
rs17225885	63764502	0.033
rs17101732	63770894	0.026
rs1256052	63787406	0.033
rs2738415	63790285	0.008
rs1256034	63814878	0.025
rs10136955	63815016	0.009
rs1256032	63815777	0.034
rs1256027	63836427	0.033
rs11625778	63843062	0.026
rs7146908	63854189	0.033
rs17101774	63863334	0.017
rs11629158	63865264	0.034
rs17226088	63873910	0.042

Proposal

Table C-7 lists 24 interesting SNPs not in HapMap. According to Entrez SNP, the missense SNP identified by FastSNP (rs1255953) is located in a SYNE2 intron. Justification for genotyping any of the SNPs in Table C-7 is weak based on information currently available. Therefore, I propose to limit *ESR2* genotyping to the 25 htSNPs listed in Table C-5.

Table C-7 *ESR2* SNPs coded as possibly functional (in FastSNP), located in coding region (in Entrez SNP), or validated in SNP500Cancer, but not represented in HapMap.

		Reason selected			
Identifier	FastSNP Possible Functional Effect	Possible Functional Effect in FastSNP	Sense coding SNPs in Entrez SNP	SNP500	MAF
rs1255953	Missense (conservative); Splicing regulation	X			0.167
rs1256054	Sense/synonymous; Splicing regulation	X	X		0.000
rs10137994	Promoter/regulatory region	X			
rs10483774	Promoter/regulatory region	X			0.000
rs17226060	Promoter/regulatory region	X			0.000
rs2738411	Promoter/regulatory region	X			
rs35945666	Promoter/regulatory region	X			
rs3832949	Promoter/regulatory region	X			
rs1541060	Sense/synonymous	X	X		
rs17225976			X		
rs45624541			X		
rs56926155			X		
rs58127193			X		
rs58256696			X		
rs60101369			X		
rs60892953			X		
rs10047818				X	0.000
rs1256030				X	0.414
rs1256041				X	0.403
rs34996860				X	0.005
rs35000350				X	0.000
rs35142532				X	0.000
rs35743760				X	0.016
rs944459				X	0.000

Literature Cited in SNP selection methodology

1. Packer, B., et al., *SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes.* Nucleic Acids Research, 2006. **34 Database issue**: p. D617-D621.
2. Yuan, H.-Y., et al., *FASTSNP: An always up-to-date and extendable service for SNP function analysis and prioritization.* Nucleic Acids Research, 2006. **34**: p. W635-W641.
3. Gennari, L., et al., *Estrogen receptor gene polymorphisms and the genetics of osteoporosis: a HuGE review.* American Journal of Epidemiology, 2005. **161**(4): p. 307-20.
4. Slattery, M.L., et al., *Associations between ERalpha, ERbeta, and AR genotypes and colon and rectal cancer.* Cancer Epidemiology, Biomarkers & Prevention, 2005. **14**(12): p. 2936-42.
5. Setiawan, V.W., et al., *Estrogen receptor beta (ESR2) polymorphisms and endometrial cancer (United States).* Cancer Causes & Control, 2004. **15**(6): p. 627-33.
6. Leigh Pearce, C., et al., *Comprehensive evaluation of ESR2 variation and ovarian cancer risk.* Cancer Epidemiology, Biomarkers & Prevention, 2008. **17**(2): p. 393-6.
7. Aschim, E.L., et al., *The RsaI polymorphism in the estrogen receptor-beta gene is associated with male infertility.* Journal of Clinical Endocrinology & Metabolism, 2005. **90**(9): p. 5343-8.
8. Sun, Y.-h., et al., *[Association between single-nucleotide polymorphisms in estrogen receptor beta gene and risk of prostate cancer]*. Chung-Hua Wai Ko Tsa Chih [Chinese Journal of Surgery], 2005. **43**(14): p. 948-51.
9. Chen, Y.-C., et al., *Sequence variants of estrogen receptor beta and risk of prostate cancer in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium.* Cancer Epidemiology, Biomarkers & Prevention, 2007. **16**(10): p. 1973-81.
10. McIntyre, M.H., et al., *Prostate cancer risk and ESR1 TA, ESR2 CA repeat polymorphisms.* Cancer Epidemiology, Biomarkers & Prevention, 2007. **16**(11): p. 2233-6.
11. Forsti, A., et al., *Polymorphisms in the estrogen receptor beta gene and risk of breast cancer: no association.* Breast Cancer Research & Treatment, 2003. **79**(3): p. 409-13.
12. Zheng, S.L., et al., *Joint effect of estrogen receptor beta sequence variants and endogenous estrogen exposure on breast cancer risk in Chinese women.* Cancer Research, 2003. **63**(22): p. 7624-9.
13. Gold, B., et al., *Estrogen receptor genotypes and haplotypes associated with breast cancer risk.* Cancer Research, 2004. **64**(24): p. 8891-900.
14. Maguire, P., et al., *Estrogen receptor beta (ESR2) polymorphisms in familial and sporadic breast cancer.* Breast Cancer Research & Treatment, 2005. **94**(2): p. 145-52.
15. Iobagiu, C., et al., *Microsatellite profile in hormonal receptor genes associated with breast cancer.* Breast Cancer Research & Treatment, 2006. **95**(2): p. 153-9.
16. Georgopoulos, N.A., et al., *Estrogen receptor polymorphisms in tamoxifen-treated women with breast cancer.* Gynecological Endocrinology, 2006. **22**(4): p. 185-9.
17. Gallicchio, L., et al., *Polymorphisms in estrogen-metabolizing and estrogen receptor genes and the risk of developing breast cancer among a cohort of women with benign breast disease.* BMC Cancer, 2006. **6**: p. 173.

18. Tsezou, A., et al., *Association of repeat polymorphisms in the estrogen receptors alpha, beta (ESR1, ESR2) and androgen receptor (AR) genes with the occurrence of breast cancer*. Breast, 2008. **17**(2): p. 159-66.
19. Breast and Prostate Cancer Cohort Consortium, et al., *Haplotypes of the estrogen receptor beta gene and breast cancer risk.* International Journal of Cancer, 2008. **122**(2): p. 387-92.
20. Thellenberg-Karlsson, C., et al., *Estrogen receptor beta polymorphism is associated with prostate cancer risk.* Clinical Cancer Research, 2006. **12**(6): p. 1936-41.
21. de Bakker, P., et al., *Efficiency and power in genetic association studies.* Nature Genetics, 2005. **37**: p. 1217-23.
22. Barrett, J., et al., *Haploview: Analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. **21**: p. 263–5.

APPENDIX D

SUPPLEMENTAL TABLES AND FIGURES FOR PROJECT#3

D.1 DESCRIPTION FOR *ESR2* GENOTYPE RESULTS

Table D-1 SNPs included in genotyping analysis

htSNP identifier		Tagged SNP Identifier				
1	rs8021944	rs8022694	rs7145919	rs12434245	rs8021944	
2	rs968257	rs1152594	rs1152592	rs1152590	rs968257	
3	rs1152589	rs2738413	rs1152591	rs1152589		
4	rs1255998	rs1152583	rs1048315	rs1256064	rs1255998	
5	rs8006145	rs2772163	rs867443	rs8006145		
6	rs4986938	rs3783736	rs17179740	rs4986938		
7	rs1256063	rs1256063				
8	rs1256061	rs4365213	rs6573549	rs12435857	rs1256061	
9	rs1952585	rs1256062	rs10144225	rs8017441	rs2274705	rs1952585
10	rs17766755	rs17766755				
11	rs1256049	rs1152596	rs1152585	rs1152580	rs1256066	rs944050
		rs944460	rs944461	rs1256060	rs953592	rs1256055
		rs1256053	rs1256049			
12	rs8003490	rs7157428	rs8003490			
13	rs12435284	rs7159462	rs12435284			
14	rs1256036	rs915057	rs1152588	rs1152582	rs928554	rs1152579
		rs1152578	rs1256065	rs1256059	rs1256056	rs1256048
		rs1256045	rs1256044	rs1256043	rs10148269	rs1271573
		rs1256038	rs1256037	rs1269056	rs960069	rs1271572
		rs3020445	rs2357479	rs1256036		
15	rs1887994	rs2987976	rs1887994			
16	rs3020450	rs7154455	rs2987983	rs3020444	rs3020450	
17	rs3020449	rs2978381	rs3020449			
18	rs10137185	rs1952586	rs10137185			
19	rs3020443	rs3020443				
20	rs1256120	rs944052	rs1256116	rs1256114	rs1256120	
21	rs10146204	rs10146204				
22	rs1256108	rs1256112	rs1256111	rs1256110	rs1256107	rs1256108

Table D- 2 SNPs selected initially but excluded in genotyping

SNPs	HapMap MAF	Forced	Tagged SNP Identifier				
rs1256031	0.420	Priority	rs1256040	rs10143616	rs960070	rs1256033	rs1256030
			rs6573553	rs1256031			
rs1273196	0.059		rs1273196				
rs8018687	0.051		rs8020646	rs1109056	rs944045	rs8018687	

Table D-3 Distribution of called number for plex 1 and 2

Plex 1 (18 SNPs)	
N non-missing	N subject
0	21
4	2
5	1
7	1
10	1
12	1
13	2
14	6
15	5
16	3
17	7
18	122
All	172

Plex 2 (4 SNPs)	
N non-missing	N subject
0	21
1	3
2	4
3	2
4	142
All	172

Table D-4 Distribution of genotype results by the source of DNA extraction

		Extraction method					
		1	2	3	4	All	p-value*
Plex1Good							<0.001
No	n	6	3	17	14	40	
Yes	n	27	47	7	51	132	
	%	82%	94%	29%	78%	77%	
All	n	33	50	24	65	172	
Plex2Good							<0.001
No	n	3	2	12	11	28	
Yes	n	30	48	12	54	144	
	%	91%	96%	50%	83%	84%	
All	n	33	50	24	65	172	
PlexAllGood							<0.001
No	n	6	3	17	16	42	
Yes	n	27	47	7	49	130	
	%	82%	94%	29%	75%	76%	
All	n	33	50	24	65	172	

*Chi-Square Test

extraction batch

1 - extracted by Maureen Lyons from tissue on slides

2 - extracted by Romkes lab, received sample at 10ng/ul concentration

3 - extracted by Jill's lab

4 - extracted by Ji

Plex1Good

NO means missing >=4 SNPs out of 18

Yes means missing <4 SNPs out of 18

Plex2Good

NO means missing >=2 SNPs out of 4

Yes means missing <2 SNPs out of 4

PlexAllGood

NO means Plex1Good=0 or Plex2Good=0

Yes means Plex1Good=1 and Plex2Good=1

Table D-5 Distribution of called rate for 22 SNPs in the study

htSNP identifier	SNP	N	Non-missing called	% Non-missing
rs1256120	1	132	130	98.5%
rs1952585	2	132	132	100.0%
rs4986938	3	132	132	100.0%
rs8006145	4	132	132	100.0%
rs3020450	5	132	132	100.0%
rs968257	6	132	132	100.0%
rs1256061	7	132	132	100.0%
rs1256063	8	132	132	100.0%
rs12435284	9	132	132	100.0%
rs17766755	10	132	131	99.2%
rs10146204	11	132	132	100.0%
rs1256049	12	132	131	99.2%
rs1887994	13	132	132	100.0%
rs1152589	14	132	125	94.7%
rs3020443	15	132	131	99.2%
rs3020449	16	132	131	99.2%
rs10137185	17	132	132	100.0%
rs1256036	18	132	132	100.0%
rs1255998	19	144	144	100.0%
rs8021944	20	144	144	100.0%
rs8003490	21	144	144	100.0%
rs1256108	22	144	142	98.6%

N= Number of subjects with plex1Good =Yes or with plex2Good=Yes, as appropriate

Table D-6 Minor allele frequency comparison with HapMap and Hardy-Weinberg Equilibrium (HWE) test result for white study subjects

					Minor Allele Frequency		
Genotype Order	Position	Forced	htSNPs	Allele [1]	Study	HapMap [2]	HWE [3]
1	63874754		rs1256120	C	0.098	0.183	0.031*
2	63782108		rs1952585	C	0.106	0.176	0.829
3	63769569	AluI	rs4986938	A	0.381	0.398	0.745
4	63769203	Priority	rs8006145	A	0.289	0.316	0.377
5	63838055	Priority	rs3020450	A	0.344	0.342	0.373
6	63750038		rs968257	G	0.404	0.400	0.373
7	63773346		rs1256061	A	0.450	0.475	0.433
8	63771970		rs1256063	T	0.050	0.083	0.579
9	63810273		rs12435284	T	0.050	0.125	0.579
10	63785526		rs17766755	A	0.352	0.331	0.317
11	63888522		rs10146204	A	0.399	0.465	0.798
12	63793804	Rsal	rs1256049	A	0.032	0.025	0.728
13	63830364		rs1887994	T	0.078	0.083	0.377
14	63753679		rs1152589	T	0.486	0.450	0.838
15	63862093		rs3020443	C	0.259	0.216	0.383
16	63843145		rs3020449	C	0.417	0.500	0.767
17	63845529		rs10137185	T	0.069	0.158	0.469
18	63813085		rs1256036	G	0.454	0.383	0.179
19	63763624		rs1255998	G	0.123	0.075	0.504
20	63749051		rs8021944	G	0.064	0.110	0.418
21	63795122		rs8003490	A	0.093	0.085	0.978
22	63891973		rs1256108	C	0.496	0.408	0.577

1. Rare allele observed in white study subjects

2. Minor allele frequency obtained from HapMap database

3. Hardy-Weinberg-Equilibirum p-value, with asterisk (*) to indicate $p<0.05$

Genotype order#1-18 is from Plex1 with 109 white subjects and #19-22 is from Plex2 with 118 white subjects

D.2 DESCRIPTION FOR PROJECT 3 STUDY SUBJECTS

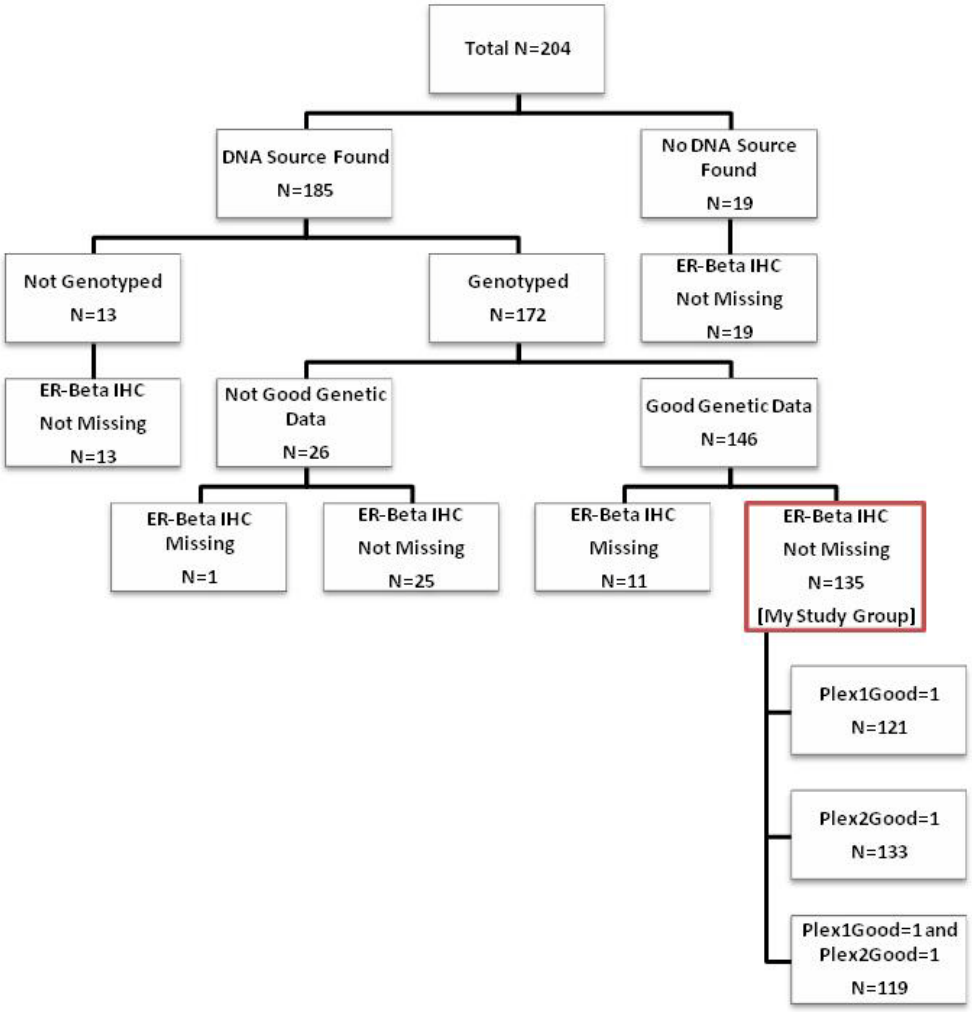


Figure D-1 Study Subject selection flow chart

*Total N=204 subjects obtained after excluding two known non-lung cancer patients (V-101 and V-102) and 1 lung cancer patients aged less than 21 years old (Subject ID=660 with age of 7)
*[Good Genetic Data] is defined as [Plex1Good=1: subjects with less than 4 missing SNPs out of 18 SNPs in Plex#1] or [Plex2Good=1: subjects with less than 2 missing SNPs out of 4 SNPs in Plex#2]

Table D-7 Characteristics of subjects excluded and included from analysis

	All n=204		Excluded n=69		Included n=135		
Characteristic	n	%	n	%	n	%	p-value*
Survival status							0.83
Dead	133	65.2	46	66.7	87	64.4	
Alive	58	28.4	18	26.1	40	29.6	
Unknown	13	6.4	5	7.2	8	5.9	
Sex							0.15
male	101	49.5	39	56.5	62	45.9	
female	103	50.5	30	43.5	73	54.1	
Race							
white	178	87.3	58	84.1	120	88.9	0.03
African-American	17	8.3	10	14.5	7	5.2	
missing	9	4.4	1	1.4	8	5.9	
Age							0.64
30-59	45	22.1	14	20.3	31	23.0	
60-69	68	33.3	26	37.7	42	31.1	
70+	91	44.6	29	42.0	62	45.9	
Smoking status							0.25
active smoker	67	32.8	24	34.8	43	31.9	
ex-smoker	81	39.7	27	39.1	54	40.0	
smoker, NOS	25	12.3	6	8.7	19	14.1	
never smoker	17	8.3	4	5.8	13	9.6	
missing	14	6.9	8	11.6	6	4.4	
Smoking dose-duration (among ever smokers)							0.72
1-25	29	16.8	9	15.8	20	17.2	
26-50	65	37.6	20	35.1	45	38.8	
51-75	38	22.0	16	28.1	22	19.0	
>76	36	20.8	11	19.3	25	21.6	
missing	5	2.9	1	1.8	4	3.4	
Stage							0.53
I	80	39.2	27	39.1	53	39.3	
II	36	17.6	12	17.4	24	17.8	
III	56	27.5	16	23.2	40	29.6	
IV	12	5.9	7	10.1	5	3.7	
recurrent	10	4.9	3	4.3	7	5.2	
missing	10	4.9	4	5.8	6	4.4	
Source of stage							0.10
pathologic	175	85.8	56	81.2	119	88.1	
clinical	9	4.4	6	8.7	3	2.2	
not applicable	20	9.8	7	10.1	13	9.6	

Table D-7 (continued)

Characteristic	All n=204 n	%	Excluded n=69 n	%	Included n=135 n	%	p-value*
Histology							
Adenocarcinoma	105	51.5	40	58.0	65	48.1	
BAC	2	1.0	1	1.4	1	0.7	
Adenosquamous	7	3.4	3	4.3	4	3.0	
Squamous cell	67	32.8	18	26.1	49	36.3	
Large cell	8	3.9	2	2.9	6	4.4	
Undifferentiated	5	2.5	2	2.9	3	2.2	
Malignant carcinoid	2	1.0	1	1.4	1	0.7	
Small cell	3	1.5	1	1.4	2	1.5	
missing	5	2.5	1	1.4	4	3.0	
Histology class							
Adenocarcinoma	105	51.5	40	58.0	65	48.1	0.32
Squamous cell	67	32.8	18	26.1	49	36.3	
Other/missing	32	15.7	11	15.9	21	15.6	

*Chi-square test

Table D-8 ER-beta IHC expression of subjects excluded and included from analysis

Characteristic		All n=204 n	% or (median)	Excluded n=69 n	% or (median)	Included n=135 n	% or (median)	p-value
ER Beta cytoplasmic	High expression, %*	73	38.0	19	33.3	54	40.0	0.38
	Median[a]	192	(7.0)	57	(6.0)	135	(7.0)	0.27
ER Beta Nuclear	High expression, %*	132	68.8	36	63.2	96	71.1	0.28
	Median[a]	192	(8.0)	57	(7.9)	135	(8.0)	0.21

*Chi-square test

[a]Wilcoxon rank sum test

High ER-beta cytoplasmic and nuclear expression defined by subject-specific averaged Allred values above 7.

Table D-9 Patients characteristics

		All n=135		Non-Missing Max N=135		
Variable	**Measure**	**No.**	**percent**	**Total N**	**No.**	**percent**
Survival status	Dead	87	64.4	127	87	68.5
Sex	Women	73	54.1	135	73	54.1
Race	African-American	7	5.2	127	7	5.5
Age	30-59 years	31	23.0	135	31	23.0
	60-69 years	42	31.1		42	31.1
	70+ years	62	45.9		62	45.9
Smoking status	never smoker	13	9.6	129	13	10.1
	ex-smoker	54	40.0		54	41.9
	active smoker	62	45.9		62	48.1
Smoking dose-duration (among ever smokers=116)	<50 pack-years	65	56.0	112	65	58.0
	50+pack-years	47	40.5		47	42.0
Stage	I	53	39.3	129	53	41.1
	II	24	17.8		24	18.6
	III	40	29.6		40	31.0
	IV	5	3.7		5	3.9
	recurrent	7	5.2		7	5.4
Histology	Adenocarcinoma	65	48.1	131	65	49.6
	BAC	1	0.7		1	0.8
	Adenosquamous	4	3.0		4	3.1
	Squamous cell	49	36.3		49	37.4
	Large cell	6	4.4		6	4.6
	Undifferentiated	3	2.2		3	2.3
	Malignant carcinoid	1	0.7		1	0.8
	Small cell	2	1.5		2	1.5
Histology Class	Adenocarcinoma	65	48.1	114	65	57.0
	Squamous cell	49	36.3		49	43.0
ERβ expression score	nuclear	135[a]	7.14 (8.0)[b]			
	cytoplasmic	135[a]	5.38 (7.0)[b]			
	total	135[a]	12.52 (14.75)[b]			

[a] Number of subjects with non-missing IHC data

[b] Mean and median of Allred score, medians in parentheses.

Table D-10 Summary statistics of nuclear, cytoplasmic, and total IHC expression scores of estrogen receptor beta in the study population by gender, N=135

		N	Mean	SD	Median	25th Percentiles	75th Percentiles
Total Subject	nuclear ERβ	135	7.13	1.80	8	7	8
	cytoplasmic ERβ	135	5.38	3.08	7	3.2	8
	total ERβ	135	12.51	4.41	14.75	10	16
Men	nuclear ERβ	62	7.25	1.66	8	7	8
	cytoplasmic ERβ	62	5.61	2.90	7	3.5	8
	total ERβ	62	12.86	4.08	15	10.5	16
Women	nuclear ERβ	73	7.03	1.92	8	7	8
	cytoplasmic ERβ	73	5.18	3.23	7	1.5	8
	total ERβ	73	12.21	4.67	14.6	8	16

D.3 ASSOCIATION BETWEEN *ESR2* SNP AND ER-BETA IHC EXPRESSION IN LUNG TUMORS

Table D-11 Association between *ESR2* SNPs and ER-beta IHC expression for all study subjects (N=135)

		Cytoplasmic ERβ					Nuclear ERβ				
SNP	Genotype	N	P25	Med	P75	p-value*	N	P25	Med	P75	p-value*
rs8021944	TT	118	3.00	7.00	8.00	0.083	118	7.00	8.00	8.00	0.028
	TG	14	6.00	7.83	8.00		14	8.00	8.00	8.00	
	GG	1	7.33	7.33	7.33		1	8.00	8.00	8.00	
	TG+GG	15	6.00	7.75	8.00	0.081	15	8.00	8.00	8.00	0.029
rs968257	AA	44	0.75	7.00	8.00	0.826	44	6.00	8.00	8.00	0.312
	AG	55	5.00	7.00	8.00		55	7.00	8.00	8.00	
	GG	22	0.00	6.00	7.70		22	7.00	8.00	8.00	
	AG+GG	77	4.00	7.00	8.00	0.576	77	7.00	8.00	8.00	0.439
rs1152589	AA	31	4.00	6.80	8.00	0.586	31	7.75	8.00	8.00	0.189
	AT	59	4.80	7.00	8.00		59	7.00	8.00	8.00	
	TT	26	0.00	6.25	7.75		26	6.00	8.00	8.00	
	AT+TT	85	3.20	7.00	8.00	0.864	85	6.50	8.00	8.00	0.064
rs1255998	CC	100	3.35	7.00	8.00	0.240	100	7.00	8.00	8.00	0.164
	CG	32	4.00	7.00	7.33		32	7.20	7.78	8.00	
	GG	1	3.50	3.50	3.50		1	6.50	6.50	6.50	
	CG+GG	33	4.00	7.00	7.25	0.259	33	7.00	7.75	8.00	0.185
rs8006145	CC	61	3.00	7.00	7.90	0.600	61	6.50	8.00	8.00	0.730
	CA	49	4.00	7.00	8.00		49	7.00	8.00	8.00	
	AA	11	0.00	6.75	8.00		11	8.00	8.00	8.00	
	CA+AA	60	3.75	7.00	8.00	0.570	60	7.00	8.00	8.00	0.068
rs4986938	GG	44	1.50	7.00	7.63	0.397	44	6.50	8.00	8.00	0.086
	GA	61	4.00	7.00	8.00		61	7.00	8.00	8.00	
	AA	16	4.75	7.00	8.00		16	7.50	8.00	8.00	
	GA+AA	77	4.00	7.00	8.00	0.462	77	7.00	8.00	8.00	0.137
rs1256063	CC	108	3.50	7.00	8.00	0.756	108	7.00	8.00	8.00	0.271
	CT	13	2.50	7.00	7.25		13	7.00	7.75	8.00	
	CT+TT	13	2.50	7.00	7.25	0.756	13	7.00	7.75	8.00	0.271
rs1256061	CC	33	0.00	6.50	7.25	0.551	33	6.00	7.67	8.00	0.632
	CA	67	4.00	7.00	8.00		67	7.00	8.00	8.00	
	AA	21	4.00	7.00	8.00		21	7.75	8.00	8.00	
	CA+AA	88	4.00	7.00	8.00	0.054	88	7.25	8.00	8.00	0.022
rs1952585	TT	96	3.35	7.00	8.00	0.133	96	7.00	8.00	8.00	0.190
	TC	24	3.25	5.80	7.23		24	6.65	7.68	8.00	
	CC	1	6.75	6.75	6.75		1	8.00	8.00	8.00	
	TC+CC	25	3.50	6.00	7.20	0.130	25	6.80	7.75	8.00	0.173
rs17766755	GG	46	0.00	7.00	7.50	0.322	46	6.50	8.00	8.00	0.097
	GA	62	3.50	7.00	8.00		62	7.00	8.00	8.00	
	AA	12	4.75	7.00	8.00		12	7.50	8.00	8.00	
	GA+AA	74	3.50	7.00	8.00	0.375	74	7.00	8.00	8.00	0.140
rs1256049	GG	112	3.35	7.00	8.00	0.421	112	7.00	8.00	8.00	0.584
	GA	8	2.00	6.00	7.20		8	6.75	7.80	8.00	
	GA+AA	8	2.00	6.00	7.20	0.421	8	6.75	7.80	8.00	0.584

Table D-11 (continued)

		Cytoplasmic ERβ					Nuclear ERβ				
SNP	Genotype	N	P25	Med	P75	p-value*	N	P25	Med	P75	p-value*
rs8003490	GG	110	4.00	7.00	8.00	0.072	110	7.00	8.00	8.00	0.062
	GA	22	0.00	5.55	7.20		22	6.50	7.50	8.00	
	AA	1	6.75	6.75	6.75		1	8.00	8.00	8.00	
	GA+AA	23	0.00	5.60	7.20	0.054	23	6.50	7.50	8.00	0.119
rs12435284	CC	109	3.00	7.00	8.00	0.073	109	7.00	8.00	8.00	0.087
	CT	12	6.50	7.95	8.00		12	8.00	8.00	8.00	
	CT+TT	12	6.50	7.95	8.00	0.073	12	8.00	8.00	8.00	0.087
rs1256036	AA	33	3.50	6.00	8.00	0.541	33	7.00	8.00	8.00	0.220
	AG	67	5.00	7.00	8.00		67	7.00	8.00	8.00	
	GG	21	0.00	6.00	7.50		21	6.00	8.00	8.00	
	AG+GG	88	3.10	7.00	8.00	0.774	88	6.90	8.00	8.00	0.434
rs1887994	GG	102	3.00	7.00	8.00	0.584	102	7.00	8.00	8.00	0.981
	GT	19	4.80	7.00	8.00		19	7.00	8.00	8.00	
	GT+TT	19	4.80	7.00	8.00	0.584	19	7.00	8.00	8.00	0.981
rs3020450	GG	52	2.25	7.00	8.00	0.582	52	6.75	8.00	8.00	0.354
	GA	53	5.00	7.00	8.00		53	7.00	8.00	8.00	
	AA	16	1.50	6.38	8.00		16	7.50	8.00	8.00	
	GA+AA	69	3.50	7.00	8.00	0.886	69	7.00	8.00	8.00	0.727
rs3020449	TT	38	0.00	6.45	7.75	0.394	38	6.00	8.00	8.00	0.092
	TC	61	5.50	7.00	8.00		61	7.40	8.00	8.00	
	CC	21	3.00	6.75	8.00		21	7.60	8.00	8.00	
	TC+CC	82	4.00	7.00	8.00	0.156	82	7.40	8.00	8.00	0.114
rs10137185	CC	106	3.00	7.00	8.00	0.087	106	6.80	8.00	8.00	0.155
	CT	14	6.80	7.13	8.00		14	7.90	8.00	8.00	
	TT	1	8.00	8.00	8.00		1	8.00	8.00	8.00	
	CT+TT	15	6.80	7.25	8.00	0.080	15	7.90	8.00	8.00	0.149
rs3020443	AA	66	3.00	7.00	7.90	0.432	66	6.50	8.00	8.00	0.140
	AC	45	4.00	7.00	8.00		45	7.00	8.00	8.00	
	CC	9	0.00	7.00	8.00		9	8.00	8.00	8.00	
	AC+CC	54	3.50	7.00	8.00	0.433	54	7.00	8.00	8.00	0.109
rs1256120	TT	100	3.00	7.00	8.00	0.805	100	6.90	8.00	8.00	0.400
	TC	16	5.75	7.00	7.58		16	7.30	7.95	8.00	
	CC	3	4.00	8.00	8.00		3	8.00	8.00	8.00	
	TC+CC	19	5.50	7.00	8.00	0.567	19	7.60	8.00	8.00	0.843
rs10146204	GG	42	0.00	6.63	7.33	0.032	42	6.00	7.75	8.00	0.258
	GA	57	5.75	7.00	8.00		57	7.50	8.00	8.00	
	AA	22	3.00	5.50	8.00		22	7.00	8.00	8.00	
	GA+AA	79	4.00	7.00	8.00	0.051	79	7.40	8.00	8.00	0.025
rs1256108	TT	30	0.00	6.20	7.75	0.494	30	5.75	8.00	8.00	0.255
	TC	67	5.60	7.00	8.00		67	7.40	8.00	8.00	
	CC	34	3.50	6.78	8.00		34	7.50	8.00	8.00	
	TC+CC	101	4.00	7.00	8.00	0.119	101	7.50	8.00	8.00	0.211

*Jonckheere-Terpstra Test

Table D-12 Association between *ESR2* SNPs and ER-beta IHC expression for only white subjects (N=120)

		Cytoplasmic ERβ					Nuclear ERβ				
SNP	Genotype	N	P25	Med	P75	p-value*	N	P25	Med	P75	p-value*
rs8021944	TT	104	3.00	7.00	8.00	0.065	104	7.00	8.00	8.00	0.041
	TG	13	6.00	7.90	8.00		13	8.00	8.00	8.00	
	GG	1	7.33	7.33	7.33		1	8.00	8.00	8.00	
	TG+GG	14	6.00	7.83	8.00	0.063	14	8.00	8.00	8.00	0.042
rs968257	AA	41	1.50	7.00	8.00	0.729	41	6.00	8.00	8.00	0.406
	AG	48	4.90	7.00	8.00		48	7.00	8.00	8.00	
	GG	20	0.00	6.00	7.35		20	7.00	8.00	8.00	
	AG+GG	68	3.35	7.00	8.00	0.681	68	7.00	8.00	8.00	0.485
rs1152589	AA	27	4.00	6.80	8.00	0.680	27	8.00	8.00	8.00	0.105
	AT	53	4.80	7.00	8.00		53	7.00	8.00	8.00	
	TT	24	0.00	6.25	7.75		24	6.00	8.00	8.00	
	AT+TT	77	3.00	7.00	8.00	0.786	77	6.50	8.00	8.00	0.042
rs1255998	CC	90	3.20	7.00	8.00	0.123	90	7.00	8.00	8.00	0.075
	CG	27	3.00	7.00	7.25		27	7.00	7.75	8.00	
	GG	1	3.50	3.50	3.50		1	6.50	6.50	6.50	
	CG+GG	28	3.25	6.90	7.23	0.134	28	6.75	7.68	8.00	0.085
rs8006145	CC	57	3.00	7.00	7.90	0.816	57	6.50	8.00	8.00	0.936
	CA	41	3.20	7.00	8.00		41	7.00	8.00	8.00	
	AA	11	0.00	6.75	8.00		11	8.00	8.00	8.00	
	CA+AA	52	3.10	6.90	8.00	0.758	52	7.00	8.00	8.00	0.095
rs4986938	GG	41	3.00	7.00	7.50	0.442	41	6.50	8.00	8.00	0.072
	GA	53	3.00	6.80	8.00		53	7.00	8.00	8.00	
	AA	15	3.50	7.00	8.00		15	8.00	8.00	8.00	
	GA+AA	68	3.10	6.90	8.00	0.539	68	7.00	8.00	8.00	0.155
rs1256063	CC	98	3.00	7.00	8.00	0.906	98	7.00	8.00	8.00	0.428
	CT	11	2.50	7.00	7.75		11	7.00	7.75	8.00	
	CT+TT	11	2.50	7.00	7.75	0.906	11	7.00	7.75	8.00	0.428
rs1256061	CC	31	0.00	6.50	7.25	0.675	31	6.00	7.67	8.00	0.997
	CA	58	3.50	7.00	8.00		58	7.00	8.00	8.00	
	AA	20	3.75	7.38	8.00		20	7.88	8.00	8.00	
	CA+AA	78	3.50	7.00	8.00	0.065	78	7.00	8.00	8.00	0.035
rs1952585	TT	87	3.00	7.00	8.00	0.128	87	7.00	8.00	8.00	0.120
	TC	21	3.00	5.60	7.20		21	6.50	7.50	8.00	
	CC	1	6.75	6.75	6.75		1	8.00	8.00	8.00	
	TC+CC	22	3.00	6.18	7.20	0.126	22	6.50	7.55	8.00	0.107
rs17766755	GG	43	0.00	7.00	7.50	0.369	43	6.50	8.00	8.00	0.085
	GA	54	3.00	6.90	8.00		54	7.00	8.00	8.00	
	AA	11	3.50	7.00	8.00		11	8.00	8.00	8.00	
	GA+AA	65	3.20	7.00	8.00	0.446	65	7.00	8.00	8.00	0.161
rs1256049	GG	101	3.00	7.00	8.00	0.580	101	7.00	8.00	8.00	0.394
	GA	7	0.00	7.00	7.40		7	6.50	7.60	8.00	
	GA+AA	7	0.00	7.00	7.40	0.580	7	6.50	7.60	8.00	0.394

Table D-12 (continued)

		Cytoplasmic ERβ					Nuclear ERβ				
SNP	Genotype	N	P25	Med	P75	p-value*	N	P25	Med	P75	p-value*
rs8003490	GG	97	3.50	7.00	8.00	0.047	97	7.00	8.00	8.00	0.042
	GA	20	0.00	5.25	7.10		20	6.50	7.50	8.00	
	AA	1	6.75	6.75	6.75		1	8.00	8.00	8.00	
	GA+AA	21	0.00	5.50	7.00	0.036	21	6.50	7.50	8.00	0.087
rs12435284	CC	98	3.00	6.90	8.00	0.058	98	6.80	8.00	8.00	0.117
	CT	11	6.00	8.00	8.00		11	8.00	8.00	8.00	
	CT+TT	11	6.00	8.00	8.00	0.058	11	8.00	8.00	8.00	0.117
rs1256036	AA	29	3.50	6.00	8.00	0.506	29	7.00	8.00	8.00	0.232
	AG	61	5.00	7.00	8.00		61	7.00	8.00	8.00	
	GG	19	0.00	6.00	7.50		19	6.00	8.00	8.00	
	AG+GG	80	3.00	7.00	8.00	0.837	80	6.90	8.00	8.00	0.373
rs1887994	GG	92	3.00	7.00	8.00	0.454	92	6.90	8.00	8.00	0.646
	GT	17	4.80	7.00	8.00		17	7.50	8.00	8.00	
	GT+TT	17	4.80	7.00	8.00	0.454	17	7.50	8.00	8.00	0.646
rs3020450	GG	49	3.00	7.00	8.00	0.506	49	7.00	8.00	8.00	0.340
	GA	45	3.50	7.00	8.00		45	6.80	8.00	8.00	
	AA	15	0.00	6.00	8.00		15	7.00	8.00	8.00	
	GA+AA	60	3.10	6.90	8.00	0.970	60	7.00	8.00	8.00	0.820
rs3020449	TT	36	0.00	6.45	7.63	0.535	36	6.25	8.00	8.00	0.188
	TC	54	5.00	7.00	8.00		54	7.40	8.00	8.00	
	CC	18	0.00	6.38	8.00		18	7.00	8.00	8.00	
	TC+CC	72	3.50	7.00	8.00	0.198	72	7.20	8.00	8.00	0.171
rs10137185	CC	95	2.50	6.75	8.00	0.070	95	6.50	8.00	8.00	0.192
	CT	13	6.80	7.25	8.00		13	7.90	8.00	8.00	
	TT	1	8.00	8.00	8.00		1	8.00	8.00	8.00	
	CT+TT	14	6.80	7.58	8.00	0.065	14	7.90	8.00	8.00	0.185
rs3020443	AA	61	3.00	7.00	7.90	0.599	61	6.50	8.00	8.00	0.148
	AC	38	3.20	7.00	8.00		38	7.00	8.00	8.00	
	CC	9	0.00	7.00	8.00		9	8.00	8.00	8.00	
	AC+CC	47	3.00	7.00	8.00	0.587	47	7.00	8.00	8.00	0.115
rs1256120	TT	89	3.00	6.75	8.00	0.813	89	6.80	8.00	8.00	0.310
	TC	15	5.50	7.00	7.90		15	7.00	7.90	8.00	
	CC	3	4.00	8.00	8.00		3	8.00	8.00	8.00	
	TC+CC	18	5.50	7.00	8.00	0.517	18	7.60	8.00	8.00	0.951
rs10146204	GG	40	0.00	6.63	7.29	0.021	40	6.25	7.75	8.00	0.288
	GA	51	5.50	7.20	8.00		51	7.50	8.00	8.00	
	AA	18	0.00	5.00	8.00		18	7.00	8.00	8.00	
	GA+AA	69	3.50	7.00	8.00	0.068	69	7.40	8.00	8.00	0.041
rs1256108	TT	28	0.00	6.20	7.63	0.657	28	5.88	8.00	8.00	0.273
	TC	61	5.60	7.00	8.00		61	7.50	8.00	8.00	
	CC	27	0.00	6.75	8.00		27	7.00	8.00	8.00	
	TC+CC	88	3.75	7.00	8.00	0.143	88	7.45	8.00	8.00	0.253

*Jonckheere-Terpstra Test

Table D-13 Association between *ESR2* SNPs and ER-beta IHC expression for subjects with adenocarcinoma of lung

		Cytoplasmic ERβ					Nuclear ERβ				
SNP	Genotype	N	P25	Med	P75	p-value*	N	P25	Med	P75	p-value*
rs8021944	TT	60	3.75	7.00	8.00	0.291	60	7.00	8.00	8.00	0.211
	TG	5	7.00	7.90	8.00		5	8.00	8.00	8.00	
	GG										
	TG+GG	5	7.00	7.90	8.00	0.291	5	8.00	8.00	8.00	0.211
rs968257	AA	18	0.00	6.25	7.90	0.404	18	5.75	7.95	8.00	0.141
	AG	31	5.50	7.00	8.00		31	7.50	8.00	8.00	
	GG	11	4.00	6.80	8.00		11	7.60	8.00	8.00	
	AG+GG	42	5.00	7.00	8.00	0.156	42	7.60	8.00	8.00	0.169
rs1152589	AA	17	5.00	6.80	8.00	0.181	17	7.60	8.00	8.00	0.701
	AT	30	5.50	7.00	8.00		30	7.75	8.00	8.00	
	TT	12	0.00	4.50	7.17		12	5.25	7.33	8.00	
	AT+TT	42	4.00	7.00	8.00	0.830	42	7.00	8.00	8.00	0.499
rs1255998	CC	47	3.50	7.00	8.00	0.211	47	7.00	8.00	8.00	0.431
	CG	18	4.00	6.40	7.00		18	7.50	7.78	8.00	
	GG										
	CG+GG	18	4.00	6.40	7.00	0.211	18	7.50	7.78	8.00	0.431
rs8006145	CC	25	3.00	6.50	7.33	0.306	25	7.00	7.75	8.00	0.266
	CA	30	4.80	7.00	8.00		30	7.60	8.00	8.00	
	AA	5	6.00	7.00	8.00		5	8.00	8.00	8.00	
	CA+AA	35	4.80	7.00	8.00	0.154	35	7.60	8.00	8.00	0.047
rs4986938	GG	19	0.00	7.00	7.33	0.093	19	5.75	7.90	8.00	0.141
	GA	34	5.00	6.90	8.00		34	7.60	8.00	8.00	
	AA	7	6.00	7.00	8.00		7	7.00	8.00	8.00	
	GA+AA	41	5.50	7.00	8.00	0.106	41	7.60	8.00	8.00	0.129
rs1256063	CC	54	4.00	7.00	8.00	0.753	54	7.00	8.00	8.00	0.439
	CT	6	0.00	6.75	8.00		6	5.50	7.88	8.00	
	CT+TT	6	0.00	6.75	8.00	0.753	6	5.50	7.88	8.00	0.439
rs1256061	CC	15	0.00	5.50	7.00	0.496	15	5.50	7.67	8.00	0.651
	CA	36	4.90	7.00	8.00		36	7.30	8.00	8.00	
	AA	9	6.00	7.00	8.00		9	8.00	8.00	8.00	
	CA+AA	45	5.50	7.00	8.00	0.023	45	7.75	8.00	8.00	0.027
rs1952585	TT	45	4.80	7.00	8.00	0.068	45	7.67	8.00	8.00	0.123
	TC	15	3.00	5.60	7.00		15	6.80	7.75	8.00	
	CC										
	TC+CC	15	3.00	5.60	7.00	0.068	15	6.80	7.75	8.00	0.123
rs17766755	GG	20	0.00	6.75	7.17	0.037	20	6.28	7.83	8.00	0.048
	GA	34	4.80	6.90	8.00		34	7.60	8.00	8.00	
	AA	6	7.00	7.50	8.00		6	8.00	8.00	8.00	
	GA+AA	40	5.25	7.00	8.00	0.076	40	7.68	8.00	8.00	0.067
rs1256049	GG	56	4.00	7.00	8.00	0.386	56	7.00	8.00	8.00	0.796
	GA	3	4.00	5.00	7.00		3	7.00	8.00	8.00	
	GA+AA	3	4.00	5.00	7.00	0.386	3	7.00	8.00	8.00	0.796

Table D-13 (continued)

		Cytoplasmic ERβ					Nuclear ERβ				
SNP	Genotype	N	P25	Med	P75	p-value*	N	P25	Med	P75	p-value*
rs8003490	GG	52	4.40	7.00	8.00	0.111	52	7.30	8.00	8.00	0.295
	GA	13	3.00	5.60	7.00		13	6.80	7.75	8.00	
	AA										
	GA+AA	13	3.00	5.60	7.00	0.111	13	6.80	7.75	8.00	0.295
rs12435284	CC	55	4.00	7.00	8.00	0.311	55	7.00	8.00	8.00	0.239
	CT	5	7.00	7.90	8.00		5	8.00	8.00	8.00	
	CT+TT	5	7.00	7.90	8.00	0.311	5	8.00	8.00	8.00	0.239
rs1256036	AA	17	5.00	6.00	8.00	0.466	17	7.60	8.00	8.00	0.375
	AG	34	5.50	7.00	8.00		34	7.50	8.00	8.00	
	GG	9	0.00	3.00	7.00		9	5.50	7.67	8.00	
	AG+GG	43	3.20	7.00	8.00	0.756	43	7.00	8.00	8.00	0.829
rs1887994	GG	53	4.00	7.00	8.00	0.897	53	7.00	8.00	8.00	0.664
	GT	7	5.50	6.00	8.00		7	7.00	7.75	8.00	
	GT+TT	7	5.50	6.00	8.00	0.897	7	7.00	7.75	8.00	0.664
rs3020450	GG	21	4.00	7.00	7.90	0.619	21	7.00	8.00	8.00	0.883
	GA	32	5.25	7.00	8.00		32	7.25	8.00	8.00	
	AA	7	3.00	7.00	8.00		7	7.00	8.00	8.00	
	GA+AA	39	4.00	7.00	8.00	0.500	39	7.00	8.00	8.00	0.436
rs3020449	TT	16	1.50	6.25	7.67	0.398	16	6.38	7.88	8.00	0.218
	TC	34	5.50	7.00	8.00		34	7.50	8.00	8.00	
	CC	10	4.00	6.90	8.00		10	7.60	8.00	8.00	
	TC+CC	44	4.50	7.00	8.00	0.285	44	7.55	8.00	8.00	0.240
rs10137185	CC	53	4.00	7.00	8.00	0.410	53	7.00	8.00	8.00	0.693
	CT	6	6.80	7.00	7.90		6	7.75	7.95	8.00	
	TT	1	8.00	8.00	8.00		1	8.00	8.00	8.00	
	CT+TT	7	6.80	7.00	8.00	0.377	7	7.75	8.00	8.00	0.664
rs3020443	AA	29	4.00	6.50	7.90	0.379	29	7.00	7.90	8.00	0.146
	AC	28	4.00	7.00	8.00		28	7.30	8.00	8.00	
	CC	3	7.00	8.00	8.00		3	8.00	8.00	8.00	
	AC+CC	31	4.00	7.00	8.00	0.241	31	7.60	8.00	8.00	0.091
rs1256120	TT	51	3.20	7.00	8.00	0.673	51	7.00	8.00	8.00	0.238
	TC	7	6.80	7.00	7.90		7	7.60	7.75	8.00	
	CC	2	4.00	6.00	8.00		2	8.00	8.00	8.00	
	TC+CC	9	6.80	7.00	7.90	0.532	9	7.75	7.90	8.00	0.831
rs10146204	GG	20	1.50	6.63	7.17	0.228	20	6.38	7.75	8.00	0.332
	GA	30	5.60	7.00	8.00		30	7.75	8.00	8.00	
	AA	10	4.00	6.00	8.00		10	7.00	8.00	8.00	
	GA+AA	40	4.50	7.00	8.00	0.144	40	7.63	8.00	8.00	0.072
rs1256108	TT	15	0.00	6.00	8.00	0.397	15	5.50	8.00	8.00	0.304
	TC	34	5.50	7.00	8.00		34	7.00	8.00	8.00	
	CC	16	4.00	6.90	8.00		16	7.68	8.00	8.00	
	TC+CC	50	5.00	7.00	8.00	0.179	50	7.50	8.00	8.00	0.272

*Jonckheere-Terpstra Test

Table D-14 Association between *ESR2* SNPs and ER-beta IHC expression for subjects with squamous cell carcinoma of lung

		Cytoplasmic ERβ					Nuclear ERβ				
SNP	Genotype	N	P25	Med	P75	p-value*	N	P25	Med	P75	p-value*
rs8021944	TT	43	3.50	7.00	8.00	0.577	43	7.40	8.00	8.00	0.454
	TG	5	6.00	7.75	8.00		5	8.00	8.00	8.00	
	GG	1	7.33	7.33	7.33		1	8.00	8.00	8.00	
	TG+GG	6	6.00	7.54	8.00	0.566	6	8.00	8.00	8.00	0.464
rs968257	AA	20	2.75	7.00	8.00	0.907	20	6.75	8.00	8.00	0.343
	AG	13	7.00	7.20	7.75		13	7.40	8.00	8.00	
	GG	7	0.00	6.75	8.00		7	8.00	8.00	8.00	
	AG+GG	20	6.38	7.10	7.88	0.701	20	7.70	8.00	8.00	0.443
rs1152589	AA	9	6.00	7.70	8.00	0.406	9	8.00	8.00	8.00	0.093
	AT	19	3.50	7.00	7.40		19	6.50	8.00	8.00	
	TT	11	4.00	7.75	8.00		11	8.00	8.00	8.00	
	AT+TT	30	4.00	7.00	8.00	0.761	30	7.40	8.00	8.00	0.231
rs1255998	CC	42	3.50	7.00	8.00	0.760	42	7.50	8.00	8.00	0.824
	CG	7	7.00	7.25	7.70		7	7.40	8.00	8.00	
	GG										
	CG+GG	7	7.00	7.25	7.70	0.760	7	7.40	8.00	8.00	0.824
rs8006145	CC	25	4.00	7.20	8.00	0.890	25	7.40	8.00	8.00	0.420
	CA	10	7.00	7.00	7.75		10	7.00	8.00	8.00	
	AA	5	6.00	6.75	8.00		5	8.00	8.00	8.00	
	CA+AA	15	6.00	7.00	8.00	0.898	15	8.00	8.00	8.00	0.326
rs4986938	GG	15	6.00	7.25	8.00	0.671	15	7.40	8.00	8.00	0.594
	GA	18	1.50	7.00	8.00		18	7.00	8.00	8.00	
	AA	7	3.50	6.75	8.00		7	8.00	8.00	8.00	
	GA+AA	25	3.50	7.00	8.00	0.670	25	7.50	8.00	8.00	0.890
rs1256063	CC	36	5.00	7.00	8.00	0.664	36	7.55	8.00	8.00	0.330
	CT	4	3.50	7.10	7.48		4	3.70	7.70	8.00	
	CT+TT	4	3.50	7.10	7.48	0.664	4	3.70	7.70	8.00	0.330
rs1256061	CC	12	5.00	7.23	7.88	0.990	12	6.95	8.00	8.00	0.925
	CA	20	6.20	7.00	7.88		20	7.55	8.00	8.00	
	AA	8	1.75	6.88	8.00		8	7.25	8.00	8.00	
	CA+AA	28	4.75	7.00	8.00	0.952	28	7.55	8.00	8.00	0.785
rs1952585	TT	35	4.00	7.00	8.00	0.576	35	7.50	8.00	8.00	0.821
	TC	4	3.60	7.23	7.48		4	6.95	7.70	8.00	
	CC	1	6.75	6.75	6.75		1	8.00	8.00	8.00	
	TC+CC	5	6.75	7.20	7.25	0.604	5	7.40	8.00	8.00	0.781
rs17766755	GG	16	6.00	7.23	7.88	0.495	16	7.50	8.00	8.00	0.974
	GA	19	1.50	7.00	8.00		19	7.00	8.00	8.00	
	AA	4	1.75	4.75	7.00		4	7.25	8.00	8.00	
	GA+AA	23	1.50	7.00	8.00	0.654	23	7.00	8.00	8.00	0.875
rs1256049	GG	38	4.00	7.00	8.00	0.925	38	7.40	8.00	8.00	0.848
	GA	2	7.00	7.20	7.40		2	7.60	7.80	8.00	
	GA+AA	2	7.00	7.20	7.40	0.925	2	7.60	7.80	8.00	0.848

Table D-14 (continued)

		Cytoplasmic ERβ					Nuclear ERβ				
SNP	Genotype	N	P25	Med	P75	p-value*	N	P25	Med	P75	p-value*
rs8003490	GG	44	5.00	7.13	8.00	0.420	44	7.50	8.00	8.00	0.122
	GA	4	0.00	3.60	7.45		4	6.95	7.45	7.75	
	AA	1	6.75	6.75	6.75		1	8.00	8.00	8.00	
	GA+AA	5	0.00	6.75	7.20	0.201	5	7.40	7.50	8.00	0.331
rs12435284	CC	36	5.00	7.00	7.88	0.855	36	7.45	8.00	8.00	0.911
	CT	4	3.00	7.00	8.00		4	6.75	8.00	8.00	
	CT+TT	4	3.00	7.00	8.00	0.855	4	6.75	8.00	8.00	0.911
rs1256036	AA	9	3.50	6.75	8.00	0.513	9	8.00	8.00	8.00	0.820
	AG	23	6.00	7.00	7.75		23	7.40	8.00	8.00	
	GG	8	5.00	7.63	8.00		8	7.00	8.00	8.00	
	AG+GG	31	6.00	7.00	8.00	0.755	31	7.40	8.00	8.00	0.603
rs1887994	GG	30	6.00	7.10	8.00	0.557	30	7.60	8.00	8.00	0.407
	GT	10	0.00	6.70	8.00		10	6.50	8.00	8.00	
	GT+TT	10	0.00	6.70	8.00	0.557	10	6.50	8.00	8.00	0.407
rs3020450	GG	23	4.00	7.25	8.00	0.658	23	7.50	8.00	8.00	0.090
	GA	11	3.50	7.00	7.75		11	6.50	8.00	8.00	
	AA	6	6.00	7.23	8.00		6	8.00	8.00	8.00	
	GA+AA	17	6.00	7.00	7.75	0.771	17	7.40	8.00	8.00	0.866
rs3020449	TT	17	4.00	7.00	7.75	0.801	17	7.50	8.00	8.00	0.314
	TC	14	6.00	7.10	8.00		14	7.40	8.00	8.00	
	CC	8	3.00	7.23	8.00		8	8.00	8.00	8.00	
	TC+CC	22	6.00	7.10	8.00	0.752	22	8.00	8.00	8.00	0.402
rs10137185	CC	35	4.00	7.00	8.00	0.787	35	7.40	8.00	8.00	0.686
	CT	5	6.00	7.25	8.00		5	8.00	8.00	8.00	
	TT										
	CT+TT	5	6.00	7.25	8.00	0.787	5	8.00	8.00	8.00	0.686
rs3020443	AA	26	4.00	7.00	7.75	0.353	26	7.40	8.00	8.00	0.431
	AC	9	7.00	7.70	8.00		9	8.00	8.00	8.00	
	CC	5	0.00	6.00	8.00		5	8.00	8.00	8.00	
	AC+CC	14	6.00	7.35	8.00	0.445	14	8.00	8.00	8.00	0.431
rs1256120	TT	33	4.00	7.00	8.00	0.828	33	7.40	8.00	8.00	0.556
	TC	6	6.00	7.13	8.00		6	8.00	8.00	8.00	
	CC										
	TC+CC	6	6.00	7.13	8.00	0.828	6	8.00	8.00	8.00	0.556
rs10146204	GG	17	4.00	7.00	7.50	0.320	17	7.00	8.00	8.00	0.653
	GA	14	7.00	7.48	8.00		14	8.00	8.00	8.00	
	AA	9	3.50	6.75	8.00		9	8.00	8.00	8.00	
	GA+AA	23	6.00	7.20	8.00	0.374	23	8.00	8.00	8.00	0.218
rs1256108	TT	13	1.50	6.40	7.75	0.475	13	6.00	8.00	8.00	0.310
	TC	21	7.00	7.25	8.00		21	7.60	8.00	8.00	
	CC	13	3.50	7.33	8.00		13	8.00	8.00	8.00	
	TC+CC	34	6.00	7.25	8.00	0.309	34	7.60	8.00	8.00	0.276

*Jonckheere-Terpstra Test

Table D-15 Crude Odds Ratios for the association between three SNPs and lung cancer characterized by cytoplasmic and nuclear ER-Beta IHC expression status

		Cytoplasmic ERβ					Nuclear ERβ				
SNP	**Genotype**	**N**	**OR**	**95% CI**		**p-value***	**N**	**OR**	**95% CI**		**p-value***
rs8021944	TT	118	1.00				118	1.00			
	TG+GG	15	2.43	0.81	7.29	0.11	15	6.15	0.78	48.52	0.09
rs1256061	CC	33	1.00				33	1.00			
	CA+AA	88	1.39	0.60	3.21	0.45	88	1.95	0.84	4.56	0.12
rs10146204	GG	42	1.00				42	1.00			
	GA+AA	79	1.99	0.89	4.44	0.09	79	1.94	0.87	4.36	0.11

High ER-beta cytoplasmic and nuclear expression defined by subject-specific averaged Allred values above 7.

Odds ratios comparing individuals with high ER-beta cytoplasmic and nuclear expression to those with low expression unless otherwise specified

*Wald Method for Testing Global Null Hypothesis: beta=0 and Wald's Chi-Square Test (p-value) for each stratified level based on analysis of maximmum likelihood estimates

Table D- 16 Crude Odds Ratios for the association between three SNPs and nuclear ER-Beta IHC expression status

		Nuclear ERβ				
SNP	**Genotype**	**N**	**OR**	**95% CI**		**p-value***
rs8021944	TT	118	1.00			
	TG+GG	15	4.78	1.03	22.12	0.05
rs1256061	CC	33	1.00			
	CA+AA	88	2.44	1.08	5.53	0.03
rs10146204	GG	42	1.00			
	GA+AA	79	2.38	1.10	5.13	0.03

High ER-beta nuclear expression defined by subject-specific averaged Allred values equals to 8 which is the median nuclear ER-beta IHC score for my study group.

Odds ratios comparing individuals with high ER-beta nuclear expression to those with low expression

*Wald Method for Testing Global Null Hypothesis: beta=0 and Wald's Chi-Square Test (p-value) for each stratified level based on analysis of maximmum likelihood estimates

Table D-17 Crude Odds Ratios for the association between three SNPs and cytoplasmic and nuclear ER-Beta IHC expression scores among all study subjects (N=135)

	ERβ cytoplasmic expression							ERβ nuclear expression						
	Allred = 0	Allred > 0 AND Allred < 8			Allred = 8			Allred ≤ 6	Allred > 6 AND Allred < 8			Allred = 8		
Genotype	n	n	OR	95% CI	n	OR	95% CI	n	n	OR	95% CI	n	OR	95% CI
rs8021944														
TT	25	60	Ref		33	Ref		19	31	Ref		68	Ref	
TG	1	7	2.92	0.34-25.0	6	4.55	0.51-40.2	1	1	0.61	0.04-10.4	12	3.35	0.41-27.4
GG	0	1			0			0	0			1		
TG+GG	1	8	3.33	0.40-28.0	6	4.55	0.51-40.2	1	1	0.61	0.04-10.4	13	3.63	0.45-29.6
rs1256061														
CC	11	17	Ref		5	Ref		9	9	Ref		15	Ref	
CA	9	37	2.66	0.93-7.61	21	5.13	1.38-19.1	9	14	1.56	0.45-5.41	44	2.93	0.98-8.76
AA	4	8	1.29	0.31-5.35	9	4.95	1.02-24.1	1	5	5.00	0.48-51.8	15	9.00	1.01-80.1
CA+AA	13	45	2.24	0.84-5.96	30	5.08	1.47-17.6	10	19	1.90	0.57-6.31	59	3.54	1.22-10.3
rs10146204														
GG	12	23	Ref		7	Ref		11	11	Ref		20	Ref	
GA	7	29	2.16	0.73-6.37	21	5.14	1.45-18.2	6	12	2.00	0.55-7.25	39	3.58	1.15-11.1
AA	5	10	1.04	0.29-3.76	7	2.40	0.55-10.5	2	5	2.50	0.40-15.7	15	4.13	0.79-21.5
GA+AA	12	39	1.70	0.66-4.39	28	4.00	1.26-12.7	8	17	2.13	0.65-6.95	54	3.71	1.31-10.6

Table D-18 *ESR2* haplotypes and ER-beta IHC expression among only white subjects

		ERβ cytoplasmic expression				ERβ nuclear expression			
		Allred > 0 AND Allred < 8 vs. Allred = 0		Allred = 8 vs. Allred = 0		Allred > 6 AND Allred < 8 vs. Allred ≤ 6		Allred = 8 vs. Allred ≤ 6	
Haplotype weight*	Freq	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
T-C-G	0.44	Ref		Ref		Ref		Ref	
T-A-A	0.25	0.668	0.12-3.83	4.07	0.54-31.0	4.15	0.32-54.0	11.43	1.06-123
T-A-G	0.15	1.46	0.12-18.5	0.92	0.04-21.6	10.19	0.45-233	1.56	0.09-27.2
T-C-A	0.10	0.992	0.08-11.9	0.11	0.00-4.48	26.87	0.61-	1.39	0.06-31.9
G-A-A	0.06	4.857	0.05-454	40.98	0.37-	1.066	0.00-479	28.42	0.34-

*Haplotype is composed of alleles in the order of rs8021944, rs1256061, and rs10146204.

D.4 SURVIVAL ANALYSIS OF LUNG CANCER PATIENTS WITH RARE VARIANT ALLELE OF *ESR2* GENE

Table D-19 Hazard ratios of the rare variant alleles of *ESR2* gene for the overall survival among lung cancer patients

SNP	Reference (common homozygous)	Genotype	HR (95% CI)	p-value*
rs8021944	TT	TG+GG	0.91 (0.47, 1.77)	0.79
rs968257	AA	AG+GG	1.23 (0.78, 1.95)	0.38
rs1152589	AA	AT+TT	1.21 (0.71, 2.05)	0.49
rs1255998	CC	CG+GG	1.05 (0.65, 1.69)	0.86
rs8006145	CC	CA+AA	1.17 (0.75, 1.82)	0.48
rs4986938	GG	GA+AA	1.31 (0.83, 2.08)	0.25
rs1256063	CC	CT+TT	1.16 (0.55, 2.43)	0.69
rs1256061	CC	CA+AA	1.16 (0.71, 1.92)	0.55
rs1952585	TT	TC+CC	0.78 (0.44, 1.37)	0.39
rs17766755	GG	GA+AA	1.20 (0.76, 1.89)	0.44
rs1256049	GG	GA+AA	0.93 (0.40, 2.15)	0.87
rs8003490	GG	GA+AA	0.89 (0.50, 1.58)	0.69
rs12435284	CC	CT+TT	0.92 (0.44, 1.91)	0.82
rs1256036	AA	AG+GG	1.26 (0.76, 2.09)	0.37
rs1887994	GG	GT+TT	1.06 (0.58, 1.92)	0.86
rs3020450	GG	GA+AA	0.99 (0.64, 1.53)	0.98
rs3020449	TT	TC+CC	0.97 (0.61, 1.54)	0.89
rs10137185	CC	CT+TT	0.88 (0.45, 1.71)	0.71
rs3020443	AA	AC+CC	1.16 (0.74, 1.80)	0.53
rs1256120	TT	TC+CC	0.93 (0.51, 1.69)	0.82
rs10146204	GG	GA+AA	1.07 (0.68, 1.70)	0.76
rs1256108	TT	TC+CC	1.17 (0.70, 1.96)	0.55

Table D-20 Hazard ratios of the rare *ESR2* genotypes for the overall survival among lung cancer patients

SNP	Reference (common homozygous)	Genotype	HR (95% CI)	p-value*
rs8021944	TT	TG	0.96 (0.48, 1.91)	0.9
		GG	0.65 (0.09, 4.72)	0.67
rs968257	AA	AG	1.44 (0.88, 2.36)	0.14
		GG	0.87 (0.45, 1.68)	0.69
rs1152589	AA	AT	1.35 (0.77, 2.36)	0.3
		TT	0.97 (0.50, 1.89)	0.93
rs1255998	CC	CG	1.01 (0.62, 1.65)	0.96
		GG	3.49 (0.47, 25.66)	0.22
rs8006145	CC	CA	1.32 (0.83, 2.09)	0.24
		AA	0.73 (0.31, 1.72)	0.47
rs4986938	GG	GA	1.48 (0.92, 2.39)	0.11
		AA	0.87 (0.41, 1.83)	0.71
rs1256063	CC	CT	1.16 (0.55, 2.43)	0.69
rs1256061	CC	CA	1.22 (0.72, 2.04)	0.46
		AA	1.01 (0.50, 2.02)	0.98
rs1952585	TT	TC	0.73 (0.41, 1.30)	0.29
		CC	12.74 (1.59, 101.96)*	0.02
rs17766755	GG	GA	1.36 (0.85, 2.18)	0.2
		AA	0.66 (0.28, 1.59)	0.36
rs1256049	GG	GA	0.93 (0.40, 2.15)	0.87
rs8003490	GG	GA	0.83 (0.46, 1.50)	0.54
		AA	11.65 (1.49, 91.03)*	0.02
rs12435284	CC	CT	0.92 (0.44, 1.91)	0.82
rs1256036	AA	AG	1.32 (0.78, 2.24)	0.31
		GG	1.10 (0.56, 2.18)	0.78
rs1887994	GG	GT	1.06 (0.58, 1.92)	0.86
rs3020450	GG	GA	0.98 (0.62, 1.57)	0.95
		AA	1.02 (0.52, 2.00)	0.95
rs3020449	TT	TC	0.98 (0.60, 1.60)	0.93
		CC	0.93 (0.48, 1.80)	0.83
rs10137185	CC	CT	0.80 (0.40, 1.60)	0.52
		TT	15.13 (1.86, 123.09)*	0.01
rs3020443	AA	AC	1.26 (0.79, 2.01)	0.33
		CC	0.78 (0.31, 1.96)	0.59
rs1256120	TT	TC	0.78 (0.40, 1.52)	0.47
		CC	2.54 (0.79, 8.18)	0.12
rs10146204	GG	GA	1.12 (0.69, 1.82)	0.65
		AA	0.96 (0.50, 1.85)	0.9
rs1256108	TT	TC	1.17 (0.68, 2.01)	0.58
		CC	1.17 (0.63, 2.17)	0.61

BIBLIOGRAPHY

1. American Cancer Society. Cancer Facts & Figures 2007. Atlanta: American Cancer Society, Inc., 2007.
2. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer Statistics, 2007. CA Cancer J Clin 2007;57(1):43-66.
3. Patel JD, Bach PB, Kris MG. Lung cancer in US women: a contemporary epidemic. Jama 2004;291(14):1763-8.
4. Belani CP, Marts S, Schiller J, Socinski MA. Women and lung cancer: epidemiology, tumor biology, and emerging trends in clinical research. Lung Cancer 2007;55(1):15-23.
5. Siegfried JM, Gubish CT, Rothstein ME, Queiroz de Oliveira PE, Stabile LP. Signaling pathways involved in cyclooxygenase-2 induction by hepatocyte growth factor in non small-cell lung cancer. Mol Pharmacol 2007;72(3):769-79.
6. Masuya D, Huang C, Liu D, Nakashima T, Kameyama K, Haba R, Ueno M, Yokomise H. The tumour-stromal interaction between intratumoral c-Met and stromal hepatocyte growth factor associated with tumour growth and prognosis in non-small-cell lung cancer patients. Br J Cancer 2004;90(8):1555-62.
7. Chow KC. The pulmonary source of hepatocyte growth factor in non-small cell lung cancer. Am J Respir Cell Mol Biol 2007;36(1):131-2; discussion 132.
8. Olivero M, Rizzo M, Madeddu R, Casadio C, Pennacchietti S, Nicotra MR, Prat M, Maggi G, Arena N, Natali PG, Comoglio PM, Di Renzo MF. Overexpression and activation of hepatocyte growth factor/scatter factor in human non-small-cell lung carcinomas. Br J Cancer 1996;74(12):1862-8.
9. Edakuni G, Sasatomi E, Satoh T, Tokunaga O, Miyazaki K. Expression of the hepatocyte growth factor/c-Met pathway is increased at the cancer front in breast carcinoma. Pathol Int 2001;51(3):172-8.
10. Kurimoto S, Moriyama N, Horie S, Sakai M, Kameyama S, Akimoto Y, Hirano H, Kawabe K. Co-expression of hepatocyte growth factor and its receptor in human prostate cancer. Histochem J 1998;30(1):27-32.
11. Ichimura E, Maeshima A, Nakajima T, Nakamura T. Expression of c-met/HGF receptor in human non-small cell lung carcinomas in vitro and in vivo and its prognostic significance. Jpn J Cancer Res 1996;87(10):1063-9.
12. Takanami I, Tanana F, Hashizume T, Kikuchi K, Yamamoto Y, Yamamoto T, Kodaira S. Hepatocyte growth factor and c-Met/hepatocyte growth factor receptor in pulmonary adenocarcinomas: an evaluation of their expression as prognostic markers. Oncology 1996;53(5):392-7.

13. Siegfried JM, Weissfeld LA, Singh-Kaw P, Weyant RJ, Testa JR, Landreneau RJ. Association of immunoreactive hepatocyte growth factor with poor survival in resectable non-small cell lung cancer. Cancer Res 1997;57(3):433-9.
14. Chen JT, Lin TS, Chow KC, Huang HH, Chiou SH, Chiang SF, Chen HC, Chuang TL, Lin TY, Chen CY. Cigarette smoking induces overexpression of hepatocyte growth factor in type II pneumocytes and lung cancer cells. Am J Respir Cell Mol Biol 2006;34(3):264-73.
15. Siegfried JM, Luketich JD, Stabile LP, Christie N, Land SR, Siegfried JM, Luketich JD, Stabile LP, Christie N, Land SR. Elevated hepatocyte growth factor level correlates with poor outcome in early-stage and late-stage adenocarcinoma of the lung. Chest 2004;125(5 Suppl):116S-9S.
16. Nakamura T, Nishizawa T, Hagiya M, Seki T, Shimonishi M, Sugimura A, Tashiro K, Shimizu S. Molecular cloning and expression of human hepatocyte growth factor. Nature 1989;342(6248):440-3.
17. Gherardi E, Stoker M. Hepatocytes and scatter factor. Nature 1990;346(6281):228.
18. Siegfried JM, Weissfeld LA, Luketich JD, Weyant RJ, Gubish CT, Landreneau RJ. The clinical significance of hepatocyte growth factor for non-small cell lung cancer. Ann Thorac Surg 1998;66(6):1915-8.
19. Nakamura Y, Niki T, Goto A, Morikawa T, Miyazawa K, Nakajima J, Fukayama M. c-Met activation in lung adenocarcinoma tissues: an immunohistochemical analysis. Cancer Sci 2007;98(7):1006-13.
20. Kuiper GG, Enmark E, Pelto-Huikko M, Nilsson S, Gustafsson JA. Cloning of a novel receptor expressed in rat prostate and ovary. Proc Natl Acad Sci U S A 1996;93(12):5925-30.
21. Kuiper GG, Gustafsson JA. The novel estrogen receptor-beta subtype: potential role in the cell- and promoter-specific actions of estrogens and anti-estrogens. FEBS Lett 1997;410(1):87-90.
22. Ogawa S, Inoue S, Watanabe T, Hiroi H, Orimo A, Hosoi T, Ouchi Y, Muramatsu M. The complete primary structure of human estrogen receptor beta (hER beta) and its heterodimerization with ER alpha in vivo and in vitro. Biochem Biophys Res Commun 1998;243(1):122-6.
23. Gronemeyer H, Laudet V. Transcription factors 3: nuclear receptors. Protein Profile 1995;2(11):1173-308.
24. Zhao C, Dahlman-Wright K, Gustafsson JA. Estrogen receptor beta: an overview and update. Nucl Recept Signal 2008;6:e003.
25. Gruber CJ, Tschugguel W, Schneeberger C, Huber JC. Production and actions of estrogens. N Engl J Med 2002;346(5):340-52.
26. Moore JT, McKee DD, Slentz-Kesler K, Moore LB, Jones SA, Horne EL, Su JL, Kliewer SA, Lehmann JM, Willson TM. Cloning and characterization of human estrogen receptor beta isoforms. Biochem Biophys Res Commun 1998;247(1):75-8.
27. Ogawa S, Inoue S, Watanabe T, Orimo A, Hosoi T, Ouchi Y, Muramatsu M. Molecular cloning and characterization of human estrogen receptor betacx: a potential inhibitor ofestrogen action in human. Nucleic Acids Res 1998;26(15):3505-12.
28. Enmark E, Pelto-Huikko M, Grandien K, Lagercrantz S, Lagercrantz J, Fried G, Nordenskjold M, Gustafsson JA. Human estrogen receptor beta-gene structure,

chromosomal localization, and expression pattern. J Clin Endocrinol Metab 1997;82(12):4258-65.
29. Zhao C D-WK, Gustafsson JA . *ESR2* (Estrogen Receptor 2 (ER beta)). Atlas Genet Cytogenet Oncol Haematol.
URL : http://AtlasGeneticsOncology.org/Genes/*ESR2*ID40500ch14q23.html
April 2008.
30. Taylor AH, Al-Azzawi F. Immunolocalisation of oestrogen receptor beta in human tissues. J Mol Endocrinol 2000;24(1):145-55.
31. Omoto Y, Kobayashi Y, Nishida K, Tsuchiya E, Eguchi H, Nakagawa K, Ishikawa Y, Yamori T, Iwase H, Fujii Y, Warner M, Gustafsson JA, Hayashi SI. Expression, function, and clinical implications of the estrogen receptor beta in human lung cancers. Biochem Biophys Res Commun 2001;285(2):340-7.
32. Fasco MJ, Hurteau GJ, Spivack SD. Gender-dependent expression of alpha and beta estrogen receptors in human nontumor and tumor lung tissue. Mol Cell Endocrinol 2002;188(1-2):125-40.
33. Mollerup S, Jorgensen K, Berge G, Haugen A. Expression of estrogen receptors alpha and beta in human lung tissue and cell lines. Lung Cancer 2002;37(2):153-9.
34. Stabile LP, Davis AL, Gubish CT, Hopkins TM, Luketich JD, Christie N, Finkelstein S, Siegfried JM. Human non-small cell lung tumors and cells derived from normal lung express both estrogen receptor alpha and beta and show biological responses to estrogen. Cancer Res 2002;62(7):2141-50.
35. Schwartz AG, Prysak GM, Murphy V, Lonardo F, Pass H, Schwartz J, Brooks S. Nuclear estrogen receptor beta in lung cancer: expression and survival differences by sex. Clin Cancer Res 2005;11(20):7280-7.
36. Wu CT, Chang YL, Shih JY, Lee YC. The significance of estrogen receptor beta in 301 surgically treated non-small cell lung cancers. J Thorac Cardiovasc Surg 2005;130(4):979-86.
37. Kawai H, Ishii A, Washiya K, Konno T, Kon H, Yamaya C, Ono I, Minamiya Y, Ogawa J. Estrogen receptor alpha and beta are prognostic factors in non-small cell lung cancer. Clin Cancer Res 2005;11(14):5084-9.
38. Slattery ML, Sweeney C, Murtaugh M, Ma KN, Wolff RK, Potter JD, Caan BJ, Samowitz W. Associations between ERalpha, ERbeta, and AR genotypes and colon and rectal cancer. Cancer Epidemiology, Biomarkers & Prevention 2005;14(12):2936-42.
39. Setiawan VW, Hankinson SE, Colditz GA, Hunter DJ, De Vivo I. Estrogen receptor beta (*ESR2*) polymorphisms and endometrial cancer (United States). Cancer Causes & Control 2004;15(6):627-33.
40. Leigh Pearce C, Near AM, Butler JL, Van Den Berg D, Bretsky P, Conti DV, Stram DO, Pike MC, Hirschhorn JN, Wu AH. Comprehensive evaluation of *ESR2* variation and ovarian cancer risk. Cancer Epidemiology, Biomarkers & Prevention 2008;17(2):393-6.
41. Aschim EL, Giwercman A, Stahl O, Eberhard J, Cwikiel M, Nordenskjold A, Haugen TB, Grotmol T, Giwercman YL. The RsaI polymorphism in the estrogen receptor-beta gene is associated with male infertility. Journal of Clinical Endocrinology & Metabolism 2005;90(9):5343-8.
42. Sun Y-h, Yang B, Wang X-h, Xu C-l, Gao X-f, Gao X, Wang L-h. [Association between single-nucleotide polymorphisms in estrogen receptor beta gene and risk of prostate cancer]. Chung-Hua Wai Ko Tsa Chih [Chinese Journal of Surgery] 2005;43(14):948-51.

43. Chen Y-C, Kraft P, Bretsky P, Ketkar S, Hunter DJ, Albanes D, Altshuler D, Andriole G, Berg CD, Boeing H, Burtt N, Bueno-de-Mesquita B, Cann H, Canzian F, Chanock S, Dunning A, Feigelson HS, Freedman M, Gaziano JM, Giovannucci E, Sanchez M-J, Haiman CA, Hallmans G, Hayes RB, Henderson BE, Hirschhorn J, Kaaks R, Key TJ, Kolonel LN, LeMarchand L, Ma J, Overvad K, Palli D, Pharaoh P, Pike M, Riboli E, Rodriguez C, Setiawan VW, Stampfer M, Stram DO, Thomas G, Thun MJ, Travis RC, Virtamo J, Trichopoulou A, Wacholder S, Weinstein SJ. Sequence variants of estrogen receptor beta and risk of prostate cancer in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium. Cancer Epidemiology, Biomarkers & Prevention 2007;16(10):1973-81.
44. McIntyre MH, Kantoff PW, Stampfer MJ, Mucci LA, Parslow D, Li H, Gaziano JM, Abe M, Ma J. Prostate cancer risk and ESR1 TA, *ESR2* CA repeat polymorphisms. Cancer Epidemiology, Biomarkers & Prevention 2007;16(11):2233-6.
45. Forsti A, Zhao C, Israelsson E, Dahlman-Wright K, Gustafsson J-A, Hemminki K. Polymorphisms in the estrogen receptor beta gene and risk of breast cancer: no association. Breast Cancer Research & Treatment 2003;79(3):409-13.
46. Zheng SL, Zheng W, Chang B-l, Shu X-O, Cai Q, Yu H, Dai Q, Xu J, Gao Y-T. Joint effect of estrogen receptor beta sequence variants and endogenous estrogen exposure on breast cancer risk in Chinese women. Cancer Research 2003;63(22):7624-9.
47. Gold B, Kalush F, Bergeron J, Scott K, Mitra N, Wilson K, Ellis N, Huang H, Chen M, Lippert R, Halldorsson BV, Woodworth B, White T, Clark AG, Parl FF, Broder S, Dean M, Offit K. Estrogen receptor genotypes and haplotypes associated with breast cancer risk. Cancer Research 2004;64(24):8891-900.
48. Maguire P, Margolin S, Skoglund J, Sun X-F, Gustafsson J-A, Borresen-Dale A-L, Lindblom A. Estrogen receptor beta (*ESR2*) polymorphisms in familial and sporadic breast cancer. Breast Cancer Research & Treatment 2005;94(2):145-52.
49. Iobagiu C, Lambert C, Normand M, Genin C. Microsatellite profile in hormonal receptor genes associated with breast cancer. Breast Cancer Research & Treatment 2006;95(2):153-9.
50. Georgopoulos NA, Adonakis GL, Fotopoulos A, Koika V, Spinos N, Saltamavros A, Keramopoulos A, Koukouras D, Decavalas G, Kourounis GS. Estrogen receptor polymorphisms in tamoxifen-treated women with breast cancer. Gynecological Endocrinology 2006;22(4):185-9.
51. Gallicchio L, Berndt SI, McSorley MA, Newschaffer CJ, Thuita LW, Argani P, Hoffman SC, Helzlsouer KJ. Polymorphisms in estrogen-metabolizing and estrogen receptor genes and the risk of developing breast cancer among a cohort of women with benign breast disease. BMC Cancer 2006;6:173.
52. Tsezou A, Tzetis M, Gennatas C, Giannatou E, Pampanos A, Malamis G, Kanavakis E, Kitsiou S. Association of repeat polymorphisms in the estrogen receptors alpha, beta (ESR1, *ESR2*) and androgen receptor (AR) genes with the occurrence of breast cancer. Breast 2008;17(2):159-66.
53. Breast and Prostate Cancer Cohort Consortium, Cox DG, Bretsky P, Kraft P, Pharoah P, Albanes D, Altshuler D, Amiano P, Berglund G, Boeing H, Buring J, Burtt N, Calle EE, Canzian F, Chanock S, Clavel-Chapelon F, Colditz GA, Feigelson HS, Haiman CA, Hankinson SE, Hirschhorn J, Henderson BE, Hoover R, Hunter DJ, Kaaks R, Kolonel L, LeMarchand L, Lund E, Palli D, Peeters PHM, Pike MC, Riboli E, Stram DO, Thun M,

Tjonneland A, Travis RC, Trichopoulos D, Yeager M. Haplotypes of the estrogen receptor beta gene and breast cancer risk. International Journal of Cancer 2008;122(2):387-92.

54. Thellenberg-Karlsson C, Lindstrom S, Malmer B, Wiklund F, Augustsson-Balter K, Adami H-O, Stattin P, Nilsson M, Dahlman-Wright K, Gustafsson J-A, Gronberg H. Estrogen receptor beta polymorphism is associated with prostate cancer risk. Clinical Cancer Research 2006;12(6):1936-41.
55. Ichikawa S, Koller DL, Peacock M, Johnson ML, Lai D, Hui SL, Johnston CC, Foroud TM, Econs MJ. Polymorphisms in the estrogen receptor beta (*ESR2*) gene are associated with bone mineral density in Caucasian men and women. J Clin Endocrinol Metab 2005;90(11):5921-7.
56. Greendale GA, Chu J, Ferrell R, Randolph JF, Jr., Johnston JM, Sowers MR. The association of bone mineral density with estrogen receptor gene polymorphisms. Am J Med 2006;119(9 Suppl 1):S79-86.
57. Wang JT, Guo Y, Yang TL, Xu XH, Dong SS, Li M, Li TQ, Chen Y, Deng HW. Polymorphisms in the estrogen receptor genes are associated with hip fractures inChinese. Bone 2008.
58. Domingues-Montanari S, Subirana I, Tomas M, Marrugat J, Senti M. Association between *ESR2* genetic variants and risk of myocardial infarction. Clin Chem 2008;54(7):1183-9.
59. Sowers MR, Symons JP, Jannausch ML, Chu J, Kardia SR. Sex steroid hormone polymorphisms, high-density lipoprotein cholesterol, and apolipoprotein A-1 from the Study of Women's Health Across the Nation (SWAN). Am J Med 2006;119(9 Suppl 1):S61-8.
60. Pirskanen M, Hiltunen M, Mannermaa A, Helisalmi S, Lehtovirta M, Hanninen T, Soininen H. Estrogen receptor beta gene variants are associated with increased risk of Alzheimer's disease in women. Eur J Hum Genet 2005;13(9):1000-6.
61. Nicolaiew N, Cancel-Tassin G, Azzouzi AR, Grand BL, Mangin P, Cormier L, Fournier G, Giordanella J-P, Pouchard M, Escary J-L, Valeri A, Cussenot O. Association between estrogen and androgen receptor genes and prostate cancer risk. European Journal of Endocrinology 2009;160(1):101-6.