Examining the Role of Genetics in Lung Cancer Survival and Risk in African Americans

By

Carissa Christine Jones

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in

Human Genetics
August 11, 2017
Nashville, Tennessee

Approved:

Melinda C. Aldrich, PhD, MPH
Digna R. Velez Edwards, PhD, MS
Scott M. Williams, PhD
Jeffrey Blume, PhD
Eric L. Grogan, MD, MPH

To my husband, Joshua, for your unending love and support.

ACKNOWLEDGEMENTS

The work presented in this dissertation would not have been possible with the financial support provided by the Training Program on Genetic Variation and Human Phenotypes, a NIH Training Grant awarded to Vanderbilt University (5T32GM080178). This work was also supported by a Department of Defense Early Investigator Synergistic Idea Award granted to M.C. Aldrich (W81XWH-12-1-0547) and E.L. Grogan (W81XWH-12-1-0544).

Much of the data utilized in this dissertation came from the Southern Community Cohort Study (SCCS). Data on SCCS cancer cases were provided by the Alabama Statewide Cancer Registry; Kentucky Cancer Registry, Lexington, KY; Tennessee Department of Health, Office of Cancer Surveillance; Florida Cancer Data System; North Carolina Central Cancer Registry, North Carolina Division of Public Health; Georgia Comprehensive Cancer Registry; Louisiana Tumor Registry; Mississippi Cancer Registry; South Carolina Central Cancer Registry; Virginia Department of Health, Virginia Cancer Registry; Arkansas Department of Health, Cancer Registry, 4815 W. Markham, Little Rock, AR 72205. The Arkansas Central Cancer Registry is fully funded by a grant from National Program of Cancer Registries, Centers for Disease Control and Prevention (CDC). Data on SCCS cancer cases from Mississippi were collected by the Mississippi Cancer Registry, which participates in the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention (CDC). The contents of this publication are solely the responsibility of the authors and do not represent the official views of the CDC or the Mississippi Cancer Registry.

During my graduate school career, I was fortunate to work alongside many individuals whose help and guidance was paramount to my success. First and foremost, I would like to express my deepest gratitude to my mentor, Melinda Aldrich, for her support during my graduate

school career. “Thank you” is not nearly enough to describe my appreciation for how she has made me a better writer, communicator, and scientist. To the members of my committee, Digna Velez-Edwards, Scott Williams, Jeffrey Blume, and Eric Grogan: thank you for your support, advice, and constructive criticism, which has undoubtedly strengthened my dissertation. I would also like to extend my gratitude to every member of the TREAT lung cancer program for the feedback provided for every poster, presentation, paper, or idea I brought before them, and to every faculty member who ever let me just stop by to ask a question or discuss a problem.

While graduate school can be a stressful and overwhelming few years, I am thankful for the friendships made during the process. Specifically to Erin Breland, Erica Shannon, and Miranda Sowder: thank you for keeping me sane, being a fantastic support system, and not letting me move to Ohio. I would also be entirely remiss to pass up the opportunity to thank my peers in the Human Genetics program, who made what could have been a difficult few years much more entertaining. I thank them for simultaneously being friends, mentors, and, occasionally, counselors.

Finally, I am most certain that I would not be where I am today without the endless love and support from my family. To my parents, Dan and Karen, whose encouragement got me here, my husband, Joshua, whose confidence in me gave me perseverance, and my in-laws, Mike and Jennie, whose company made my time outside of graduate school so much more enjoyable: words cannot adequately express the love and gratitude I have for all of you. Thank you.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Epidemiology of lung cancer

Incidence

Cancer is a major public health burden worldwide, with 14.1 million new cases diagnosed in 2012 and is expected to increase in number to an estimated 20 million new diagnoses yearly by 2030 (Stewart et al. 2014). Globally, lung cancer accounts for approximately 13% of all cancer cases (Ferlay et al. 2010, Stewart et al. 2014). In females alone, lung cancer is the third leading cause of cancer worldwide (8.7% of cancer cases) behind breast and colorectal cancers (25.2% and 9.2%, respectively), while it remains the leading cause of cancer in men (16.7%), followed closely by prostate cancer (15.0%) (Stewart et al. 2014).

In the United States, lung cancer is responsible for 13% of all new cancer cases (Figure 1), second only to prostate cancer incidence in men and breast cancer in women (American Cancer Society 2017). Cumulatively, men and women have a 6-7% probability of being diagnosed with lung cancer during their lifetime, with greater risk at increasing ages (American Cancer Society 2017). Specifically, 16.4 in 100,000 individuals under the age of 65 are diagnosed with lung cancer, compared to 327.9 in 100,000 individuals above 65 years of age (Howlader N). The majority (57%) of lung cancer cases are diagnosed at distant stage disease; regional and localized disease diagnoses account for 22% and 16% of cases, respectively (Figure 2) (Siegel et al. 2017).

Lung cancer incidence rates also differ by race ethnicity, with more black/African American cases occurring than any other race/ethnicity (Howlader N , American Cancer Society 2017). Specifically, the North American Association of Central Cancer Registries reported 77.7 in 100,000 white male cases in 2009-2013 verses 90.8 in 100,000 black male cases during the same time (American Cancer Society 2017). The number of female lung cancer cases is slightly higher in whites compared to blacks, with 58.2 and 51.0 in 100,000 lung cancer cases in 2009-2013, respectively (Howlader N , American Cancer Society 2017). While overall incidence rates are similar for black and white individuals (males and females combined) over the age of 65 (341.0 vs. 339.9 per 100,000, respectively), blacks, particularly black males, are diagnosed at younger ages compared to whites (Figure 3) (Howlader N). Furthermore, blacks are more commonly diagnosed at distant stage disease (61% vs. 56%) and less commonly diagnosed at localized stage disease (13% vs. 16%) compared to whites (Figure 2) (Howlader N , Siegel et al. 2017).

Surveillance, Epidemiology, and End Results (SEER) Program data show lung cancer incidence rates peaked for all races in 1992, and have been declining ever since (Howlader N). However, it is important to note that incidence rates have peaked separately for males and females, in large part due to differing smoking habits between the sexes, which will be discussed later in this chapter. For black males, incidence rates were highest in 1984 at 158.8 per 100,000 individuals, while rates among black females did not peak until 2002, with 60.6 per 100,000 individuals (Figure 4) (Howlader N). Similarly, incidence rates for white males peaked in 1987 (100.7 per 100,000), nearly 20 years before white females in 2007 (56.0 per 100,000, Figure 4) (Howlader N).

Mortality

Worldwide, cancer accounted for 8.2 million deaths in 2012, with approximately 13% of those deaths attributable to lung cancer, making lung cancer the leading cause of cancer-related death across the world (Ferlay et al. 2010, Stewart et al. 2014). Within the United States, 155,870 lung cancer death are estimated for 2017, which translates to 26.6% of cancer-related deaths in males and 25.2% of cancer-related deaths in females (American Cancer Society 2017). Lung cancer accounts for more deaths than the next three leading cancers combined (Figure 1), which can largely be attributed to the high number of distant stage diagnoses (Howlader N , American Cancer Society 2017). Only 4% of individuals diagnosed with distant stage disease survive 5 years (Howlader N , American Cancer Society 2017). Survival rates increase with earlier stage diagnosis: individuals diagnosed at regional stage disease have a 28% 5-year survival rate and individuals diagnosed with localized disease have 55% 5-year survival rates (Howlader N , American Cancer Society 2017).

While 5-year mortality rates have increased in recent years, blacks have consistently had poorer survival compared to whites (16% vs. 19% 5-year survival rates, respectively). (Howlader N , American Cancer Society 2017). Death rates for black and white males peaked in 1990 (125.2 and 86.0 per 100,000, respectively) and have steadily declined since then (Figure 5) (Howlader N). In 2014, the most recent year of SEER data, male death rates were 62.5 and 52.1 per 100,000 for blacks and whites, respectively (Howlader N). In females, death rates did not begin to decline until nearly 10 years after males, a trend similar to female incidence rates (Figure 5).

Lung cancer histology

The two most frequent histological types of lung cancer are non-small cell and small cell. Non-small cell lung cancer cases can be further divided into multiple subtypes, the most common being adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. The following sections will discuss common characteristics and epidemiologic patterns of the four main lung cancer subtypes.

Non-small cell lung cancer

Non-small cell lung cancer comprises 84% of all lung cancer diagnoses, of which adenocarcinoma accounts for almost half (46.6% of all lung cancer cases) (Howlader N). Squamous cell and large cell carcinomas account for 23.2% and 1.6% of lung cancers, respectively (Table 1). There is relatively little difference in the distribution of histologic subtypes between blacks and whites (Howlader N). There are, however, differences in subtype presentation between males and females: a greater percentage of females (51% compared to 43% of males) are diagnosed with adenocarcinoma, while a greater percentage of males (28% compared to 18% of females) are diagnosed with squamous cell carcinoma (Howlader N). While adenocarcinoma is the most common lung cancer histologic type in men and women, it has only risen to that position in recent years; until the early 1990s squamous cell carcinoma was the most common lung cancer subtype, accounting for nearly 50% of lung cancers in males in 1973 (Devesa et al. 2005, Meza et al. 2015). Interestingly, squamous cell carcinoma is more commonly associated with smoking-related lung cancers, while adenocarcinoma includes more never-smoking lung cancer cases than any other histologic subtype (William D. Travis 2004, Stewart et al. 2014).

Distinct morphologic and histologic characteristics separate non-small cell lung cancer subtypes. Lung adenocarcinomas, for example, are most commonly diagnosed in the peripheral lung, while most squamous cell carcinomas are centralized (though the proportion of periphery tumors is growing) (Stewart et al. 2014). Morphologically, adenocarcinomas are glandular with acinar, papillary, bronchioloalveolar or solid with mucin growth patterns (William D. Travis 2004, Stewart et al. 2014). By contrast, squamous cell carcinomas display intercellular bridges arising from the bronchial epithelium and keratinization (William D. Travis 2004, Stewart et al. 2014). Large cell carcinoma is histologically diagnosed through the exclusion of granular or squamous cell carcinoma (William D. Travis 2004, Stewart et al. 2014). Non-small cell lung cancers also differentiate by genetic mutation profiles.

While most non-small cell lung cancers are diagnosed at late stage disease, there is some variation among subtypes. Specifically, squamous cell carcinoma tends to be diagnosed at earlier disease stages compared to adenocarcinomas and large cell carcinomas (Table 1) (William D. Travis 2004). Five-year relative survival rates for localized, regional, and distant non-small cell lung cancer are 59.5%, 32.3%, and 5.2%, respectively (Howlader N). However, these rates differ between blacks and whites in both sexes and across all stages, with only 17.5% of black non-small cell lung cancer cases (all stages) surviving 5 years compared to 22.6% percent of whites (Howlader N).

Small cell lung cancer

Small cell carcinoma accounts for 15% of all lung cancers and typically occurs in heavy smokers (Howlader N , William D. Travis 2004, Stewart et al. 2014). Morphologically, small cell lung cancer is defined by small round, oval or spindle shaped cells, with scant cytoplasm,

granular chromatin, and absent or inconspicuous nucleoli (William D. Travis 2004, Stewart et al. 2014). Seventy-four percent of small cell lung cancers occur in a centralized location within the lung (William D. Travis 2004).

The relative proportion of small cell lung cancer cases has remained steady over time (Meza et al. 2015). The overall 5-year relative survival rate for small cell lung cancer (all stages, all races) is substantially reduced compared to non-small cell lung cancers (6.5%, vs. 22.1%, respectively) (Howlader N). The majority (53%) of small cell lung cancers are diagnosed at distant stage disease (Table 1), which is associated with a 2.9% 5-year relative survival rate (Howlader N , William D. Travis 2004). Regionally staged disease, which accounts for 26% of small cell lung cancers, and locally staged disease, which accounts for 8% of small cell lung cancers (Table 1), have 5-year survival rates of 15.3% and 28.9%, respectively (Howlader N , William D. Travis 2004). There is little difference between stage distributions between blacks and whites, though a slightly greater proportion of females are diagnosed at regional stage disease compared to males (22% vs. 18%, respectively), which in turn gives females slightly better 5-year survival rates (7.8% vs. 5.2% for all stages combined, respectively) (Howlader N).

Risk Factors

Smoking

Smoking is by far the greatest risk factor for lung cancer, accounting for 80-90% of all lung cancer cases (Centers for Disease Control and Prevention , American Cancer Society 2017). Some of the first evidence for an association between tobacco smoking and lung cancer came in 1950 when Wynder *et al.* examined a population of 684 confirmed lung cancer cases, of which 95% were heavy smokers compared to 75% of controls (Wynder and Graham 1950). Similarly,

Doll and Hill (1950) observed significantly increased smoking rates and quantities among lung cancer cases in central London. However, it wasn't until the release of the Surgeon General's report on smoking and health in 1964 (Wellmann 1964) that widespread tobacco control measures began, resulting in a dramatic decrease in tobacco consumption. In fact, it is estimated that tobacco consumption would be more than 370% higher than it is today in the absence of the surgeon general's report and the resulting tobacco control efforts (Warner et al. 2014). Similarly, Holford et al. estimate that approximately 60% of males and 45% of females would be current smokers today without these counter tobacco efforts (compared to the approximately 25% and 20% actual current smokers today, respectively), and as a result more than 8 million lives have been saved from smoking-related deaths (Holford et al. 2014). With specific regard to lung cancer, approximately 800,000 lung cancer related deaths were avoided due to tobacco control efforts and another 1.7 million could have been avoided had control efforts resulted in complete smoking cessation in 1965 (Moolgavkar et al. 2012).

The increase in tobacco control efforts has undoubtedly contributed to the observed changes in histological prevalence over time. As stated previously, squamous cell carcinomas are predominantly observed in ever smokers, while adenocarcinomas are observed more frequently in never smokers (William D. Travis 2004, Stewart et al. 2014). As tobacco cessation efforts have increased, the frequency of squamous cell carcinoma diagnoses has decreased, while adenocarcinoma diagnoses have increased (Meza et al. 2015).

Historical differences in smoking habits and behaviors between men and women have also influenced lung cancer trends. Examination of SEER data from 1975-2014, reveals that incidence rates among men began declining nearly 20 years earlier than in women (Howlader N , Siegel et al. 2017). Similarly, women were also approximately 20 years delayed in smoking

initiation and began smoking at much older ages compared to men (Harris 1983, Shopland 1995, Moolgavkar et al. 2012). Furthermore, smoking cessation rates have declined much more rapidly among men than among women (Shopland 1995, Moolgavkar et al. 2012).

Occupational and environmental exposures

While smoking is by far the strongest risk factor for lung cancer, occupational and environmental exposures account for a substantial number of lung cancer cases. A list of common exposures is provided in Table 2. Below I will describe a few examples of such exposures.

Ionizing radiation (e.g. gamma-rays, x-rays, and alpha-particles such as radon and plutonium) is a group 1 carcinogen as classified by the International Agency for Research on Cancer (IARC), meaning it is a known human carcinogen (Field and Withers 2012). Exposure to x-rays and gamma-rays occurs primarily in the medical industry, while plutonium exposure occurs in occupations related to nuclear weapon or nuclear power production (William D. Travis 2004, Field and Withers 2012). Radon is perhaps one of the oldest known occupational lung carcinogen and most commonly affects miners, though radon exposure can occur in other environments (William D. Travis 2004, Samet et al. 2009, Field and Withers 2012). Among the first evidence of the association between radon and its decay products was observed in a population of Navajo uranium miners (Samet et al. 1984, Roscoe et al. 1995). Within a non-smoking mining population, the lung cancer mortality ratio was 12.7 (8.0-20.1), indicating a strong association between radon exposure and lung cancer even in the absence of tobacco exposure (Roscoe et al. 1989). A causal link between radon exposure and lung cancer in miners has been confirmed in numerous studies (1999, Schubauer-Berigan et al. 2009, Lane et al. 2010).

Another commonly cited occupational exposure is asbestos, a fibrous silicate material. Asbestos possesses many desirable qualities that made it a product commonly used in industry, such as heat stability, thermal and electrical insulation, and resistance to degradation (National Toxicology Program). At the peak of asbestos consumption, the primary uses for it included cement pipes, flooring, roofing, and automobile brakes and clutches, though production of asbestos-containing products has drastically decreased since it was associated with increased lung cancer risk, particularly among smokers (National Toxicology Program , Selikoff et al. 1968). According to a population-based study of lung cancer risk in Sweden, the relative risk for lung cancer with asbestos exposure is 1.68 (95% Confidence Interval (CI): 1.15-2.46) (Gustavsson et al. 2000). While a nearly two-fold increased risk of death was observed among asbestos textile workers compared to non-asbestos exposed textile workers (Loomis et al. 2009), with similar observations across several studies (Loomis et al. 2010, Elliott et al. 2012, Loomis et al. 2012).

Several of the occupational exposures listed in Table 2 are also classified as environmental exposures. While individuals working in such industries are undoubtedly exposed to higher concentrations of the carcinogenic compound, risk still remains for individuals in the general population as particulates are released into the air. For example, diesel exhaust is classified as an IARC group 1 known carcinogen and is associated with lung cancer risk (Attfield et al. 2012, Garshick et al. 2012, Silverman et al. 2012, Silverman 2017); while truckers, miners, and other related occupations are at higher risk given their repeated exposures (Pronk et al. 2009, Attfield et al. 2012, Silverman et al. 2012), the release of diesel exhaust into the air also acts as an environmental exposure to the general population.

Chronic Obstructive Pulmonary Disease

Chronic obstructive pulmonary disease (COPD) describes a group of lung diseases including emphysema and chronic bronchitis and is diagnosed through measures of lung function such as forced expiratory volume in one second (FEV_1). Like lung cancer, one of the most common risk factors for COPD is tobacco smoking. Approximately 40-70% of lung cancer patients also have COPD and several studies have noted a strong association between lung cancer and COPD (Wasswa-Kintu et al. 2005, Dela Cruz et al. 2011, Houghton 2013). Specifically, a meta-analysis of 39 studies reported a relative risk for lung cancer of 1.80 (95% CI: 1.60-2.11) for individuals with a previous diagnosis of COPD, emphysema, or chronic bronchitis (Brenner et al. 2011) and results from the National Lung Cancer Screening Trial (NLST) show a 2-fold increased risk for lung cancer among individuals with confirmed COPD (Young et al. 2015). Other studies have suggested that a COPD diagnosis may also be associated with certain histologic subtypes (Papi et al. 2004, Young et al. 2015). Despite clear evidence of an association between lung cancer and COPD, the causal relationship between the two diseases remains poorly understood (Houghton 2013). Given the increased risk for both diseases imposed by smoking, several have suggested a connection simply through shared etiologic pathways (Houghton 2013). However, in a Caucasian population of ever smokers, the prevalence of COPD was six times greater among lung cancer cases compared to controls matched on smoking status, age, and sex, suggesting COPD alone is a risk factor for lung cancer (Young et al. 2009).

Genetic contributions to lung cancer

Somatic mutations

Somatic mutations in tumor cells have been heavily studied in lung cancer. Across all cancer types, mutations within *TP53* or the retinoblastoma pathway and loss of heterozygosity at 3p are among the most common, occurring in 50-80% of lung cancers (William D. Travis 2004). However, within lung cancer subtypes, specific mutations can dominate. For example, both non-small cell and small cell lung cancers display mutations in the retinoblastoma pathway. Small cell cancers, however, have mutations that result in loss of Rb protein, while non-small cell lung cancers rarely display such loss, but instead have a high frequency of inactivating *INK4* mutations (Sherr and McCormick 2002, William D. Travis 2004). While *TP53* mutations are frequent in both non-small cell and small cell lung cancers, they are more common among smokers than never smokers (William D. Travis 2004).

A number of other genes have also been shown to influence lung cancer treatment response. Specifically, mutations within the epidermal growth factor receptor gene (*EGFR*) are sensitive to tyrosine kinase (TK) inhibitors such as gefitinib and erlotinib and result in increased progression-free survival rates compared to traditional chemotherapeutics (Lynch et al. 2004, Paez et al. 2004, Pao et al. 2004, Mok et al. 2009). However, some individuals with *EGFR* mutations develop resistance to TK inhibitors as a result of additional *EGFR* mutations, including T790M, and stop responding to treatment (Pao et al. 2005). As a result, next generation TK inhibitors that are sensitive even in the presence of acquired resistance mutations are being tested (Cross et al. 2014). *EGFR* mutations are most commonly observed in lung adenocarcinoma and are observed at higher frequencies in Asian populations compared to European-descent populations (Dearden et al. 2013). Furthermore, *EGFR* mutations are

estimated to be a much as twice as frequent in never/light smokers compared to ever/heavy smokers (Dearden et al. 2013).

Mutations in *KRAS* are also common among lung cancer cases, and together with *EGFR* mutations, are estimated to occur in nearly 50% of all lung cancers (Dearden et al. 2013). Like *EGFR*, *KRAS* mutations are also observed more frequently in adenocarcinoma lung cancers; however, mutations in *KRAS* are more frequent in individuals of European ancestry compared to Asian ancestry. *KRAS* mutations have also been shown to have prognostic value: mutations in this gene are associated with poor response to TK inhibitors and may predict organ-specific metastasis (Pao et al. 2005, Lohinai et al. 2017).

Somatic mutations in several other genes, including *HER2*, *PIK3CA*, *BRAF*, *MEK1*, *ALK*, and *PTEN*, have also been identified in lung cancer and may influence lung cancer outcomes (Cappuzzo et al. 2006, Ludovini et al. 2011, Marchetti et al. 2011, Cui et al. 2014, Lovly et al. 2014, Caunt et al. 2015). In turn, assays capturing such information are being developed for the clinic as a means of driving genetically informed clinical care for the improvement of lung cancer outcomes (Pao and Chmielecki 2010, Su et al. 2011, Levy et al. 2012).

Heritability and Family Studies

While somatic variation inarguably plays a large role in lung cancer development, progression, and treatment, evidence for an inherited component to lung cancer also exists. Studies of lung cancer risk in families show a two- to three-times increased risk with at least one first-degree relative with lung cancer (Tokuhata and Lilienfeld 1963, Ooi et al. 1986, Cannon-Albright et al. 1994, Schwartz et al. 1996, Etzel et al. 2003, Cote et al. 2005, Matakidou et al. 2005, Nitadori et al. 2006, Chen and Kaphingst 2011, Cote et al. 2012). These associations

persist even after adjustment for smoking (Lorenzo Bermejo and Hemminki 2005). Furthermore, a twin study identified a heritability estimate of ~14% (Hemminki et al. 2001), while analysis of heritability in genome-wide association studies reveals as much as 20% heritability for lung cancer (Sampson et al. 2015).

The first family-based linkage study for lung cancer was conducted in 2004 and identified a significant association on chromosome 6q23-25 (Bailey-Wilson et al. 2004). The authors observed that the strength of the association was directly proportional to number of affected individuals within each pedigree (Bailey-Wilson et al. 2004). Fine mapping of the 6q23-25 region localized the signal to the *RGS17* gene, which showed increased expression in sporadic lung tumors (You et al. 2009). Haplotypes within this region were associated with increased risk among never and light smokers, suggesting an interaction between smoking status and inherited risk at this region (Amos et al. 2010).

Genome-wide association studies

The first genome-wide association studies (GWAS) for lung cancer risk were published in 2008 in a series of three papers (Amos et al. 2008, Hung et al. 2008, Thorgeirsson et al. 2008). All three groups reported an association between chromosome 15q25 and increased lung cancer risk in European-descent populations. Thorgeirsson *et al.* also showed an association with smoking quantity and nicotine dependence, suggesting gene-environment interactions in the etiology of lung cancer. Hung *et al.*, on the other hand, observed similarly increased lung cancer risk at this locus even among never smokers and estimated an attributable risk of approximately 14%. The 15q25 locus is a region of strong linkage disequilibrium containing numerous genes, including several nicotinic receptor subunit genes. Several studies have attempted to clarify the

relationship between the 15q25 locus, smoking behaviors, and lung cancer risk, and have subsequently confirmed a direct relationship between genetic variation at 15q25 and lung cancer risk (Spitz et al. 2008, Lips et al. 2010, Saccone et al. 2010, Wang et al. 2010, VanderWeele et al. 2012).

Since the initial lung cancer GWAS a number of additional risk loci have been identified. Perhaps one of the most consistently replicated (other than 15q25, of course) is 5p15.33, which was first reported by McKay *et al.* (2008). The study identified two variants in this region in high linkage disequilibrium, both of which had odds ratios near 1.2 per risk allele (McKay et al. 2008). Numerous studies have reported similar associations (Broderick et al. 2009, Landi et al. 2009, Hsiung et al. 2010, Yoon et al. 2010, Hu et al. 2011, Lan et al. 2012, Li et al. 2013, Zhao et al. 2014). The two genes most commonly implicated in this region are *TERT* (telomerase reverse transcriptase) and *CLPTM1L* (cleft lip and palate transmembrane protein 1 like protein). TERT combines with TERC (telomerase RNA component) to form telomerase, an enzyme responsible for maintaining telomere structure through the addition of TTAGGG repeats to chromosome ends (Feng et al. 1995, Harrington et al. 1997, Lingner et al. 1997). Telomerase is usually absent in somatic tissues, but has been shown to be active in rapidly dividing cancer cells. Telomerase and telomere-related proteins have been heavily implicated in cancer (Shay and Wright 2011, Jafri et al. 2016).

Genome-wide association studies have identified additional lung cancer risk loci on chromosomes 3q28-9, 6p21, 6q22.2, 9p21.3, 10q25.2, 12q13.13, 12q23.1, 13q12.12, 18p11.22, and 22q12.2 (Broderick et al. 2009, Landi et al. 2009, Yoon et al. 2010, Hu et al. 2011, Ahn et al. 2012, Lan et al. 2012, Timofeeva et al. 2012, Dong et al. 2013, Wang et al. 2016). Other studies

have reported increased lung cancer risk for variants associated with smoking-related metabolites (Park et al. 2015, Patel et al. 2016).

Several studies have also sought to identify genetic variants associated with lung cancer survival, with particular emphasis on response to lung cancer treatments. Niu et al. (2012) performed a genome-wide association study for cytotoxicity induced by paclitaxel and docetaxel, two common chemotherapeutics used in lung cancer treatment, in lymphoblastic cell lines. Significant SNPs were then examined in a population of small and non-small cell lung cancer patients. Functional interrogation of the most significant associations identified multiple SNPs associated with survival response, including variation that resulted in desensitization to paclitaxel in cell lines (Niu et al. 2012). Additional genetic variants associated with survival have been identified for both small cell (Wu et al. 2010, Han et al. 2014)and non-small cell lung cancers (Sato et al. 2011, Wu et al. 2011, Hu et al. 2012, Lee et al. 2013, Tang et al. 2015).

Pleiotropy studies of lung cancer

Genetic studies of lung cancer risk have recently turned to investigating pleiotropy, which occurs when a single variant or locus is associated with more than one phenotype (Solovieff et al. 2013). A 2012 study consisting of ~18,000 lung cancer cases and ~60,000 controls of European descent tested 165 SNPs previously associated with at least one other type of cancer for an association with lung cancer risk (Park et al. 2014). Their results confirmed risk loci at 5p15.33 (*TERT* gene) and 9p21 (*CDKN2B-AS1* gene) and identified one novel association: a variant in the *LSP1* gene previously associated with breast cancer risk (Park et al. 2014). A genome-wide association study of lung, gastric, and esophageal cancers found evidence for pleiotropy at chromosomes 6p21.1 and 7p15.2, suggesting these loci are important in

multiple cancer types (Jin et al. 2012). Finally, a large study of lung, ovary, breast, prostate, and colorectal cancer cases and controls, reported a shared risk locus at 1q22 for lung and breast cancers (Fehringer et al. 2016). Furthermore, Fehringer et al. (2016) also confirmed known lung cancer risk loci in the *BRCA2* gene (shared between lung and ovarian cancers) and in the *CDKN2B-AS1* gene (shared between lung and prostate cancers), which was also reported as pleiotropic by Park *et al.*

Genetics of lung cancer in African Americans

Family studies comparing European- and African-descent populations revealed a nearly two-fold increased risk for African American first-degree relatives of lung cancer patients compared to their white counterparts (Cote et al. 2005). Despite evidence for increased risk and poorer survival compared to other racial/ethnic groups, relatively few genetic studies of lung cancer have been conducted in African Americans, and most have focused on candidate genes or confirming associations previously identified in European or Asian descent populations (Spitz et al. 2013, Walsh et al. 2013). Schwartz et al. (2009) confirmed 15q25 as a lung cancer risk locus in African Americans, which persisted even when controlling for smoking status. Walsh *et al.* (2012) similarly confirmed 15q25, but also identified novel risk associations in *CHRNA1* on chromosome 2, while Walsh *et al.* (2013) provided evidence for histology-specific associations for the previously identified 5p15.33 and 6p22.1-21.31 risk loci. A genome-wide admixture mapping study reported associations with lung cancer risk for a region with excess European ancestry on chromosome 1 among female non-small cell lung cancer cases and a region with excess African ancestry on chromosome 3q among ever smoking non-small cell lung cancers (Schwartz et al. 2011). However, the first African American genome-wide association study of

lung cancer was conducted in 2016 (Zanetti et al. 2016). The largest study of African American lung cancer to date, Zanetti *et al.* failed to identify novel lung cancer risk loci, but confirmed associations with 15q25 and 5p15.33. Furthermore, while genetic studies of lung cancer risk in African Americans are few, they are completely lacking for studies of lung cancer survival.

Admixed Populations

Genetic Admixture

The vast majority of genetic epidemiologic studies tend to group individuals by race or race/ethnicity, with the most common classifications in the United States being white (or non-Hispanic white), black (or non-Hispanic black), Asian/Pacific Islander, American Indian/Alaska Native, and Hispanic. Such classifications do an adequate job accounting for differing social and cultural norms observed between populations. However, given recent reports that have estimated as many as 94-98% of individuals today have at least some degree of mixed ancestry, predetermined race classifications poorly represent underlying genetic architecture within each racial group (Shriner et al. 2014, Baker et al. 2017).

Genetic admixture occurs when two previously isolated populations begin mating (Winkler et al. 2010). Multiple generations of random mating and recombination have left modern day descendants with mosaic chromosomes, in which chromosomal segments of one ancestry are interspersed between chromosomal segments of the other ancestry (Figure 6) (Winkler et al. 2010). Within the United States, most individuals, including individuals of European ancestry, exhibit some degree of genetic admixture (Bryc et al. 2015). The admixture event(s) responsible for the African American population is estimated to have occurred in the South prior to the start of the Civil war (Baharian et al. 2016). Admixture estimates for self-

identified African Americans have consistently approximated 80% African ancestry and 20% European ancestry, though these estimates can range from less than 1% to nearly 99% European ancestry (Tishkoff et al. 2009, Bryc et al. 2010). Thus a high degree of genetic diversity exists even within self-identified racial/ethnic groups.

Importance of ancestry in biomedical research

The discovery of ancestry informative markers, genetic variants that have substantially different allele frequencies between populations, has allowed for the estimation of ancestry at the individual level. While there are many methods for estimating ancestry, they can be grouped into two categories: those that estimate global ancestry (Pritchard et al. 2000, Price et al. 2006, Alexander et al. 2009) and those that estimate local ancestry (Pasaniuc et al. 2009, Maples et al. 2013). Local ancestry estimation, which assigns ancestry to each position in the genome, has shown utility through admixture mapping studies to identify genetic variants associated with racial disparities in disease prevalence (Schwartz et al. 2011, Molineros et al. 2013, Velez Edwards et al. 2014, Ruiz-Narvaez et al. 2016).

While global ancestry is most commonly used in controlling for confounding by population substructure in genetic association studies, several studies have shown strong associations between global ancestry estimates and biomedical phenotypes (Signorello et al. 2010, Goetz et al. 2014, Dumitrescu et al. 2015). In a population of African Americans, Kumar *et al.* showed that global African ancestry was inversely associated with lung function through FEV_1 and forced vital capacity (2010). Associations between population structure and lung function were also observed in Mexican populations (Moreno-Estrada et al. 2014). Similarly, Yang et al. showed that increased proportions of Native American ancestry are associated with

acute lymphoblastic leukemia relapse in children (2011). Such associations with genetic ancestry have also been noted for breast cancer (Fejerman et al. 2008, Fejerman et al. 2013) and coronary artery disease (Peralta et al. 2010), to name a few. The role of genetic ancestry in lung cancer survival remains to be examined.

Motivation for the research

Racial disparity in lung cancer

As stated earlier, a racial disparity exists in both lung cancer risk and survival, such that blacks have increased incidence and poorer survival compared to whites (Howlader N , American Cancer Society 2017). Black males are diagnosed at earlier ages and later stages compared to their white counterparts (Howlader N , Efird et al. 2014, Ganti et al. 2014). Furthermore, studies show blacks are less likely to receive stage-appropriate treatments compared to whites (Hardy et al. 2009, Ganti et al. 2014) and have poorer 5-year lung cancer survival rates (Howlader N , Aizer et al. 2014, Zeng et al. 2015, American Cancer Society 2017). However, several studies suggest that the observed racial disparity in lung cancer mortality would diminish with equal access to healthcare between blacks and whites (Zheng et al. 2012, Caposole et al. 2014, Ganti et al. 2014, Hua et al. 2014). Other published findings demonstrate there is no difference in lung cancer mortality rates between blacks and whites when stage and treatment are accounted for (Bach et al. 1999, Aldrich et al. 2013), though these results are in contrast to a recent study of SEER data consisting of >150,00 individuals in which African Americans continued to have poorer survival even after controlling for stage and treatment (Aizer et al. 2014). Thus continued research is necessary to better understand the observed racial disparity in lung cancer survival.

Importance of racially diverse populations in genetic research

The present work is made relevant by the need to include racially diverse populations in biomedical research. A 2016 analysis of genome-wide association studies revealed that only 19% of studies are conducted in individuals of non-European ancestry, of which only 3% were conducted in African-descent populations (Popejoy and Fullerton 2016). This is especially concerning given the recent studies highlighting issues with assuming equivalent genetic risk across racial/ethnic groups. Specifically, Manrai *et al.* examined diagnostic sequence variants for hypertrophic cardiomyopathy, which were originally detected in white populations, and found that benign mutations misdiagnosed as pathogenic were most common among African Americans and resulted in incorrect positive diagnoses (2016). Furthermore, Martin *et al.* recently calculated polygenic risk scores in diverse populations for several well-studied phenotypes and found that risk is most accurately estimated in populations matching the population from which the score was derived (2017). For several traits, calculated polygenic risk scores were in direct opposition to clinical and anthropological evidence (Martin et al. 2017). While non-transferability of genetic risk factors across diverse populations can result for multiple reasons, including differing linkage disequilibrium patterns or allele frequencies, these studies highlight the need for including racially diverse populations such as African Americans in genetic research, or, at the very least, conducting cross-population validation studies to confirm known associations. Furthermore, despite strong evidence that blacks are at increased risk and have poorer outcomes, few lung cancer genetics studies have been conducted in African American populations. We aim to address this dearth of lung cancer research conducted among African Americans

Research aims to be addressed

This dissertation seeks to examine the role of genetics in lung cancer survival and risk in African Americans through three independent studies:

1. Examine the role of African genetic ancestry compared to other known contributors to lung cancer survival in a population of whites and blacks (Chapter 2).
2. Examine variants previously associated with lung cancer survival and identify novel common and rare variants associated with lung cancer survival in an African American population (Chapter 3).
3. Investigate cross-cancer pleiotropic associations for lung cancer risk in African Americans (Chapter 4).

Only by increasing our understanding of how genetic ancestry impacts lung cancer survival compared to known social, environmental, or clinical risk factors, will we be able to adequately address (and ultimately remove) the observed racial disparity. Furthermore, previous studies have revealed the non-transferability of genetic risk factors across racial/ethnic groups (Manrai et al. 2016, Martin et al. 2017); thus, it is of utmost importance to examine genetic risk factors of lung cancer survival and risk directly in African Americans. Such efforts may ultimately lead to better treatments (in the case of survival-associated variants) or improved cancer detection or prevention (in the case of risk-associated variants).

Table 1. Overall frequency of lung carcinoma by histologic subtype and stage at Modified from William D. Travis (2004) and (Howlader N).

Stage	Adenocarcinoma	Squamous Cell	Large Cell	Small Cell
Overall, %	46.6	23.2	1.6	13.1
Localized, %	22.2	21.5	15.2	8.2
Regional, %	33.1	38.5	31.5	26.1
Distant, %	35.9	25.2	40.3	52.8
Unstaged, %	8.8	14.8	12.9	12.8

Table 2. Known lung cancer carcinogens in humans with exposure type and examples of the industry in which the agent is used.

Agent	Exposure Type	Industry/Use**
Ionizing radiation		
Radon	E, O	Mining
Plutonium	O	Nuclear
X- and gamma-rays	E, O	Medical, nuclear
Chemicals and Mixtures		
Bis(chloromethyl)ether; chloromethyl methyl ether	O	Chemical Production
Coal-tar pitch	O	Construction, electrodes
Soot	O	Pigments, chimney sweeps
Sulfur mustard	O	Military
Diesel exhausts	E, O	Mining, trucking, construction
Metals		
Arsenic and arsenic compounds	E, O	Glass, metals, pesticides, textiles
Beryllium and beryllium compounds	O	Aerospace, automotive, biomedical
Cadmium and cadmium compounds	O	Dye/pigment, welding, batteries
Chromium (VI) compounds	O	Metal planting, dye/pigment
Nickel compounds	O	Metallurgy, alloy, catalyst
Dust and fibers		
Asbestos	E, O	Insulation, construction, textiles
Sillica dust, crystalline	E, O	Stone cutting, mining, glass, paper
Occupations		
Aluminum production	O	
Coal gasification	O	
Coke production	O	
Hematite mining	O	
Iron and steel founding	O	
Painting	O	
Rubber production industry	O	

Adapted from (William D. Travis 2004, Field and Withers 2012); E=environmental exposure; O=occupational exposure

*Examples; not intended to be inclusive of all possible uses or industries.

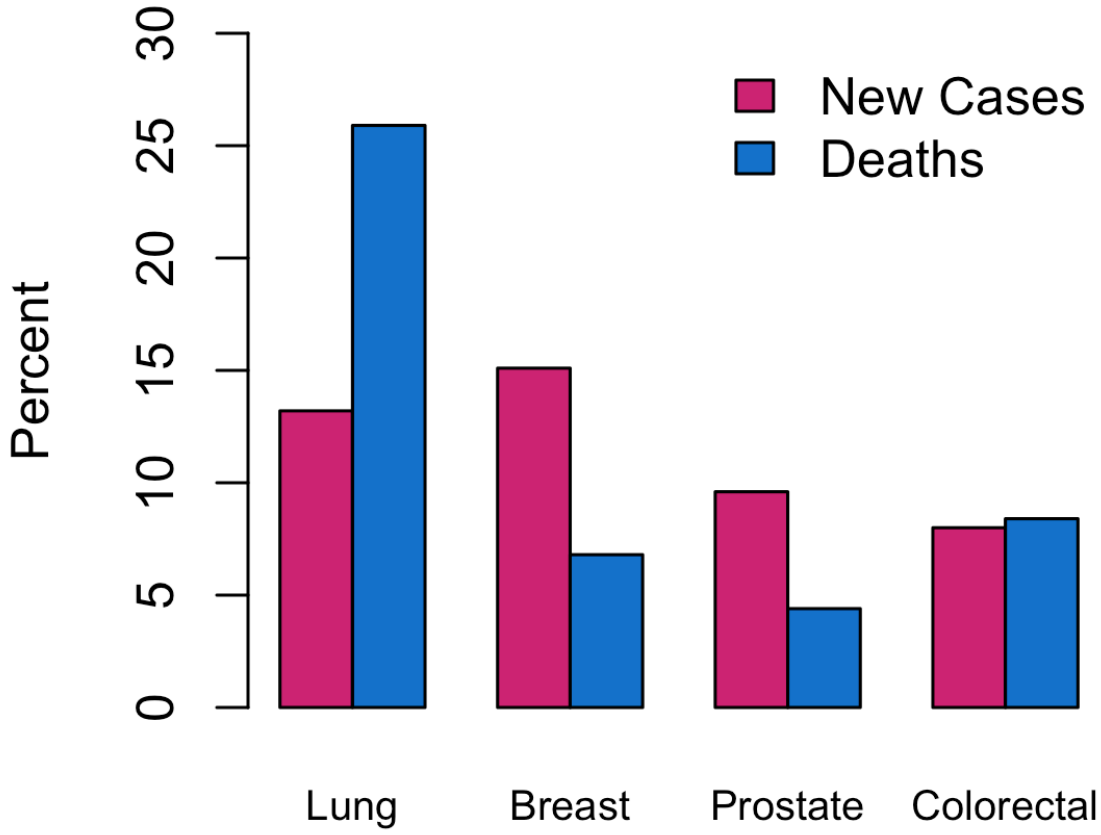

Figure 1. Estimated frequency of new lung cancer cases and deaths by cancer site for 2017.

Data from American Cancer Society (2017).

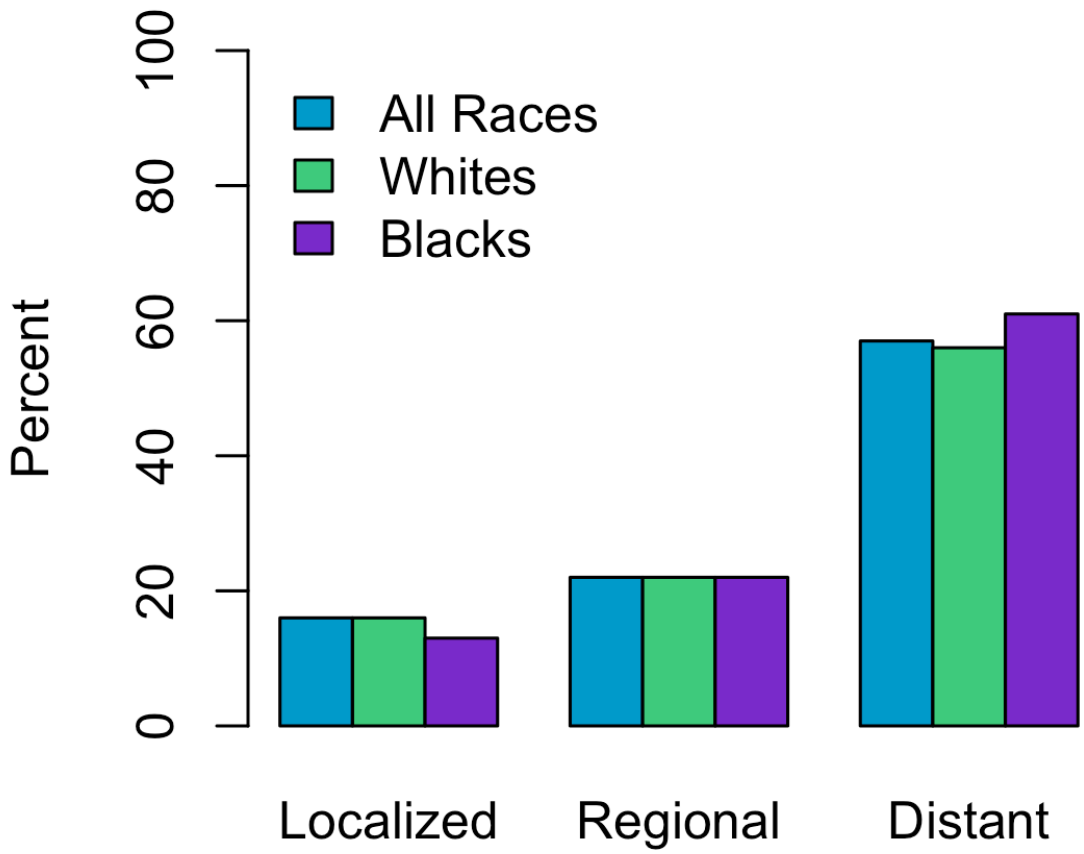

Figure 2. Stage distribution by race, (2007-2013).

Data from Howlader N (2017).

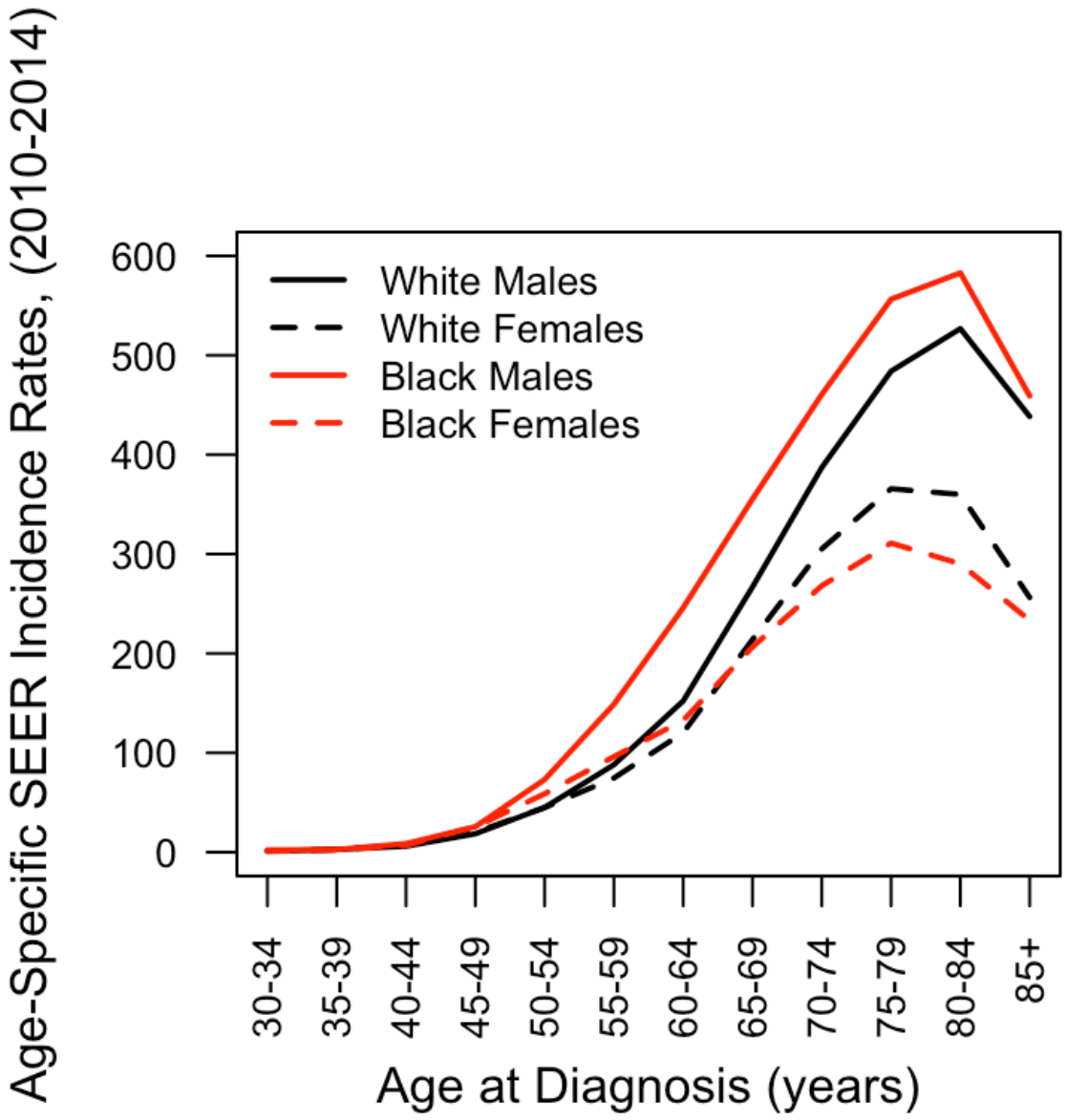

Figure 3. Age-specific SEER Incidence rates (2010-2014), by race and sex.

Rates per 100,000. Data from Howlader N (2017).

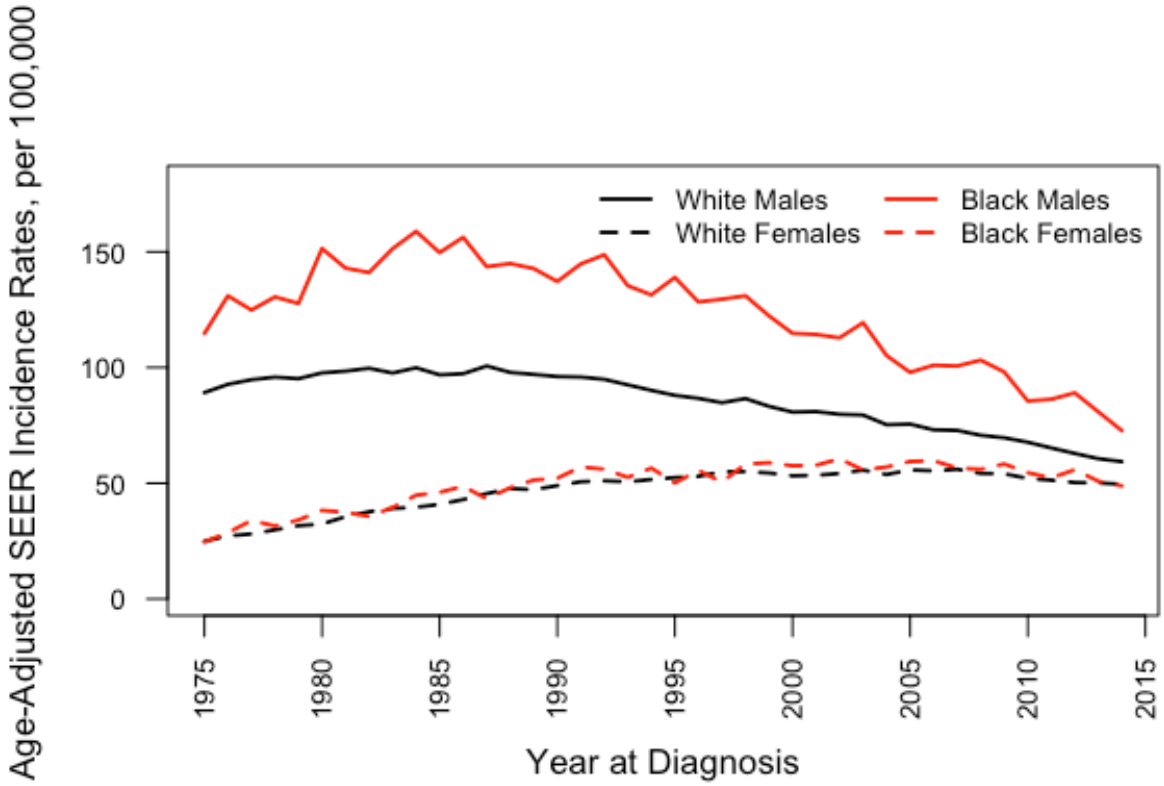

Figure 4. Age-adjusted SEER incidence rates by race and sex.

Rates per 100,000. Data from Howlader N (2017).

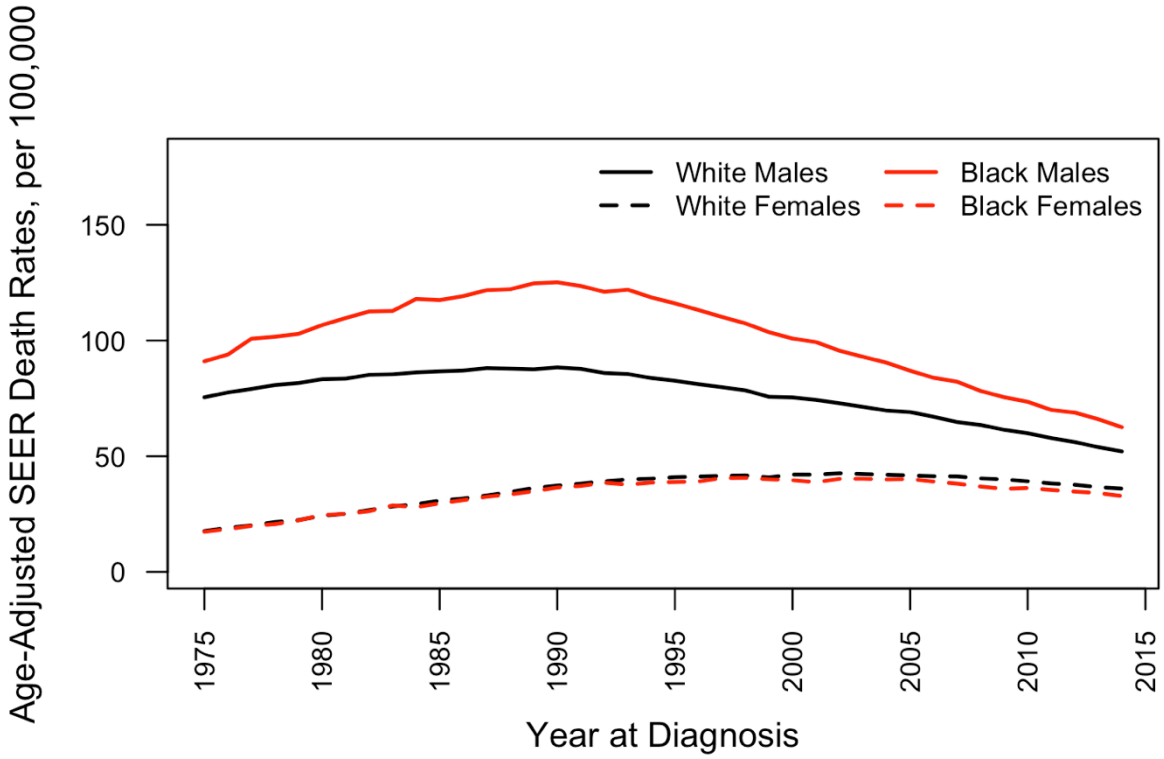

Figure 5. Age-adjusted U.S. death rates by race and sex.

Rate per 100,000. Data from Howlader N (2017).

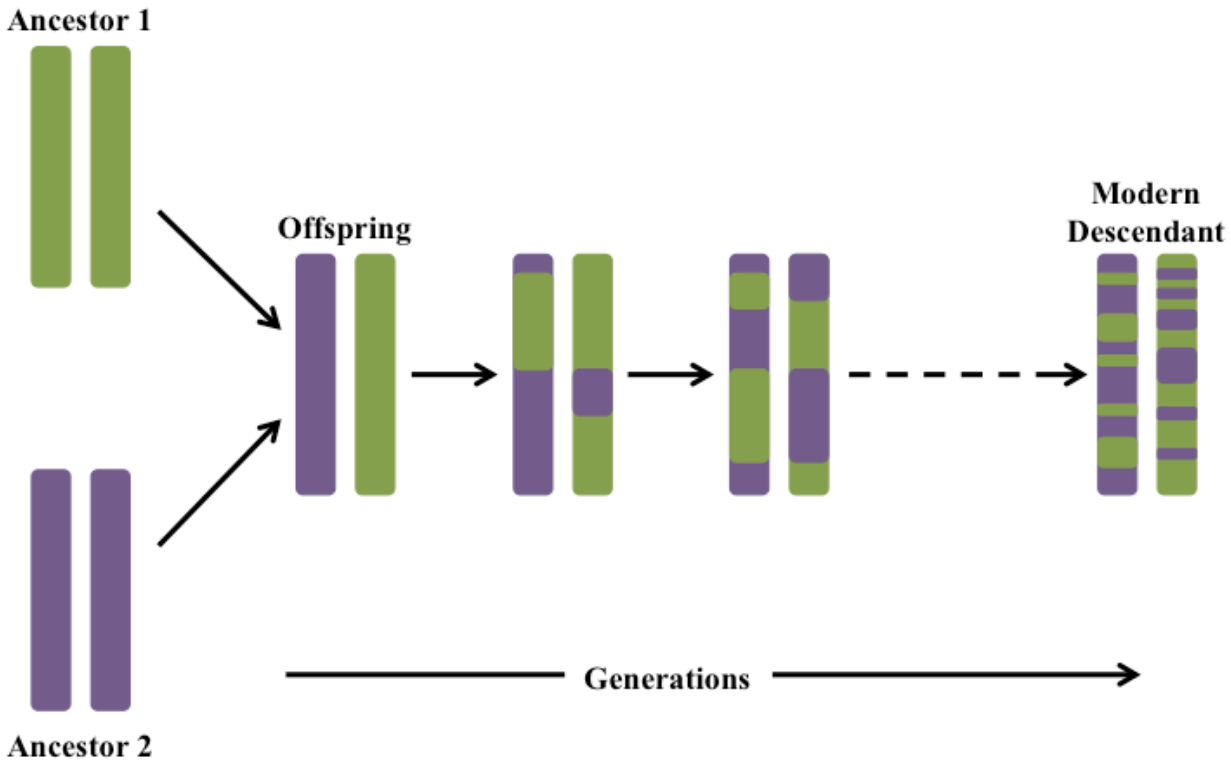

Figure 6. Mating between two previously isolated ancestors results in admixed offspring, the effects of which are observable in modern day descendants.

CHAPTER 2

NO EVIDENCE FOR AN ASSOCIATION BETWEEN AFRICAN GENETIC ANCESTRY AND LUNG CANCER SURVIVAL

Introduction

Epidemiology of lung cancer survival

Lung cancer is the leading cause of cancer death among both men and women in the United States and worldwide, with a 5-year relative survival of 18% (Howlader N , Stewart et al. 2014, Henley et al. 2015). Several factors can influence lung cancer survival rates, but among the most important is stage of diagnosis. Individuals diagnosed with local stage disease have a 55% five-year survival rate, while only 4% of individuals diagnosed with distant stage disease survive up to 5 years (Howlader N , American Cancer Society 2017). These statistics are made even more harrowing by the low frequency of early stage diagnoses compared to late stage diagnoses: 16% of lung cancers are diagnosed at local stage disease, 22% at regional stage, and 57% at distant stage (Howlader N , American Cancer Society 2017).

Racial disparity in lung cancer survival

While lung cancer mortality has decreased in all races in recent years in large part due to greater smoking cessation efforts, a racial disparity continues to exist such that blacks experience poorer survival compared to whites (Moolgavkar et al. 2012, Aizer et al. 2014, Henley et al. 2015, American Cancer Society 2017). Specifically, the national five-year survival is 19% among white individuals and 16% among blacks (Howlader N). Blacks are more frequently

diagnosed at late stage disease compared to whites and less likely to receive the recommended course of treatment based on disease stage (Howlader N , Halpern et al. 2008, Hardy et al. 2009, Efird et al. 2014). Several recent studies have suggested that controlling for differential access to healthcare results in no difference in survival outcomes among blacks and whites (Zheng et al. 2012, Caposole et al. 2014, Ganti et al. 2014). We and others have demonstrated blacks and whites experience no difference in lung cancer survival after controlling for stage and socioeconomic factors (Aldrich et al. 2013, Hua et al. 2014). A recent analysis of Surveillance, Epidemiology, and End Results (SEER) Program data also suggests that blacks have similar lung cancer survival compared to whites (Zeng et al. 2015).

Admixture in African Americans

Blacks in the United States are an admixed population with varying proportions of African ancestry (Tishkoff et al. 2009, Bryc et al. 2010) and self-identified whites can carry African ancestry (Bryc et al. 2015). Identification of ancestry informative markers, genetic variants that differ in frequency between ancestral populations, allows us to distinguish individual-level ancestral origins at the genetic level, i.e. genetic ancestry. Prior studies have shown important associations between genetic ancestry and biomedical phenotypes (Peralta et al. 2010, Goetz et al. 2014, Dumitrescu et al. 2015) such as lung function (Salari et al. 2005, Kumar et al. 2010, Moreno-Estrada et al. 2014) and breast cancer risk (Fejerman et al. 2008, Fejerman et al. 2013).

Rationale

While the association between race and lung cancer survival has been described in several studies, including large, population-based cohorts, the association between genetic ancestry and survival after a diagnosis of lung cancer has yet to be examined. We examined the effect of African ancestry on lung cancer survival in blacks and whites with non-small cell lung cancer in the Southern Community Cohort Study (SCCS), a cohort with the largest representation of blacks in the United States. Black and white SCCS participants were largely recruited from community health centers and thus have similar access to healthcare. Analyses were replicated in a population of black lung cancer cases ascertained from the population-based Metropolitan Detroit Cancer Surveillance System.

Methods

Study population: the Southern Community Cohort Study

Study participants were selected from the SCCS, a prospective cohort study of ~86,000 adults. Participants were enrolled between March 2002 and September 2009 from a 12-state region across the Southeastern United States (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia). Approximately 15% of participants were recruited through mail-in questionnaires, which were sent to a random subset of adults across the 12-state region. The remaining 85% of participants were enrolled at community health centers throughout the region. Demographic characteristics, family history of disease, insurance coverage, tobacco use and other information were collected via in-person interviews by a trained interviewer upon enrollment at the health centers. Individuals self-reported race/ethnicity by selecting any of the following investigator-defined

racial/ethnic groups: white, black/African American, Hispanic/Latino, Asian or Pacific Islander, American Indian or Alaska Native, or Other racial or ethnic group. Approximately two-thirds of participants self-identified as "Black/African American". Upon enrollment, all individuals were asked to donate a biologic specimen (blood, urine, saliva, or buccal cell), of which ~90% of participants agreed. Individuals were eligible to participate if there were between the ages of 40-79 years. A detailed description of study design and recruitment has been previously published (Signorello et al. 2005, Signorello et al. 2010). The SCCS was approved by institutional review boards at Vanderbilt University and Meharry Medical College. Written informed consent was obtained from all participants.

Case Identification and Mortality Assessment

All incident non-small cell lung cancer (NSCLC) cases occurring within the SCCS between 2002 and 2010 were identified through linkage with the 12 state cancer registries. Individuals with a lung cancer diagnosis prior to study enrollment were excluded. Histology, stage at diagnosis, and treatment information were obtained from individual state cancer registries. Stage was derived using the American Joint Committee on Cancers (AJCC) TNM System staging guidelines (6^{th} and 7^{th} Editions). Due to small sample size, we combined individuals with stage II and III disease. For individuals missing stage information, we used the Surveillance, Epidemiology, and End Results (SEER) Summary Stage guidelines, assuming local disease was equivalent to stage I, regional disease was equivalent to stage II/III, and distant disease was equivalent to stage IV. Treatment information describing the administration of chemotherapy, radiation therapy, hormone therapy, immunotherapy, surgery, or other cancer-directed treatment was summarized into a design variable with five levels: no treatment,

chemotherapy only, radiation only, surgery only, and multi-modality (patients receiving any combination of the above treatment options). Participants were followed for all-cause mortality. Vital status was determined at end-of-follow-up (December 31, 2011) through linkage with the Social Security Administration or the National Death Index. Survival time was defined as the time from the date of diagnosis to the date of death, loss to follow up, or censoring.

Genotyping and Quality Control

SCCS individuals were genotyped on the Illumina HumanExome BeadChip v1.1, which contains a panel of >3,000 ancestry informative markers for distinguishing between African and European ancestries. Briefly, quality control removed individuals with sex inconsistencies, <98% genotyping, or self-reported race other than "Black/African American" or "White". Relatedness among individuals was also examined and the individual with the lowest call rate in each relationship pair was removed. Variants were removed during quality control if they were non-autosomal, had <98% genotyping efficiency, or <5% minor allele frequency. Mendelian errors were examined using HapMap trio controls. All quality control measures described were applied using PLINK (Purcell et al. 2007) (version 1.07).

Ancestry Estimation

Global ancestry estimates describe the proportion of an individual's total genome inherited from each contributing ancestral population. To distinguish European and African ancestry, we used a panel of ancestry informative markers. After standard quality control filtering (minor allele frequency > 0.05, genotyping efficiency > 98%, Hardy-Weinberg equilibrium $p > 0.00001$) and pruning based on linkage disequilibrium (window size = 50 SNPs,

step size = 10 SNPs, r2 < 0.4), a total of N=553 markers were available for whites and N=1137 for blacks for global ancestry estimation. Supervised admixture analysis was performed using the software program ADMIXTURE (Alexander et al. 2009) to estimate individual ancestry proportions assuming two ancestral populations with CEU (Utah residents with ancestry from northern and western Europe) and YRI (Yoruban in Ibadan, Nigeria) HapMap (International HapMap Consortium 2003) populations as representative ancestral populations. The resulting output contains the estimated proportions of African and European ancestry for each individual; proportion of African ancestry was then converted to a percent and used for survival analyses (hereafter referred to as "African ancestry").

Statistical Analysis

The impact of African ancestry on overall survival among self-reported black and white lung cancer cases was examined using Cox proportional hazards models. Individuals with less than 30-day survival time were excluded from analysis to remove potential bias related to treatment effects. We first estimated the time-dependent area under the curve (AUC) for an unadjusted Cox proportional hazards model assessing the impact of African ancestry on overall survival. We then derived a multivariable model (hereafter called the "main effects model"), examining African ancestry and overall survival, adjusting for age at diagnosis, sex, BMI (kg/m^2), number of cigarettes per day, stage at diagnosis, treatment, highest education level, and family history of lung cancer. Covariates were selected based on *a priori* knowledge. Using a flexible parametric additive model, it was determined that self-reported race could be predicted from African ancestry and the other model covariates, so we only included the continuous measurement of African genetic ancestry in the Cox regression models. All continuous

variables, including African ancestry, were modeled with restricted cubic splines using three knots. Missing data was multiply imputed ten times using predictive mean matching among the eight confounding variables. The survival models were fit with each of the 10 completed data sets, and results were pooled using Rubin's rules (Rubin 2009). A time-dependent receiver operating characteristic (ROC) curve (Heagerty et al. 2000) and AUC was calculated for every time point between 30 days and 4 years to estimate how well the model predicts survival. The time-dependent AUCs were then averaged over the time interval to obtain the average time-dependent AUC. We then compared the time-dependent AUC of our main effects model to an over-fit Cox proportional hazards model with two-way interaction terms (i.e. African ancestry x treatment, African ancestry x stage, African ancestry x education, education x treatment, education x stage, treatment x stage, sex x age, sex x BMI and sex x cigarettes per day) to assess how well our main effects model performed. We examined the impact of African ancestry, stage, and treatment by removing these variables from the model and compared the time-dependent AUC with and without these variables. A subgroup analysis was performed among self-reported black individuals only. All statistical analyses were carried out in R version 3.2.2 with packages Hmisc and survivalROC.

Replication

Incident black NSCLC cases were identified from three lung cancer case-control studies (Family Health Study III; Women's Epidemiology of Lung Disease Study; and Exploring Health, Ancestry and Lung Epidemiology Study) conducted at the Barbara Ann Karmanos Cancer Institute (KCI) affiliated with Wayne State University (WSU) in Detroit, MI. These studies have been previously described (Schwartz et al. 2009). Rapid case ascertainment was

used to identify cases in the population-based Metropolitan Detroit Cancer Surveillance System, an NCI-funded SEER registry. The institutional review board at Wayne State University approved this study and written informed consent was provided by all participants. Stage, treatment, and vital status were obtained through linkage with the Detroit SEER registry. Treatment and stage variables were summarized in the same manner as the SCCS using both AJCC and SEER staging information. Individuals were previously genotyped on the Illumina 1M-Duo BeadChip. Supervised analysis (K=2) was performed with the use of genome-wide single nucleotide polymorphisms, CEU and YRI reference populations, and the software program ADMIXTURE. Time-dependent AUCs were estimated using the same methods described above for the KCI/WSU population alone and combined with the SCCS population.

Results

Descriptive characteristics

Among the SCCS individuals, 450 incident NSCLC cases occurred with 425 (286 black and 139 white) remaining after quality control procedures. Individuals had a median survival time of 0.7 years (range: 0.003-8.6 years), during which 359 deaths occurred (248 blacks and 111 whites). Forty-two individuals (32 black and 10 white) were excluded for having survival times less than 30 days. Forty-seven percent of blacks with lung cancer had less than 12 years of education compared to 35% of whites (Table 3). The mean age at lung cancer diagnosis was 60 years for blacks and 63 years for whites. More males were diagnosed with lung cancer among blacks than among whites (60% vs. 42%). Smoking status (current/former/never) did not differ between blacks and whites, with 94% of blacks and 96% of whites having smoked cigarettes. Twenty-three percent of whites reported a first-degree relative with a history of lung cancer

compared to 9% of blacks. A greater percentage of blacks were diagnosed at stage IV disease compared to whites (52% vs. 43%). Although similar numbers of blacks and whites were diagnosed with stage I disease (16% vs. 21%), almost twice as many whites received a surgery only course of treatment compared to blacks (11% vs. 21%) (Figure 7). Median African ancestry for self-reported blacks was 85.6%, and was 1.3% for self-reported whites (Table 3 and Figure 8).

Survival analysis

Cox proportional hazards models were implemented to determine the impact of African ancestry on overall survival. The unadjusted Cox model for percent African ancestry had an average time-dependent AUC of 0.54 (Table 4). In the main effects model, African ancestry was not associated with overall survival, with or without stage and treatment (Figure 9A-B), although at smaller values of African ancestry a reduction in mortality was observed. We then estimated the area under the curve to assess the predictive ability of each model. With an average time-dependent AUC of 0.79, the main effects multivariable model examining the association between African ancestry and overall survival, performed dramatically better than the univariate model with African ancestry alone (Figure 10). The inclusion of interaction terms to the main effects model to create an "over fit" model increased the average time-dependent AUC slightly to 0.83, indicating that the main effects model had high predictive ability. Removing African ancestry had no impact on the average time-dependent AUC for either the main effects model (Table 4 and Figure 10) or the interactions model (Table 4). Removing stage and treatment from the main effects model substantially decreased the average time-dependent AUC to 0.65. Further removal of African ancestry from the main effects model without stage and treatment resulted in little

change in the average time-dependent AUC (Figure 10). We then removed whites from the SCCS and examined each of the Cox proportional hazards models among blacks only. Observations were similar to those of the overall NSCLC population and are presented in Figure 11 and Table 5.

Replication analysis

We evaluated the impact of African ancestry on overall survival in a population of 316 black NSCLC cases ascertained from the KCI/WSU. Briefly, the mean age of diagnosis was 60 years and 41% of lung cancer cases were male. Thirty-four percent of cases were diagnosed at stage I disease, 39% were diagnosed at stage II/III disease, and 26% were diagnosed at stage IV disease (Table 6 and Figure 7). The median African ancestry was 83.3% (Figure 12). Additional descriptive characteristics of this population are provided in Table 6. Similar findings for the impact of African ancestry on overall survival were observed in the blacks in the SCCS (Table 5 and Figure 13). Removal of African ancestry from both the main effects model and the interaction model had a negligible effect on the average time-dependent AUC (0.74 and 0.76, respectively, Table 5). Similar to the SCCS, the removal of stage and treatment from the main effects model resulted in a dramatic decrease in the average time-dependent AUC (0.63) and removal of African ancestry in addition to stage and treatment had little impact (average time-dependent AUC = 0.61).

We further examined the impact of African ancestry on survival in a combined SCCS and KCI/WSU population. Average time-dependent AUCs are presented in Table 7 and Figure 14. There was no difference in average time-dependent AUCs for any model compared to SCCS or WSU populations alone.

Discussion

We found that African ancestry was not associated with NSCLC survival in the SCCS, using Cox proportional hazard models adjusted for stage and treatment. Since African ancestry did not have a linear relationship with survival, we chose to model the predictor as a restricted cubic spline. As such, there is no single hazard ratio and p-value to describe the magnitude and significance of the association. Instead, we present the hazard ratio and confidence interval as a function of African ancestry in Figure 9, using the median African ancestry value as the referent. Furthermore, we examined the impact of African ancestry on survival by comparing the average time-dependent area under the curve for Cox proportional hazards models with and without African ancestry or stage and treatment. This allows us to examine the clinical utility of such predictors, rather than simply the strength of the association. Using these methods, we find that stage and treatment were strongly predictive of overall survival and are more important predictors of survival than African ancestry. This result was recapitulated in an independent population of blacks with NSCLC from KCI/WSU, providing further evidence of the null association between African ancestry and overall survival.

In this study we used genetic ancestry as a continuous proxy for race. While genetic ancestry and race are highly correlated, race can be viewed as a social construct that captures both genetic and non-genetic factors, such as culture and social perception. Given the high variability of genetic admixture that is not captured by categorically-defined race, adjusting for race alone while examining diseases with established racial disparities may not entirely account for underlying population substructure. Here, the utilization of African ancestry instead of race allows us to attempt to disentangle the genetic and social disparities associated with lung cancer survival. We find that African ancestry is not associated with overall survival when adjusting for

stage and treatment (Figure 9A), with hazard ratios and confidence intervals encompassing 1.0 for all values of African ancestry. When stage and treatment are removed from the model (Figure 9B), we observed a slightly, although not statistically significant, greater survival for individuals with smaller proportions of African ancestry in the SCCS. We hypothesize that this observed reduction in mortality is due to the influence of unmeasured treatment, social, environmental, or other factors not captured by genetic ancestry alone. It is possible that these observed differences are the result of cultural differences in the perception of disease and the willingness to seek treatment (Margolis et al. 2003, Lathan et al. 2010).

Herein we observed a striking decrease in clinical validity when stage and treatment are excluded from overall survival models for lung cancer. Previous work by our group and others has shown that racial disparities in lung cancer survival disappear when controlling for stage at diagnosis or treatment (Zheng et al. 2012, Aldrich et al. 2013, Ganti et al. 2014). Together these studies along with the present analysis suggest that the observed disparity in survival between blacks and whites can be attributed to differences in stage at diagnosis or receipt of treatment rather than race. It is important to note that while this study shows no association between genetic ancestry and overall survival, it does not eliminate the potential for a genetic contribution to lung cancer survival or the possibility of race-specific genetic risk factors.

To our knowledge, this is the first study to examine the relationship between genetic ancestry and lung cancer survival in blacks and whites with NSCLC. By utilizing the unique SCCS, we were able to control for multiple factors potentially influencing lung cancer survival, including socioeconomic status, family history of lung cancer, cigarette smoking, disease stage, and treatment received. While our study was limited by small sample size, our confidence intervals are narrow and the consistency of our findings between two independent study

populations, despite the different demographic and epidemiologic characteristics, emphasizes the robustness of our findings. Furthermore, we acknowledge that the staging and treatment information obtained through linkage with cancer registries may not depict the most accurate and up to date clinical and staging information. Future studies should focus on a clinical population with carefully annotated clinical information for examining ancestral differences in lung cancer survival.

In summary, we find that African ancestry is not associated with NSCLC survival and that stage and treatment are robust predictors of lung cancer survival. These findings suggest that efforts to increase the early detection of lung cancer will improve lung cancer outcomes for both blacks and whites.

Table 3. Descriptive characteristics by race of incident non-small cell lung cancer (NSCLC) cases in the Southern Community Cohort Study.

	Blacks (N=286) N (%)	Whites (N=139) N (%)	Total (N=425) N (%)
Sex			
Male	171 (59.8)	58 (41.7)	229 (53.9)
Female	115 (40.2)	81 (58.3)	196 (46.1)
Vital status			
Alive	38 (13.3)	28 (20.1)	66 (15.5)
Dead	248 (86.7)	111 (79.9)	359 (84.5)
Median African ancestry, % (range)	85.6 (<0.01-98.7)	1.3 (<0.01-91.1)	80.1 (<0.01-98.7)
Lung cancer stage at diagnosis			
I	44 (15.7)	29 (21.3)	73 (17.5)
II/III	90 (32.0)	49 (36.0)	139 (33.3)
IV	147 (52.3)	58 (42.6)	205 (49.2)
Unknown	5	3	8
Treatment			
No treatment	75 (26.8)	36 (27.3)	111 (26.9)
Surgery only	31 (11.1)	27 (20.5)	58 (14.1)
Chemotherapy only	51 (18.2)	18 (13.6)	69 (16.7)
Radiation only	36 (12.9)	12 (9.1)	48 (11.7)
Multi-modality	87 (31.1)	39 (29.5)	126 (30.6)
Unknown	6	7	13
Histology			
Adenocarcinoma	113 (39.5)	51 (36.7)	164 (38.6)
NSCLC-NOS	78 (27.3)	34 (24.5)	112 (26.4)
Squamous	72 (25.2)	41 (29.5)	113 (26.6)
Other NSCLC	22 (7.7)	12 (8.6)	34 (8.0)
Multiple histologies	1 (0.3)	1 (0.7)	2 (0.5)
Mean age at enrollment, yr (SD)	56.5 (9.0)	60.4 (8.6)	57.8 (9.1)
Mean age at diagnosis, yr (SD)	59.6 (9.1)	62.8 (8.7)	60.6 (9.1)
Median observed duration of disease among those who died, yr (range)	0.50 (0.003-8.61)	0.54 (0.01-5.4)	0.52 (0.003-8.61)
Median observed duration of disease among those alive at last follow-up, yr (range)	3.8 (1.5-8.2)	4.0 (1.5-7.7)	3.5 (1.5-8.2)
Highest education level, yr			
<12	134 (47.2)	48 (34.8)	182 (43.1)
≥12	150 (52.8)	90 (65.2)	240 (56.9)
Unknown	2	1	3
Household income in last year			
<$15,000	190 (67.6)	87 (64.0)	277 (66.4)
≥$15,000	91 (32.4)	49 (36.0)	140 (33.6)
Unknown	5	3	8

Table 3 Continued. Descriptive characteristics by race of incident non-small cell lung cancer (NSCLC) cases in the Southern Community Cohort Study.

Smoking status at cohort entry			
Current	206 (72.8)	93 (68.4)	299 (71.4)
Former	59 (20.8)	37 (27.2)	96 (22.9)
Never	18 (6.4)	6 (4.4)	24 (5.7)
Unknown	3	3	6
Mean cigarettes per day (SD)	15.7 (13.1)	23.2 (14.7)	18.1 (14.1)
Smokes menthol cigarettes			
Yes	182 (69.2)	23 (17.8)	205 (52.3)
No	81 (30.8)	106 (82.2)	187 (47.7)
Unknown	23	10	33
Self-reported doctor diagnosis of emphysema or chronic bronchitis			
Yes	28 (9.9)	41 (29.7)	69 (16.4)
No	255 (90.1)	97 (70.3)	352 (83.6)
Unknown	3	1	4
First-degree relative with lung cancer			
Yes	21 (9.3)	26 (22.6)	47 (13.8)
No	205 (90.7)	89 (77.4)	294 (86.2)
Unknown	60	24	84
Mean BMI, kg/m^2 (SD)	26.8 (6.1)	27.1 (6.0)	26.9 (6.1)
Health insurance status			
Yes	174 (61.7)	98 (71.5)	272 (64.9)
No	108 (38.3)	39 (28.5)	147 (35.1)
Unknown	4	2	6
Enrollment source			
Community Health Center	268 (93.7)	113 (81.3)	381 (89.6)
General Population	18 (6.3)	26 (18.7)	44 (10.4)

NOS=Not otherwise specified
Yr=Year
SD=Standard deviation
BMI=Body mass index

Table 4. Average time-dependent AUCs for Cox proportional hazards models in the Southern Community Cohort Study

Model	Average time-dependent AUC
African ancestry only (unadjusted)	0.54
Main effects[a]	0.79
Interactions[b]	0.83
Main effects without African ancestry	0.79
Interactions without African ancestry	0.82
Main effects without stage and treatment	0.65
Main effects without stage, treatment, and African ancestry	0.63

[a]Main effects model is a Cox proportional hazards model examining association between African ancestry and lung cancer survival, adjusting for age at diagnosis, sex, BMI (kg/m^2), cigarettes per day, stage at diagnosis, treatment, highest education level, and family history of lung cancer.

[b]Interaction model is a Cox proportional hazards model examining association between African ancestry and lung cancer survival, adjusting for the same variables in the main effects model but also including the following two-way interactions (percent African ancestry x treatment, African ancestry x stage, African ancestry x education, education x treatment, education x stage, treatment x stage, sex x age, sex x BMI, sex x cigarettes per day).

Table 5. Average time-dependent AUCs for each Cox proportional hazards model among black non-small cell lung cancer cases from the Southern Community Cohort Study (SCCS) and Wayne State University (WSU).

	Average time-dependent AUC	
Model	**SCCS blacks**	**WSU**
African ancestry only	0.53	0.53
Main effects[1]	0.82	0.75
Interactions[2]	0.85	0.80
Main effects without African ancestry	0.81	0.74
Interactions without African ancestry	0.83	0.76
Main effects without stage and treatment	0.67	0.63
Main effects without stage, treatment, and African ancestry	0.67	0.61

Main effects model is a Cox proportional hazards model examining association between African ancestry and lung cancer survival, adjusting for age at diagnosis, sex, BMI (kg/m^2), cigarettes per day, stage at diagnosis, treatment, highest education level, and family history of lung cancer.

Interaction model is a Cox proportional hazards model examining association between African ancestry and lung cancer survival, adjusting for the same variables in the main effects model but also including the following two-way interactions (percent African ancestry x treatment, African ancestry x stage, African ancestry x education, education x treatment, education x stage, treatment x stage, sex x age, sex x BMI, sex x cigarettes per day).

Table 6. Descriptive characteristics of non-small cell lung cancer (NSCLC) cases from the Karmanos Cancer Institute/Wayne State University (KCI/WSU).

	Blacks (N=316) N (%)
Sex	
Male	129 (40.8)
Female	187 (59.2)
Vital status	
Alive	107 (33.9)
Dead	209 (66.1)
Median African ancestry, % (range)	83.3 (33.6-100.0)
Lung cancer stage at diagnosis	
I	107 (34.3)
II/III	123 (39.4)
IV	82 (26.3)
Unknown	4
Treatment	
No treatment	27 (8.5)
Surgery only	85 (26.9)
Chemotherapy only	32 (10.1)
Radiation only	26 (8.2)
Multi-modality	146 (46.2)
Histology	
Adenocarcinoma	159 (50.3)
NSCLC-NOS	79 (25.0)
Squamous	71 (22.5)
Other	7 (2.2)
Mean age at diagnosis, yr (SD)	60.3 (11.02)
Median observed duration of disease among those who died, yr (range)	2.1 (0.33-11.5)
Median observed duration of disease among those alive at last follow-up, yr (range)	6.5 (3.8-13.7)
Highest education level, yr	
<12	89 (28.2)
≥12	227 (71.8)
Smoking status at cohort entry	
Current	177 (56.0)
Former	123 (38.9)
Never	16 (5.1)
Unknown	0
Mean cigarettes per day (SD)	18.3 (13.3)
First-degree relative with lung cancer	
Yes	62 (19.6)
No	254 (80.4)
Mean BMI, kg/m^2 (SD)	26.6 (6.6)

NOS=Not otherwise specified, Yr=Year, SD=Standard deviation, BMI=Body mass index

Table 7. Average time-dependent AUCs for Cox proportional hazards models in the combined SCCS and KCI/WSU population

Model	Average time-dependent AUC
African ancestry only (unadjusted)	0.55
Main effects[a]	0.80
Interactions[b]	0.85
Main effects without African ancestry	0.80
Interactions without African ancestry	0.85
Main effects without stage and treatment	0.63
Main effects without stage, treatment, and African ancestry	0.61

[a]Main effects model is a Cox proportional hazards model examining association between African ancestry and lung cancer survival, adjusting for age at diagnosis, sex, BMI (kg/m^2), cigarettes per day, stage at diagnosis, treatment, highest education level, and family history of lung cancer.

[b]Interaction model is a Cox proportional hazards model examining association between African ancestry and lung cancer survival, adjusting for the same variables in the main effects model but also including the following two-way interactions (percent African ancestry x treatment, African ancestry x stage, African ancestry x education, education x treatment, education x stage, treatment x stage, sex x age, sex x BMI, sex x cigarettes per day).

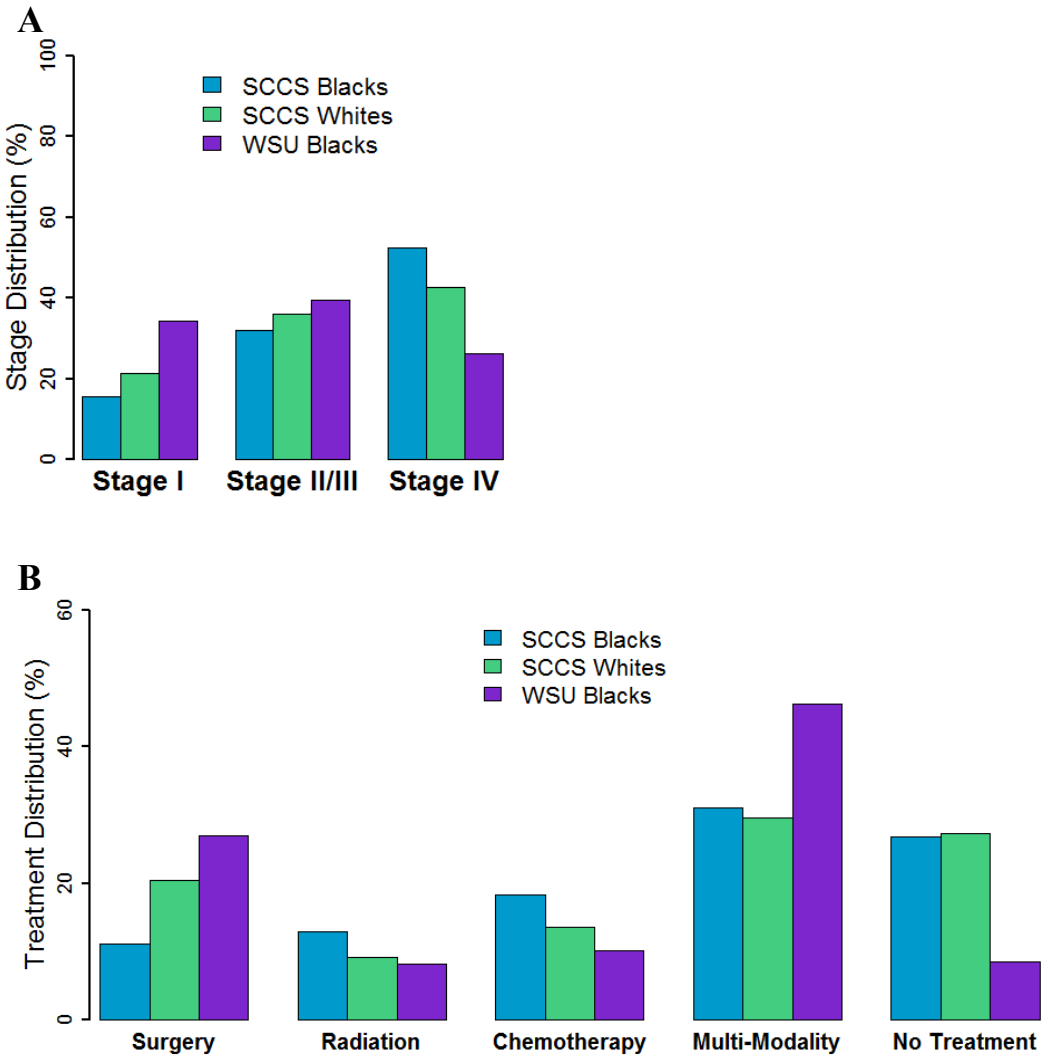

Figure 7. Distribution of stage (A) and treatment (B) for blacks and whites in the Southern Community Cohort Study (SCCS) and blacks in the Karmanos Cancer Institute(KCI)/Wayne State University (WSU) studies.

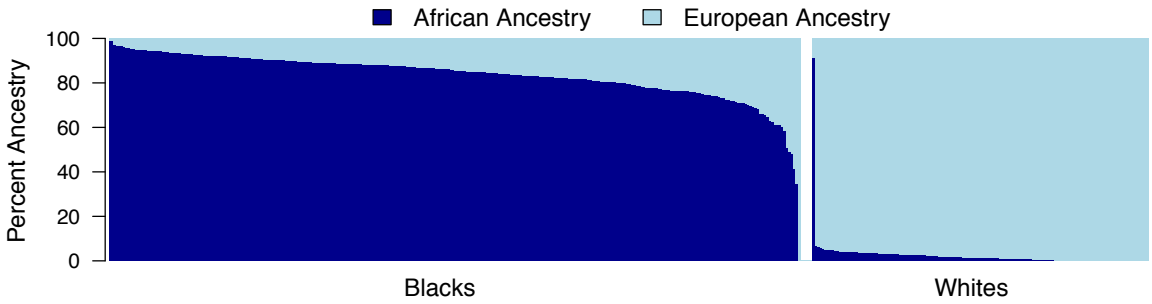

Figure 8. Genetic ancestry estimates for blacks (N=286) and whites (N=139) with non-small cell lung cancer in the Southern Community Cohort Study.

Global ancestry was estimated using ADMIXTURE software and ancestry informative markers (N=1137 in blacks and N=553 in whites) using a supervised method including CEU (European) and YRI (African) reference populations from the International HapMap Project. Individuals are plotted along the x-axis and along the y-axis is the percent African ancestry (dark blue) and percent European ancestry (light blue) for each individual.

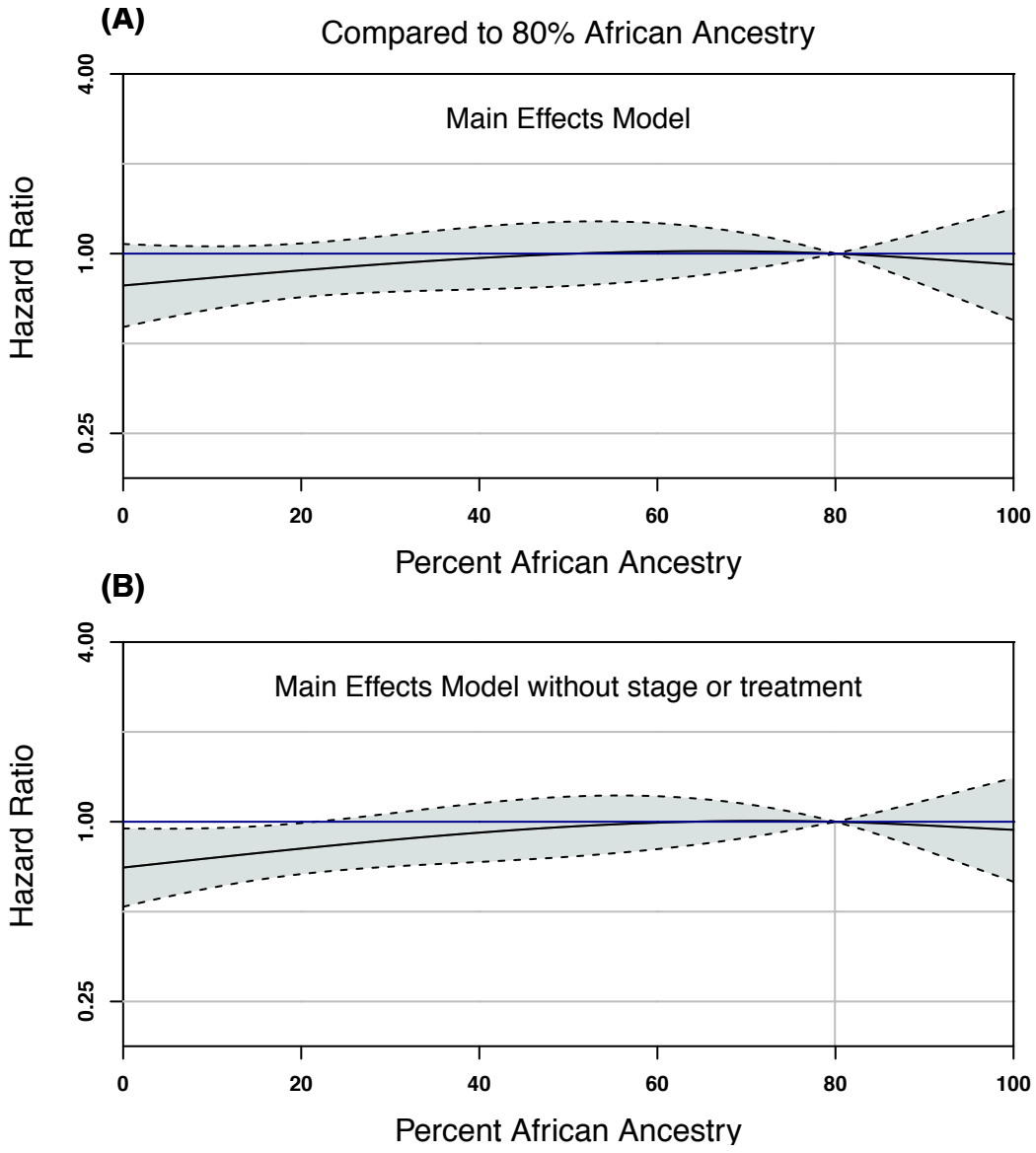

Figure 9. African ancestry is not associated with lung cancer survival, with or without stage and treatment included in the model.

Hazard ratios and 95% confidence intervals are plotted on the y-axis for the association between splined percent African ancestry (x-axis) and lung cancer survival in the Southern Community Cohort Study for the (**A**) main effects model and the (**B**) main effects model without stage and treatment (see Methods). Median African ancestry (80%) is the referent (vertical grey line).

Main effects model is a Cox proportional hazards model examining the association between African ancestry and lung cancer survival, adjusting for age at diagnosis, sex, BMI (kg/m^2), cigarettes per day, stage at diagnosis, treatment, highest education level, and family history of lung cancer.

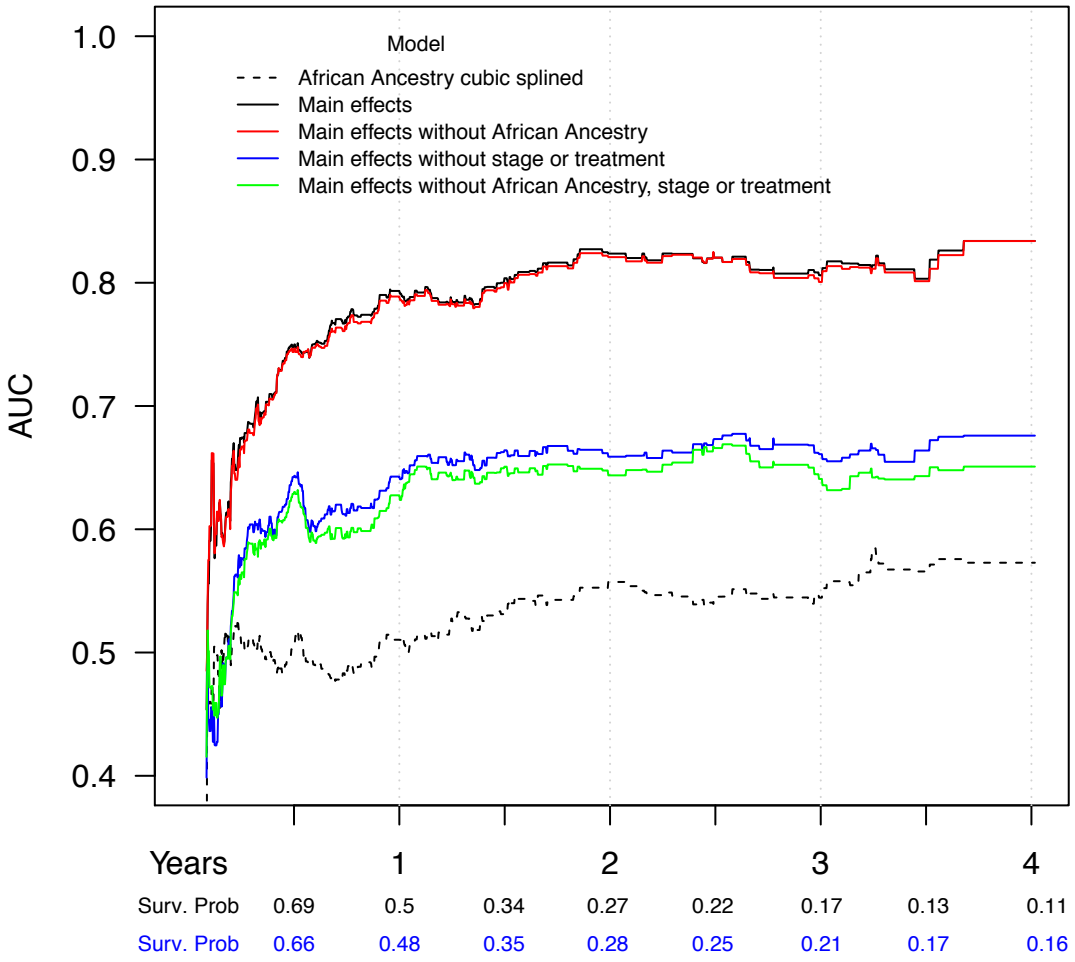

Figure 10. Time-dependent AUCs for each Cox proportional hazards model in the Southern Community Cohort Study.

Removal of stage and treatment, not African ancestry, resulted in a reduction in the predictive ability of the main effects model. African ancestry had no effect on the time-dependent AUC of the main effects model, with or without stage and treatment.

The Kaplan-Meier survival probability estimates below the x-axis represent the probability of surviving to the indicated 6 month intervals as estimated from the main effects model (black) and from the main effects model without stage or treatment (blue).

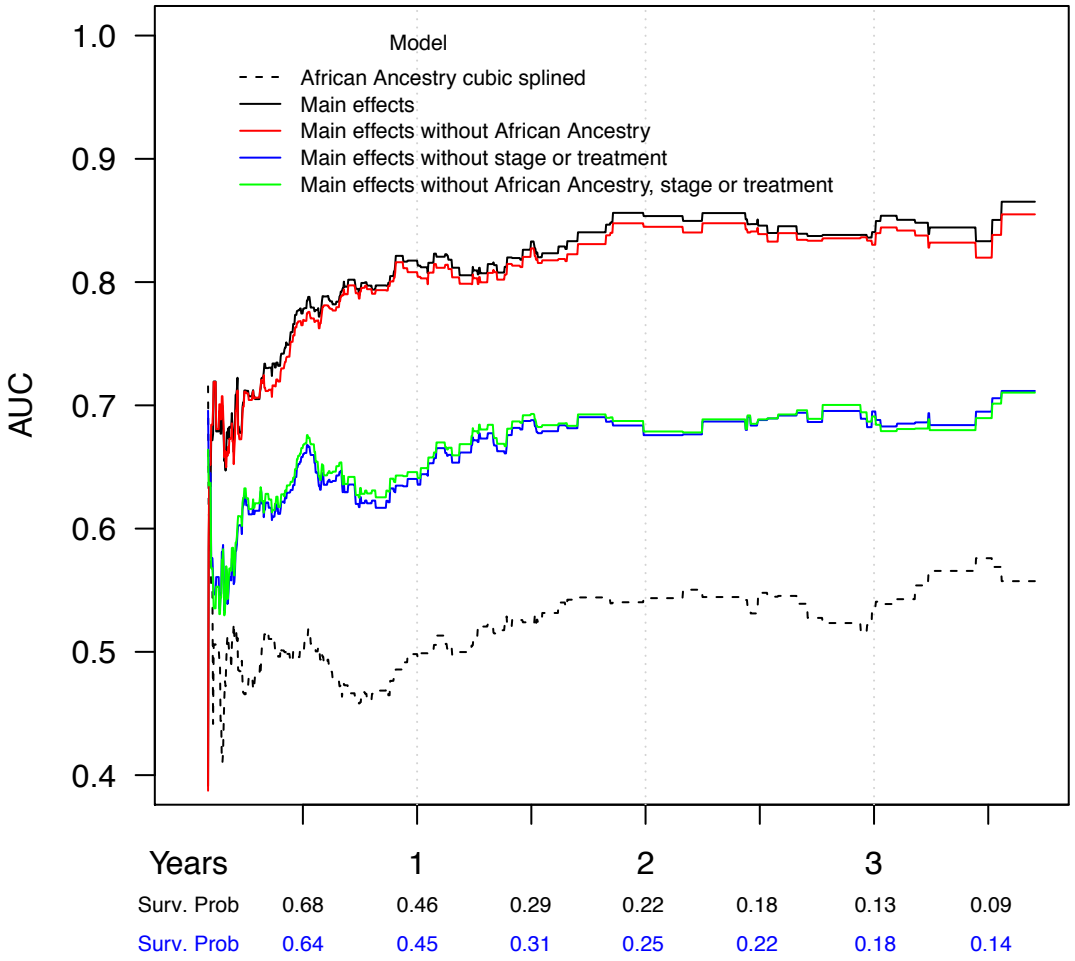

Figure 11. Time-dependent AUCs for each Cox proportional hazards model among self-reported blacks in the Southern Community Cohort Study (SCCS).

The Kaplan-Meier survival probability estimates below the x-axis are the represent the probability of surviving to the indicated 6 month intervals as estimated from the main effects model (black) and from the main effects model without stage or treatment (blue).

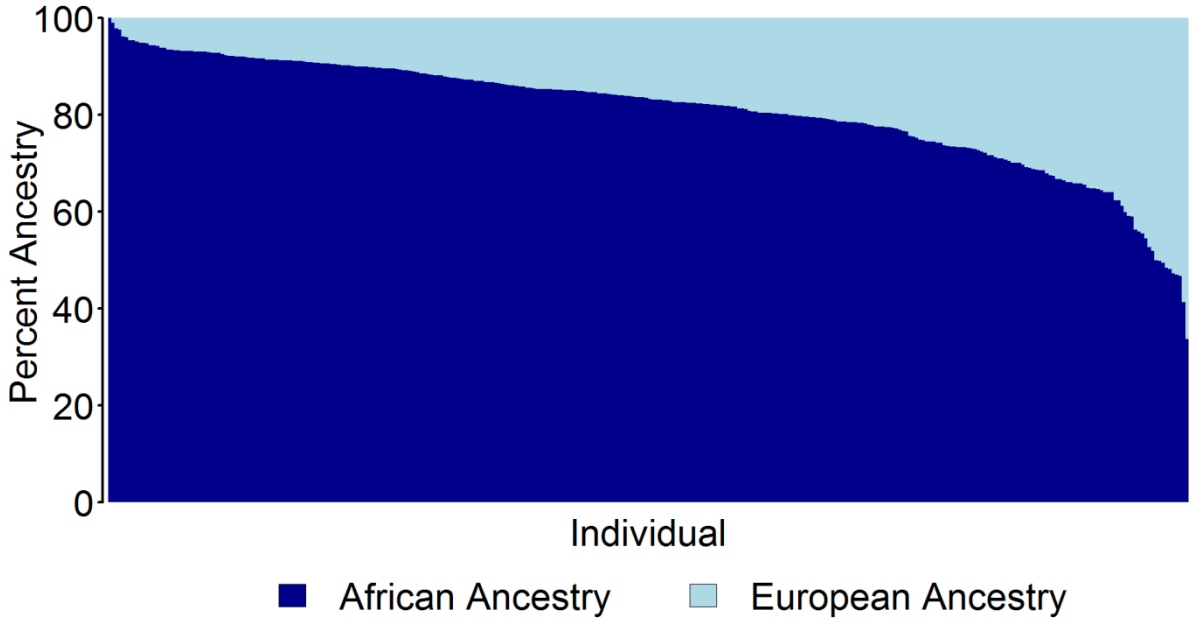

Figure 12. Genetic ancestry estimates for non-small cell lung cancer cases in the Karmanos Cancer Institute/Wayne State University study population (N=316).

Global ancestry was estimated using ADMIXTURE software and genome-wide single nucleotide polymorphisms in a supervised manor using CEU (European) and YRI (African) reference populations from the International HapMap Project. Individuals are plotted along the x-axis and along the y-axis is the percent African ancestry (dark blue) and percent European ancestry (light blue) for each individual.

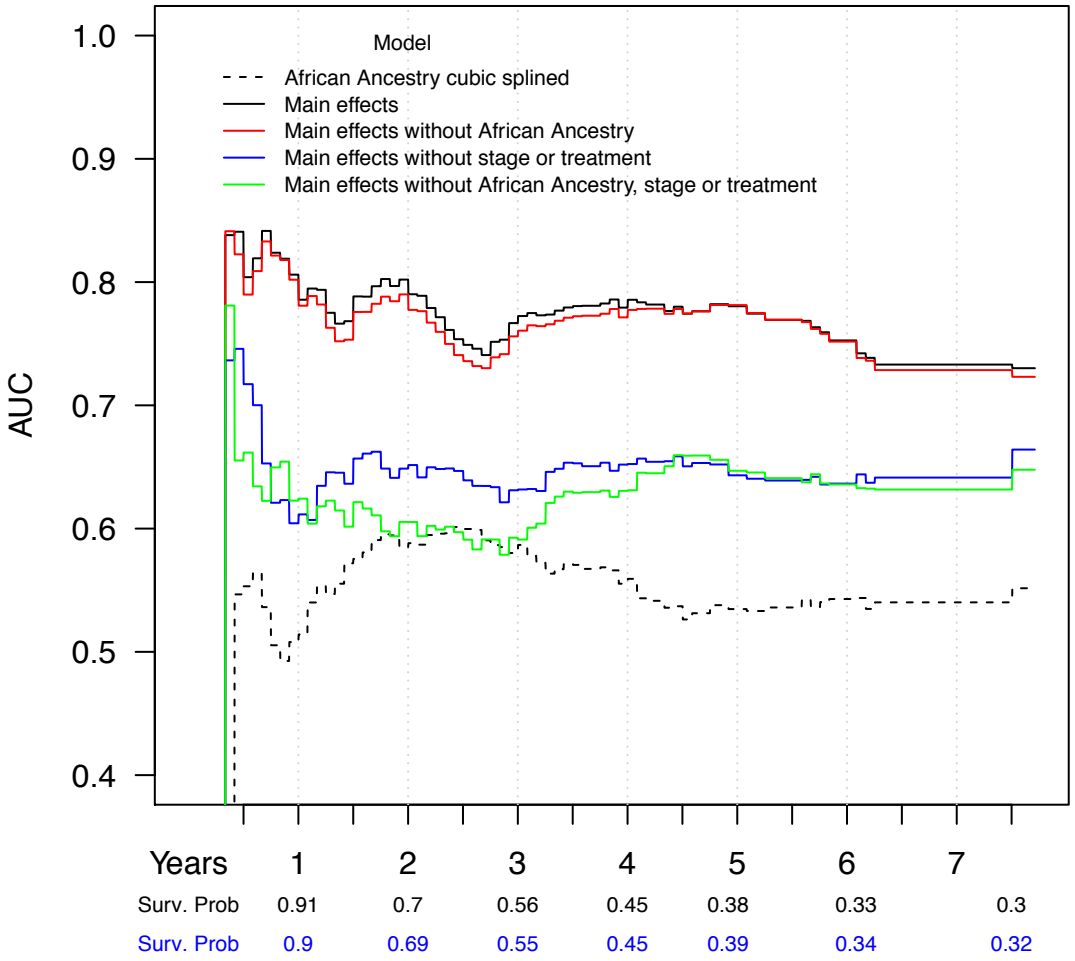

Figure 13. Time-dependent AUCs for each Cox proportional hazards model among black non-small cell lung cancer cases in the Karmanos Cancer Institute/Wayne State University study population.

The Kaplan-Meier survival probability estimates below the x-axis are the represent the probability of surviving to the indicated 6 month intervals as estimated from the main effects model (black) and from the main effects model without stage or treatment (blue).

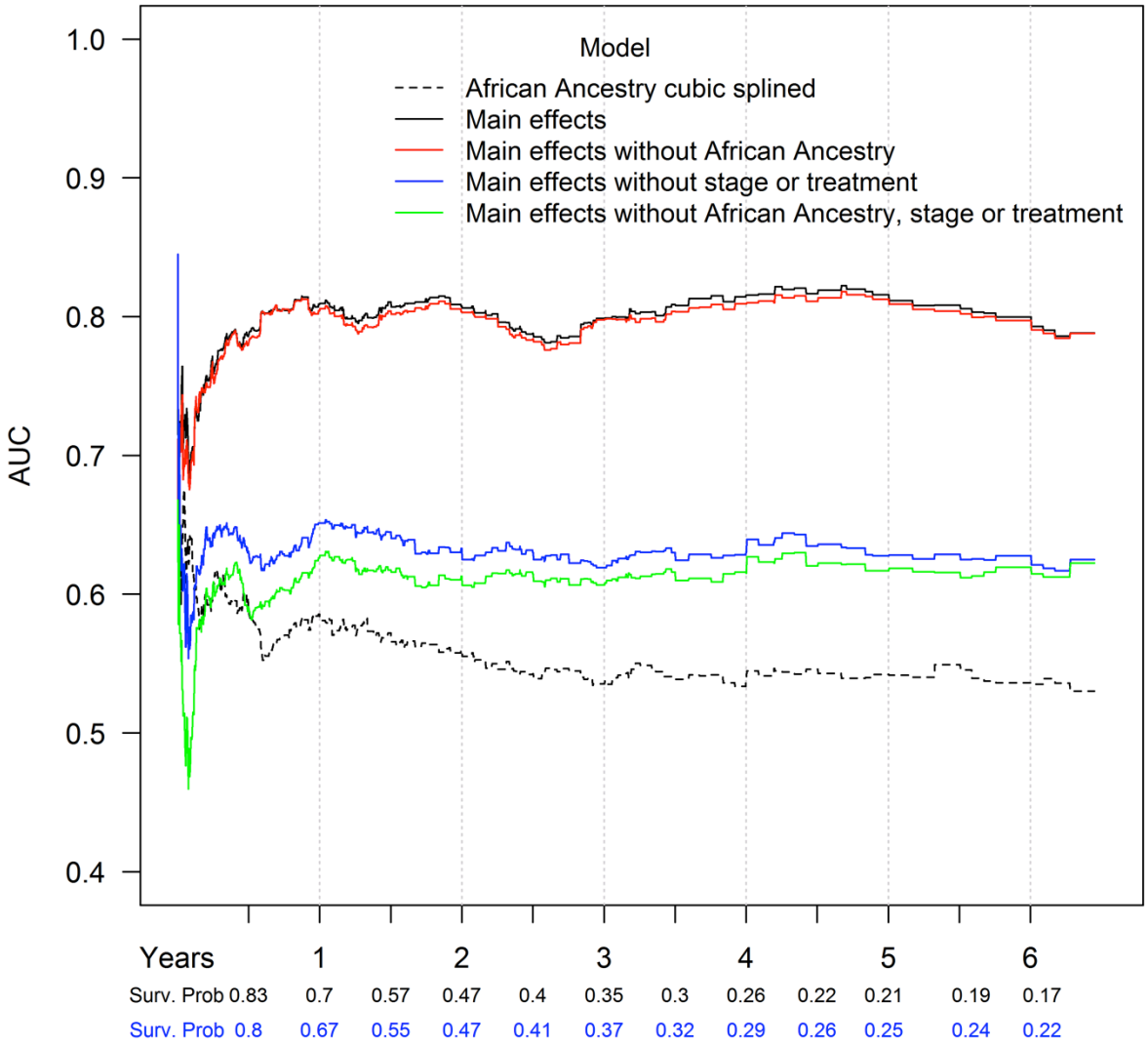

Figure 14. Time-dependent AUCs for each Cox proportional hazards model in the combined SCCS and KCI/WSU population.

The Kaplan-Meier survival probability estimates below the x-axis are the represent the probability of surviving to the indicated 6 month intervals as estimated from the main effects model (black) and from the main effects model without stage or treatment (blue).

CHAPTER 3

GERMLINE GENETIC VARIANTS AND LUNG CANCER SURVIVAL IN AFRICAN AMERICANS[1]

Introduction

Epidemiology of lung cancer survival

Lung cancer is the leading cause of cancer-related death in the United States (American Cancer Society 2017). Notably African Americans have poorer lung cancer survival compared to whites, namely a 16% 5-year survival in African Americans compared to 19% in whites (Howlader N). This poor survival can largely be attributed to presentation at a later stage in African Americans, which is associated with reduced lung cancer survival (Efird et al. 2014). In the clinical setting, an individual's treatment plan is tailored to the stage of lung cancer diagnosis and can impact lung cancer survival (National Comprehensive Cancer Network 2017). Other factors associated with poor lung cancer survival include low socioeconomic status (SES), smoking, older age, male sex, site of origin, tumor grade, histologic subtype and somatic mutation profile (William D. Travis 2004, Islami et al. 2015).

[1] Adapted from: Jones, C. C., W. S. Bush, D. C. Crawford, A. S. Wenzlaff, A. G. Schwartz, J. K. Wiencke, M. R. Wrensch, W. J. Blot, S. J. Chanock, E. L. Grogan and M. C. Aldrich (2017). "Germline genetic variants and lung cancer survival in African Americans." Cancer Epidemiol Biomarkers Prev.

Genetic studies of lung cancer survival

Tumor mutation status is frequently incorporated into genetic medicine to guide therapeutic decisions (Pao et al. 2005, Pao and Chmielecki 2010). However, germline variation could also influence treatment response (Bell et al. 2005, Gregorc et al. 2008). Genetic association studies have successfully identified germline variants that contribute to lung cancer survival, but genome-wide association studies (GWAS) of lung cancer survival have been conducted only in populations of European or Asian descent (Wu et al. 2010, Sato et al. 2011, Tan et al. 2011, Wu et al. 2011, Hu et al. 2012, Lee et al. 2013, Wu et al. 2013, Han et al. 2014). To improve the precision of approaches based on population genetic history, genetic studies in diverse racial/ethnic populations need to be conducted. Overall, African descent populations have shorter segments of linkage disequilibrium compared to European descent populations, thus examining genetic associations of disease in African American populations may allow for improved fine mapping of causal variants with potential therapeutic benefits applicable across racial/ethnic groups.

Rationale

Using a prospective cohort of African Americans, we sought to validate lung cancer survival variants previously reported in the National Human Genome Research Institute GWAS catalog (Burdett T (EBI)), accounting for known factors influencing survival. We also sought to identify novel and potentially functional genome-wide germline variants associated with lung cancer survival in African Americans. Though the present study is underpowered to detect genome-wide associations, this is the first study to our knowledge to examine genome-wide germline genetic associations of lung cancer survival in African Americans.

Methods

Southern Community Cohort Study

The Southern Community Cohort Study (SCCS) is a large prospective cohort study of ~86,000 individuals designed to examine racial disparities in cancer. Adults between the ages of 40-79 years were enrolled primarily at community health centers throughout a 12-state region across the Southeastern United States (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia) between March 2002 and September 2009. Approximately 90% of enrolled individuals agreed to donate a biologic specimen. Nearly two-thirds of participants self-identified as African American. Additional details regarding the study design and recruitment have been previously published (Signorello et al. 2005, Signorello et al. 2010). The SCCS was approved by Institutional Review Boards at Vanderbilt University and Meharry Medical College. Written informed consent was obtained from all participants.

Case identification and mortality assessment

Incident African American lung cancer cases occurring in the SCCS through September 2012 were identified via linkage with the 12 state cancer registries throughout the study catchment area. Epidemiologic data such as demographic characteristics, self-reported medical history, and tobacco use, were ascertained at study enrollment by trained personnel via in-person computer-assisted interviews at community health centers or by questionnaire mailed to a random subset of the general population in the 12-state region (Signorello et al. 2005, Signorello et al. 2010). Histology, stage at diagnosis, and clinical treatment were collected from state cancer registry data. Lung cancer stage was determined using the American Joint Committee on Cancers

(AJCC) staging system and Surveillance, Epidemiology, and End Results (SEER) summary stage. Treatment information was identified from cancer registries and summarized as a 5-level design variable: no treatment, chemotherapy only, radiation only, surgery only, or multi-modality (any combination of the above therapy regimens). Participants were followed for all-cause mortality through 2013, which was assessed via linkage with the Social Security Administration and the National Death Index. Survival time (years) was defined as the time between date of diagnosis and date of death, loss to follow-up, or end of follow-up period for the study.

Genotyping and quality control

Individuals with an available blood or buccal cell biospecimen from which germline DNA could be extracted as of September 2012 were genotyped on the Illumina HumanExome BeadChip v1.1 (San Diego, CA). The HumanExome genotyping array contains >240,000 exonic variants, as well as a panel of >3,000 Ancestry Informative Markers (AIMs) for distinguishing African and European ancestries, and >4,700 disease-associated tag markers identified from GWAS. A detailed description of the quality control (QC) process is presented in Figure 15. Briefly, QC procedures removed individuals having mismatched genetic sex, <98% genotyping efficiency and related individuals. Autosomal variants were excluded that had Mendelian errors, a call rate <98% or were monomorphic. A total of 14 samples were used to verify genotype reproducibility, with a genotype concordance of >99.99%. Analysis of previous lung cancer survival associations from the NHGRI-EBI GWAS catalog and common variants was restricted to variants with a minor allele frequency (MAF) >5%. QC filters were applied using PLINK and

R (version 3.0.3) (Purcell et al. 2007). Variants were mapped to dpSNP Build 137/GRCH37.p5 based on chromosome and position to obtain reference SNP (rs) identifiers.

Admixture estimation

Global ancestry was estimated using the ADMIXTURE software (Alexander et al. 2009). The CEU (Utah residents with ancestry from northern and western Europe) and YRI (Yoruba from Ibadan, Nigeria) HapMap populations were included as ancestral reference populations to inform the admixture estimation process (International HapMap Consortium 2003). After merging with the HapMap reference populations and pruning based on linkage disequilibrium (r^2>0.4), 1,137 AIMs remained among the 2,587 AIMs in the final QC dataset for estimation of genetic admixture. Supervised (k=2) admixture analysis was performed to estimate the percent African (YRI) and European (CEU) ancestry for each African American individual.

Statistical Analysis

Replication of Lung Cancer Survival Variants from the NHGRI-EBI GWAS Catalog

Variants previously associated with lung cancer survival and genotyped on the ExomeChip (MAF ≥ 5%) were identified from the NHGRI-EBI GWAS catalog (rs1878022, rs1209950, rs9981861, rs1656402, and rs716274, Table 8) (Burdett T (EBI)). For each additively coded variant, we ran a Cox proportional hazards model, controlling for age at diagnosis, sex, disease stage, treatment, and percent African ancestry. Statistical significance was determined using Bonferroni correction based on the number of *a priori* variants tested (five variants, α=0.05/5=0.01). We also examined whether smoking pack-years, education, and self-reported

COPD status modified our genetic associations. All statistical analyses were conducted using the R survival package (survival, version 2.37-7).

Discovery of novel common associations

Cox proportional hazards models were used to estimate hazard ratios assessing the association between common (MAF ≥ 5%) variants and lung cancer survival. Genotypes were coded additively and models were adjusted for age at diagnosis, sex, disease stage, treatment, and percent African ancestry. We further adjusted our models to assess the impact of smoking, education, and self-reported COPD status on our genetic associations. Functional predictions were conducted using PolyPhen-2 and SIFT gathered from the ENSEMBL website (Kumar et al. 2009, Adzhubei et al. 2010). A Bonferroni-corrected type I error rate of α=0.05/28,041 variants = 1.78×10^{-6} was set for inference of statistical 'significance'. For comparison, the number of unique tests was calculated using simpleM (Gao et al. 2008, Gao et al. 2010) and a secondary type I error rate was examined.

Rare variant analysis

Survival analysis of rare variants was conducted using the seqMeta of R (version 1.5). We examined the association of variants with lung cancer survival using both the sequence kernel association test (SKAT) and burden test. Both tests perform gene-level association tests by collapsing variants into predefined regions. For the present study, variants were aggregated into gene regions based on hg 19 transcription start and stop sites. Variants that did not map to gene boundaries were excluded from analysis. For the SKAT, variants within each gene region are weighted based on their minor allele frequency, while the burden test assumes a constant

weight for all variants. Both tests were performed for all variants and for variants with a MAF < 5%. Cox proportional hazards models were adjusted for age, sex, and percent African ancestry.

Replication

Two independent African American lung cancer study populations were available for replication of common variants: 1) the Karmanos Cancer Institute (KCI) at Wayne State University (WSU) and 2) University of California San Francisco (UCSF). Non-small cell lung cancer cases from both study populations (N=316 and N=298, respectively) were previously genotyped on the Illumina HumanHap 1M Duo array (San Diego, CA) and filtered based on standard quality control measures. Four variants meeting our p-value threshold for replication that were not directly genotyped were imputed with IMPUTE2 using a cosmopolitan reference population from 1000 Genomes (phase 1, b37, June 2014 release), with phasing performed using SHAPEIT (Howie et al. 2009). KCI/WSU African American lung cancer cases were included from three WSU studies: the Family Health Study (FHS study III), the Women's Epidemiology of Lung Disease (WELD) Study, and the EXHALE (Exploring Health, Ancestry and Lung Epidemiology) Study, and have been previously described (Schwartz et al. 2009). Cases were ascertained using rapid case ascertainment through the population-based Metropolitan Detroit Cancer Surveillance System, an NCI-funded SEER registry. Stage, treatment, and vital status were ascertained from the Detroit SEER registry. African Americans participating in the UCSF lung cancer study were identified using rapid case ascertainment methods from September 1998 to March 2003 (Wrensch et al. 2005). Specifically, cancer histology and stage were determined using ICD-O codes abstracted from SEER data from the California Cancer Registry (CCR). The CCR also provided data for whether surgery, radiation, and chemotherapy were given to the

patients, their last known vital status, and date of death or date of last contact. Stage and treatment information for both replication populations were summarized using the same algorithm as implemented within the SCCS. Percent African ancestry was estimated using genome-wide data and a supervised analysis (k=2) implemented with ADMIXTURE (Alexander et al. 2009). A fixed effects meta-analysis of the discovery (SCCS) and replication (KCI/WSU and UCSF) cohorts was performed using METAL (Willer et al. 2010).

One variant, rs1878022, was also examined in a population of self-reported whites from the SCCS. Incident lung cancer cases were identified and genotyped in the same manner as described for African Americans. In total, 129 white incident lung cancer cases met quality control criteria and were included in the analysis.

Results

Descriptive Characteristics

Using linkage with the state cancer registries, we identified a total of 336 incident African American lung cancer cases in the SCCS cohort with an available biospecimen. After implementation of QC procedures, 286 African Americans with non-small cell lung cancer (NSCLC) were included in our study (Figure 15). The mean follow-up time was 1.3 years (min=0.003, max=8.6), and 87% of individuals were deceased at the end of follow-up. Lung cancer cases were diagnosed at a mean age of 60 years and were primarily (60%) male (Table 9). Ninety-four percent of individuals were recruited from community health centers resulting in a primarily low socioeconomic status (SES) population with 68% having an annual household income less than $15,000, and nearly half (47%) of participants had less than a high school education. Fifty-two percent of African American lung cancer cases in the SCCS were diagnosed

with stage IV disease (Table 9, Figure 16). Median African ancestry was 86% (Figure 17A) among SCCS African American non-small cell lung cancer cases. African American NSCLC cases in the KCI/WSU (N=316) and UCSF (N=298) studies had lower African ancestry (83% and 82%, respectively; Figure 17B and Figure 17C), higher SES and fewer stage IV diagnoses (26% KCI/WSU and 29% UCSF; Table 10 and Figure 16) than African Americans in the SCCS. The mean follow-up time for the KCI/WSU and UCSF populations was 4.0 years (min=0.3, max=13.7) and 2.8 years (min=0.2, max=9.0), respectively. Our final sample size for complete case analysis was N=275 in the SCCS, N=312 in KCI/WSU, and N=284 in UCSF.

NHGRI-EBI GWAS Catalog Variants

Of the variants previously associated with lung cancer survival in the NHGRI-EBI GWAS Catalog, only five were genotyped on the Illumina HumanExome BeadChip and passed quality control metric (Wu et al. 2010, Sato et al. 2011, Wu et al. 2011). Variants rs1209950, rs9981861, rs1656402, and rs716274 were previously identified in East Asian populations, while rs1878022 was previously reported in a European descent population (Table 8 and Table 11). Variants had similar allele frequencies as those observed in the 1000 Genomes Americans of African Ancestry in the Southwestern US (ASW) population (Table 11). We observed that the C allele for rs1878022 in the *CMKLR1* gene was significantly associated with lower mortality after correction for multiple testing (Hazard Ratio, HR = 0.70, 95% Confidence Interval (CI): 0.54-0.92), adjusting for age, sex, stage, treatment, and African ancestry (Table 13). Though hazard ratios were in the same direction as previously reported for the remaining four variants, confidence intervals were wide and results were not significantly associated with lung cancer survival in our SCCS discovery population of African Americans. Inclusion of pack years,

education, and self-reported COPD status in our Cox proportional hazards models did not appreciably alter (<10% change) our genetic associations (Table 12). Associations of rs1878022 in the KCI/WSU and UCSF study populations had hazard ratios in the same direction as the SCCS, although not statistically significant (KCI/WSU: HR = 0.87, 95% CI: 0.66-1.14; UCSF: HR = 0.94, 95% CI: 0.72-1.24, Table 12). A fixed effects meta-analysis of the three populations revealed a summary hazard ratio of 0.83 (95% CI: 0.71-0.97, p = 0.02, I^2=0.30, Figure 18). Stratification of the SCCS cohort by stage revealed a significant association among African Americans with stage IV lung cancer (HR = 0.62, 95% CI: 0.44-0.88, p = 0.008, Table 14). Pooled analysis of stage IV individuals from KCI/WSU and UCSF resulted in a HR of 1.22 (95% CI: 0.86-1.72, Table 14) and a stage-stratified meta-analysis of the three populations (N=308) found an attenuated and non-significant association among stage IV individuals (HR = 0.83, 95% CI: 0.65-1.07). Genotype frequencies and ancestry estimates by study population and stage at diagnosis are presented in Table 15 and Table 16, respectively.

Since the C allele at rs1878022 was previously associated with increased lung cancer mortality in a European-descent population (Wu et al. 2011), we also examined the effects of rs1878022 in a population of whites from the SCCS. In total, 129 white incident NSCLC cases passed quality control and were included in our analysis. While the hazard ratio was reduced (HR=0.88), the confidence intervals were wide (95% CI=0.64-1.19), providing no for an association between rs1878022 and lung cancer survival in whites.

ExomeChip Common Variant Associations

We examined common variants associated with lung cancer survival among African Americans. After quality control, 28,041 common variants remained for analysis, of which

simpleM identified 23,653 unique tests. No variant was considered significant based on a Bonferroni correction for the number of tests (N=28,041) or for the number of unique tests (N=23,653). However, thirteen variants had p-values < $1.0x10^{-4}$, Table 17 and Figure 19. Of the 13 variants with the smallest p-values, seven were associated with greater all-cause mortality and six were associated with reduced all-cause mortality in additive models with increasing copies of the minor allele. All SNPs had similar allele frequencies to those observed in the 1000 Genomes ASW reference population (Table 18). SNP rs1133358, a variant in the *SUN5* gene on chromosome 20q11, was the most strongly associated with mortality (HR = 0.61, 95% CI: 0.49-0.76) and predicted by SIFT to be deleterious to protein function, though with low confidence (Table 17 and Table 19). Four of the 13 variants (rs2072633, rs1505229, rs7502216, rs537160) were located within gene introns and an additional five (rs1133358, rs8176785, rs7626962, rs35761244, rs1639122) were exonic. SNP rs7626962 was also predicted to have deleterious effects by SIFT (Table 19). A peak on chromosome 6p21.33 revealed three variants (rs605203, rs2072633, rs537160, $r^2 < 0.5$) associated with reduced mortality (HR=0.46, 95% CI: 0.33-0.66; HR=0.66, 95% CI: 0.54-0.81; HR=0.61, 95% CI: 0.48-0.78, respectively). Conditional analysis, which examined all two- and three-way combinations of the three SNPs at chromosome 6p21.33, identified no single variant solely responsible for the association at this locus. P-values were attenuated for all variants; however, rs605203 remained consistently significant (Table 20). Since the three reported SNPs are in relatively weak LD ($r^2<0.5$) with each other it is unsurprising that conditional analysis did not reveal the 'causal' signal. We hypothesize that all three variants are in LD with the true causal variant, which is driving the observed association of these three variants in the present analysis.

Adding smoking pack-years, education, and self-reported COPD status to our statistical models did not substantially alter the observed hazard ratios (Table 21). We further examined these 13 variants in our KCI/WSU and UCSF replication populations (Table 17). Within the KCI/WSU study population, the G allele at rs7302017 was associated with increased mortality (HR: 1.29, 95% CI: 1.01-1.63), while the G allele at rs605203 was associated with reduced mortality in the UCSF study population (HR: 0.55, 95% CI: 0.38-0.78). Both of these associations have the same magnitude and direction of effect as the discovery cohort.

Rare Variant Analysis

After quality control, 85,018 variants, including 62,077 with MAF < 5%, mapped to gene boundaries and were non-monomorphic in the final testing population of 275 African Americans with incident NSCLC. A total of 15,279 genes were examined by both the SKAT and burden test when no restriction based on MAF was applied. The mean number of variants per gene was 5.6, with a maximum of 341. When analysis was restricted to rare variants (MAF <5%), 14,015 genes were tested, with the mean of 4.4 SNPs per gene (max. 283). The distribution of the number of variants per gene is presented in Figure 20. Cumulative allele frequencies of all variants within a gene region ranged from 0.0018 to 16.04 (median=0.1855) when all variants were examined and 0.0018 to 2.075 (median=0.0255) when the variants were restricted to those with a MAF<5%.

Gene-level SKAT results for all variants and for variants with a MAF<5% are presented in Figure 21 and Figure 22, respectively. No gene surpassed a Bonferroni corrected significance threshold and restriction based on MAF did not drastically alter the results, with eight of the ten most significant genes identified in both analyses. Burden tests results for all variants and for variants with a MAF<5% are presented in Figure 23 and Figure 24. There was substantially less

consistency between burden tests results with and without restriction based on MAF, with only three genes identified in both analyses. The burden test inclusive of all variants identified one gene, *PLIN3*, surpassing a Bonferroni significance p-value threshold ($p=7.5x10^{-7}$, HR=0.65, 95% CI: 0.55-0.77). Restriction to rare variants reduced the cumulative allele frequency from 0.46 to 0.02 and resulted a near complete attenuation of the signal at *PLIN3* (p=0.04).

Discussion

We sought to validate lung cancer survival variants previously identified in Asian or European populations in an incident study of African Americans. We additionally sought to identify novel common variants associated with lung cancer survival in African Americans and performed a preliminary examination of rare variants. Our evaluation of five previously identified lung cancer survival GWAS SNPs identified one variant in the *CMKLR1* gene significantly associated with reduced mortality in African Americans with NSCLC. The remaining four variants were not significantly associated. The observed reduced mortality for rs1878022 is in contrast to a prior study by Wu and colleagues conducted in European Americans in which they reported that the minor allele of rs1878022 was associated with increased mortality (Wu et al. 2011). All individuals in the study by Wu *et al*. were white ever-smokers with stage III or IV NSCLC who received platinum-based chemotherapy and no surgery; whereas, in the present study of African Americans no restriction was made on stage, treatment, or smoking status for study inclusion. Our analysis of white incident cases from the SCCS revealed no significant association between rs1878022 and survival. Given the reduced power resulting from our small population of whites, we are unable to verify the population specific effects. Of note, the frequency of the C allele in the 1000 Genomes populations differs

between European and African populations (35% versus 13%). The observed allele frequency of 14% in the present study of SCCS African Americans is similar to the reported frequency (18%) in the 1000 Genomes African American (ASW) population and in the UCSF and WSU lung cancer cases (17.4% and 18.8%, respectively). While the observed population-specific effects of rs1878022 may simply be the result of a spurious association, it could also result from differences in linkage disequilibrium (LD) patterns between Africans and Europeans as a reflection of distinct backbone haplotypes that could have different causal variants. Examination of LD structure surrounding the *CMKLR1* gene revealed differences between the YRI and CEU populations, with the CEU having larger blocks of strong LD (Figure 25). Using LDlink (Machiela and Chanock 2015) and data from an African (YRI) population, we find five variants are in strong LD with rs1878022 (r^2>0.8) and an additional eight variants in moderate LD (r^2>0.3 and r^2 < 0.8) while six variants are in moderate LD in whites (CEU) and none in strong LD. RegulomeDB (Boyle et al. 2012) reveals two of the variants in moderate LD in the YRI (rs4964244 and rs4964245, r^2=0.44 and 0.38, respectively) and one variant in the CEU (rs4964242, r^2=0.55) have high regulatory potential and are likely to affect binding. Variant rs4964242 is not in LD with rs4964244 or rs4964245 in either the CEU or YRI population (r^2≤0.1), indicating that different regulatory variants could account for the observed population-specific effects. It is possible that rs1878022 is tagging a different causal variant in the YRI population than in the CEU population and thus it remains necessary to fine map the *CMKLR1* gene to determine the causal variant influencing lung cancer survival in African Americans compared to European Americans.

SNP rs1878022 is located within the second intron within the 5'-untranslated region of the chemokine-like receptor 1 (*CMKLR1*) gene on chromosome 12q23.3. *CMKLR1* encodes a

seven transmembrane G-protein coupled receptor that has been associated with adiposity, glucose intolerance, and inflammation (Rama et al. 2011, Ernst et al. 2012). It has also been shown to play a role in the immune response to cigarette smoke in murine models of chronic obstructive pulmonary disease, an established risk factor for lung cancer risk (Demoor et al. 2011). Furthermore, in response to binding of its ligand, chemerin, *CMKLR1* has been shown to activate MAPK, ERK1/2, and Akt signaling cascades involved in cell cycle regulation (Kaur et al. 2010). Examination of ENCODE data at rs1878022 identified a small peak of H3K27Ac and H3K4me1 in normal human lung fibroblasts (Figure 26). Epigenetic markers found surrounding rs1878022 are commonly found at active regulatory elements and suggest that this variant might play a role in the regulation of *CMKLR1* activity.

It is important to note that the present analyses were limited by the design of the genotyping array used in the discovery population. The Illumina HumanExome BeadChip was designed to capture biologically-relevant coding variation. As a result, variants were not evenly distributed across the genome and were predominantly rare in frequency (MAF < 5%). This array also includes >4,500 variants identified in recent genome-wide association studies. However, while there were >20 variants associated with lung cancer survival in the NHGRI-EBI GWAS catalog at the time of our data extraction, only 5 were present on the genotyping array for examination in the current study. This discrepancy correlates with the date variants were deposited into the NHGRI-EBI GWAS Catalog verses the creation of the array. Furthermore, the unique design of the exome array precluded us from analyzing variants in LD with the five previously reported variants examined here. Examination of LD patterns surrounding the five NHGRI-EBI GWAS variants in the relevant 1000 Genomes reference population revealed four variants in LD (+/-100 kb, r^2>0.6, CEU reference population) with rs1878022, the only variant

previously identified in a European descent population, and 57 variants in LD (+/-100 kb, r^2>0.6, Han Chinese in Beijing, China (CHB) reference population) with one of the four variants identified in an Asian descent population; none of these variants were present on the exome array. Future analyses should perform a thorough examination of LD structure surrounding these variants and include present in the NHGRI-EBI GWAS catalog but not present here.

While our analysis of common variants did not identify any significant variants, several variants had promising p-values for associations with mortality, including multiple protein coding variants. Three variants on chromosome 6p21.33 in weak LD (rs605203, rs2072633, and rs537160, r^2<0.5) were associated with a reduction in mortality. All three variants were non-protein coding. Variant rs605203, which had a similar hazard ratio in the UCSF population, is located 70bp upstream of the *SLC44A4* gene and downstream of *EHMT2*. Variants rs2072633 and rs537160 are located within the 29th and 19th intron of the complement factor B (*CFB*) gene, respectively. Chromosome 6p21.33 is part of the gene-rich human leukocyte antigen (HLA) region. The 6p22.1-p21.31 chromosome region has been previously associated with lung cancer risk in both European and African Americans, though these results have inconsistently replicated (Truong et al. 2010, Zhang et al. 2010, Bae et al. 2012, Timofeeva et al. 2012, Walsh et al. 2013). Furthermore, rs4324798 located at 6p22.1 was previously associated with increased survival in a European descent population of never-smoking small cell lung cancer cases (Yang et al. 2010, Xun et al. 2011). Lung cancer susceptibility within this region has been largely attributed to the *BAT3* and *MSH5* genes, which play an important role in DNA damage response and potentially response to cancer therapeutics (Sasaki et al. 2007, Tompkins et al. 2009, Wu et al. 2013). Our findings support an association of the 6p22.1-p21.33 region with overall survival in African American NSCLC cases.

We identified variant rs1639122 on chromosome 12, located within the chromodomain helicase DNA binding protein 4 (*CHD4*). A change from the A to C allele results in a missense mutation that changes the amino acid from glutamate to aspartate. *CHD4* is involved in nucleosome remodeling and transcriptional silencing as part of the nucleosome remodeling and deacetylation (NuRD) complex (Tong et al. 1998, Xue et al. 1998). Additionally, *CHD4* plays a role in DNA damage response and cell cycle progression (Polo et al. 2010, Smeenk et al. 2010). Somatic mutations in *CHD4* were identified in endometrial tumors and a germline protein coding variant was associated with increased risk of overall cancer, malignant lymphoma, rectal cancer, and lung cancer, providing further evidence for a potential role of this gene in lung cancer survival (Le Gallo et al. 2012, Zhao et al. 2013, Yamada et al. 2015). While the majority of these common variants failed to replicate in the KCI/WSU and UCSF populations, the totality of evidence from previously published associations with lung cancer risk and biological plausibility suggest our findings should be further investigated in additional African American lung cancer populations.

Analysis of rare variants utilized two common collapsing methods: the SKAT and burden test (Li and Leal 2008, Wu et al. 2011). Both tests attempt to increase statistical power by aggregating variants based on gene boundaries, which reduces the number of overall tests and increases the frequency of variation for a gene region. However, despite the employment of these methods we acknowledge that the present study is still underpowered to detect rare variant associations and, as such, the analysis of rare variants presented in this chapter is solely exploratory in nature. While the SKAT and burden test are both collapsing methods, the two tests differ in several key assumptions. Specifically, the burden test assumes that all variants within a gene region have the same magnitude and direction of effect. SKAT, on the other hand,

allows different effects for variants within the same gene region. Among all rare variant tests, only one gene was significantly associated with reduced mortality, *PLIN3,* though this association was most certainly driven by common variants, as the signal was almost completely attenuated upon restriction to rare variants (MAF<5%). *PLIN3,* or perilipin 3, is located on chromosome 19p13.3 and is involved in lipid oxidation in skeletal muscle (Covington et al. 2015). *PLIN3* has also been observed to be involved in endosomal localization of *RAB9*, a member of the RAS oncogene family (Aivazian et al. 2006).

To our knowledge, this study is the first to examine genome-wide genetic variants associated with all-cause mortality among African American lung cancer cases. While we acknowledge our sample size is small, these analyses serve as a starting point for further investigation. With our limited sample size of African American lung cancer cases we have reduced statistical power to detect true associations and thus replication in a larger population of African American lung cancer cases is necessary. However, we had *a priori* evidence to assess variants previously associated with lung cancer survival, thus reducing the burden of multiple testing typical of genetic association studies. Importantly, assessing trans-ethnic replication of the NHGRI-EBI GWAS Catalog lung cancer survival variants provides genetic information about lung cancer desperately needed among diverse racial/ethnic populations. Moreover, meta-analysis of the discovery and replication cohorts provides additional support for this association. Our findings provide greater evidence of an association between the *CMKLR1* gene and survival among lung cancer cases.

While the two replication populations (KCI/WSU and UCSF) were similar, we note key differences between the discovery (SCCS) and replication study populations likely due to differences in the composition of the study base from which the lung cancer cases arose. The

discovery and the replication lung cancer cases were all sampled using cancer registries within primary study bases; however, the study base of the SCCS is defined by its prospective cohort design of primarily individuals seeking medical care at community health centers across the Southeastern United States and KCI/WSU and UCSF are defined by geographic regions encompassing the Detroit and the San Francisco Bay Area metropolitan areas, respectively. These different geographic samplings likely contributed to important differences across the three studies in education level, current smoking prevalence, and stage distribution. Specifically, a greater percentage of SCCS lung cancer cases were diagnosed at later stages of disease compared to KCI/WSU and UCSF cases (Figure 16). This difference in stage distribution between discovery and replication studies could be attributed to differences in cohort vs. case-control ascertainment since case-control designs often miss the sickest individuals. The observed difference in results across the discovery and replication studies among stage IV lung cancer cases for the association between rs1878022 and survival may be due to unique gene-environment interactions or simply a matter of chance.

In summary, this study identified several variants associated with survival in African Americans lung cancer cases and controls, a high-risk and underrepresented population in lung cancer genetics research. By examining variants previously associated with lung cancer survival, we observed that rs1878022 is significantly associated with survival in African Americans. However, the direction of effect observed is in contrast to a previous study in Europeans, suggesting rs1878022 may have populations-specific effects on lung cancer survival due to possible differing causal variants between European descent populations and African Americans due to distinct LD substructure (Wu et al. 2011). Additionally, we identified several potential novel variants, both protein coding and non-coding, associated with survival in African

Americans, including a region on chromosome 6p21.33-p22.1 that has been previously associated with lung cancer risk and small cell lung cancer survival. Future studies fine-mapping *CMKLR1* and the 6p21-22 region and conducting functional studies could lead to potential therapeutic interventions to improve lung cancer survival, especially in African Americans.

Table 8. Summary of the five previously reported lung cancer survival GWAS variants examined in the present study.

SNP	Study	PubMed ID	Population	N	Risk Allele	HR	95% CI	p value
rs1878022	Wu X, et al. JNCI 2011	21483023	European	327	C	1.59	(1.32-1.92)	$1.4x10^{-6}$
rs1209950	Sato Y, et al. JTO 2010	21079520	East Asian	105	T	4.96	(2.52-9.76)	$2.8x10^{-7}$
rs9981861	Sato Y, et al. JTO 2010	21079520	East Asian	105	G	16.1	(5.38-51.2)	$3.5x10^{-6}$
rs1656402	Sato Y, et al. JTO 2010	21079520	East Asian	105	G	4.22	(2.32-7.66)	$8.4x10^{-8}$
rs716274	Wu C, et al. CR 2010	21118971	East Asian	245	G	1.92	(1.51-2.45)	$1.3x10^{-7}$

Table 9. Descriptive characteristics of African Americans with incident non-small cell lung cancer participating in the Southern Community Cohort Study (N=286)

	N (%)
Sex	
Male	171 (59.8)
Female	115 (40.2)
Vital status	
Alive	38 (13.3)
Dead	248 (86.7)
Median African ancestry, %	85.6
Lung cancer stage at diagnosis	
I	44 (15.7)
II/III	90 (32.0)
IV	147 (52.3)
Unknown	5
Treatment	
No treatment	75 (26.8)
Surgery only	31 (11.1)
Chemotherapy only	51 (18.2)
Radiation only	36 (12.9)
Multi-modality	87 (31.1)
Unknown	6
Histology	
Adenocarcinoma	113 (39.5)
Non-small cell lung cancer-NOS[a]	78 (27.3)
Squamous	72 (25.2)
Other NSCLC	22 (7.7)
Multiple histologies	1 (0.3)
Mean age at diagnosis, yr (SD)	59.6 (9.1)
Mean duration of disease among those who died, yr (SD)	0.88 (1.1)
Mean duration of disease among those alive at last follow-up, yr (SD)	4.1 (1.7)
Highest education level, yr	
<12	134 (47.2)
≥12	150 (52.8)
Unknown	2
Smoking status at cohort entry	
Current	206 (72.8)
Former	59 (20.8)
Never	18 (6.4)
Unknown	3
Mean cigarettes per day (SD)	15.7 (13.1)
Mean smoking pack-years (SD)	37.5 (30.7)
First-degree relative with lung cancer	
Yes	21 (9.3)
No	205 (90.7)
Unknown	60

NOS = not otherwise specified; SD = standard deviation; Yr = years

Table 10. Descriptive characteristics of African Americans with NCSLC participating in the KCI/WSU and UCSF studies.

	KCI/WSU (N=316) N (%)	UCSF (N=298) N (%)
Sex		
Male	129 (40.8)	139 (46.6)
Female	187 (59.2)	159 (53.4)
Vital status		
Alive	107 (33.9)	102 (34.2)
Dead	209 (66.1)	196 (65.8)
Median African ancestry, %	83.3	82.1
Lung cancer stage at diagnosis		
I	107 (34.3)	106 (37.7)
II/III	123 (39.4)	95 (33.5)
IV	82 (26.3)	83 (28.9)
Unknown	4	14
Treatment		
No treatment	27 (8.5)	37 (12.4)
Surgery only	85 (26.9)	89 (29.9)
Chemotherapy only	32 (10.1)	35 (11.7)
Radiation only	26 (8.2)	44 (14.8)
Multi-modality	146 (46.2)	93 (31.2)
Histology		
Adenocarcinoma	159 (50.3)	127 (42.6)
Non-small cell lung cancer-NOS[a]	79 (25.0)	74 (24.8)
Squamous	71 (22.5)	77 (25.8)
Other NSCLC	7 (2.2)	15 (5.0)
Multiple histologies		5 (1.7)
Mean age at diagnosis, yr (SD)	60.3 (11.0)	63.1 (10.8)
Mean duration of disease among those who died, yr (SD)	2.6 (1.8)	1.9 (1.6)
Mean duration of disease among those alive at last follow-up, yr (SD)	6.8 (2.1)	4.5 (2.4)
Highest education level, yr		
<12	89 (28.2)	72 (25.2)
≥12	227 (71.8)	214 (74.8)
Unknown	0	12
Smoking status		
Current	177 (56.0)	95 (31.9)
Former	123 (38.9)	183 (61.4)
Never	16 (5.1)	20 (6.7)
Mean cigarettes per day (SD)	18.4 (13.3)	20.5 (26)
Mean smoking pack years (SD)	36.9 (28.6)	7.5 (15.2)
First-degree relative with lung cancer		
Yes	62 (19.6)	55 (18.5)
No	254 (80.4)	242 (81.5)
Unknown	0	1

NOS = not otherwise specified; SD = standard deviation; Yr = years

Table 11. Allele frequency for the five variants previously associated with lung cancer survival in the NHGRI-EBI GWAS catalog in the SCCS (N=275), KCI/WSU (N=312), and UCSF (N=284) populations. Allele frequencies are also presented for the ASW 1000 Genomes African American reference population and for the 1000 Genomes reference population that matches the racial/ethnic group of the original discovery population (CEU or CHB).

SNP	Allele[a]	SCCS MAF	KCI/WSU MAF	UCSF MAF	ASW[b] MAF	Discovery GWAS population	CEU/CHB[b] MAF[c]
rs1878022	C/T	0.14	0.19	0.17	0.19	European	0.30
rs1209950	T/C	0.14	0.14	0.16	0.16	East Asian	0.04
rs9981861	G/A	0.29	0.35	0.35	0.34	East Asian	0.16
rs1656402	A/G	0.42	0.41	0.36	0.39	East Asian	0.28
rs716274	G/A	0.43	0.47	0.49	0.54	East Asian	0.20

Table 12. Multivariable Cox proportional hazards results for African Americans with incident non-small cell lung cancer and the five SNPs previously associated with lung cancer survival in the NHGRI-EBI GWAS catalog assayed by the Illumina HumanExome BeadChip v1.1.

				SCCS (N=275)			KCI/WSU (N=312)			UCSF (N=284)		
SNP	**Chr.**	**bp Position[a]**	**Gene[b]**	**HR**	**95% CI**	***p* value**	**HR**	**95% CI**	***p* value**	**HR**	**95% CI**	***p* value**
rs1878022	12q23.3	108699032	*CMKLR1*	0.70	(0.54-0.92)	$9.8x10^{-3}$	0.87	(0.66-1.14)	0.30	0.94	(0.71-1.24)	0.66
rs1209950	21q22.2	40173528	|| *ETS2*	1.16	(0.90-1.50)	0.24	0.93	(0.70-1.25)	0.65	0.91	(0.68-1.22)	0.53
rs9981861	21q22.2	41415044	*DSCAM*	1.08	(0.88-1.33)	0.46	1.13	(0.91-1.39)	0.27	1.15	(0.91-1.45)	0.24
rs1656402	2q37.1	233426526	*EIF4E2*	0.93	(0.76-1.14)	0.48	0.90	(0.72-1.11)	0.30	1.11	(0.86-1.43)	0.41
rs716274[c]	11q22.3	103418158	*DYNC2H1* ||	1.02	(0.84-1.23)	0.86	0.97	(0.80-1.17)	0.72	1.21	(0.97-1.50)	0.08

Models are adjusted for age, sex, treatment, stage, and percent African ancestry.
[a]dbSNP build 137/GRCH37.p5
[b]For variants outside of gene boundaries, || denotes the location of the variant relative to the closest gene.
[c]Imputed in KCI/WSU and USCF populations. Imputation score=0.99.
HR = hazard ratio, CI = confidence interval

Table 13. Adjusted multivariable Cox proportional hazards model results for African Americans with incident non-small cell lung cancer participating in the Southern Community Cohort Study (N=264) for the five candidate SNPs previously associated with lung cancer survival from the NHGRI GWAS catalog.

SNP	HR	95% CI	*p* value
rs1878022	0.68	(0.51-0.89)	$5.5x10^{-3}$
rs1209950	1.17	(0.90-1.51)	0.24
rs9981861	1.10	(0.89-1.36)	0.40
rs1656402	0.93	(0.76-1.14)	0.50
rs716274	1.00	(0.82-1.22)	0.99

Models are adjusted for smoking pack-years, education status, self-reported COPD, age, sex, treatment, stage, and percent African ancestry.
HR = hazard ratio, CI = confidence interval

Table 14. Stage-stratified multivariable Cox proportional hazards results for the association of rs1878022 and overall survival in the SCCS and pooled KCI/WSU and UCSF populations.

	SCCS (N=275)					Pooled KCI/WSU and UCSF (N=596)				
Stage	N	MAF	HR	95% CI	p value	N	MAF	HR	95% CI	p value
I	44	0.15	0.65	(0.24-1.74)	0.39	214	0.19	0.99	(0.68-1.46)	0.98
II/III	87	0.13	1.18	(0.67-2.05)	0.57	218	0.19	0.79	(0.59-1.07)	0.13
IV	144	0.15	0.62	(0.44-0.88)	7.6x10-3	164	0.16	1.22	(0.86-1.72)	0.27

Models are adjusted for age, sex, treatment, stage, and percent African ancestry. Pooled KCI/WSU and UCSF models are further adjusted for study.
HR = hazard ratio, CI = confidence interval

Table 15. Genotype frequencies for rs1878022 by stage and cohort.

	TT N (%)	TC N (%)	CC N (%)
Stage I			
SCCS	31 (70.5)	13 (29.5)	0 (0)
WSU	70 (65.4)	32 (29.9)	5 (4.8)
UCSF	72 (67.9)	31 (29.2)	3 (2.8)
Stage II/III			
SCCS	66 (75.9)	20 (23.0)	1 (1.1)
WSU	82 (67.2)	36 (29.5)	4 (3.3)
UCSF	61 (64.2)	28 (29.5	6 (6.3)
Stage IV			
SCCS	104 (72.2)	38 (26.4)	2 (1.4)
WSU	52 (63.4)	29 (35.4)	1 (1.2)
UCSF	62 (74.7)	20 (24.1)	1 (1.2)

Table 16. Median African ancestry (%, IQR) for each study by stage at diagnosis.

	SCCS % [IQR]	WSU % [IQR]	UCSF % [IQR]
Stage I	85.9 [76.9, 92.5]	83.1 [73.4, 90.5]	79.7 [70.7, 88.4]
Stage II/III	85.6 [80.4, 89.3]	83.1 [77.5, 89.8]	84.0 [75.2, 88.8]
Stage IV	86.4 [80.2, 98.8]	83.6 [72.5, 88.4]	83.3 [74.4, 89.3]

IQR=interquartile range

Table 17. Multivariable Cox proportional hazard results for African Americans with incident non-small cell lung cancer and common variants assayed by the Illumina HumanExome BeadChip v1.1. Results are show for associations with a *p* value <1.0x10^{-4}.

				SCCS (N=275)			KCI/WSU (N=312)			UCSF (N=284)		
SNP	**Chr**	**bp Position[a]**	**Gene[b]**	**HR**	**95% CI**	***p* value**	**HR**	**95% CI**	***p* value**	**HR**	**95% CI**	***p* value**
rs1133358	20q11.21	31590686	*SUN5*	0.61	(0.49-0.76)	8.12x10^{-06}	1.01	(0.82-1.24)	0.94	1.08	(0.87-1.35)	0.49
rs8176785	11p15.1	20805286	*NELL1*	1.65	(1.32-2.07)	1.20x10^{-05}	0.98	(0.79-1.21)	0.82	1.11	(0.87-1.42)	0.39
rs7626962	3p22.2	38620907	*SCN5A*	1.95	(1.44-2.64)	1.44x10^{-05}	1.05	(0.75-1.46)	0.78	0.80	(0.53-1.22)	0.30
rs605203	6p21.33	31847012	*SLC44A4* || *EHMT2*	0.46	(0.33-0.66)	1.63x10^{-05}	1.13	(0.82-1.57)	0.43	0.55	(0.38-0.79)	0.001
rs35761244[c]	19p13.2	13873698	*CCDC130*	1.66	(1.32-2.09)	1.68x10^{-05}	0.98	(0.70-1.36)	0.90	0.94	(0.68-1.32)	0.73
rs6959964[c]	7q11.22	68905738	|| *AUTS2*	1.56	(1.27-1.91)	2.21x10^{-05}	0.84	(0.67-1.04)	0.10	0.88	(0.71-1.10)	0.27
rs1639122	12p13.31	6711147	*CHD4*	1.80	(1.36-2.38)	3.38x10^{-05}	0.93	(0.74-1.19)	0.58	1.05	(0.78-1.40)	0.76
rs7138803	12q13.12	50247468	*BCDIN3D* || *FAIM2*	1.70	(1.32-2.18)	3.78x10^{-05}	0.96	(0.74-1.24)	0.74	0.90	(0.69-1.19)	0.48
rs2072633	6p21.33	31919578	*CFB*	0.66	(0.54-0.81)	4.07x10^{-05}	1.00	(0.81-1.23)	0.98	0.91	(0.73-1.13)	0.40
rs1505229	2p12	77589901	*LRRTM4*	0.67	(0.56-0.82)	5.44x10^{-05}	0.96	(0.78-1.19)	0.72	1.04	(0.84-1.30)	0.72
rs7502216	17q12	36612948	*ARHGAP23*	0.67	(0.55-0.81)	5.78x10^{-05}	0.85	(0.69-1.04))	0.11	1.10	(0.89-1.36)	0.40
rs537160	6p21.33	31916400	*CFB*	0.61	(0.48-0.78)	7.44x10^{-05}	0.98	(0.74-1.29)	0.87	0.81	(0.61-1.06)	0.13
rs7302017[c]	12q14.1	63004583	*MIRLET7I* || *PPM1H*	1.53	(1.24-1.89)	9.28x10^{-05}	1.29	(1.01-1.63)	0.04	0.87	(0.67-1.13)	0.30

Models are adjusted for age, sex, treatment, stage, and percent African ancestry.
[a]dbSNP build 137/GRCH37.p5
[b]For variants outside of gene boundaries, || denotes the location of the variant relative to the closest gene(s).
[c]Imputed in KCI/WSU and USCF populations. Imputation score=0.90 for rs35761244, =0.98 for rs6959964, and =0.99 for rs7302017.
HR = hazard ratio, CI = confidence interval

Table 18. Allele frequency for common variants associated with survival in the SCCS (N=275), KCI/WSU (N=312), and UCSF (N=284) populations. Allele frequencies are also presented for the ASW 1000 Genomes African American reference population.

SNP	Allele[a]	SCCS MAF	KCI/WSU MAF	UCSF MAF	ASW[b] MAF
rs1133358	A/C	0.27	0.26	0.30	0.24
rs8176785	A/G	0.27	0.31	0.29	0.30
rs7626962	T/G	0.09	0.10	0.07	0.06
rs605203	G/T	0.10	0.11	0.11	0.14
rs35761244	C/G	0.17	0.13	0.14	0.14
rs6959964	T/C	0.40	0.39	0.42	0.48
rs1639122	C/A	0.15	0.20	0.18	0.19
rs7138803	A/G	0.17	0.19	0.19	0.19
rs2072633	T/C	0.35	0.40	0.42	0.36
rs1505229	T/C	0.47	0.46	0.49	0.48
rs7502216	C/A	0.33	0.33	0.35	0.31
rs537160	T/C	0.17	0.18	0.20	0.22
rs7302017	G/A	0.27	0.26	0.24	0.30

[a]Effect/referent allele
[b]Minor allele frequency in 1000 Genomes Phase 3 African American samples.

Table 19. Functional predictions (PolyPhen-2, SIFT) of amino acid changes for protein coding variants with a *p* value < 1.0x10-4 from the common variant analysis among African Americans with incident NSCLC participating in the SCCS (N=275).

SNP	Gene[a]	Amino Acid Change	PolyPhen-2 Prediction	SIFT Prediction
rs1133358	*SUN5*	E > D	Benign	Deleterious
rs8176785	*NELL1*	R > Q	Benign	Tolerated
rs7626962	*SCN5A*	S > Y	Benign	Deleterious
rs35761244	*CCDC130*	C > S	Benign	Tolerated
rs1639122	*CHD4*	E > D	Benign	Tolerated

Table 20. Conditional analysis results for variants on chromosome 6q21.33. Results are presented for the Cox proportional hazards model that includes all three variants.

SNP	HR	95% CI	p-value
rs605203	0.56	(0.36-0.87)	$1.0x10^{-2}$
rs2072633	0.77	(0.59-1.00)	$4.7x10^{-2}$
rs537160	0.98	(0.67-1.42)	0.9

Model is adjusted for age at diagnosis, sex, African ancestry, stage, and treatment.

Table 21. Adjusted multivariable Cox proportional hazard results for African Americans with incident NSCLC in the SCCS (N=264) for common variants assayed by the Illumina HumanExome BeadChip v1.1.

SNP	HR	95% CI	*p* value[a]
rs1133358	0.60	(0.48-0.75)	$5.8x10^{-6}$
rs8176785	1.69	(1.33-2.14)	$1.3 x10^{-5}$
rs7626962	1.96	(1.43-2.68)	$3.0x10^{-5}$
rs605203	0.46	(0.32-0.66)	$2.5x10^{-5}$
rs35761244	1.74	(1.35-2.24)	$1.4x10^{-5}$
rs6959964	1.60	(1.29-1.98)	$1.4x10^{-5}$
rs1639122	1.83	(1.38-2.44)	$2.7x10^{-5}$
rs7138803	1.68	(1.29-2.19)	$1.1x10^{-4}$
rs2072633	0.64	(0.52-0.79)	$2.2x10^{-5}$
rs1505229	0.65	(0.53-0.80)	$4.1x10^{-5}$
rs7502216	0.66	(0.54-0.81)	$5.8x10^{-5}$
rs537160	0.60	(0.47-0.77)	$6.6x10^{-5}$
rs7302017	1.56	(1.25-1.95)	$9.6x10^{-5}$

Models are adjusted for smoking pack-years, education status, self-reported COPD status, age, sex, treatment, stage, and percent African ancestry.
[a]Sorted based on *p* value presented in Table 3
HR = hazard ratio, CI = confidence interval

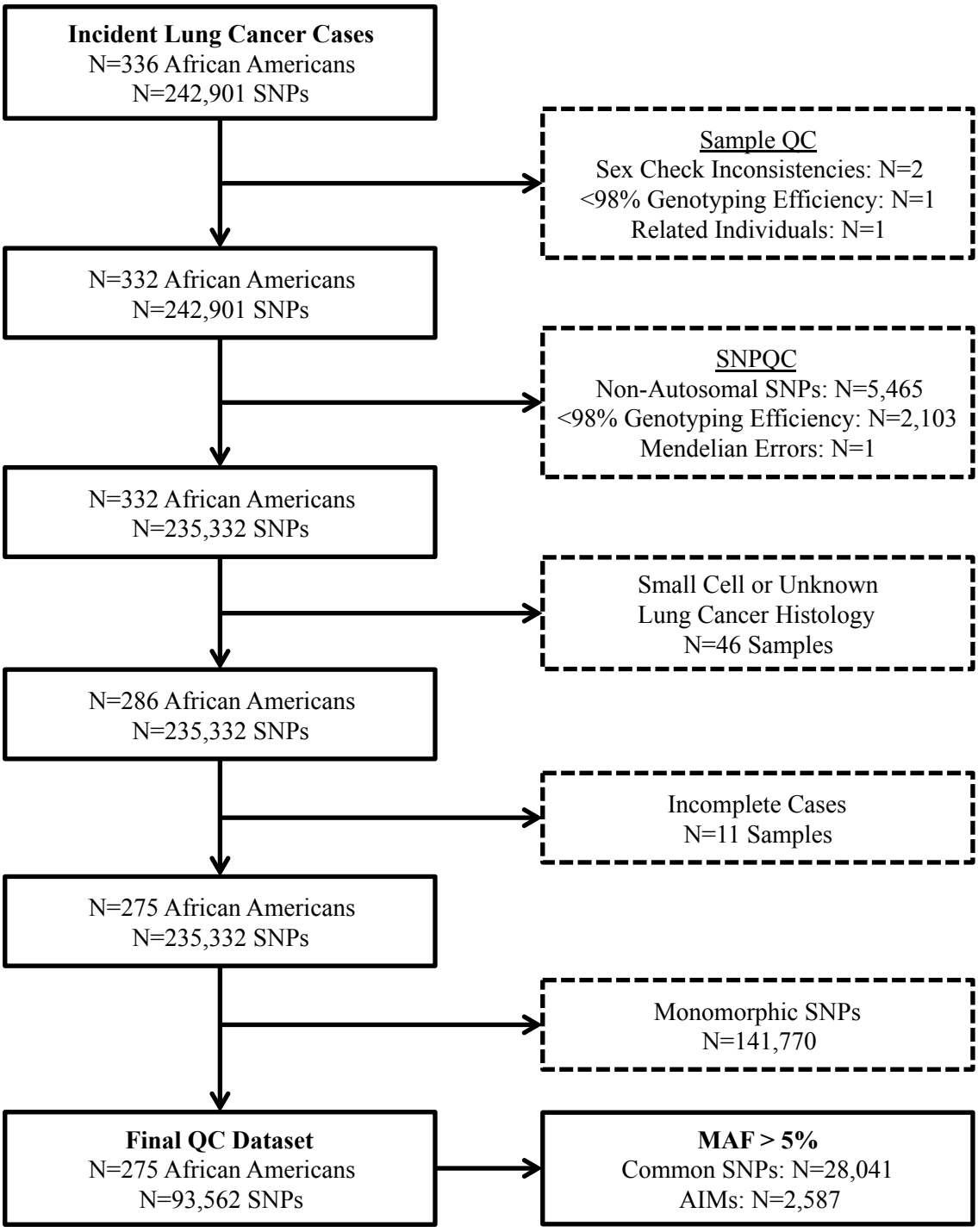

Figure 15. Flow chart of quality control (QC) procedures performed on individuals and variants in the SCCS and genotyped on the Illumina HumanExome BeadChip.

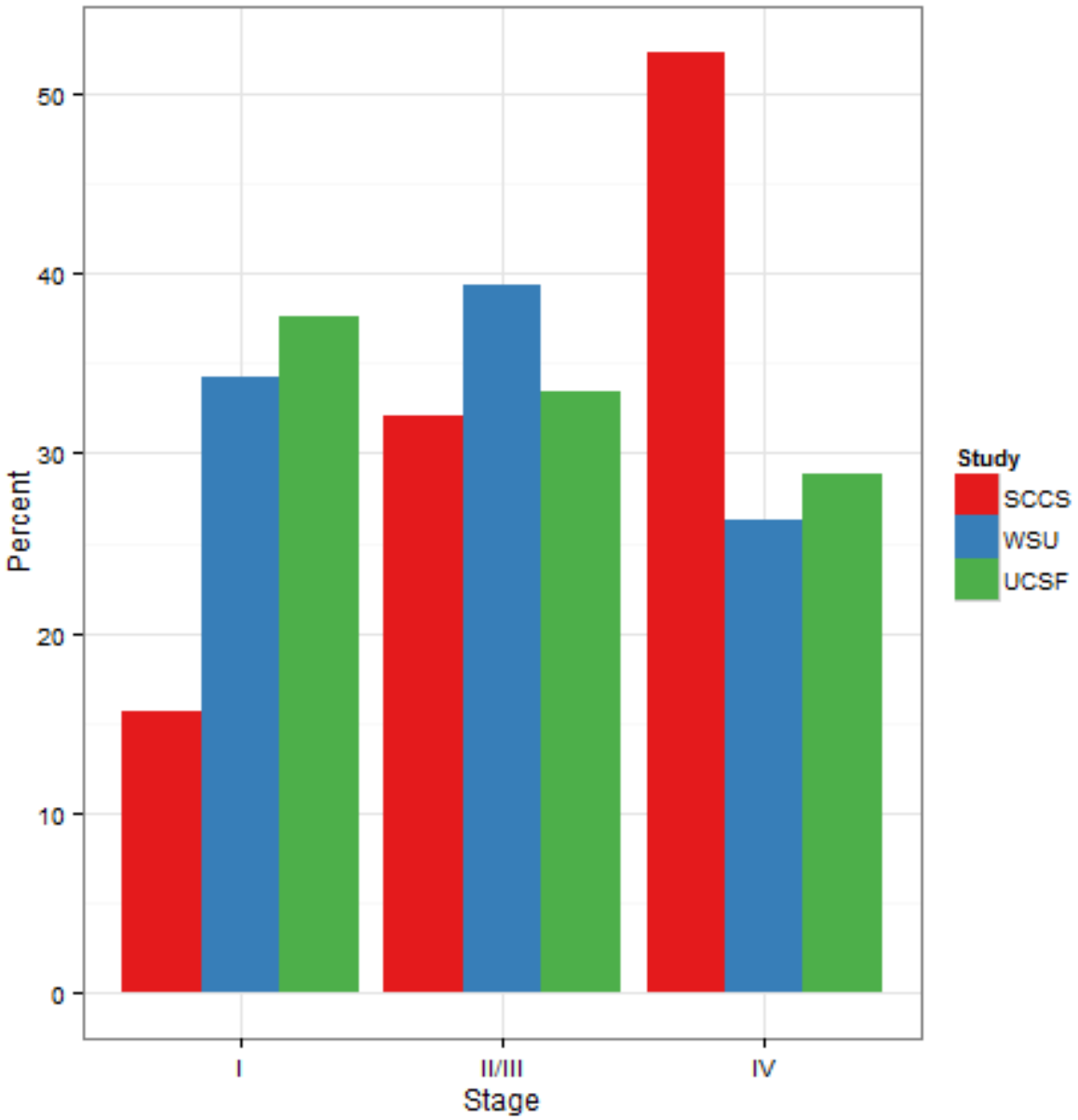

Figure 16. Distribution of stage at diagnosis across the SCCS (N=286), KCI/WSU (N=316), and UCSF (N=298).

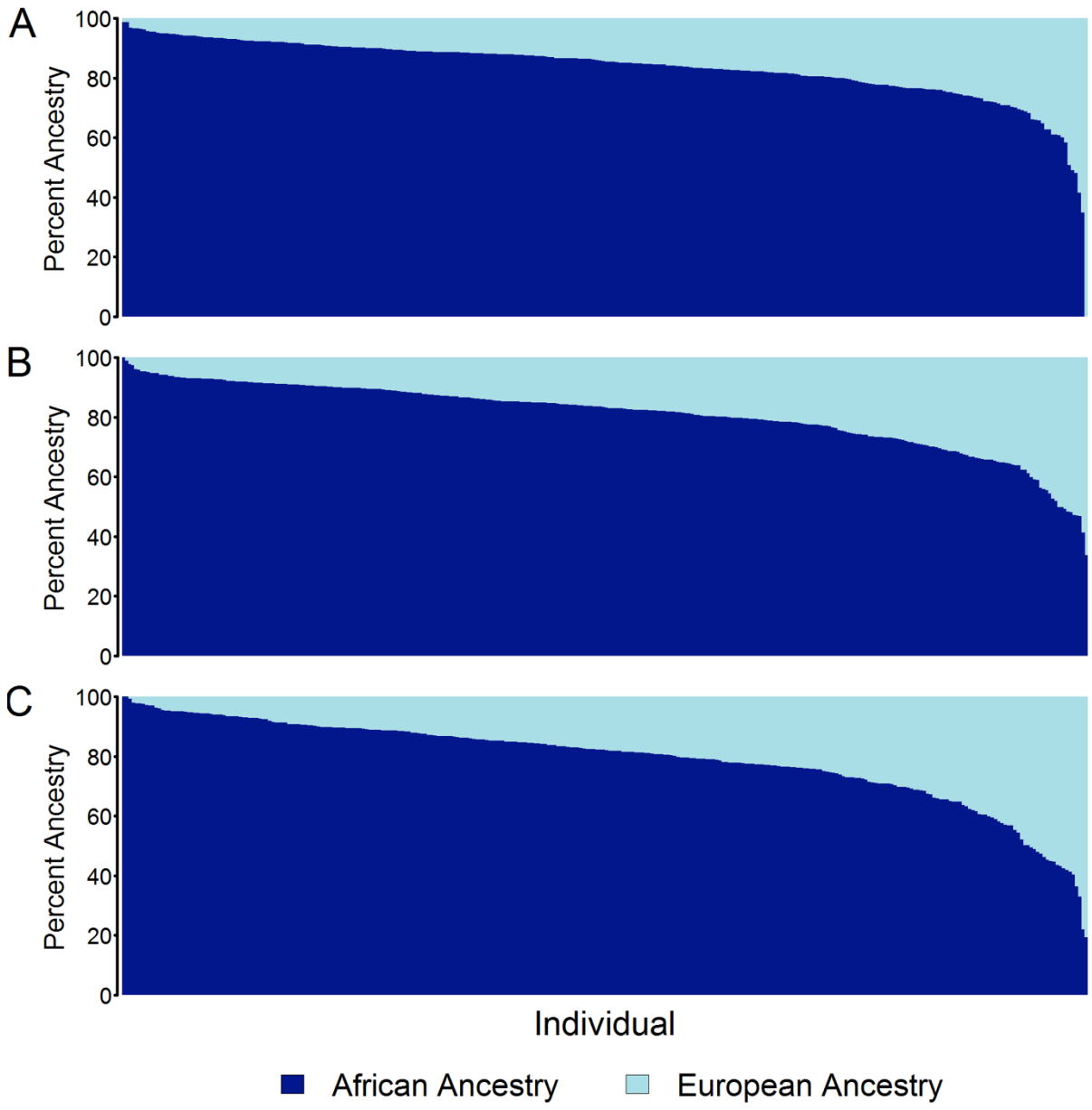

Figure 17. Global genetic ancestry estimates for African Americans with NSCLC participating in the (A) SCCS, (B) KCI/WSU, and (C) UCSF study populations.

African and European ancestry estimated using ADMIXTURE software and ancestry informative markers (SCCS) or genome-wide single nucleotide polymorphisms (KCI/WSU and UCSF). Supervised analysis (K=2) was performed with the use of YRI and CEU 1000 Genomes (phase 3) reference populations. Individuals are sorted from highest African ancestry to least.

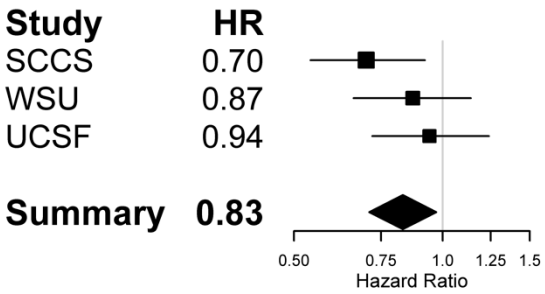

Figure 18. Meta-analysis of Cox proportional hazard results for rs1878022 in the discovery (SCCS, N=275) and replication (KCI/WSU and UCSF, N=312 and 284, respectively) study populations.

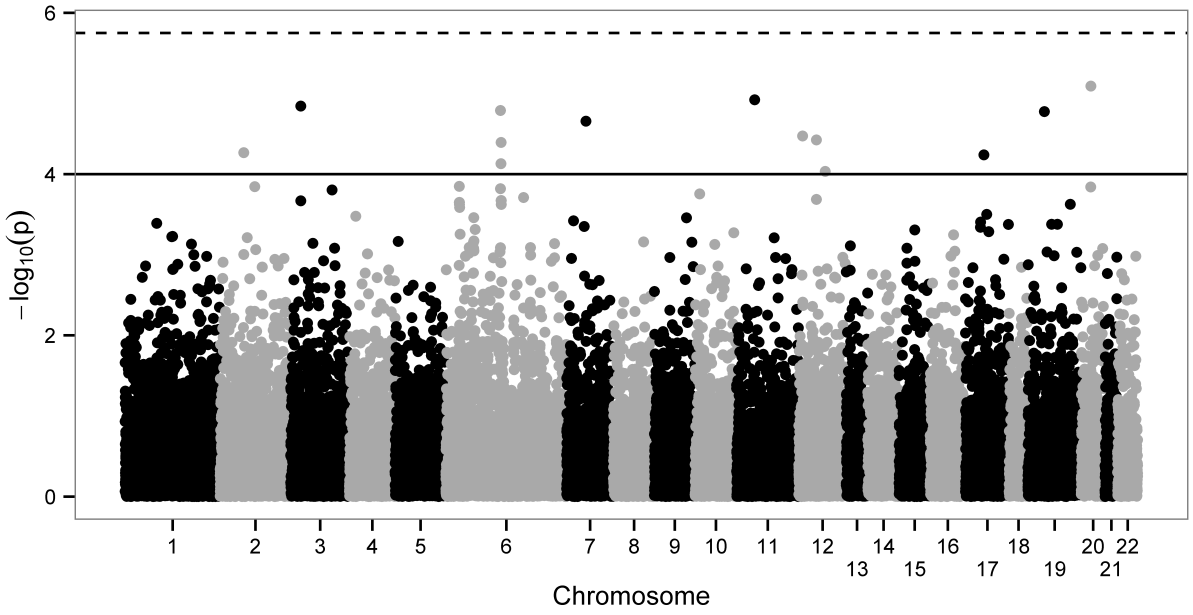

Figure 19. Multivariable Cox proportional hazard results for common SNPs assayed on the Illumina HumanExome BeadChip v1.1 and lung cancer survival among African Americans with NSCLC in the SCCS (N=275).

Models are adjusted for age, sex, treatment, stage, and percent African ancestry.
Solid black line represents our suggestive significance threshold (p value < 1.0×10^{-4}).
Dashed black line represents a Bonferroni corrected significance threshold (p value < 1.78×10^{-6})

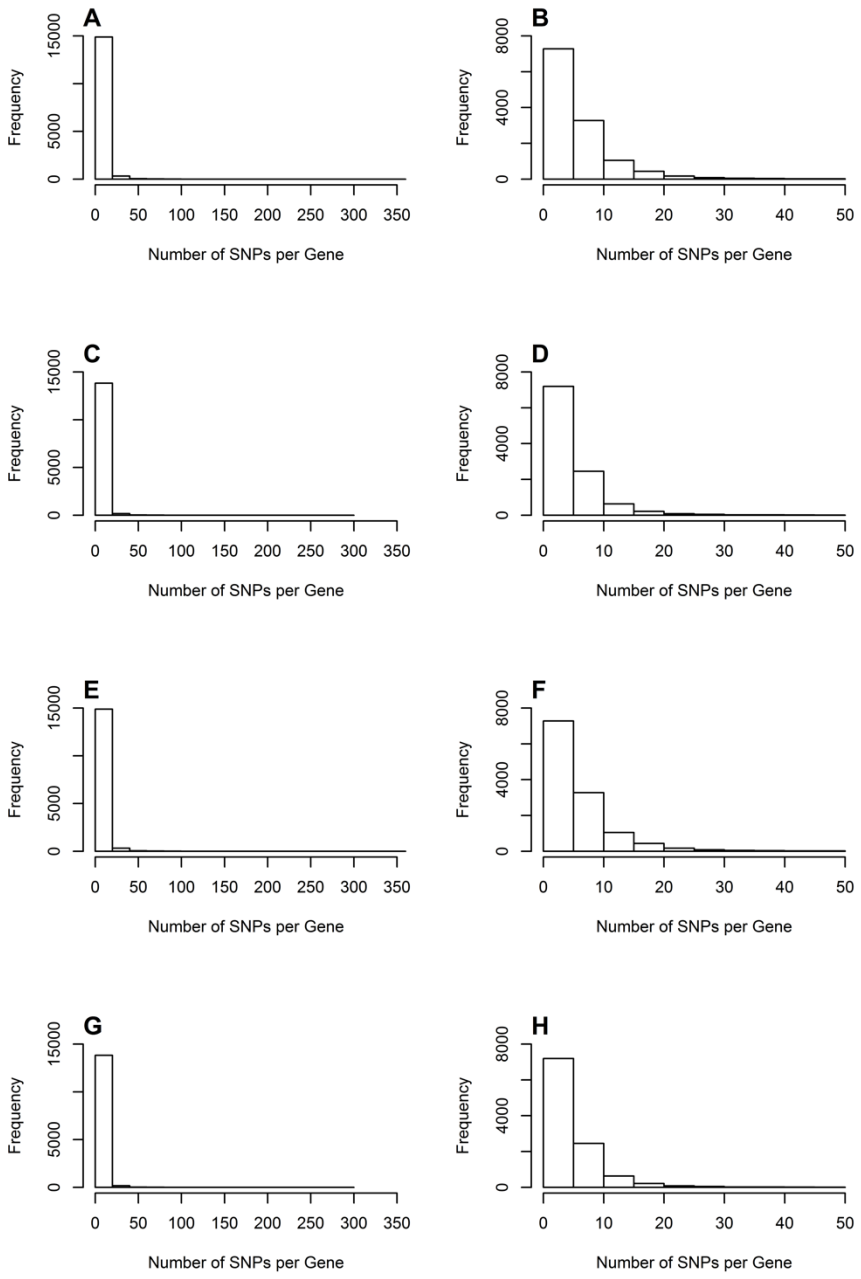

Figure 20. Distribution of the number of SNPs per gene for each test. A, C, E, and G show the distribution of SNPs per gene for the SKAT (all variants), SKAT (MAF<5%), burden test (all variants), and burden test (MAF <5%), respectively. B, D, F, and G are zoomed in to provide a better view of the frequency of genes with 1-50 variants for the SKAT (all variants), SKAT (MAF <5%), burden test (all variants), and burden test (MAF <5%), respectively.

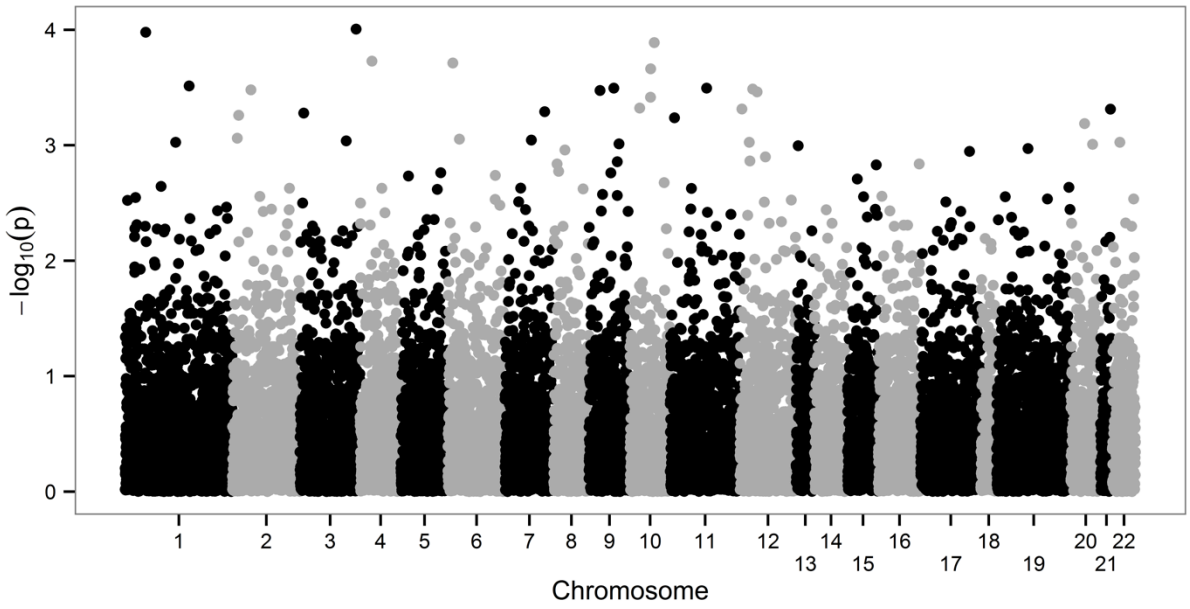

Figure 21. Gene-level association results for all variants using the Sequence Kernel Association Test (SKAT).

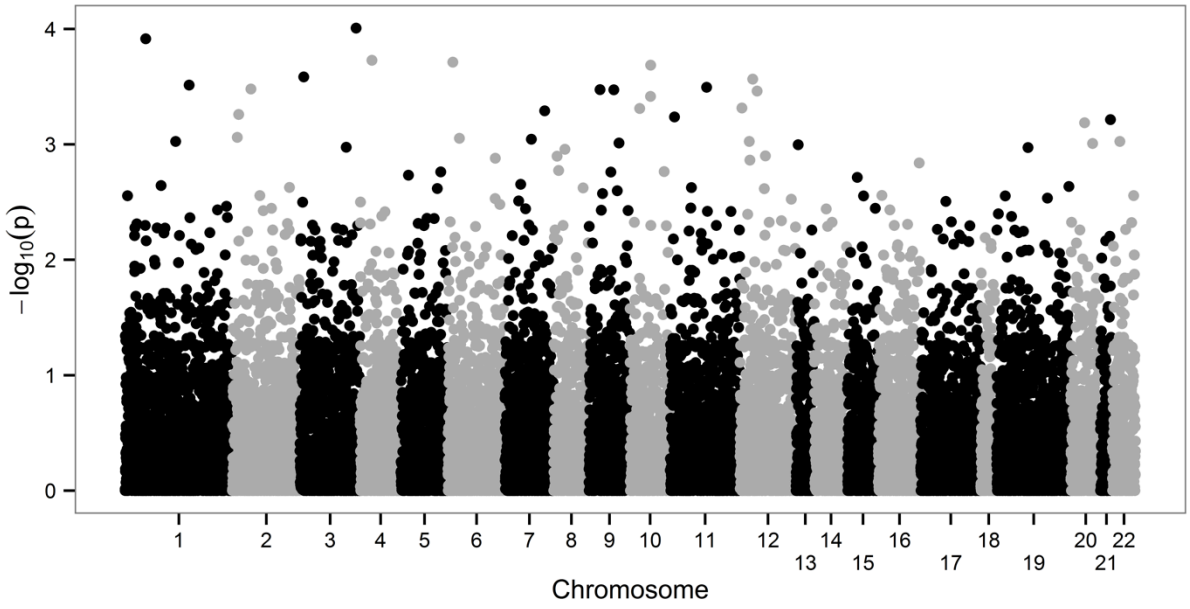

Figure 22. Gene-level association results for variants with a MAF<5% using the Sequence Kernel Association Test (SKAT).

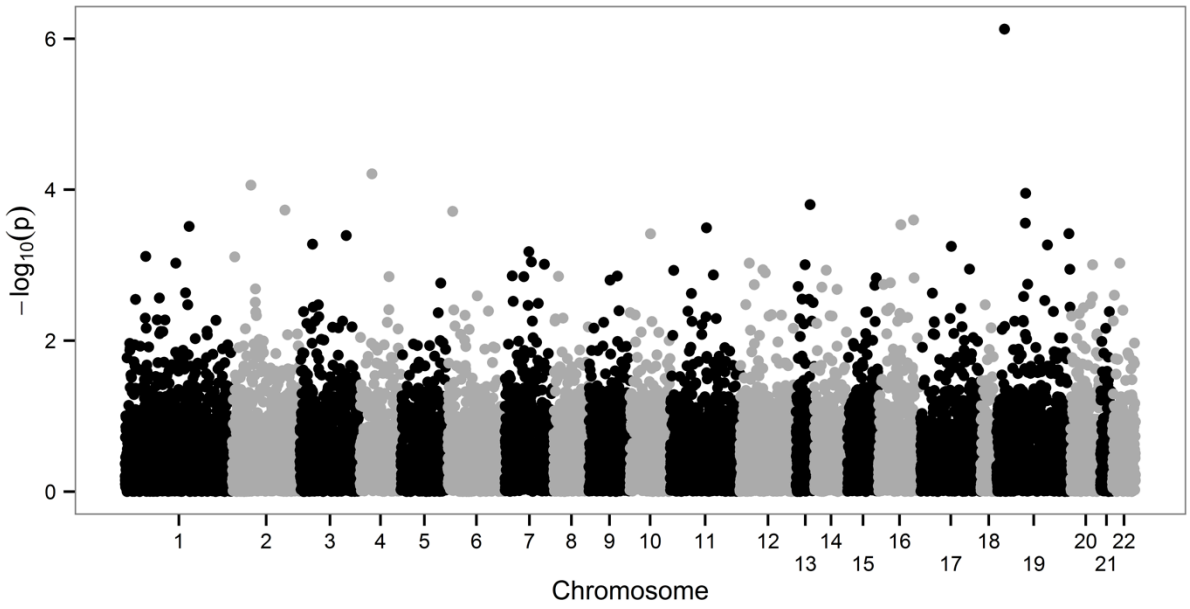

Figure 23. Gene-level association results for all variants using the burden test.

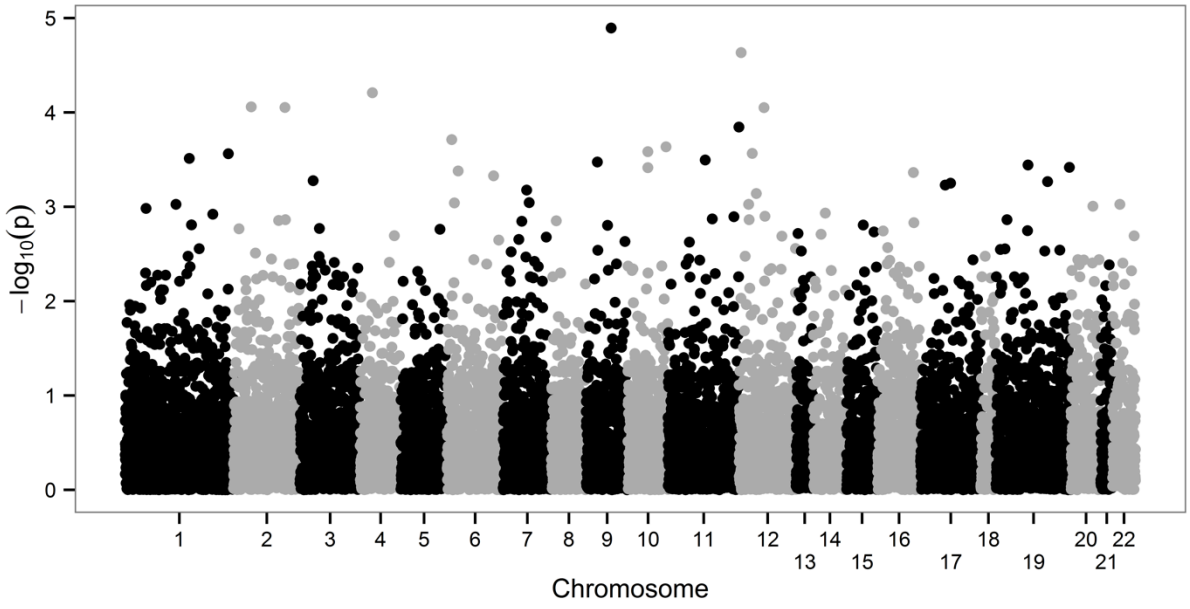

Figure 24. Gene-level association results for variants with a MAF<5% using the burden test.

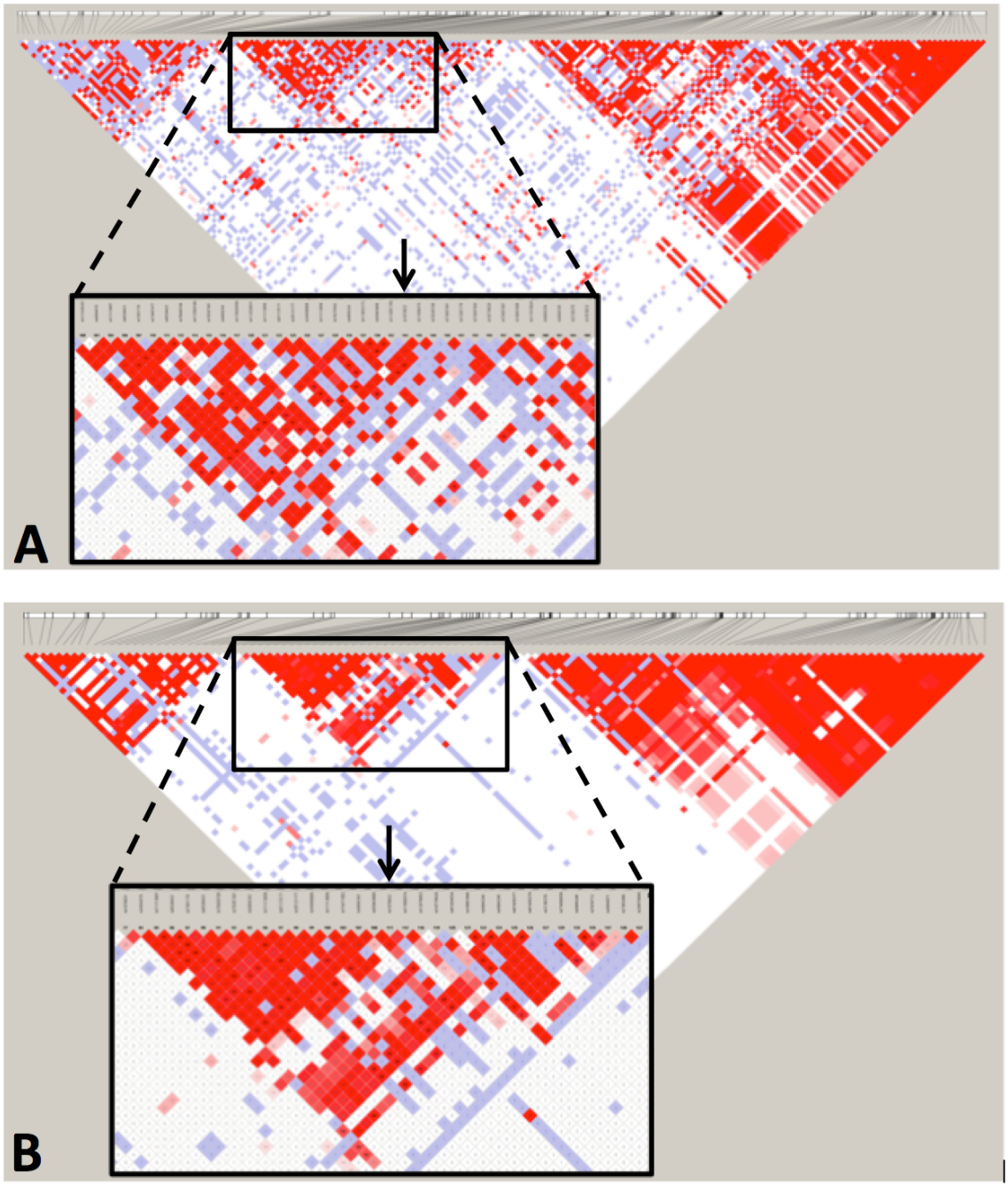

Figure 25. Linkage disequilibrium patterns +/-10kb surrounding the CMKLR1 gene region in the YRI (A) and CEU (B) 1000 Genomes (phase 3) population.

LD patterns surrounding rs1878022 (denoted by black arrow) are presented as inserts. 1000 Genomes data and Haploview software were utilized to display r2 values (filters: Hardy-Weinberg p-value > 0.001, genotyping efficiency > 90%, and minor allele frequency > 0.05).

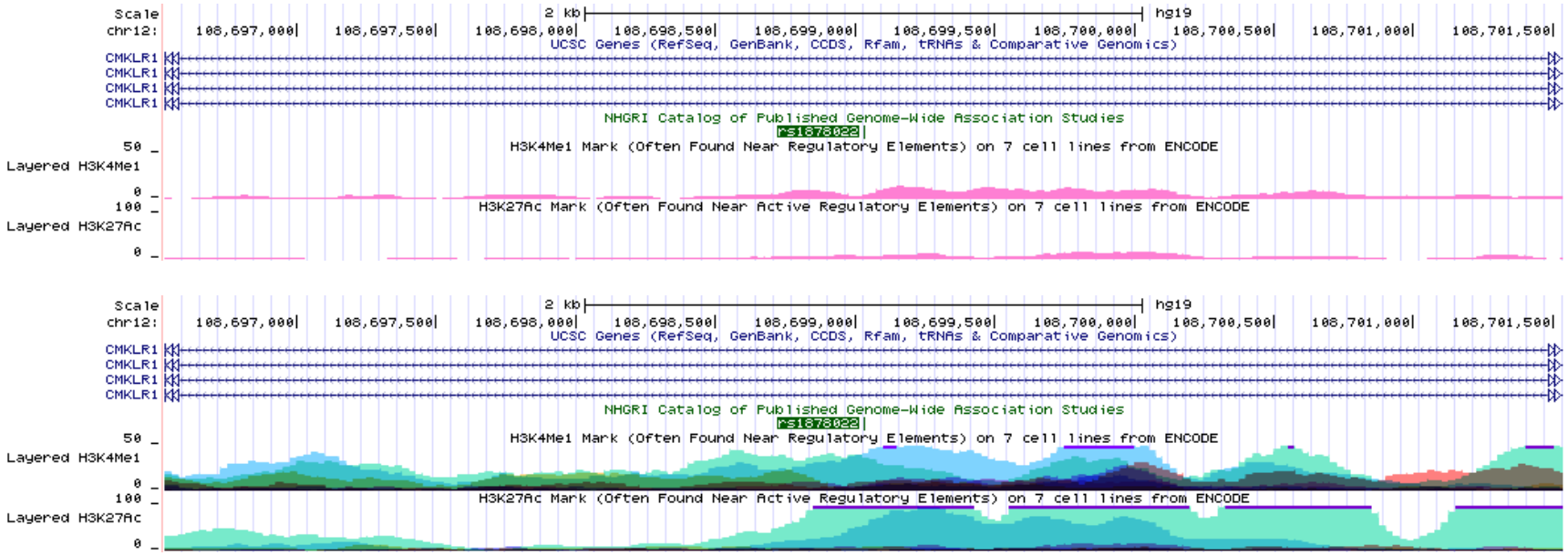

Figure 26. H3K4me1 and H3K27Ac histone modification marks surrounding rs1878022 in the *CMKLR1* gene in (A) normal human lung fibroblasts and (B) all cell lines (GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK, and NHLF), visualized using UCSC Genome Browser.

CHAPTER 4

CROSS-CANCER PLEIOTROPIC ANALYSIS OF LUNG CANCER RISK IN AFRICAN AMERICANS

Introduction

Epidemiology of lung cancer risk

Lung cancer is a leading cause of cancer worldwide, accounting for an estimated 13% of all new cancer cases in 2017 (American Cancer Society 2017). While lung cancer incidence rates are second only to breast cancer in women and prostate cancer in men, it is the leading cause of cancer-related mortality and is predicted to account for 26% of all cancer-related deaths in the United States in 2017 (American Cancer Society 2017). Between 2006 and 2012, the five-year survival rate for all races was 19% and 16% for African Americans (American Cancer Society 2017). Incidence rates also vary by race/ethnicity, with black males accounting for 83.7 new cases per 100,000 persons compared to 65.9 and 46.4 new cases per 100,000 persons for white and Asian males, respectively (Howlader N). Incidence rates for black, white, and Asian females were 49.0, 50.8, and 27.9, respectively (Howlader N).

Lung cancer cases can be divided into two main histologic types: small cell and non-small cell. Small cell lung cancer (SCLC) arises from neuroendocrine cells within the lungs and accounts for 13% of all lung cancer cases (Howlader N). The vast majority (84%) of lung cancer cases are non-small lung cancer (NSCLC) and can be further divided based on histological subtypes, with adenocarcinoma, squamous cell carcinoma, and large cell carcinoma being the most prevalent subtypes at 47%, 23%, and 2% of NSCLC cases, respectively (Howlader N).

Lung cancer is most commonly diagnosed in males after 65 years of age (American Cancer Society 2017). Additional risk factors include environmental and occupational exposures to carcinogens such as radon gas, pollution, gamma-radiation, arsenic, asbestos, and many more (Field and Withers 2012). However, smoking remains the highest risk factor for lung cancer and accounts for approximately 90% of all lung cancer cases (Prevention 2005), though incidence rates have steadily decreased since the mid-1980s for men and 2000s for women in large part due to smoking cessation efforts (Howlader N , American Cancer Society 2017).

Genetic studies of lung cancer risk

Despite the strong association with tobacco use, 10-15% of lung cancer cases occur in nonsmokers (Prevention 2005, Thun et al. 2006, Sisti and Boffetta 2012). In a study of nonsmoking lung cancer cases, individuals with a first-degree relative with lung cancer had a 7.2-fold increased risk for the disease (Schwartz et al. 1996) and heritability estimates range from 10-19% (Hemminki et al. 2001, Lorenzo Bermejo and Hemminki 2005, Sampson et al. 2015). Linkage analysis in 52 pedigrees with multiple affected individuals identified a lung cancer susceptibility locus on chromosome 6q23-25 (Bailey-Wilson et al. 2004). The heavily replicated lung cancer risk locus is chromosome 15q25.1, which contains nicotinic acetylcholine receptor subunit genes *CHRNA5*, *CHRNA3*, and *CHRNB4* (Amos et al. 2008, Hung et al. 2008, Thorgeirsson et al. 2008, Zanetti et al. 2016). Since the discovery of the 15q25.1 risk locus, more than 20 additional lung cancer risk loci have been reported through genome-wide association studies, though all but one study were conducted in individuals of European- or Asian-descent populations (Burdett T (EBI)).

Problems with traditional genome-wide association studies

Despite the known increased lung cancer incidence in African Americans, very few lung cancer genetics studies have been conducted in African-descent individuals, a pattern common across genetics research (Popejoy and Fullerton 2016). Furthermore, genome-wide association studies are plagued by a high burden of multiple testing and frequently used multiple test corrections can be too stringent, often resulting in true associations being rejected as false positives. While study sample sizes are increasing in large part due to decreased genotyping costs, power to detect such associations remains limited. One method of identifying previously undetected associations is by reducing the number of SNPs being examined to those with *a priori* biological evidence.

Pleiotropy

Pleiotropy occurs when a genetic locus is associated with more than one trait (Solovieff et al. 2013). Pleiotropy has been observed across multiple phenotypes, including cancer (Cheng et al. 2014, Panagiotou et al. 2014, Park et al. 2014, Fehringer et al. 2016). For example, mutations in the telomerase reverse transcriptase (*TERT*) gene have been associated with telomere length (Liu et al. 2014), pulmonary fibrosis (Fingerlin et al. 2013), red blood cell count (Kamatani et al. 2010), breast cancer (Couch et al. 2016), lung cancer (McKay et al. 2008, Landi et al. 2009, Hsiung et al. 2010), glioma (Kinnersley et al. 2015), pancreatic cancer (Wolpin et al. 2014), and prostate cancer (Berndt et al. 2015), to name a few (Mocellin et al. 2012). Cross-cancer pleiotropic effects in lung cancer have previously been examined in a European descent population and identified a significant association with a single variant in *LSP1* gene, which had been previously associated with breast cancer risk (Park et al. 2014). A 2016 study by Fehringer

et al. identified additional cross-cancer pleiotropic associations for lung cancer with *ADAM15/THBS3*, *CDKN2B-AS1*, and *BRCA2* genes. It remains necessary, however, to examine these associations in African American individuals.

Rationale

The present study sought to identify novel pleiotropic associations for lung cancer risk in African Americans by examining variants previously associated with cancer risk and reported in the National Human Genome Research Institute (NHGRI) - European Bioinformatics Institute (EBI) Genome-Wide Association Study (GWAS) Catalog. By selecting variants with *a priori* evidence for an association with cancer, we are able to reduce the high burden of multiple testing incurred by traditional genome-wide association studies, thereby increasing the possibility of detecting novel lung cancer risk associations. Unlike previous studies of pleiotropy across cancer types, the present analysis also accounts for differing patterns of linkage disequilibrium between racial/ethnic populations, further expanding our ability to detect cross-cancer and cross-population associations.

Methods

Study population

African American lung cancer cases and controls were selected from six studies participating in the African American Lung Cancer Consortium: MD Anderson (MDA) Lung Cancer Epidemiology Study, Project CHURCH (Creating a Higher Understanding of Cancer Research & Community Health, an MD Anderson Cancer Center study), NCI-MD Lung Cancer Case Control Study, Northern California Lung Cancer Study from the University of California,

San Francisco (UCSF), the Southern Community Cohort Study (SCCS), and the Karmanos Cancer Institute at Wayne State University (KCI/WSU) (Zanetti et al. 2016). A detailed description of each study has been previously reported (Zanetti et al. 2016). All studies were approved by the Institutional Review Board at each respective institution and written informed consent was obtained for all participants.

Genotyping, quality control, and imputation

All samples were genotyped on the Illumina Human Hap 1M Duo array at the NCI Cancer Genomics Research Laboratory (CGR) in the Division of Cancer Epidemiology and Genetics (DCEG) at the National Cancer Institute. A detailed description of the quality control (QC) process is provided (Figure 27). Briefly, SNPs were excluded if they were non-autosomal, had a MAF <1%, <95% genotyping efficiency, or a Hardy-Weinberg Equilibrium (HWE) p-value <0.000001. Individuals were excluded if they had a <95% genotyping efficiency. Pairwise identity-by-descent was also examined to identify related individuals; for each genetically related pair, the individual with the lowest genotyping efficiency was excluded (N=146). No individuals were excluded due to inconsistencies between reported and genetic sex, but four individuals with an "unknown" reported sex were filled in based on the calculated genetic sex. All quality control filtering was applied using PLINK (Purcell et al. 2007).

Missing genotypes were imputed using IMPUTE2 (Howie et al. 2009), with pre-phasing performed using SHAPEIT (Delaneau et al. 2008, Delaneau et al. 2011, Delaneau et al. 2013). Haplotypes from the cosmopolitan 1000 Genomes phase 3 population consisting of 2,504 individuals from 27 countries were used as a reference population. SNPs imputed with low certainty were excluded based on an info score <0.4, MAF<0.01, and HWE p-value <0.00001.

Ancestry estimation

Supervised admixture analysis was performed using ADMIXTURE software (Alexander et al. 2009) to obtain global estimates of African and European ancestry for all individuals. Admixture analysis was performed on pre-imputation genotypes merged with the CEU (CEPH Utah residents with Northern and Western European ancestry) and YRI (Yoruba from Ibadan, Nigeria) HapMap reference populations (International HapMap Consortium 2003), and pruned to a set of unlinked variants (window size = 50, step size = 10, $r^2 > 0.1$). A total of 140,591 variants remained for supervised (k=2) admixture analysis.

Selection of variants for pleiotropic analysis

Cancer-associated variants were identified from the NHGRI-EBI GWAS catalog. The catalog was queried based on the search term "neoplasm" on April 27, 2016. Manual review excluded studies in which the outcome was not risk (i.e. outcomes such as survival, prognosis, toxicity, relapse, or another disease). Interaction studies were also excluded. SNPs from each of the remaining studies were aggregated into a single list, hereafter referred to as "reported SNPs."

We further expanded our list of SNPs based on linkage disequilibrium patterns of the 1000 Genomes (phase 3) reference population (Genomes Project et al. 2015, Sudmant et al. 2015) matching the race/ethnicity of the population used in the reported GWAS Catalog study. For example, the CHB (Han Chinese in Beijing, China) reference population was used to identify SNPs in LD with variants reported in an Asian-descent population, while the CEU (CEPH Utah Residents with Northern and Western European Ancestry) reference population was used for SNPs reported in European-descent populations. For the admixed Latino/a and African American populations we used all relevant reference populations: CEU, CHB, YRI (Yoruba in Ibadan,

Nigeria), and MXL (Mexican Ancestry from Los Angeles) for Latino/a and CEU, YRI, and ASW (Americans of African Ancestry in SW United States) for African American. We tested four different methods for selecting variants based on LD on five randomly selected SNPs previously identified in a European population: 1) the PLINK tag-SNP selection function with default r^2 and window settings (--show-tags, default settings of r^2>0.8, +/-250kb), 2) the PLINK tag-SNP selection with specified r^2 and window range (r^2>0.6, +/-100kb), 3) the PLINK pairwise LD estimation (--ld-snp, r^2>0.6, +/-100kb), and 4) the online NIH/NIEHS LD Tag SNP selection tool (https://snpinfo.niehs.nih.gov/snpinfo/snptag.php). These results were compared to a SNP-selection method that included all SNPs +/-100kb of the reported SNP. A final list of SNPs for analysis was generated using the PLINK pairwise LD estimation method (r^2>0.6, +/-100kb), hereafter referred to as the "selected SNPs." Selected SNPs were extracted from the imputed genotyping data based on chromosome and position.

Statistical analysis

Logistic regression was performed for each additively coded variant using SNPTest to account for imputation probabilities (Marchini et al. 2007). Age, sex, smoking status (current/former/never), global African ancestry, and study site were included as covariates in the logistic regression models. A Benjamini- Hochberg false discovery rate (FDR) correction was applied to p-values to account for multiple testing. Each study site was then individually examined using SNPTest (adjusted for age, sex, smoking status, and global African ancestry) and meta-analyzed using METAL. Samples were additionally stratified by histology, smoking status, and sex to evaluate strata-specific associations. A formal test of interaction was then performed for SNPs with FDR-corrected p-values ≤ 0.10 in either sex- or smoking status-stratified analysis.

Results

Descriptive Characteristics

A total of 4,253 African American individuals remained following quality control, including 1,410 cases and 2,843 controls. Forty-five percent of all samples were male and the mean age at diagnosis was 58 years (standard deviation = 12.5). Never smokers accounted for 36% of all individuals, with current and former smokers accounting for 31% and 32% of individuals, respectively. The median global African ancestry was 83%. Among cases, 45% of individuals had adenocarcinoma, followed by 24% of individuals with squamous cell carcinoma. Small cell carcinoma represented 6% of cases. Descriptive characteristics by study and case/control status are presented in Table 22.

Variant Selection

A total of 266 unique studies were extracted from the NHGRI-EBI GWAS Catalog based on the search term "neoplasm." Forty-six studies were excluded after manual review, resulting in 220 studies reporting risk associations for 959 unique SNPs ("reported SNPs"). Seventy-four percent (163 out of 220) of studies were conducted in European-descent populations, followed by 26% (57 of 220) in Asian-descent populations[2] (Table 23). The admixed Latino/a and African American populations accounted for only 3% and 5% of included studies[2] (Table 23). Of all reported SNPs, 629 were directly observed in the genotype data and an additional 294 were imputed. Thirty-six reported SNPs were not

[2] The sum of percentages is greater than 100 because 11 studies used multiple racial/ethnic populations.

present in the 1000 Genomes reference populations used for LD-based selection of SNPs and were dropped from analysis.

Because genotyping arrays are designed to capture genome-wide variation using as few SNPs as possible, the majority of reported associations are rarely causal, but rather correlated (or in linkage disequilibrium) with the true causal variant. Additionally, we know that linkage disequilibrium (LD) patterns differ between racial/ethnic groups. Since the vast majority of the reported SNPs are identified in European- or Asian-descent populations, we selected variants in LD with the reported SNP in the 1000 Genomes reference population matching the race/ethnicity of the population reported in the NHGRI-EBI GWAS Catalog. We evaluated five methods for extracting variants based on LD for five randomly selected SNPs previously identified in a European-descent population using the CEU reference population (Table 24). Only one method, NIH/NIEHS LD Tag SNP Selection, failed to identify any additional SNPs to examine. Both PLINK tag SNP selection methods (default and user-specified settings) and the PLINK pairwise LD estimation method produced similar, though non-identical results (Table 24). For all PLINK LD-based methods tested, the number of additional SNPs identified ranged from 0 to 45 for the five tested SNPs. This was substantially less than the 784-979 SNPs identified when LD was not considered (Table 24). Given the discrepancy between the PLINK tag SNP selection and PLINK pairwise LD estimation methods even when the same parameters (r^2>0.6, +/-100kb) were set, we chose to apply the later method to all reported variants, in large part because we were most familiar with the underlying methodology. Additionally, the PLINK pairwise LD estimation method allowed us to still perform hypothesis-based testing, whereas selecting all variants within the +/-100kb region regardless of LD would

undoubtedly result in testing SNPs that have zero correlation with the reported SNP, thereby negating our hypothesis driven approach. Application of the PLINK pairwise LD estimation method to all reported SNPs increased our number of SNPS to be examined to 39,010 (Table 23). For the 20 and 49 SNPs originally identified in admixed Latino or African American populations, we applied the PLINK pairwise LD estimation method to all relevant reference populations (Table 23) and examined the overlap between selected SNPs in all populations (Figure 28 and Figure 29).

Logistic Regression Analysis

Approximately 1,700 selected SNPs were neither observed nor imputed in our African American population and 280 SNPs failed to meet post-imputation quality control filtering, resulting in 36,958 selected SNPs for analysis. Logistic regression analysis revealed 28 SNPs with an FDR-corrected p-value less than 0.05 and an additional 12 with a p-value less than 0.10 (Figure 30). The most statistically significant peak of association was chromosome 15q25 (odds ratio (OR)=1.41, 95% confidence interval (CI): 1.26-1.57), followed by chromosome 5p15 (OR=0.79, 95% CI: 0.71-0.88, Table 25). Three additional SNPs on chromosomes 5q14.3, 16q22.2, and 17q12 also had significant associations with lung cancer risk. The C allele at rs336958 on 5q14.3 was associated with reduced risk with an OR of 0.68 and 95% CI of 0.56-0.82. The peak on chromosome 16q22.2 consisted of four SNPs with similar effect sizes, though only one, rs7186207, surpassed a 10% FDR correction threshold (OR=1.24, 95% CI=1.12-1.38, FDR p-value = 0.04). Similarly, the G allele at rs11658063 on chromosome 17q12 had an OR of 1.24 and 95% CI of 1.11-1.39. One additional SNP in the 17q12 region had a similar effect size,

but did not surpass the 10% FDR threshold. Similar results were observed when study sites were meta-analyzed (Figure 32).

Stratified Analyses

To identify subtype-specific associations for lung cancer risk, we examined adenocarcinoma cases and squamous cell carcinoma cases separately. No genetic variant was significantly associated with lung cancer risk in either histological subtype (Figure 32 and Figure 33). The smallest FDR-corrected p-value was 0.18 in adenocarcinoma cases for a SNP on chromosome15q25 (Table 26). Among squamous cell carcinoma cases, the most significant SNP was rs1950005 on chromosome 6q25.2, though the FDR-corrected p-value was 0.48 (Table 27).

Stratification by sex and smoking status revealed the chromosome 15q25 association was among females only and ever smokers only (Figure 34 and Figure 36). Beyond the peak on 15q25, no other SNP passed a 10% FDR correction threshold among women only, though the risk (G) allele at rs7486184 on chromosome 12q21.32 was associated with an OR of 1.35 (95% CI=1.17-1.55) at an FDR-corrected p-value of 0.10, just above our statistical significance threshold (Figure 34 and Table 28). No SNPs were significant after an FDR correction in males (Table 29). Among ever smoking individuals, the association at chromosome 15q25 was by far the most significant signal. However, two other peaks also surpassed a 10% FDR threshold: chromosome 5p15.33 and 16q22.2 (Figure 36 and Table 30). While no p-values were statistically significant after FDR correction in never smokers, peaks at chromosome 5p15.33 and 15q25 had p-values of suggestive significance (FDR-corrected p-value = 0.13, Figure 37 and Table 31).

The 33 SNPs among ever smokers and 12 SNPs among females with an FDR-corrected p-value ≤ 0.10 were then examined for an interaction with smoking and sex, respectively. No

significant SNP*smoking interactions were present among the 33 SNPs associated with lung cancer risk among ever smokers (Table 32). Of the 12 SNPs associated with lung cancer risk among females, 9 showed significant evidence of a SNP*sex interaction (Table 33).

Discussion

The present analysis sought to identify cross-cancer pleiotropic genetic associations for lung cancer risk in African Americans. The two most statistically significant peaks were on chromosome 15q25 and 5p15, both of which have been previously associated with lung cancer in African Americans (Schwartz et al. 2009, Walsh et al. 2012, Walsh et al. 2013) and recently validated in a large African American population that included all cases and controls utilized in the present study (Zanetti et al. 2016). Stratification by sex and smoking status revealed strong signals of association among women and ever smokers, suggesting the observed association among all individuals may be driven by these two subgroups. However, it should not go unnoticed that, despite not meeting our threshold for statistical significance, there was evidence for an association of 5p15 and 15q25 among never smokers (FDR-corrected p-values = 0.13, Table 31). Furthermore, this confirms previous research noting the direct association between gene region 15q25 and lung cancer risk in addition to its association mediated by nicotine dependence (Hung et al. 2008, Thorgeirsson et al. 2008).

In addition to the previously described associations on 5p15 and 15q25, our study also identified three additional significant associations on chromosomes 5q14.3, 16q22.2, and 17q12 among all individuals. Chromosome 16q22.2 was also observed among ever smokers and an additional peak on 12q21.32 was observed among women. Excluding 5p15 and 15q25, the most significant association was for rs7186207 on chromosome 16q22.2 (FDR-corrected p-value =

0.04). All four SNPs in this region (rs7186207, rs8051239, rs7195958, and rs3213422) had risk alleles with odds ratios of approximately 1.24 and 95% confidence intervals from 1.12 to 1.38 and were in strong LD ($r^2 > 0.68$.) with each other in both African (YRI and ASW) and European (CEU) 1000 Genomes Reference Populations. Given the high degree of correlation between variants, it is unsurprising that risk allele frequencies were similar, ranging from 0.56 to 0.62. None of the four SNPs were among the reported SNPs extracted from the GWAS Catalog, but were selected because of their strong linkage disequilibrium ($r^2 = 0.75$-0.78) with rs12597458, a variant previously associated with prostate cancer risk (Berndt et al. 2015). Additional SNPs in the 16q22.2 region have also been associated with prostate cancer (Al Olama et al. 2014), though none of those SNPs were identified here. SNPs rs7186207, rs8051239, and rs7195958 are intergenic and located between *PKD1L3* and *DHODH* and do not appear to be located at sites with regulatory potential (Figure 38). The remaining SNP at this locus, rs3213422, is located within the first intron of *DHODH* and encodes a missense mutation, though SIFT and PolyPhen both predict the mutation to be tolerated/benign (Figure 38). Given its close proximity to the exon boundary, variation at rs3213422 could also affect exon splicing. ENCODE data reveal H3K4me3 and H3K27ac markers surrounding rs3213422, indicative of active promoters and regulatory elements, as well as evidence for transcription factor binding (Figure 38). Given the lack of regulatory evidence for rs7186207, rs8051239, and rs7195958, it is possible that their association with lung cancer risk in the present study is due to their high correlation with rs3213422, which has clear regulatory potential.

The *DHODH*, or dihydroorotate dehydrogenase, gene encodes a 43-kDa enzymatic protein localized to the inner mitochondrial membrane, where it interacts with the mitochondrial respiratory chain (Evans and Guy 2004, Khutornenko et al. 2010, Fang et al. 2013). Mutations

within *DHODH* have been linked with Miller Syndrome, a recessive disorder characterized by malformations of the limbs and eyes, among other symptoms (Ng et al. 2010, Kinoshita et al. 2011, Fang et al. 2012, Rainger et al. 2012). *DHODH* is also being investigated for a role in cancer (White et al. 2011). Most recently, decreased expression of *DHODH* was associated with breast cancer risk (Hoffman et al. 2017). Several other studies have examined the utility of *DHODH* inhibitors in inducing apoptosis in cancer cells (Baumann et al. 2009, Zhu et al. 2013, He et al. 2014). While *DHODH* has not been previously associated with lung cancer risk, the abundance of biological evidence for its role in cancer gives credibility to the association and stresses the importance of performing a replication analysis in an independent population of African American lung cancer cases and controls and fine mapping the region to better understand the underlying mechanism.

Analysis of all individuals identified significant associations on chromosomes 5q14.3 and 17q12. Chromosome 5q14.3 SNP rs336958 is an intronic variant for *HAPLN1*, hyaluronan and proteoglycan link protein 1, which has been shown to play a role in cell adhesion and extracellular matrix structure. rs336958 is in LD (r^2=0.97) with rs4466137, which is reported in the GWAS catalog for an association with prostate cancer risk (Murabito et al. 2007). The larger 5q14.3 region has also been associated with prostate cancer, breast cancer and Wilms tumors (Turnbull et al. 2012, Cai et al. 2014, Berndt et al. 2015). The most significant SNP on chromosome 17q12 is rs11658063, a variant located in the first intron of *HNF1B*. HNF1 homeobox B (*HNF1B*) encodes a transcription factor and has been shown to play a role in cell development. Studies have also linked *HNF1B* to hepatocellular cancer prognosis and have investigated its utility as a biomarker in breast cancer (Yu et al. 2015, Huang et al. 2016, Bubancova et al. 2017). The final notable peak of association was on chromosome 12q21.32, in

which rs7486184 was associated with lung cancer in females. The intergenic variant rs7486184 is located approximately 50 kb downstream of the previously reported variant rs995030, which has been previously associated with testicular cancers (Rapley et al. 2009, Chung et al. 2013, Ruark et al. 2013).

While previous studies have examined pleiotropy in lung cancer (Park et al. 2014, Fehringer et al. 2016), few have considered differences in LD structure between racial/ethnic groups. Such considerations are especially important given recent publications noting the non-transferability of genetic risk predictions across diverse populations (Manrai et al. 2016, Martin et al. 2017). In the present analysis, we expanded the list of reported SNPs by considering the LD structure of the racial/ethnic population each SNP was discovered in, thus removing the assumption that the reported SNP has the same correlation structure, and therefore, tagging ability, with the causal SNP in all racial/ethnic groups. Of the six methods for expanding our list of reported SNPs examined here, only one (NIH/NIEHS LD Tag SNP Selection) identified no additional SNPs. This is likely due to the fact that the selection method uses genotyping data from HapMap reference populations, which provide sparser information compared to whole genome sequencing data from the 1000 Genomes reference populations that utilized in our other approaches. Importantly, it was only through consideration of LD structure that the present study was able to identify novel lung cancer risk associations, as none of the most significant SNPs were among the list of reported SNPs extracted from the GWAS catalog.

The first African American genome-wide association study was published by Zanetti *et al.* in 2016 and was also the largest study of lung cancer in African Americans. It is of utmost importance to note that the present study is not independent from the former as our cases and controls were a subset of the individuals utilized by Zanetti *et al.* However, our study restricted

analysis to SNPs with *a priori* evidence to examine cross-cancer pleiotropic associations. Our study was able to identify novel lung cancer risk loci that may have been missed due to stringent multiple test corrections required in genome-wide association studies.

Given our large sample size, we are adequately powered to detect the associations reported here (Figure 39). While our primary analysis lacked a replication population, meta-analysis of all five studies revealed results consistent to the pooled analysis, confirming our results. Despite the internal replication, future analyses should confirm these associations in an independent population of African American lung cancer cases and controls. Future analyses should confirm these associations in other racial/ethnic groups and focus on fine mapping of the chromosome 16q22.2 region to better understand its impact on lung cancer risk.

Table 22. Descriptive characteristics of African American lung cancer cases and controls.

Characteristic	MDA N=1374	NCI N=583	SCCS N=506	UCSF N=986	WSU N=804	Total Cases N=1,410	Total Controls N=2,843	Total Combined N=4,253
Status								
Cases, N (%)	373 (27.1)	208 (35.7)	168 (33.2)	325 (33.0)	336 (41.8)			1,410 (33.2)
Controls, N (%)	1001 (72.9)	375 (64.3)	338 (66.8)	661 (67.0)	468 (58.2)			2,843 (66.8)
Sex								
Male	525 (38.2)	308 (52.8)	300 (59.3)	450 (45.6)	317 (39.4)	698 (49.5)	1,202 (42.3)	1,900 (44.7)
Female	849 (61.8)	275 (47.2)	206 (40.7)	536 (54.4)	487 (60.6)	712 (50.5)	1,641 (57.7)	2,353 (55.3)
Mean age at diagnosis, yr (SD)	52.2 (13.5)	64.4 (9.6)	55.9 (8.9)	63.5 (11.2)	60.3 (11.3)	61.5 (10.5)	56.9 (13.2)	58.4 (12.6)
Smoking Status								
Current, N (%)	339 (24.8)	196 (33.7)	276 (54.8)	364 (38.1)	358 (44.6)	778 (55.3)	755 (27.0)	1,533 (36.4)
Former, N (%)	379 (27.7)	250 (43.0)	111 (22.0)	361 (37.8)	257 (32.0)	520 (36.9)	838 (29.9)	1,358 (32.3)
Never, N (%)	649 (47.5)	135 (23.2)	117 (23.2)	230 (24.1)	187 (23.3)	110 (7.8)	1,208 (43.1)	1,318 (31.3)
Median African Ancestry, %	82.6	82.4	88.3	81.9	82.9	83.6	83.1	83.2
Histology								
Adenocarcinoma, N (%)	173 (46.4)	97 (47.1)	47 (31.3)	145 (44.6)	170 (50.6)	632 (45.5)	-	632 (45.5)
Squamous Cell, N (%)	103 (27.6)	53 (25.7)	28 (18.7)	82 (25.2)	71 (21.1)	337 (24.2)	-	337 (24.2)
Large Cell, N (%)	2 (0.5)	5 (2.4)	11 (7.3)	4 (1.2)	12 (3.6)	34 (2.4)	-	34 (2.4)
Small Cell, N (%)	23 (6.2)	2 (1.0)	14 (9.3)	20 (6.0)	22 (6.8)	81 (5.8)	-	81 (5.8)
Other, N (%)	72 (19.3)	49 (23.8)	50 (33.3)	72 (22.2)	63 (18.8)	306 (22.0)	-	306 (22.0)

Table 23. Number of studies and variants reported in the NHGRI-EBI GWAS Catalog and the number of SNPs identified after selecting based on LD.

Race/Ethnicity in GWAS Catalog	Number of Studies	Number of SNPs	1000 Genomes Population Used for LD Selection	Number of SNPs after LD Selection
European	163	743	CEU	29,727
Asian	57	217	CHB	10,625
Latino	6	20	CEU, YRI, MXL, CHB	968
African American	11	49	CEU, YRI, ASW	1,490
Total	**220**	**959**		**39,010**

Table 24. Number of additional SNPs extracted by each method for five randomly selected SNPs previously identified in European-descent populations.

Reported SNP	All SNPs (r^2>0, +/- 100kb)	PLINK pairwise LD estimation (r^2>0.6, +/- 100kb)	PLINK tag SNP selection (default settings)	PLINK tag SNP selection (r^2>0.6, +/- 100kb)	NIH/NIEHS LD Tag SNP Selection
rs12413624	960	18	10	16	0
rs6772209	949	1	0	0	NA
rs34479159	869	12	9	10	NA
rs7632500	979	45	36	43	0
rs7580717	784	34	15	32	0

Table 25. SNPs significantly associated with lung cancer risk among African American lung cancer cases and controls (N=4,253).

SNP	Chr.	BP	Risk/ Ref. Allele	Risk Allele Freq	Info Score	OR	95% CI	Unadjusted P-value	FDR P-value
rs17486278	15	78867482	C/A	0.3	1	1.41	(1.26-1.57)	6.27E-10	$2.32x10^{-5}$
rs55781567	15	78857986	G/C	0.28	1	1.37	(1.23-1.53)	1.54E-08	$2.84x10^{-4}$
rs2036527	15	78851615	A/G	0.23	1	1.36	(1.21-1.54)	2.87E-07	$2.89x10^{-4}$
rs58365910	15	78849034	C/T	0.27	0.98	1.35	(1.2-1.51)	3.13E-07	$2.89x10^{-4}$
rs147144681	15	78900908	T/C	0.18	1	1.36	(1.2-1.55)	2.14E-06	0.01
rs576982	15	78870803	T/C	0.29	1	0.77	(0.69-0.86)	2.74E-06	0.01
rs664172	15	78862762	A/G	0.28	1	0.76	(0.68-0.85)	2.17E-06	0.01
rs667282	15	78863472	C/T	0.29	1	0.76	(0.68-0.85)	1.91E-06	0.01
rs938682	15	78896547	A/G	0.72	1	1.3	(1.17-1.46)	2.56E-06	0.01
rs569207	15	78873119	T/C	0.28	1	0.77	(0.69-0.86)	3.71E-06	0.01
rs637137	15	78873976	A/T	0.29	1	0.77	(0.69-0.86)	4.15E-06	0.01
rs11637630	15	78899719	A/G	0.71	0.99	1.29	(1.16-1.44)	5.35E-06	0.01
rs2456020	15	78868398	T/C	0.4	1	0.79	(0.71-0.87)	5.52E-06	0.01
rs55676755	15	78898932	G/C	0.17	1	1.35	(1.19-1.54)	5.69E-06	0.01
rs7183604	15	78899213	C/T	0.72	0.99	1.3	(1.16-1.45)	4.87E-06	0.01
rs12440014	15	78926726	G/C	0.22	0.91	0.75	(0.66-0.85)	6.39E-06	0.01
rs3825845	15	78910258	T/C	0.23	1	0.76	(0.68-0.86)	8.24E-06	0.02
rs503464	15	78857896	A/T	0.27	0.98	0.77	(0.69-0.87)	1.00E-05	0.02
rs189218934	15	78903987	C/T	0.73	0.99	1.29	(1.15-1.44)	1.08E-05	0.02
rs113931022	15	78901113	T/C	0.19	1	1.32	(1.16-1.5)	2.18E-05	0.04
rs138544659	15	78900701	G/T	0.19	1	1.32	(1.16-1.49)	2.21E-05	0.04
rs112878080	15	78900647	G/A	0.19	1	1.32	(1.16-1.49)	2.21E-05	0.04
rs2853677	5	1287194	A/G	0.71	1	0.79	(0.71-0.88)	2.28E-05	0.04
rs111704647	15	78900650	T/C	0.19	1	1.31	(1.16-1.49)	2.50E-05	0.04
rs2735940	5	1296486	G/A	0.47	0.98	0.81	(0.73-0.89)	2.60E-05	0.04
rs7186207	16	72035359	C/T	0.57	1	1.24	(1.12-1.38)	2.66E-05	0.04
rs2853672	5	1292983	A/C	0.47	1	0.81	(0.73-0.89)	2.85E-05	0.04
rs56077333	15	78899003	A/C	0.19	1	1.31	(1.15-1.49)	3.24E-05	0.04
rs7170068	15	78912943	A/G	0.23	0.99	0.78	(0.69-0.88)	3.93E-05	0.05
rs1051730	15	78894339	A/G	0.12	1	1.37	(1.18-1.6)	5.00E-05	0.06
rs28491218	15	78267947	C/T	0.22	0.91	0.77	(0.68-0.87)	5.43E-05	0.06
rs951266	15	78878541	A/G	0.11	0.99	1.39	(1.18-1.63)	5.28E-05	0.06
rs12914385	15	78898723	T/C	0.2	1	1.29	(1.14-1.46)	5.69E-05	0.06
rs336958	5	82973396	C/T	0.92	1	0.68	(0.56-0.82)	5.90E-05	0.06
rs7172118	15	78862453	A/C	0.11	0.99	1.38	(1.18-1.62)	6.73E-05	0.07
rs28360704	15	78268603	T/C	0.21	0.94	0.77	(0.68-0.88)	8.70E-05	0.08
rs56390833	15	78877381	A/C	0.11	1	1.38	(1.17-1.62)	8.10E-05	0.08
rs7180002	15	78873993	T/A	0.11	1	1.38	(1.17-1.61)	8.42E-05	0.08
rs905739	15	78845110	G/A	0.25	0.98	0.79	(0.7-0.89)	8.77E-05	0.08
rs11658063	17	36103872	G/C	0.61	0.88	1.24	(1.11-1.39)	1.04E-04	0.10
rs16969968	15	78882925	A/G	0.07	1	1.49	(1.22-1.83)	1.14E-04	0.10

Table 25 continued. SNPs significantly associated with lung cancer risk among African American lung cancer cases and controls

SNP	Chr.	BP	Risk/ Ref. Allele	Risk Allele Freq	Info Score	OR	95% CI	Unadjusted P-value	FDR P-value
rs8192482	15	78886198	T/C	0.07	1	1.49	(1.22-1.83)	1.18E-04	0.10
rs8051239	16	72036257	A/T	0.62	1	1.23	(1.1-1.36)	1.31E-04	0.11
rs11633958	15	78862064	T/C	0.07	0.99	1.47	(1.2-1.79)	1.57E-04	0.13
rs140330585	15	78866445	A/G	0.13	0.99	1.34	(1.15-1.56)	1.64E-04	0.13
rs17486195	15	78865197	G/A	0.13	0.99	1.34	(1.15-1.56)	1.73E-04	0.13
rs3734131	5	95278679	A/G	0.4	1	0.82	(0.75-0.91)	1.72E-04	0.13
rs3777175	5	95277555	G/A	0.4	1	0.82	(0.75-0.91)	1.68E-04	0.13
rs4243084	15	78911672	C/G	0.2	0.99	1.27	(1.12-1.45)	1.60E-04	0.13
rs7195958	16	72036577	G/A	0.62	1	1.22	(1.1-1.36)	1.53E-04	0.13
rs72740964	15	78868636	A/G	0.07	1	1.48	(1.21-1.81)	1.68E-04	0.13
rs28534575	15	78923845	G/T	0.25	0.93	0.8	(0.71-0.9)	1.93E-04	0.14
rs146009840	15	78906177	T/A	0.07	0.99	1.47	(1.2-1.79)	2.05E-04	0.14
rs518425	15	78883813	G/A	0.49	1	0.83	(0.75-0.92)	2.00E-04	0.14
rs55853698	15	78857939	G/T	0.09	0.99	1.41	(1.17-1.68)	2.01E-04	0.14
rs4887072	15	78925435	G/A	0.25	0.93	0.8	(0.71-0.9)	2.10E-04	0.14
rs11263763	17	36103565	A/G	0.61	0.89	1.23	(1.1-1.37)	2.29E-04	0.15
rs3213422	16	72042682	C/A	0.56	1	1.17	(1.06-1.3)	1.91E-03	0.51

Table 26. Top SNPs associated with lung cancer risk among African American lung adenocarcinoma cases and controls (N=3,475).

SNP	Chr.	BP	Risk/ Ref. Allele	Risk Allele Freq.	Info Score	OR	95% CI	Unadjusted P-value	FDR P-value
rs17486278	15	78867482	C/A	0.29	1	1.39	(1.21-1.6)	4.75E-06	0.18
rs55781567	15	78857986	G/C	0.27	1	1.31	(1.14-1.51)	2.19E-04	0.46
rs10509798	10	107503741	T/C	0.26	1	0.78	(0.67-0.9)	8.19E-04	0.46
rs111493981	10	107501561	T/C	0.26	1	0.78	(0.67-0.9)	8.17E-04	0.46
rs111964481	10	107495801	G/A	0.26	1	0.78	(0.67-0.9)	7.50E-04	0.46

Table 27. Top SNPs associated with lung cancer risk among African American lung squamous cell carcinoma cases and controls (N=3,180).

SNP	Chr.	BP	Risk/ Ref. Allele	Risk Allele Freq.	Info Score	OR	95% CI	Unadjusted P-value	FDR P-value
rs55781567	15	78857986	G/C	0.26	1	1.5	(1.23-1.82)	4.45E-05	0.48
rs1950005	6	154481394	A/G	0.5	1	0.7	(0.59-0.83)	4.17E-05	0.48
rs9383692	6	154471682	G/A	0.53	0.99	0.7	(0.59-0.83)	5.68E-05	0.48
rs9397689	6	154467342	A/C	0.58	0.98	0.7	(0.59-0.83)	6.54E-05	0.48
rs9479767	6	154476528	C/T	0.5	1	0.7	(0.59-0.83)	5.23E-05	0.48

Table 28. SNPs significantly associated with lung cancer risk among female African American lung cancer cases and controls (N=2,353).

SNP	Chr.	BP	Risk/ Ref. Allele	Risk Allele Freq.	Info Score	OR	95% CI	Unadjusted P-value	FDR P-value
rs17486278	15	78867482	C/A	0.3	1	1.51	(1.3-1.76)	1.16E-07	4.29E-03
rs113931022	15	78901113	T/C	0.18	1	1.57	(1.31-1.88)	1.38E-06	7.53E-03
rs138544659	15	78900701	G/T	0.18	1	1.57	(1.3-1.88)	1.42E-06	7.53E-03
rs112878080	15	78900647	G/A	0.18	1	1.57	(1.3-1.88)	1.42E-06	7.53E-03
rs147144681	15	78900908	T/C	0.18	1	1.58	(1.32-1.9)	1.03E-06	7.53E-03
rs111704647	15	78900650	T/C	0.18	1	1.56	(1.3-1.88)	1.43E-06	7.53E-03
rs55676755	15	78898932	G/C	0.17	1	1.59	(1.32-1.92)	9.97E-07	7.53E-03
rs56077333	15	78899003	A/C	0.18	1	1.55	(1.29-1.86)	2.37E-06	0.01
rs12914385	15	78898723	T/C	0.2	1	1.51	(1.26-1.8)	5.49E-06	0.02
rs55781567	15	78857986	G/C	0.28	1	1.42	(1.21-1.66)	1.50E-05	0.06
rs58365910	15	78849034	C/T	0.27	0.98	1.41	(1.2-1.66)	3.20E-05	0.10
rs7486184	12	88847001	G/A	0.53	0.98	1.35	(1.17-1.55)	3.33E-05	0.10

Table 29. Top SNPs associated with lung cancer risk among male African American lung cancer cases and controls (N=1,900).

SNP	Chr.	BP	Risk/ Ref. Allele	Risk Allele Freq.	Info Score	OR	95% CI	Unadjusted P-value	FDR P-value
rs3825845	15	78910258	T/C	0.23	1	0.7	(0.59-0.83)	4.57E-05	0.31
rs569207	15	78873119	T/C	0.28	1	0.72	(0.62-0.85)	5.73E-05	0.31
rs576982	15	78870803	T/C	0.29	1	0.72	(0.61-0.84)	3.25E-05	0.31
rs637137	15	78873976	A/T	0.29	1	0.72	(0.62-0.85)	5.77E-05	0.31
rs664172	15	78862762	A/G	0.28	1	0.72	(0.61-0.84)	4.24E-05	0.31

Table 30. SNPs significantly associated with lung cancer risk among ever smoking African American lung cancer cases and controls (N=2,891).

SNP	Chr.	BP	Risk/ Ref. Allele	Risk Allele Freq.	Info Score	OR	95% CI	Unadjusted P-value	FDR P-value
rs17486278	15	78867482	C/A	0.31	1	1.41	(1.26-1.58)	3.84E-09	1.42E-04
rs55781567	15	78857986	G/C	0.28	1	1.38	(1.22-1.54)	7.01E-08	1.30E-03
rs147144681	15	78900908	T/C	0.19	1	1.39	(1.22-1.59)	1.60E-06	8.45E-03
rs2036527	15	78851615	A/G	0.23	1	1.36	(1.2-1.55)	1.24E-06	8.45E-03
rs576982	15	78870803	T/C	0.28	1	0.75	(0.67-0.84)	1.43E-06	8.45E-03
rs664172	15	78862762	A/G	0.27	1	0.74	(0.66-0.84)	8.06E-07	8.45E-03
rs667282	15	78863472	C/T	0.28	1	0.75	(0.66-0.84)	9.47E-07	8.45E-03
rs11637630	15	78899719	A/G	0.72	0.99	1.32	(1.18-1.49)	2.34E-06	9.94E-03
rs637137	15	78873976	A/T	0.28	1	0.76	(0.67-0.85)	2.42E-06	9.94E-03
rs12440014	15	78926726	G/C	0.22	0.9	0.73	(0.64-0.83)	2.75E-06	0.01
rs3825845	15	78910258	T/C	0.23	0.99	0.74	(0.65-0.84)	3.21E-06	0.01
rs938682	15	78896547	A/G	0.72	1	1.32	(1.17-1.49)	3.31E-06	0.01
rs503464	15	78857896	A/T	0.26	0.98	0.76	(0.67-0.85)	5.74E-06	0.01
rs55676755	15	78898932	G/C	0.18	1	1.37	(1.2-1.57)	5.59E-06	0.01
rs569207	15	78873119	T/C	0.28	1	0.76	(0.68-0.86)	5.37E-06	0.01
rs7183604	15	78899213	C/T	0.73	0.99	1.31	(1.17-1.48)	5.84E-06	0.01
rs58365910	15	78849034	C/T	0.27	0.98	1.32	(1.17-1.49)	6.45E-06	0.01
rs2456020	15	78868398	T/C	0.39	1	0.78	(0.71-0.87)	9.66E-06	0.02
rs189218934	15	78903987	C/T	0.73	0.99	1.31	(1.16-1.47)	1.10E-05	0.02
rs113931022	15	78901113	T/C	0.19	1	1.34	(1.18-1.53)	1.40E-05	0.02
rs138544659	15	78900701	G/T	0.19	1	1.34	(1.18-1.53)	1.41E-05	0.02
rs112878080	15	78900647	G/A	0.19	1	1.34	(1.18-1.53)	1.41E-05	0.02
rs7170068	15	78912943	A/G	0.22	0.99	0.75	(0.66-0.85)	1.25E-05	0.02
rs111704647	15	78900650	T/C	0.19	1	1.34	(1.17-1.53)	1.71E-05	0.03
rs56077333	15	78899003	A/C	0.19	1	1.33	(1.17-1.52)	2.34E-05	0.03
rs2735940	5	1296486	G/A	0.46	0.98	0.8	(0.72-0.89)	3.69E-05	0.05
rs12914385	15	78898723	T/C	0.21	1	1.31	(1.15-1.49)	4.19E-05	0.06
rs905739	15	78845110	G/A	0.25	0.98	0.77	(0.68-0.87)	4.13E-05	0.06
rs2853672	5	1292983	A/C	0.46	1	0.8	(0.72-0.89)	5.06E-05	0.06
rs2853677	5	1287194	A/G	0.71	1	0.79	(0.7-0.89)	5.88E-05	0.07
rs7186207	16	72035359	C/T	0.58	1	1.25	(1.12-1.39)	5.76E-05	0.07
rs951266	15	78878541	A/G	0.12	0.99	1.41	(1.19-1.66)	6.13E-05	0.07
rs8051239	16	72036257	A/T	0.62	1	1.25	(1.12-1.39)	8.21E-05	0.09
rs7195958	16	72036577	G/A	0.62	1	1.24	(1.11-1.39)	9.77E-05	0.10
rs9928425	16	53854498	G/A	0.93	0.99	1.51	(1.23-1.86)	9.77E-05	0.10

Table 31. Top SNPs of suggestive significance among never smoking African American lung cancer cases and controls (N=1,318).

SNP	Chr.	BP	Risk/ Ref. Allele	Risk Allele Freq.	Info Score	OR	95% CI	Unadjusted P-value	FDR P-value
rs115936429	5	638087	A/G	0.03	0.99	7.19	(3-17.22)	9.67E-06	0.13
rs200570688	15	78265899	T/C	0.38	0.93	1.99	(1.47-2.7)	1.02E-05	0.13
rs2937593	5	629879	T/C	0.03	0.96	7.47	(3.12-17.89)	6.50E-06	0.13

Table 32. SNP*sex interaction p-values for SNPs significantly associated with lung cancer risk in sex-stratified analyses.

SNP	Unadjusted P-value	FDR P-value
rs17486278	1.98E-01	2.37E-01
rs113931022	5.97E-03	1.43E-02
rs138544659	5.97E-03	1.43E-02
rs112878080	5.97E-03	1.43E-02
rs147144681	1.99E-02	2.66E-02
rs111704647	5.47E-03	1.43E-02
rs55676755	9.73E-03	1.57E-02
rs56077333	7.33E-03	1.47E-02
rs12914385	1.05E-02	1.57E-02
rs55781567	5.87E-02	5.87E-01
rs58365910	2.34E-01	2.55E-01
rs7486184	8.51E-06	1.02E-04

Table 33. SNP*Smoking interaction p-values for SNPs significantly associated with lung cancer risk in smoking status-stratified analyses.

SNP	Unadjusted P-value	FDR P-value
rs17486278	3.08 E-01	6.23 E-01
rs55781567	7.74 E-01	8.40 E-01
rs147144681	2.24 E-01	6.23 E-01
rs2036527	8.80E-01	9.07E-01
rs576982	3.40 E-01	6.23 E-01
rs664172	2.85 E-01	6.23 E-01
rs667282	3.48 E-01	6.23 E-01
rs11637630	2.80 E-01	6.23 E-01
rs637137	3.79 E-01	6.26 E-01
rs12440014	1.21 E-01	6.23 E-01
rs3825845	2.72 E-01	6.23 E-01
rs938682	5.50 E-01	7.44 E-01
rs503464	3.48E-01	6.23E-01
rs55676755	2.96 E-01	6.23 E-01
rs569207	6.06 E-01	7.69 E-01
rs7183604	5.39 E-01	7.44 E-01
rs58365910	3.59E-01	6.23E-01
rs2456020	6.47 E-01	7.91 E-01
rs189218934	4.53 E-01	7.12 E-01
rs113931022	2.12 E-01	6.23 E-01
rs138544659	2.12 E-01	6.23 E-01
rs112878080	2.12 E-01	6.23 E-01
rs7170068	1.86 E-01	6.23 E-01
rs111704647	2.18 E-01	6.23 E-01
rs56077333	2.47 E-01	6.23 E-01
rs2735940	7.89E-01	8.40E-01
rs12914385	3.41 E-01	6.23 E-01
rs905739	3.38E-01	6.23E-01
rs2853672	7.89E-01	8.40E-01
rs2853677	7.38E-01	8.40E-01
rs7186207	9.92 E-01	9.92 E-01
rs951266	5.63 E-01	7.44 E-01
rs8051239	5.52 E-01	7.44 E-01

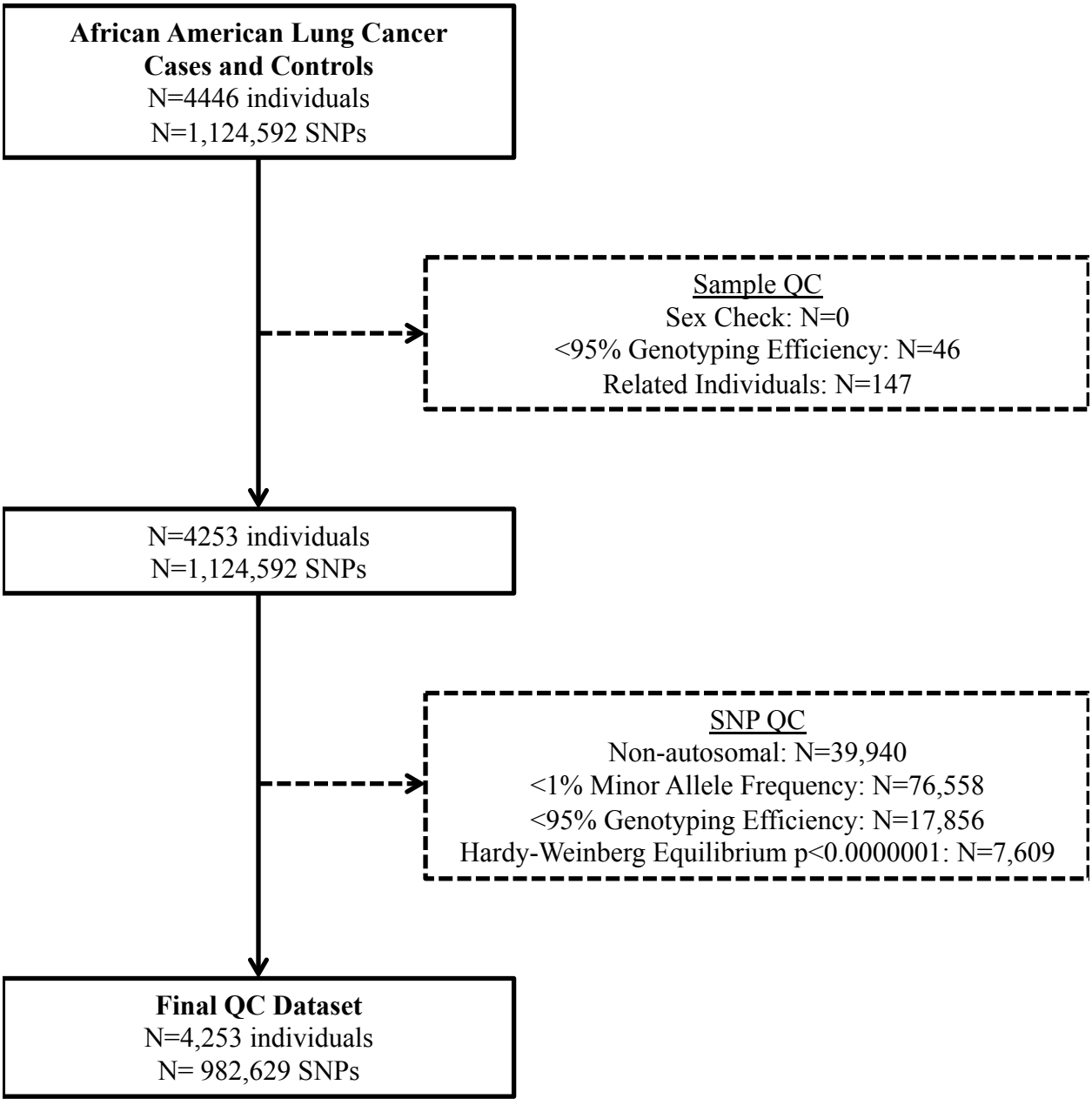

Figure 27. Quality control of African American lung cancer cases and controls genotyped on the Illumina Human Hap 1M Duo array.

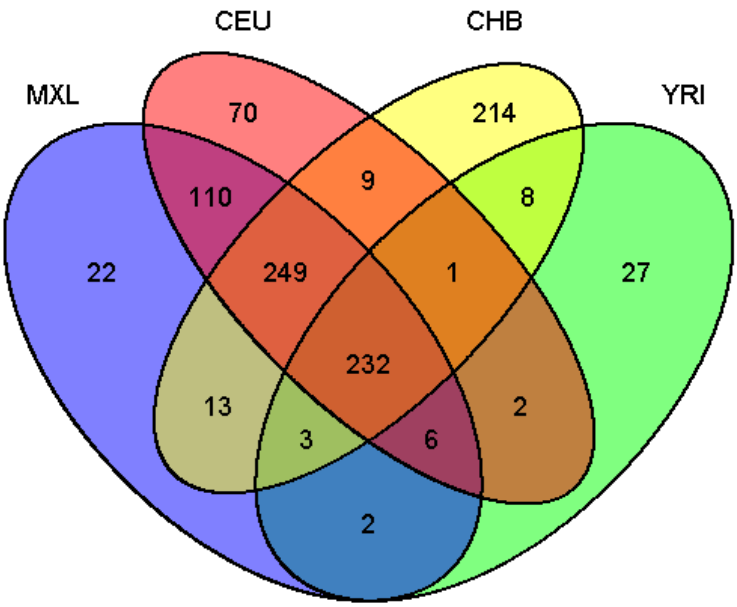

Figure 28. Overlap between the number of SNPs in LD with a reported SNP in the MXL, CEU, CHB, and YRI 1000 Genomes reference populations for variants previously reported in Latin populations.

MXL= Mexican ancestry from Los Angeles, CA; CEU=Utah residents (CEPH) with Northern an Western European ancestry; CHB=Han Chinese in Beijing, China; YRI=Yoruba in Ibadan, Nigeria

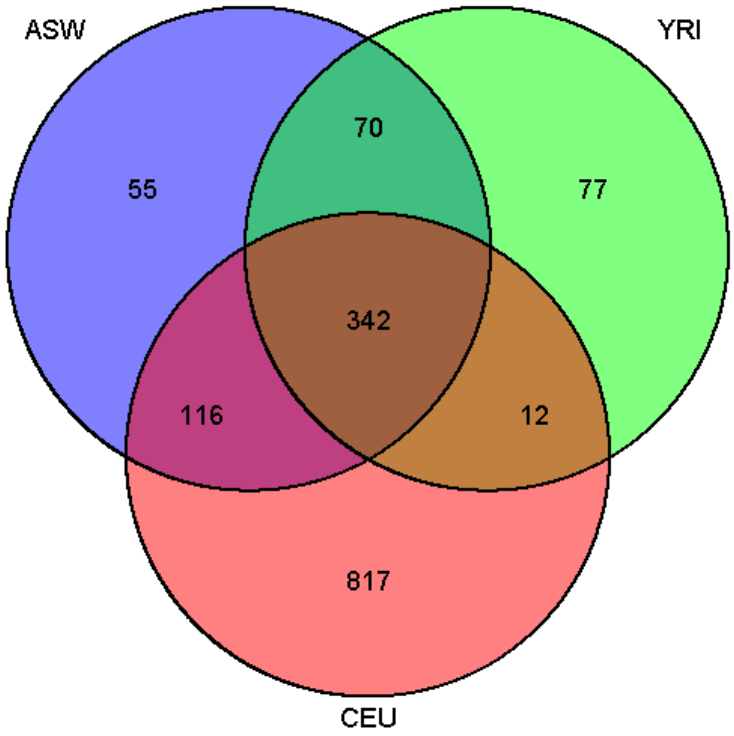

Figure 29. Overlap between the number of SNPs in LD with a reported SNP in the ASW, YRI, and CEU 1000 Genomes reference populations for variants previously reported in African-descent populations.

ASW=Americans of African ancestry in SW USA; CEU=Utah residents (CEPH) with Northern an Western European ancestry; YRI=Yoruba in Ibadan, Nigeria

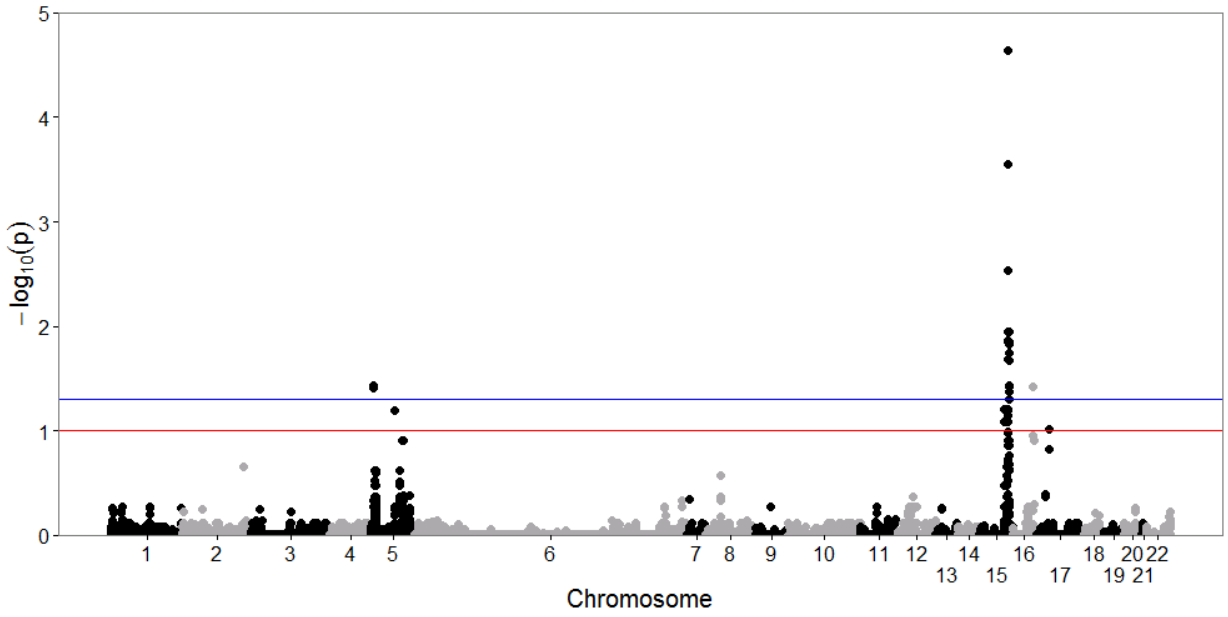

Figure 30. Logistic regression results for pleiotropic genetic associations with lung cancer risk among all African American lung cancer cases (N=4,253).

Red line = 10% false discover rate (FDR); blue line = 5% FDR.

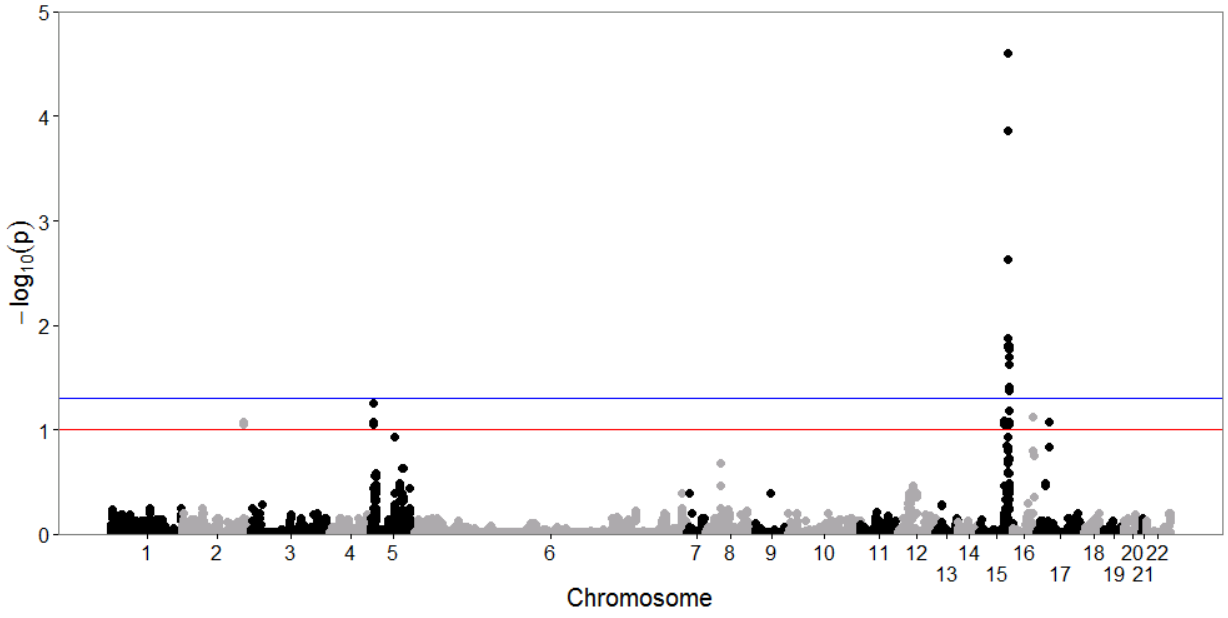

Figure 31. Meta-analysis logistic regression results for pleiotropic genetic associations with lung cancer risk

Red line = 10% false discover rate (FDR); blue line = 5% FDR.

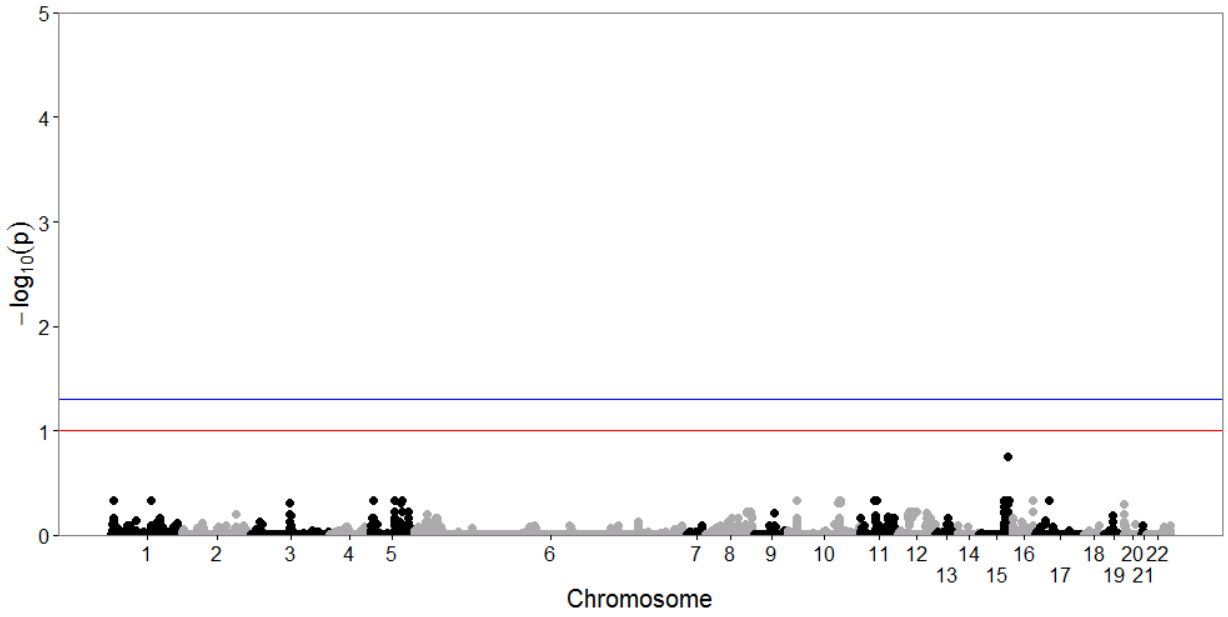

Figure 32. Logistic regression results for pleiotropic genetic associations with lung cancer risk among African American adenocarcinoma cases and controls (N=3,475).

Red line = 10% FDR; blue line = 5% FDR.

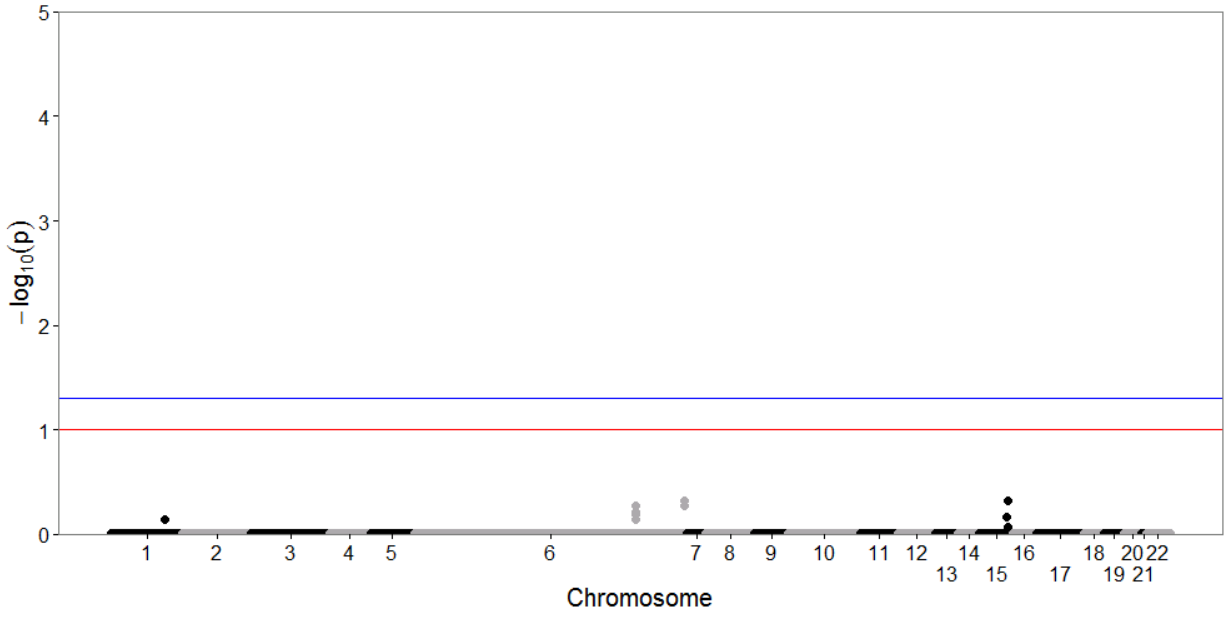

Figure 33. Logistic regression results for pleiotropic genetic associations with lung cancer risk among African American squamous cell carcinoma cases and controls (N=3,180).

Red line = 10% FDR; blue line = 5% FDR.

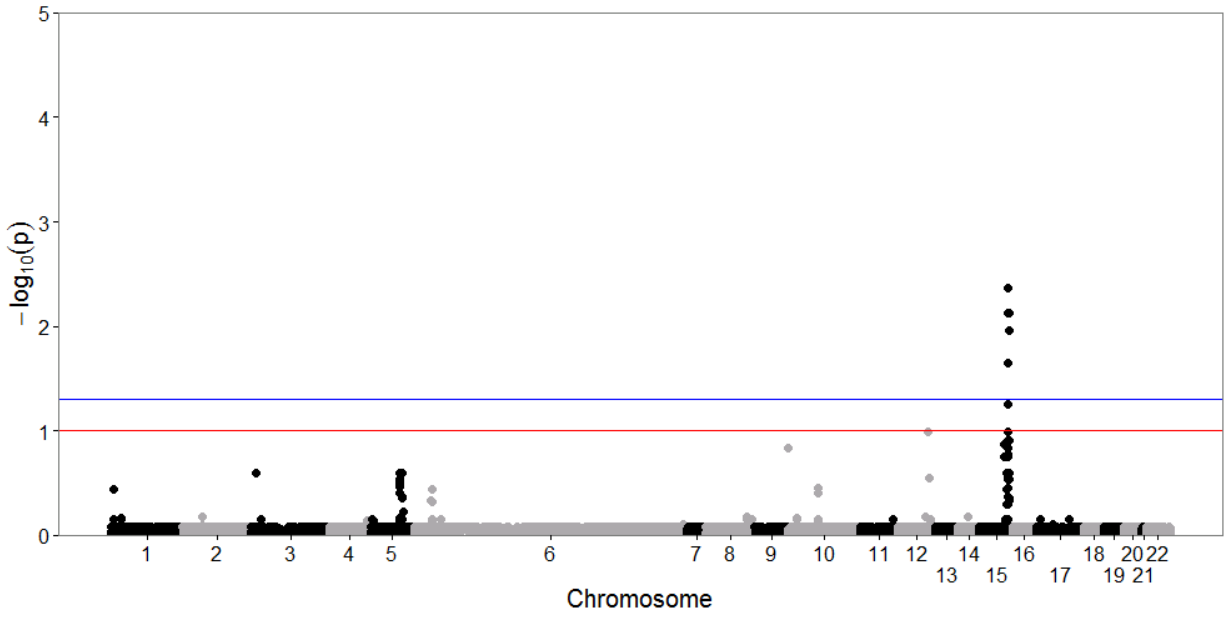

Figure 34. Logistic regression results for pleiotropic genetic associations with lung cancer risk among African American female cases and controls (N=2,353).

Red line = 10% FDR; blue line = 5% FDR.

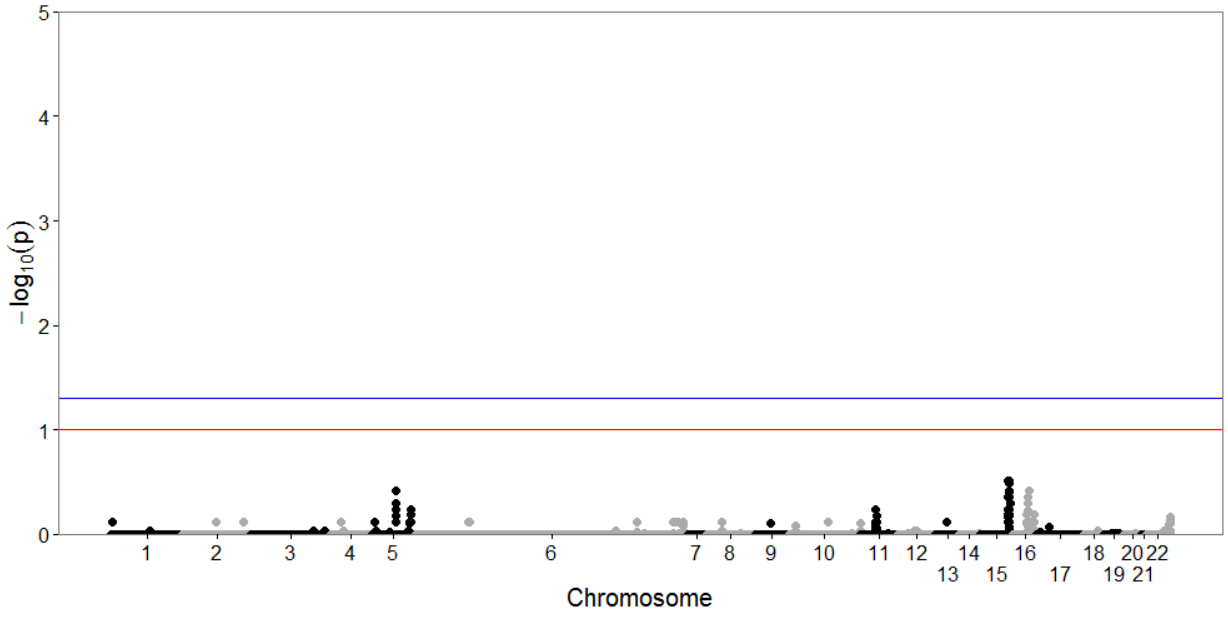

Figure 35. Logistic regression results for pleiotropic genetic associations with lung cancer risk among African American male cases and controls (N=1,900).

Red line = 10% FDR; blue line = 5% FDR.

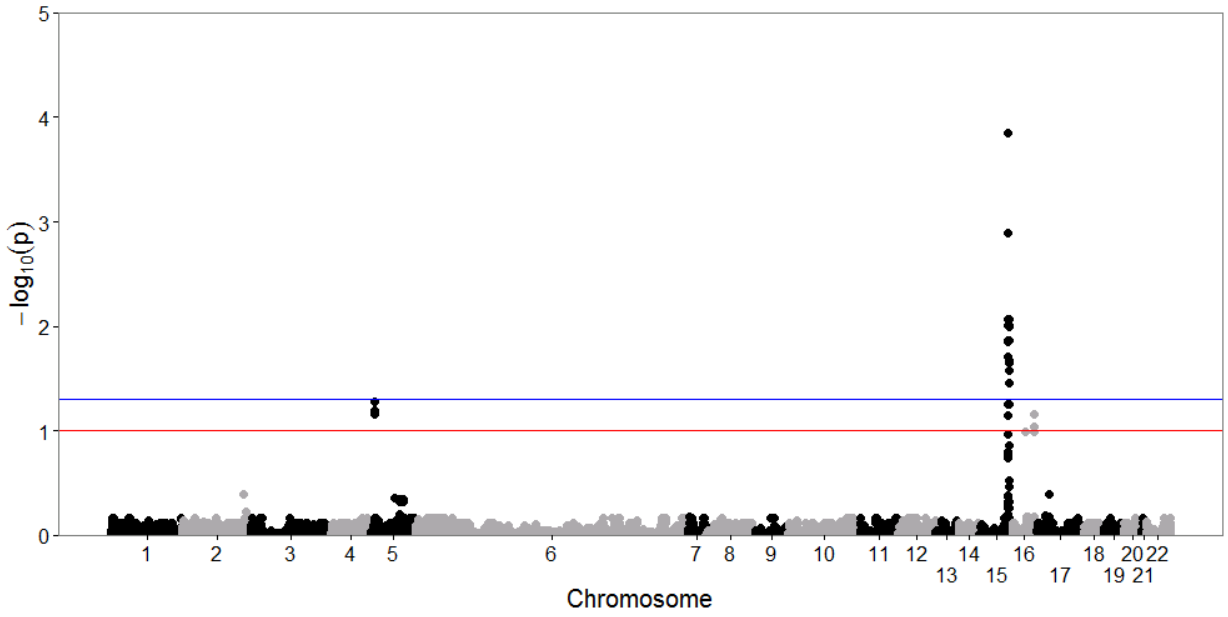

Figure 36. Logistic regression results for pleiotropic genetic associations with lung cancer risk among African American ever smoking cases and controls (N=2,891).

Red line = 10% FDR; blue line = 5% FDR.

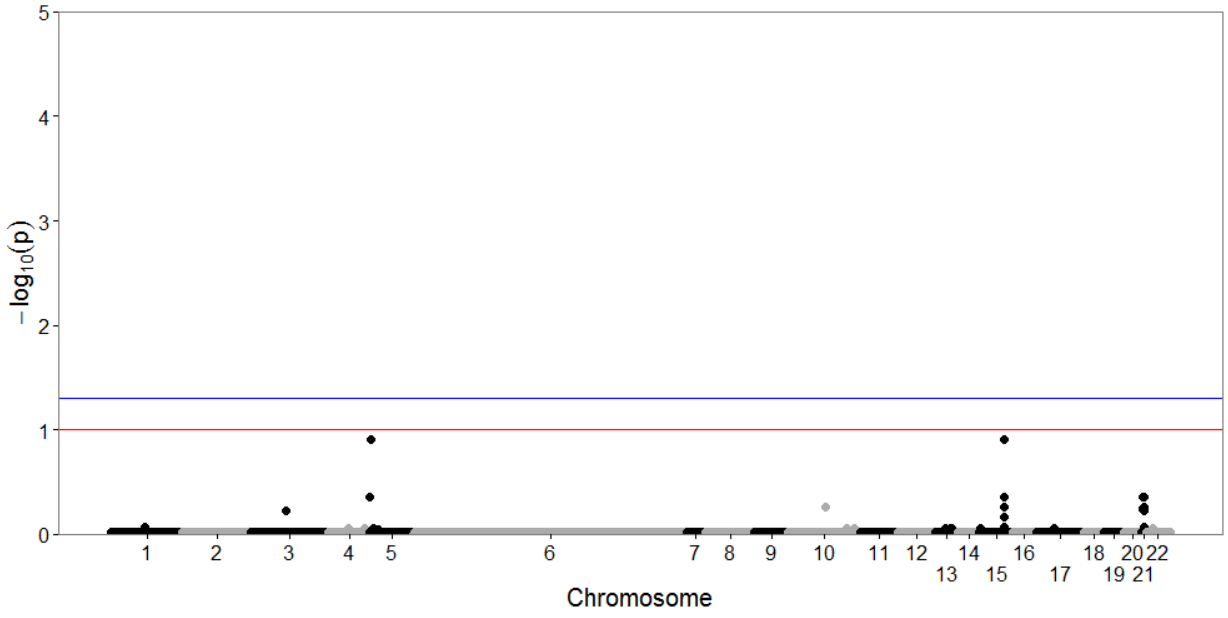

Figure 37. Logistic regression results for pleiotropic genetic associations with lung cancer risk among African American never smoking cases and controls (N=1,318).

Red line = 10% FDR; blue line = 5% FDR.

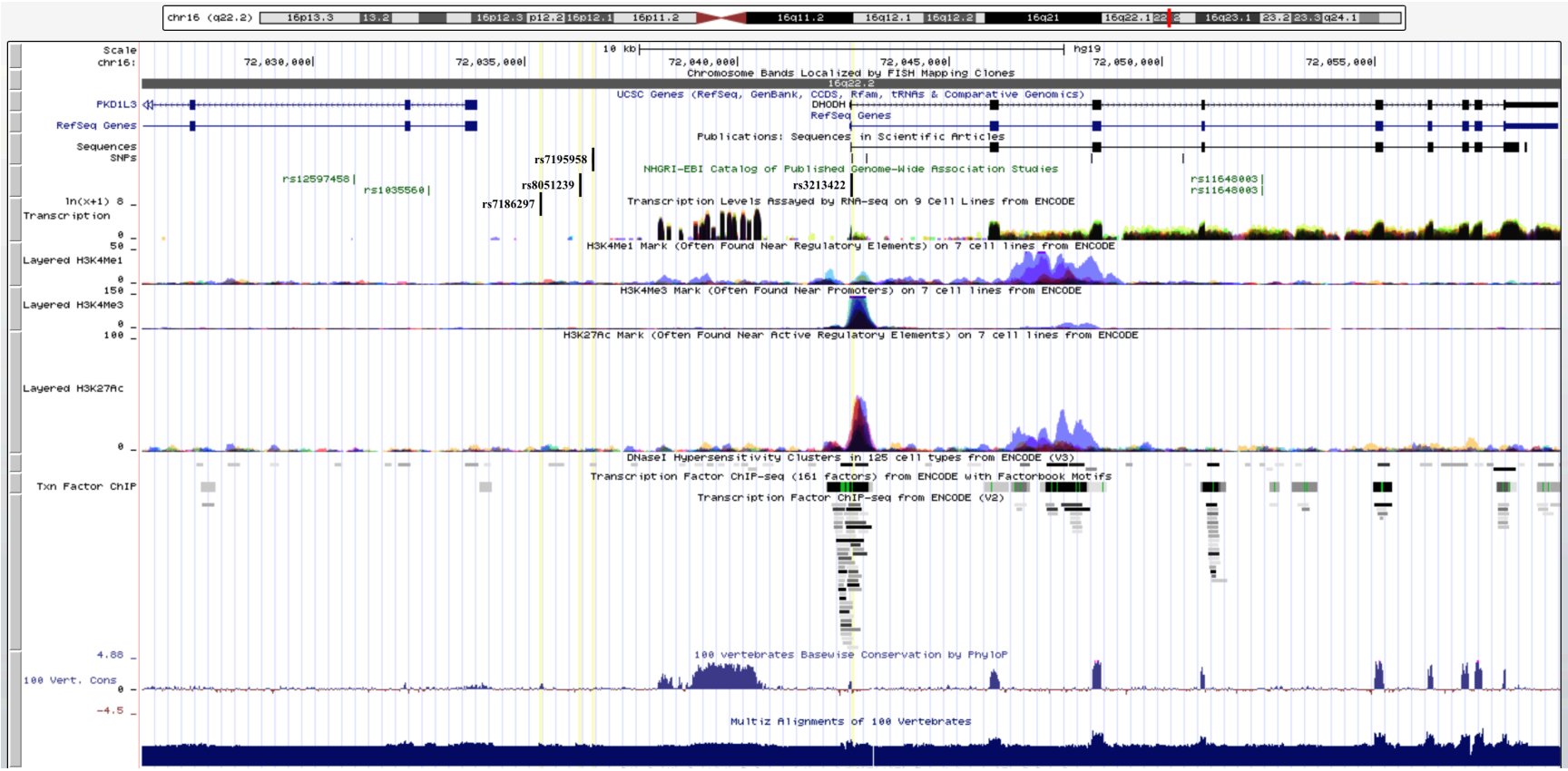

Figure 38. Chromosome 16q22.2 as visualized in the UCSC Genome Browser.

Variants rs7186207, rs8051239, rs7195958, and rs3213422 are denoted by faint yellow lines and labeled in black under the "NHGRI-EBI Catalog of Published Genome-Wide Association Studies" track.

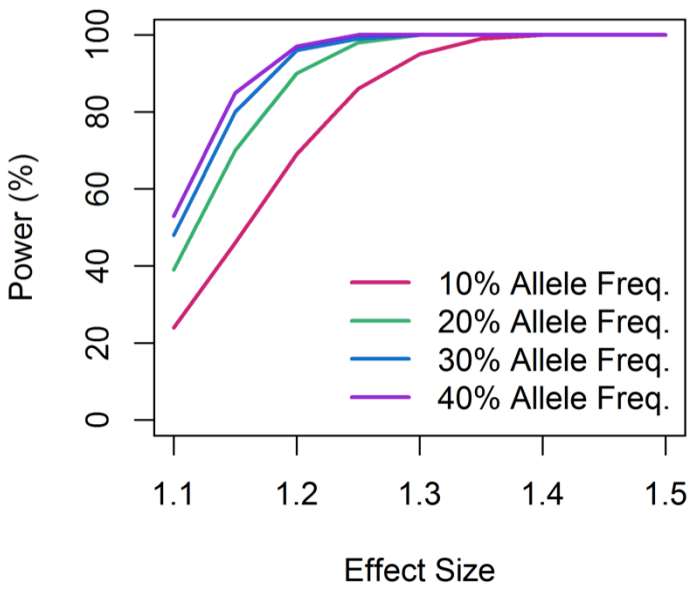

Figure 39. Power of the present study to detect associations at various allele frequencies and effect sizes.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

The goal of this dissertation was to examine genetic contributors to lung cancer survival and risk in African Americans. The three aims were as follows:

1. Examine the role of African genetic ancestry compared to other known contributors to lung cancer survival in a population of whites and blacks (Chapter 2).
2. Examine variants previously associated with lung cancer survival and identify novel common and rare variants associated with lung cancer survival in an African American population (Chapter 3).
3. Investigate cross-cancer pleiotropic associations for lung cancer risk in African Americans (Chapter 4).

Our analysis of lung cancer survival utilized individuals from the Southern Community Cohort Study (SCCS), a prospective study of ~86,000 individuals recruited in the Southeastern Untied States. The SCCS is a unique study population in that the majority of individuals are black/African American. Furthermore, the primary method of recruitment was through community health centers, resulting in a low-income population with relatively equal access to healthcare (though exceptions exist for the ~15% of individuals recruited from the general population) (Signorello et al. 2005, Signorello et al. 2010). Using a population of white and black/African American incident lung cancer cases from the SCCS, we sought to provide a better understanding of the impact of genetic ancestry on lung cancer survival, while adjusting for known clinical, environmental, and socioeconomic confounders. We find that global African

ancestry is not associated with lung cancer survival, nor does it increase accuracy of lung cancer survival prediction models. In contrast, we find that stage at diagnosis and the type of treatment received are highly predictive of lung cancer survival.

This study contributes evidence against a racial disparity in lung cancer survival due to genetics alone and instead suggests that the observed reduced survival time in blacks compared to whites is the result of differing stage and treatment distributions. Even in our study population of relatively equal access to healthcare, we observe that stage at diagnosis and treatment received differ between blacks and whites. Additionally, stage and treatment distributions vary dramatically between the SCCS discovery population and WSU replication population from the Metropolitan Detroit area. Future research should focus on providing a better understanding of the interplay between stage, treatment, and race/ancestry in lung cancer survival. Clinically, emphasis should be placed on shifting the stage distribution toward earlier stage diagnoses for all races, which will naturally increase lung cancer survival rates. Increasing the frequency of early stage diagnoses could result from increased lung cancer screening efforts among all racial/ethnic groups. Furthermore, increased knowledge of known risk factors will also improve our ability to detect and diagnose lung cancer at earlier stages.

The second aim of this dissertation sought to identify genetic variants associated with lung cancer survival in the same population of African Americans from the SCCS used in Aim 1. Our primary analysis focused first on replicating variants previously associated with lung cancer survival, then on identifying novel common and rare variation associated with lung cancer survival. Our primary analysis identified one SNP, rs1878022 on chromosome 12q23.3, associated with lung cancer survival, though the direction of effect was in contrast to previous literature in a European population (Wu et al. 2011), suggesting possible population-specific

effects. Continued research is required to confirm this association and its opposing effects in European- vs. African-descent populations. Fine-mapping of this region is also necessary to identify the causal allele and understand the functional impact of variation and its effect on lung cancer survival.

While we examined rare variant associations through multiple aggregation-based methods, we are underpowered to draw any conclusions about their association with lung cancer survival. Substantially increased sample sizes are required in order to confirm any associations suggested in the present analysis. Our analysis of common variants, on the other hand, identified several variants with suggestive significance, including a peak on chromosome 6p21.33. While previous literature on this region provides biological plausibility (Sasaki et al. 2007, Tompkins et al. 2009, Yang et al. 2010, Xun et al. 2011, Wu et al. 2013), we lacked an appropriate replication population to confirm the association. Thus, replication remains necessary. Combined with the primary analysis of Aim 2, we identified multiple gene regions that may be associated with lung cancer survival. Should these regions be confirmed through replication, future analysis should focus on understanding the mechanism through which they influence survival and their potential as therapeutic targets in lung cancer treatment.

Finally, the third aim of this dissertation sought to identify genetic variants associated with lung cancer risk. Using a large population of African American lung cancer case and controls from five different study populations, we identified several genetic loci associated with lung cancer risk, including previous associations on chromosomes 15q25 and 5p15. Significant peaks on chromosomes 5q14.3, 16q22.2, and 17q12 were observed among all individuals and a SNP on 12q21.32 fell just below our significance threshold. SNPs in all of these regions have been previously implicated in cancer risk, suggesting plausibility for a role in lung cancer.

However, no replication population was available to confirm these novel associations. Future analysis requires replicating significant associations and investigating LD structure for the purpose of identifying causal alleles conferring increased lung cancer risk. Positive identification of lung cancer risk alleles has the potential to increase prevention and detection methods, though such applications would require extensive continued epidemiologic and molecular research.

The final aim also provided evidence of the utility in using pleiotropy to perform hypothesis-driven research. Unlike traditional genome-wide association studies, we only examined SNPs with *a priori* evidence for an association with cancer risk, reducing the total number of tests performed. Because we were not hindered by a high multiple testing correction, we were able to detect novel lung cancer risk loci, likely rejected as false positives in previous studies. Furthermore, our study shows the importance of considering population structure differences across races: none of our most significant associations were among the list of SNPs reported in the NHGRI-EBI GWAS catalog, but were examined because of their strong correlation to a reported SNP. Future studies of pleiotropy should include such considerations.

Taken together, the aims of this dissertation expand our understanding of the genetic components of lung cancer survival and risk in African Americans. While survival and risk are often considered separately, the high frequency of late stage diagnoses combined with extremely poor survival rates for such diagnoses makes knowledge of lung cancer risk factors also an important part of the fight to increase lung cancer survival rates. Only by understanding the social, environmental, and genetic risk factors and their combined impact, will we be able to fully address the observed racial disparity in lung cancer incidence and survival.

REFERENCES

Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov and S. R. Sunyaev (2010). "A method and server for predicting damaging missense mutations." Nat Methods **7**(4): 248-249.

Ahn, M. J., H. H. Won, J. Lee, S. T. Lee, J. M. Sun, Y. H. Park, J. S. Ahn, O. J. Kwon, H. Kim, Y. M. Shim, J. Kim, K. Kim, Y. H. Kim, J. Y. Park, J. W. Kim and K. Park (2012). "The 18p11.22 locus is associated with never smoker non-small cell lung cancer susceptibility in Korean populations." Hum Genet **131**(3): 365-372.

Aivazian, D., R. L. Serrano and S. Pfeffer (2006). "TIP47 is a key effector for Rab9 localization." J Cell Biol **173**(6): 917-926.

Aizer, A. A., T. J. Wilhite, M. H. Chen, P. L. Graham, T. K. Choueiri, K. E. Hoffman, N. E. Martin, Q. D. Trinh, J. C. Hu and P. L. Nguyen (2014). "Lack of reduction in racial disparities in cancer-specific mortality over a 20-year period." Cancer **120**(10): 1532-1539.

Al Olama, A. A., Z. Kote-Jarai, S. I. Berndt, D. V. Conti, F. Schumacher, Y. Han, S. Benlloch, D. J. Hazelett, Z. Wang, E. Saunders, D. Leongamornlert, S. Lindstrom, S. Jugurnauth-Little, T. Dadaev, M. Tymrakiewicz, D. O. Stram, K. Rand, P. Wan, A. Stram, X. Sheng, L. C. Pooler, K. Park, L. Xia, J. Tyrer, L. N. Kolonel, L. Le Marchand, R. N. Hoover, M. J. Machiela, M. Yeager, L. Burdette, C. C. Chung, A. Hutchinson, K. Yu, C. Goh, M. Ahmed, K. Govindasami, M. Guy, T. L. Tammela, A. Auvinen, T. Wahlfors, J. Schleutker, T. Visakorpi, K. A. Leinonen, J. Xu, M. Aly, J. Donovan, R. C. Travis, T. J. Key, A. Siddiq, F. Canzian, K. T. Khaw, A. Takahashi, M. Kubo, P. Pharoah, N. Pashayan, M. Weischer, B. G. Nordestgaard, S. F. Nielsen, P. Klarskov, M. A. Roder, P. Iversen, S. N. Thibodeau, S. K. McDonnell, D. J. Schaid, J. L. Stanford, S. Kolb, S. Holt, B. Knudsen, A. H. Coll, S. M. Gapstur, W. R. Diver, V. L. Stevens, C. Maier, M. Luedeke, K. Herkommer, A. E. Rinckleb, S. S. Strom, C. Pettaway, E. D. Yeboah, Y. Tettey, R. B. Biritwum, A. A. Adjei, E. Tay, A. Truelove, S. Niwa, A. P. Chokkalingam, L. Cannon-Albright, C. Cybulski, D. Wokolorczyk, W. Kluzniak, J. Park, T. Sellers, H. Y. Lin, W. B. Isaacs, A. W. Partin, H. Brenner, A. K. Dieffenbach, C. Stegmaier, C. Chen, E. L. Giovannucci, J. Ma, M. Stampfer, K. L. Penney, L. Mucci, E. M. John, S. A. Ingles, R. A. Kittles, A. B. Murphy, H. Pandha, A. Michael, A. M. Kierzek, W. Blot, L. B. Signorello, W. Zheng, D. Albanes, J. Virtamo, S. Weinstein, B. Nemesure, J. Carpten, C. Leske, S. Y. Wu, A. Hennis, A. S. Kibel, B. A. Rybicki, C. Neslund-Dudas, A. W. Hsing, L. Chu, P. J. Goodman, E. A. Klein, S. L. Zheng, J. Batra, J. Clements, A. Spurdle, M. R. Teixeira, P. Paulo, S. Maia, C. Slavov, R. Kaneva, V. Mitev, J. S. Witte, G. Casey, E. M. Gillanders, D. Seminara, E. Riboli, F. C. Hamdy, G. A. Coetzee, Q. Li, M. L. Freedman, D. J. Hunter, K. Muir, H. Gronberg, D. E. Neal, M. Southey, G. G. Giles, G. Severi, Breast, C. Prostate Cancer Cohort, P. Consortium, C. Consortium, G.-O. E. Consortium, M. B. Cook, H. Nakagawa, F. Wiklund, P. Kraft, S. J. Chanock, B. E. Henderson, D. F. Easton, R. A. Eeles and C. A. Haiman (2014). "A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer." Nat Genet **46**(10): 1103-1109.

Aldrich, M. C., E. L. Grogan, H. M. Munro, L. B. Signorello and W. J. Blot (2013). "Stage-Adjusted Lung Cancer Survival Does Not Differe between Low-Income Blacks and Whites." Journal of Thoracic Oncology **8**(10): 1248-1454.

Alexander, D. H., J. Novembre and K. Lange (2009). "Fast model-based estimation of ancestry in unrelated individuals." Genome Res **19**(9): 1655-1664.

American Cancer Society (2017). Cancer Facts & Figures 2017. Atlanta, American Cancer Society.

Amos, C. I., S. M. Pinney, Y. Li, E. Kupert, J. Lee, M. A. de Andrade, P. Yang, A. G. Schwartz, P. R. Fain, A. Gazdar, J. Minna, J. S. Wiest, D. Zeng, H. Rothschild, D. Mandal, M. You, T. Coons, C. Gaba, J. E. Bailey-Wilson and M. W. Anderson (2010). "A susceptibility locus on chromosome 6q greatly increases lung cancer risk among light and never smokers." Cancer Res **70**(6): 2359-2367.

Amos, C. I., X. Wu, P. Broderick, I. P. Gorlov, J. Gu, T. Eisen, Q. Dong, Q. Zhang, X. Gu, J. Vijayakrishnan, K. Sullivan, A. Matakidou, Y. Wang, G. Mills, K. Doheny, Y. Y. Tsai, W. V. Chen, S. Shete, M. R. Spitz and R. S. Houlston (2008). "Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1." Nat Genet **40**(5): 616-622.

Attfield, M. D., P. L. Schleiff, J. H. Lubin, A. Blair, P. A. Stewart, R. Vermeulen, J. B. Coble and D. T. Silverman (2012). "The Diesel Exhaust in Miners study: a cohort mortality study with emphasis on lung cancer." J Natl Cancer Inst **104**(11): 869-883.

Bach, P. B., L. D. Cramer, J. L. Warren and C. B. Begg (1999). "Racial differences in the treatment of early-stage lung cancer." N Engl J Med **341**(16): 1198-1205.

Bae, E. Y., S. Y. Lee, B. K. Kang, E. J. Lee, Y. Y. Choi, H. G. Kang, J. E. Choi, H. S. Jeon, W. K. Lee, S. Kam, K. M. Shin, G. Jin, S. S. Yoo, J. Lee, S. I. Cha, C. H. Kim, T. H. Jung and J. Y. Park (2012). "Replication of results of genome-wide association studies on lung cancer susceptibility loci in a Korean population." Respirology **17**(4): 699-706.

Baharian, S., M. Barakatt, C. R. Gignoux, S. Shringarpure, J. Errington, W. J. Blot, C. D. Bustamante, E. E. Kenny, S. M. Williams, M. C. Aldrich and S. Gravel (2016). "The Great Migration and African-American Genomic Diversity." PLoS Genet **12**(5): e1006059.

Bailey-Wilson, J. E., C. I. Amos, S. M. Pinney, G. M. Petersen, M. de Andrade, J. S. Wiest, P. Fain, A. G. Schwartz, M. You, W. Franklin, C. Klein, A. Gazdar, H. Rothschild, D. Mandal, T.

Coons, J. Slusser, J. Lee, C. Gaba, E. Kupert, A. Perez, X. Zhou, D. Zeng, Q. Liu, Q. Zhang, D. Seminara, J. Minna and M. W. Anderson (2004). "A major lung cancer susceptibility locus maps to chromosome 6q23-25." Am J Hum Genet **75**(3): 460-474.

Baker, J. L., C. N. Rotimi and D. Shriner (2017). "Human ancestry correlates with language and reveals that race is not an objective genomic classifier." Sci Rep **7**(1): 1572.

Baumann, P., S. Mandl-Weber, A. Volkl, C. Adam, I. Bumeder, F. Oduncu and R. Schmidmaier (2009). "Dihydroorotate dehydrogenase inhibitor A771726 (leflunomide) induces apoptosis and diminishes proliferation of multiple myeloma cells." Mol Cancer Ther **8**(2): 366-375.

Bell, D. W., I. Gore, R. A. Okimoto, N. Godin-Heymann, R. Sordella, R. Mulloy, S. V. Sharma, B. W. Brannigan, G. Mohapatra, J. Settleman and D. A. Haber (2005). "Inherited susceptibility to lung cancer may be associated with the T790M drug resistance mutation in EGFR." Nat Genet **37**(12): 1315-1316.

Berndt, S. I., Z. Wang, M. Yeager, M. C. Alavanja, D. Albanes, L. Amundadottir, G. Andriole, L. Beane Freeman, D. Campa, G. Cancel-Tassin, F. Canzian, J. N. Cornu, O. Cussenot, W. R. Diver, S. M. Gapstur, H. Gronberg, C. A. Haiman, B. Henderson, A. Hutchinson, D. J. Hunter, T. J. Key, S. Kolb, S. Koutros, P. Kraft, L. Le Marchand, S. Lindstrom, M. J. Machiela, E. A. Ostrander, E. Riboli, F. Schumacher, A. Siddiq, J. L. Stanford, V. L. Stevens, R. C. Travis, K. K. Tsilidis, J. Virtamo, S. Weinstein, F. Wilkund, J. Xu, S. Lilly Zheng, K. Yu, W. Wheeler, H. Zhang, G. C. African Ancestry Prostate Cancer, J. Sampson, A. Black, K. Jacobs, R. N. Hoover, M. Tucker and S. J. Chanock (2015). "Two susceptibility loci identified for prostate cancer aggressiveness." Nat Commun **6**: 6889.

Boyle, A. P., E. L. Hong, M. Hariharan, Y. Cheng, M. A. Schaub, M. Kasowski, K. J. Karczewski, J. Park, B. C. Hitz, S. Weng, J. M. Cherry and M. Snyder (2012). "Annotation of functional variation in personal genomes using RegulomeDB." Genome Res **22**(9): 1790-1797.

Brenner, D. R., J. R. McLaughlin and R. J. Hung (2011). "Previous lung diseases and lung cancer risk: a systematic review and meta-analysis." PLoS One **6**(3): e17479.

Broderick, P., Y. Wang, J. Vijayakrishnan, A. Matakidou, M. R. Spitz, T. Eisen, C. I. Amos and R. S. Houlston (2009). "Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study." Cancer Res **69**(16): 6633-6641.

Bryc, K., A. Auton, M. R. Nelson, J. R. Oksenberg, S. L. Hauser, S. Williams, A. Froment, J. M. Bodo, C. Wambebe, S. A. Tishkoff and C. D. Bustamante (2010). "Genome-wide patterns of population structure and admixture in West Africans and African Americans." Proc Natl Acad Sci U S A **107**(2): 786-791.

Bryc, K., E. Y. Durand, J. M. Macpherson, D. Reich and J. L. Mountain (2015). "The genetic ancestry of African Americans, Latinos, and European Americans across the United States." Am J Hum Genet **96**(1): 37-53.

Bubancova, I., H. Kovarikova, J. Laco, E. Ruszova, O. Dvorak, V. Palicka and M. Chmelarova (2017). "Next-Generation Sequencing Approach in Methylation Analysis of HNF1B and GATA4 Genes: Searching for Biomarkers in Ovarian Cancer." Int J Mol Sci **18**(2).

Burdett T (EBI), H. P. N., Hastings E (EBI), Hindorff LA (NHGRI), Junkins HA (NHGRI), Klemm AK (NHGRI), MacArthur J (EBI), Manolio TA (NHGRI), Morales J (EBI), Parkinson H (EBI) and Welter D (EBI). "The NHGRI-EBI Catalog of published genome-wide association studies." Retrieved Februray 21, 2014, from www.ebi.ac.uk/gwas.

Cai, Q., B. Zhang, H. Sung, S. K. Low, S. S. Kweon, W. Lu, J. Shi, J. Long, W. Wen, J. Y. Choi, D. Y. Noh, C. Y. Shen, K. Matsuo, S. H. Teo, M. K. Kim, U. S. Khoo, M. Iwasaki, M. Hartman, A. Takahashi, K. Ashikawa, K. Matsuda, M. H. Shin, M. H. Park, Y. Zheng, Y. B. Xiang, B. T. Ji, S. K. Park, P. E. Wu, C. N. Hsiung, H. Ito, Y. Kasuga, P. Kang, S. Mariapun, S. H. Ahn, H. S. Kang, K. Y. Chan, E. P. Man, H. Iwata, S. Tsugane, H. Miao, J. Liao, Y. Nakamura, M. Kubo, D. G.-O. Consortium, R. J. Delahanty, Y. Zhang, B. Li, C. Li, Y. T. Gao, X. O. Shu, D. Kang and W. Zheng (2014). "Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1." Nat Genet **46**(8): 886-890.

Cannon-Albright, L. A., A. Thomas, D. E. Goldgar, K. Gholami, K. Rowe, M. Jacobsen, W. P. McWhorter and M. H. Skolnick (1994). "Familiality of cancer in Utah." Cancer Res **54**(9): 2378-2385.

Caposole, M. Z., K. Miller, J. N. Kim, N. A. Steward and T. L. Bauer (2014). "Elimination of socioeconomic and racial disparities related to lung cancer: closing the gap at a high volume community cancer center." Surg Oncol **23**(2): 46-52.

Cappuzzo, F., L. Bemis and M. Varella-Garcia (2006). "HER2 mutation and response to trastuzumab therapy in non-small-cell lung cancer." N Engl J Med **354**(24): 2619-2621.

Caunt, C. J., M. J. Sale, P. D. Smith and S. J. Cook (2015). "MEK1 and MEK2 inhibitors and cancer therapy: the long and winding road." Nat Rev Cancer **15**(10): 577-592.

Centers for Disease Control and Prevention Annual smoking-attributable mortality, years of potential life lost, and productivity losses—United States, 1997–2001. Morb Mortal Wkly Rep 2005 ; 54 : 625 –8.

Chen, L. S. and K. A. Kaphingst (2011). "Risk perceptions and family history of lung cancer: differences by smoking status." Public Health Genomics **14**(1): 26-34.

Cheng, I., J. M. Kocarnik, L. Dumitrescu, N. M. Lindor, J. Chang-Claude, C. L. Avery, C. P. Caberto, S. A. Love, M. L. Slattery, A. T. Chan, J. A. Baron, L. A. Hindorff, S. L. Park, F. R. Schumacher, M. Hoffmeister, P. Kraft, A. M. Butler, D. J. Duggan, L. Hou, C. S. Carlson, K. R. Monroe, Y. Lin, C. L. Carty, S. Mann, J. Ma, E. L. Giovannucci, C. S. Fuchs, P. A. Newcomb, M. A. Jenkins, J. L. Hopper, R. W. Haile, D. V. Conti, P. T. Campbell, J. D. Potter, B. J. Caan, R. E. Schoen, R. B. Hayes, S. J. Chanock, S. I. Berndt, S. Kury, S. Bezieau, J. L. Ambite, G. Kumaraguruparan, D. M. Richardson, R. J. Goodloe, H. H. Dilks, P. Baker, B. W. Zanke, M. Lemire, S. Gallinger, L. Hsu, S. Jiao, T. A. Harrison, D. Seminara, C. A. Haiman, C. Kooperberg, L. R. Wilkens, C. M. Hutter, E. White, D. C. Crawford, G. Heiss, T. J. Hudson, H. Brenner, W. S. Bush, G. Casey, L. Le Marchand and U. Peters (2014). "Pleiotropic effects of genetic risk variants for other cancers on colorectal cancer risk: PAGE, GECCO and CCFR consortia." Gut **63**(5): 800-807.

Chung, C. C., P. A. Kanetsky, Z. Wang, M. A. Hildebrandt, R. Koster, R. I. Skotheim, C. P. Kratz, C. Turnbull, V. K. Cortessis, A. C. Bakken, D. T. Bishop, M. B. Cook, R. L. Erickson, S. D. Fossa, K. B. Jacobs, L. A. Korde, S. M. Kraggerud, R. A. Lothe, J. T. Loud, N. Rahman, E. C. Skinner, D. C. Thomas, X. Wu, M. Yeager, F. R. Schumacher, M. H. Greene, S. M. Schwartz, K. A. McGlynn, S. J. Chanock and K. L. Nathanson (2013). "Meta-analysis identifies four new loci associated with testicular germ cell tumor." Nat Genet **45**(6): 680-685.

Cote, M. L., S. L. Kardia, A. S. Wenzlaff, J. C. Ruckdeschel and A. G. Schwartz (2005). "Risk of lung cancer among white and black relatives of individuals with early-onset lung cancer." JAMA **293**(24): 3036-3042.

Cote, M. L., M. Liu, S. Bonassi, M. Neri, A. G. Schwartz, D. C. Christiani, M. R. Spitz, J. E. Muscat, G. Rennert, K. K. Aben, A. S. Andrew, V. Bencko, H. Bickeboller, P. Boffetta, P. Brennan, H. Brenner, E. J. Duell, E. Fabianova, J. K. Field, L. Foretova, S. Friis, C. C. Harris, I. Holcatova, Y. C. Hong, D. Isla, V. Janout, L. A. Kiemeney, C. Kiyohara, Q. Lan, P. Lazarus, J. Lissowska, L. Le Marchand, D. Mates, K. Matsuo, J. I. Mayordomo, J. R. McLaughlin, H. Morgenstern, H. Mueller, I. Orlow, B. J. Park, M. Pinchev, O. Y. Raji, H. S. Rennert, P. Rudnai, A. Seow, I. Stucker, N. Szeszenia-Dabrowska, M. Dawn Teare, A. Tjonnelan, D. Ugolini, H. F. van der Heijden, E. Wichmann, J. K. Wiencke, P. J. Woll, P. Yang, D. Zaridze, Z. F. Zhang, C. J. Etzel and R. J. Hung (2012). "Increased risk of lung cancer in individuals with a family history of the disease: a pooled analysis from the International Lung Cancer Consortium." Eur J Cancer **48**(13): 1957-1968.

Couch, F. J., K. B. Kuchenbaecker, K. Michailidou, G. A. Mendoza-Fandino, S. Nord, J. Lilyquist, C. Olswold, E. Hallberg, S. Agata, H. Ahsan, K. Aittomaki, C. Ambrosone, I. L. Andrulis, H. Anton-Culver, V. Arndt, B. K. Arun, B. Arver, M. Barile, R. B. Barkardottir, D.

Barrowdale, L. Beckmann, M. W. Beckmann, J. Benitez, S. V. Blank, C. Blomqvist, N. V. Bogdanova, S. E. Bojesen, M. K. Bolla, B. Bonanni, H. Brauch, H. Brenner, B. Burwinkel, S. S. Buys, T. Caldes, M. A. Caligo, F. Canzian, J. Carpenter, J. Chang-Claude, S. J. Chanock, W. K. Chung, K. B. Claes, A. Cox, S. S. Cross, J. M. Cunningham, K. Czene, M. B. Daly, F. Damiola, H. Darabi, M. de la Hoya, P. Devilee, O. Diez, Y. C. Ding, R. Dolcetti, S. M. Domchek, C. M. Dorfling, I. Dos-Santos-Silva, M. Dumont, A. M. Dunning, D. M. Eccles, H. Ehrencrona, A. B. Ekici, H. Eliassen, S. Ellis, P. A. Fasching, J. Figueroa, D. Flesch-Janys, A. Forsti, F. Fostira, W. D. Foulkes, T. Friebel, E. Friedman, D. Frost, M. Gabrielson, M. D. Gammon, P. A. Ganz, S. M. Gapstur, J. Garber, M. M. Gaudet, S. A. Gayther, A. M. Gerdes, M. Ghoussaini, G. G. Giles, G. Glendon, A. K. Godwin, M. S. Goldberg, D. E. Goldgar, A. Gonzalez-Neira, M. H. Greene, J. Gronwald, P. Guenel, M. Gunter, L. Haeberle, C. A. Haiman, U. Hamann, T. V. Hansen, S. Hart, S. Healey, T. Heikkinen, B. E. Henderson, J. Herzog, F. B. Hogervorst, A. Hollestelle, M. J. Hooning, R. N. Hoover, J. L. Hopper, K. Humphreys, D. J. Hunter, T. Huzarski, E. N. Imyanitov, C. Isaacs, A. Jakubowska, P. James, R. Janavicius, U. B. Jensen, E. M. John, M. Jones, M. Kabisch, S. Kar, B. Y. Karlan, S. Khan, K. T. Khaw, M. G. Kibriya, J. A. Knight, Y. D. Ko, I. Konstantopoulou, V. M. Kosma, V. Kristensen, A. Kwong, Y. Laitman, D. Lambrechts, C. Lazaro, E. Lee, L. Le Marchand, J. Lester, A. Lindblom, N. Lindor, S. Lindstrom, J. Liu, J. Long, J. Lubinski, P. L. Mai, E. Makalic, K. E. Malone, A. Mannermaa, S. Manoukian, S. Margolin, F. Marme, J. W. Martens, L. McGuffog, A. Meindl, A. Miller, R. L. Milne, P. Miron, M. Montagna, S. Mazoyer, A. M. Mulligan, T. A. Muranen, K. L. Nathanson, S. L. Neuhausen, H. Nevanlinna, B. G. Nordestgaard, R. L. Nussbaum, K. Offit, E. Olah, O. I. Olopade, J. E. Olson, A. Osorio, S. K. Park, P. H. Peeters, B. Peissel, P. Peterlongo, J. Peto, C. M. Phelan, R. Pilarski, B. Poppe, K. Pylkas, P. Radice, N. Rahman, J. Rantala, C. Rappaport, G. Rennert, A. Richardson, M. Robson, I. Romieu, A. Rudolph, E. J. Rutgers, M. J. Sanchez, R. M. Santella, E. J. Sawyer, D. F. Schmidt, M. K. Schmidt, R. K. Schmutzler, F. Schumacher, R. Scott, L. Senter, P. Sharma, J. Simard, C. F. Singer, O. M. Sinilnikova, P. Soucy, M. Southey, D. Steinemann, M. Stenmark-Askmalm, D. Stoppa-Lyonnet, A. Swerdlow, C. I. Szabo, R. Tamimi, W. Tapper, M. R. Teixeira, S. H. Teo, M. B. Terry, M. Thomassen, D. Thompson, L. Tihomirova, A. E. Toland, R. A. Tollenaar, I. Tomlinson, T. Truong, H. Tsimiklis, A. Teule, R. Tumino, N. Tung, C. Turnbull, G. Ursin, C. H. van Deurzen, E. J. van Rensburg, R. Varon-Mateeva, Z. Wang, S. Wang-Gohrke, E. Weiderpass, J. N. Weitzel, A. Whittemore, H. Wildiers, R. Winqvist, X. R. Yang, D. Yannoukakos, S. Yao, M. P. Zamora, W. Zheng, P. Hall, P. Kraft, C. Vachon, S. Slager, G. Chenevix-Trench, P. D. Pharoah, A. A. Monteiro, M. Garcia-Closas, D. F. Easton and A. C. Antoniou (2016). "Identification of four novel susceptibility loci for oestrogen receptor negative breast cancer." Nat Commun **7**: 11375.

Covington, J. D., R. C. Noland, R. C. Hebert, B. S. Masinter, S. R. Smith, A. C. Rustan, E. Ravussin and S. Bajpeyi (2015). "Perilipin 3 Differentially Regulates Skeletal Muscle Lipid Oxidation in Active, Sedentary, and Type 2 Diabetic Males." J Clin Endocrinol Metab **100**(10): 3683-3692.

Cross, D. A., S. E. Ashton, S. Ghiorghiu, C. Eberlein, C. A. Nebhan, P. J. Spitzler, J. P. Orme, M. R. Finlay, R. A. Ward, M. J. Mellor, G. Hughes, A. Rahi, V. N. Jacobs, M. Red Brewer, E. Ichihara, J. Sun, H. Jin, P. Ballard, K. Al-Kadhimi, R. Rowlinson, T. Klinowska, G. H. Richmond, M. Cantarini, D. W. Kim, M. R. Ranson and W. Pao (2014). "AZD9291, an

irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer." Cancer Discov **4**(9): 1046-1061.

Cui, M., A. Augert, M. Rongione, K. Conkrite, S. Parazzoli, A. Y. Nikitin, N. Ingolia and D. MacPherson (2014). "PTEN is a potent suppressor of small cell lung cancer." Mol Cancer Res **12**(5): 654-659.

Dearden, S., J. Stevens, Y. L. Wu and D. Blowers (2013). "Mutation incidence and coincidence in non small-cell lung cancer: meta-analyses by ethnicity and histology (mutMap)." Ann Oncol **24**(9): 2371-2376.

Dela Cruz, C. S., L. T. Tanoue and R. A. Matthay (2011). "Lung cancer: epidemiology, etiology, and prevention." Clin Chest Med **32**(4): 605-644.

Delaneau, O., C. Coulonges and J. F. Zagury (2008). "Shape-IT: new rapid and accurate algorithm for haplotype inference." BMC Bioinformatics **9**: 540.

Delaneau, O., J. Marchini and J. F. Zagury (2011). "A linear complexity phasing method for thousands of genomes." Nat Methods **9**(2): 179-181.

Delaneau, O., J. F. Zagury and J. Marchini (2013). "Improved whole-chromosome phasing for disease and population genetic studies." Nat Methods **10**(1): 5-6.

Demoor, T., K. R. Bracke, L. L. Dupont, M. Plantinga, B. Bondue, M. O. Roy, V. Lannoy, B. N. Lambrecht, G. G. Brusselle and G. F. Joos (2011). "The role of ChemR23 in the induction and resolution of cigarette smoke-induced inflammation." J Immunol **186**(9): 5457-5467.

Devesa, S. S., F. Bray, A. P. Vizcaino and D. M. Parkin (2005). "International lung cancer trends by histologic type: male:female differences diminishing and adenocarcinoma rates rising." Int J Cancer **117**(2): 294-299.

Doll, R. and A. B. Hill (1950). "Smoking and carcinoma of the lung; preliminary report." Br Med J **2**(4682): 739-748.

Dong, J., G. Jin, C. Wu, H. Guo, B. Zhou, J. Lv, D. Lu, Y. Shi, Y. Shu, L. Xu, M. Chu, C. Wang, R. Zhang, J. Dai, Y. Jiang, D. Yu, H. Ma, X. Zhao, Z. Yin, L. Yang, Z. Li, Q. Deng, S. Cao, Z. Qin, J. Gong, C. Sun, J. Wang, W. Wu, G. Zhou, H. Chen, P. Guan, Y. Chen, X. Liu, L. Liu, P. Xu, B. Han, C. Bai, Y. Zhao, H. Zhang, Y. Yan, J. Liu, C. I. Amos, F. Chen, W. Tan, L. Jin, T. Wu, Z. Hu, D. Lin and H. Shen (2013). "Genome-wide association study identifies a novel

susceptibility locus at 12q23.1 for lung squamous cell carcinoma in han chinese." PLoS Genet **9**(1): e1003190.

Dumitrescu, L., N. A. Restrepo, R. Goodloe, J. Boston, E. Farber-Eger, S. A. Pendergrass, W. S. Bush and D. C. Crawford (2015). "Towards a phenome-wide catalog of human clinical traits impacted by genetic ancestry." BioData Min **8**: 35.

Efird, J. T., H. Landrine, K. Y. Shiue, W. T. O'Neal, T. Podder, J. G. Rosenman and T. Biswas (2014). "Race, insurance type, and stage of presentation among lung cancer patients." Springerplus **3**: 710.

Elliott, L., D. Loomis, J. Dement, M. J. Hein, D. Richardson and L. Stayner (2012). "Lung cancer mortality in North Carolina and South Carolina chrysotile asbestos textile workers." Occup Environ Med **69**(6): 385-390.

Ernst, M. C., I. D. Haidl, L. A. Zuniga, H. J. Dranse, J. L. Rourke, B. A. Zabel, E. C. Butcher and C. J. Sinal (2012). "Disruption of the chemokine-like receptor-1 (CMKLR1) gene is associated with reduced adiposity and glucose intolerance." Endocrinology **153**(2): 672-682.

Etzel, C. J., C. I. Amos and M. R. Spitz (2003). "Risk for smoking-related cancer among relatives of lung cancer patients." Cancer Res **63**(23): 8531-8535.

Evans, D. R. and H. I. Guy (2004). "Mammalian pyrimidine biosynthesis: fresh insights into an ancient pathway." J Biol Chem **279**(32): 33035-33038.

Fang, J., T. Uchiumi, M. Yagi, S. Matsumoto, R. Amamoto, T. Saito, S. Takazaki, T. Kanki, H. Yamaza, K. Nonaka and D. Kang (2012). "Protein instability and functional defects caused by mutations of dihydro-orotate dehydrogenase in Miller syndrome patients." Biosci Rep **32**(6): 631-639.

Fang, J., T. Uchiumi, M. Yagi, S. Matsumoto, R. Amamoto, S. Takazaki, H. Yamaza, K. Nonaka and D. Kang (2013). "Dihydro-orotate dehydrogenase is physically associated with the respiratory complex and its loss leads to mitochondrial dysfunction." Biosci Rep **33**(2): e00021.

Fehringer, G., P. Kraft, P. D. Pharoah, R. A. Eeles, N. Chatterjee, F. R. Schumacher, J. M. Schildkraut, S. Lindstrom, P. Brennan, H. Bickeboller, R. S. Houlston, M. T. Landi, N. Caporaso, A. Risch, A. Amin Al Olama, S. I. Berndt, E. L. Giovannucci, H. Gronberg, Z. Kote-Jarai, J. Ma, K. Muir, M. J. Stampfer, V. L. Stevens, F. Wiklund, W. C. Willett, E. L. Goode, J. B. Permuth, H. A. Risch, B. M. Reid, S. Bezieau, H. Brenner, A. T. Chan, J. Chang-Claude, T. J. Hudson, J. K. Kocarnik, P. A. Newcomb, R. E. Schoen, M. L. Slattery, E. White, M. A. Adank,

H. Ahsan, K. Aittomaki, L. Baglietto, C. Blomquist, F. Canzian, K. Czene, I. Dos-Santos-Silva, A. H. Eliassen, J. D. Figueroa, D. Flesch-Janys, O. Fletcher, M. Garcia-Closas, M. M. Gaudet, N. Johnson, P. Hall, A. Hazra, R. Hein, A. Hofman, J. L. Hopper, A. Irwanto, M. Johansson, R. Kaaks, M. G. Kibriya, P. Lichtner, J. Liu, E. Lund, E. Makalic, A. Meindl, B. Muller-Myhsok, T. A. Muranen, H. Nevanlinna, P. H. Peeters, J. Peto, R. L. Prentice, N. Rahman, M. J. Sanchez, D. F. Schmidt, R. K. Schmutzler, M. C. Southey, R. Tamimi, R. C. Travis, C. Turnbull, A. G. Uitterlinden, Z. Wang, A. S. Whittemore, X. R. Yang, W. Zheng, D. D. Buchanan, G. Casey, D. V. Conti, C. K. Edlund, S. Gallinger, R. W. Haile, M. Jenkins, L. Le Marchand, L. Li, N. M. Lindor, S. L. Schmit, S. N. Thibodeau, M. O. Woods, T. Rafnar, J. Gudmundsson, S. N. Stacey, K. Stefansson, P. Sulem, Y. A. Chen, J. P. Tyrer, D. C. Christiani, Y. Wei, H. Shen, Z. Hu, X. O. Shu, K. Shiraishi, A. Takahashi, Y. Bosse, M. Obeidat, D. Nickle, W. Timens, M. L. Freedman, Q. Li, D. Seminara, S. J. Chanock, J. Gong, U. Peters, S. B. Gruber, C. I. Amos, T. A. Sellers, D. F. Easton, D. J. Hunter, C. A. Haiman, B. E. Henderson, R. J. Hung, C. Ovarian Cancer Association, P. Consortium, B. Hereditary, N. Ovarian Cancer Research Group, S. Colorectal Transdisciplinary, C. African American Breast Cancer and C. African Ancestry Prostate Cancer (2016). "Cross-Cancer Genome-Wide Analysis of Lung, Ovary, Breast, Prostate, and Colorectal Cancer Reveals Novel Pleiotropic Associations." Cancer Res **76**(17): 5103-5114.

Fejerman, L., D. Hu, S. Huntsman, E. M. John, M. C. Stern, C. A. Haiman, E. J. Perez-Stable and E. Ziv (2013). "Genetic ancestry and risk of mortality among U.S. Latinas with breast cancer." Cancer Res **73**(24): 7243-7253.

Fejerman, L., E. M. John, S. Huntsman, K. Beckman, S. Choudhry, E. Perez-Stable, E. G. Burchard and E. Ziv (2008). "Genetic ancestry and risk of breast cancer among U.S. Latinas." Cancer Res **68**(23): 9723-9728.

Feng, J., W. D. Funk, S. S. Wang, S. L. Weinrich, A. A. Avilion, C. P. Chiu, R. R. Adams, E. Chang, R. C. Allsopp, J. Yu and et al. (1995). "The RNA component of human telomerase." Science **269**(5228): 1236-1241.

Ferlay, J., H. R. Shin, F. Bray, D. Forman, C. Mathers and D. M. Parkin (2010). "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008." Int J Cancer **127**(12): 2893-2917.

Field, R. W. and B. L. Withers (2012). "Occupational and environmental causes of lung cancer." Clin Chest Med **33**(4): 681-703.

Fingerlin, T. E., E. Murphy, W. Zhang, A. L. Peljto, K. K. Brown, M. P. Steele, J. E. Loyd, G. P. Cosgrove, D. Lynch, S. Groshong, H. R. Collard, P. J. Wolters, W. Z. Bradford, K. Kossen, S. D. Seiwert, R. M. du Bois, C. K. Garcia, M. S. Devine, G. Gudmundsson, H. J. Isaksson, N. Kaminski, Y. Zhang, K. F. Gibson, L. H. Lancaster, J. D. Cogan, W. R. Mason, T. M. Maher, P. L. Molyneaux, A. U. Wells, M. F. Moffatt, M. Selman, A. Pardo, D. S. Kim, J. D. Crapo, B. J. Make, E. A. Regan, D. S. Walek, J. J. Daniel, Y. Kamatani, D. Zelenika, K. Smith, D. McKean,

B. S. Pedersen, J. Talbert, R. N. Kidd, C. R. Markin, K. B. Beckman, M. Lathrop, M. I. Schwarz and D. A. Schwartz (2013). "Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis." Nat Genet **45**(6): 613-620.

Ganti, A. K., S. P. Subbiah, A. Kessinger, W. I. Gonsalves, P. T. Silberstein and F. R. Loberiza, Jr. (2014). "Association between race and survival of patients with non--small-cell lung cancer in the United States veterans affairs population." Clin Lung Cancer **15**(2): 152-158.

Gao, X., L. C. Becker, D. M. Becker, J. D. Starmer and M. A. Province (2010). "Avoiding the high Bonferroni penalty in genome-wide association studies." Genet Epidemiol **34**(1): 100-105.

Gao, X., J. Starmer and E. R. Martin (2008). "A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms." Genet Epidemiol **32**(4): 361-369.

Garshick, E., F. Laden, J. E. Hart, M. E. Davis, E. A. Eisen and T. J. Smith (2012). "Lung cancer and elemental carbon exposure in trucking industry workers." Environ Health Perspect **120**(9): 1301-1306.

Genomes Project, C., A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean and G. R. Abecasis (2015). "A global reference for human genetic variation." Nature **526**(7571): 68-74.

Goetz, L. H., L. Uribe-Bruce, D. Quarless, O. Libiger and N. J. Schork (2014). "Admixture and clinical phenotypic variation." Hum Hered **77**(1-4): 73-86.

Gregorc, V., M. Hidalgo, A. Spreafico, G. Cusatis, V. Ludovini, R. G. Ingersoll, S. Marsh, S. M. Steinberg, M. G. Vigano, D. Ghio, E. Villa, A. Sparreboom and S. D. Baker (2008). "Germline polymorphisms in EGFR and survival in patients with lung cancer receiving gefitinib." Clin Pharmacol Ther **83**(3): 477-484.

Gustavsson, P., R. Jakobsson, F. Nyberg, G. Pershagen, L. Jarup and P. Scheele (2000). "Occupational exposure and lung cancer risk: a population-based case-referent study in Sweden." Am J Epidemiol **152**(1): 32-40.

Halpern, M. T., E. M. Ward, A. L. Pavluck, N. M. Schrag, J. Bian and A. Y. Chen (2008). "Association of insurance status and ethnicity with cancer stage at diagnosis for 12 cancer sites: a retrospective analysis." Lancet Oncol **9**(3): 222-231.

Han, J. Y., Y. S. Lee, E. S. Shin, J. A. Hwang, S. Nam, S. H. Hong, H. Y. Ghang, J. Y. Kim, S. J. Yoon and J. S. Lee (2014). "A genome-wide association study of survival in small-cell lung cancer patients treated with irinotecan plus cisplatin chemotherapy." Pharmacogenomics J **14**(1): 20-27.

Hardy, D., C. C. Liu, R. Xia, J. N. Cormier, W. Chan, A. White, K. Burau and X. L. Du (2009). "Racial disparities and treatment trends in a large cohort of elderly black and white patients with nonsmall cell lung cancer." Cancer **115**(10): 2199-2211.

Harrington, L., W. Zhou, T. McPhail, R. Oulton, D. S. Yeung, V. Mar, M. B. Bass and M. O. Robinson (1997). "Human telomerase contains evolutionarily conserved catalytic and structural subunits." Genes Dev **11**(23): 3109-3115.

Harris, J. E. (1983). "Cigarette smoking among successive birth cohorts of men and women in the United States during 1900-80." J Natl Cancer Inst **71**(3): 473-479.

He, T., S. Haapa-Paananen, V. O. Kaminskyy, P. Kohonen, V. Fey, B. Zhivotovsky, O. Kallioniemi and M. Perala (2014). "Inhibition of the mitochondrial pyrimidine biosynthesis enzyme dihydroorotate dehydrogenase by doxorubicin and brequinar sensitizes cancer cells to TRAIL-induced apoptosis." Oncogene **33**(27): 3538-3549.

Heagerty, P. J., T. Lumley and M. S. Pepe (2000). "Time-dependent ROC curves for censored survival data and a diagnostic marker." Biometrics **56**(2): 337-344.

Hemminki, K., I. Lonnstedt, P. Vaittinen and P. Lichtenstein (2001). "Estimation of genetic and environmental components in colorectal and lung cancer and melanoma." Genet Epidemiol **20**(1): 107-116.

Henley, S. J., S. D. Singh, J. King, R. Wilson, M. E. O'Neil and A. B. Ryerson (2015). "Invasive cancer incidence and survival - United States, 2011." MMWR Morb Mortal Wkly Rep **64**(9): 237-242.

Hoffman, J. D., R. E. Graff, N. C. Emami, C. G. Tai, M. N. Passarelli, D. Hu, S. Huntsman, D. Hadley, L. Leong, A. Majumdar, N. Zaitlen, E. Ziv and J. S. Witte (2017). "Cis-eQTL-based trans-ethnic meta-analysis reveals novel genes associated with breast cancer risk." PLoS Genet **13**(3): e1006690.

Holford, T. R., R. Meza, K. E. Warner, C. Meernik, J. Jeon, S. H. Moolgavkar and D. T. Levy (2014). "Tobacco control and the reduction in smoking-related premature deaths in the United States, 1964-2012." JAMA **311**(2): 164-171.

Houghton, A. M. (2013). "Mechanistic links between COPD and lung cancer." Nat Rev Cancer **13**(4): 233-245.

Howie, B. N., P. Donnelly and J. Marchini (2009). "A flexible and accurate genotype imputation method for the next generation of genome-wide association studies." PLoS Genet **5**(6): e1000529.

Howlader N, N. A., Krapcho M, Miller D, Bishop K, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). "SEER Cancer Statistics Review, 1975-2014, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2014/, based on November 2016 SEER data submission, posted to the SEER web site, April 2017.".

Hsiung, C. A., Q. Lan, Y. C. Hong, C. J. Chen, H. D. Hosgood, I. S. Chang, N. Chatterjee, P. Brennan, C. Wu, W. Zheng, G. C. Chang, T. Wu, J. Y. Park, C. F. Hsiao, Y. H. Kim, H. Shen, A. Seow, M. Yeager, Y. H. Tsai, Y. T. Kim, W. H. Chow, H. Guo, W. C. Wang, S. W. Sung, Z. Hu, K. Y. Chen, J. H. Kim, Y. Chen, L. Huang, K. M. Lee, Y. L. Lo, Y. T. Gao, J. H. Kim, L. Liu, M. S. Huang, T. H. Jung, G. Jin, N. Caporaso, D. Yu, C. H. Kim, W. C. Su, X. O. Shu, P. Xu, I. S. Kim, Y. M. Chen, H. Ma, M. Shen, S. I. Cha, W. Tan, C. H. Chang, J. S. Sung, M. Zhang, T. Y. Yang, K. H. Park, J. Yuenger, C. L. Wang, J. S. Ryu, Y. Xiang, Q. Deng, A. Hutchinson, J. S. Kim, Q. Cai, M. T. Landi, C. J. Yu, J. Y. Park, M. Tucker, J. Y. Hung, C. C. Lin, R. P. Perng, P. Boffetta, C. Y. Chen, K. C. Chen, S. Y. Yang, C. Y. Hu, C. K. Chang, J. F. Fraumeni, Jr., S. Chanock, P. C. Yang, N. Rothman and D. Lin (2010). "The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia." PLoS Genet **6**(8).

Hu, L., C. Wu, X. Zhao, R. Heist, L. Su, Y. Zhao, B. Han, S. Cao, M. Chu, J. Dai, J. Dong, Y. Shu, L. Xu, Y. Chen, Y. Wang, F. Lu, Y. Jiang, D. Yu, H. Chen, W. Tan, H. Ma, J. Chen, G. Jin, T. Wu, D. Lu, D. C. Christiani, D. Lin, Z. Hu and H. Shen (2012). "Genome-wide association study of prognosis in advanced non-small cell lung cancer patients receiving platinum-based chemotherapy." Clin Cancer Res **18**(19): 5507-5514.

Hu, Z., C. Wu, Y. Shi, H. Guo, X. Zhao, Z. Yin, L. Yang, J. Dai, L. Hu, W. Tan, Z. Li, Q. Deng, J. Wang, W. Wu, G. Jin, Y. Jiang, D. Yu, G. Zhou, H. Chen, P. Guan, Y. Chen, Y. Shu, L. Xu, X. Liu, L. Liu, P. Xu, B. Han, C. Bai, Y. Zhao, H. Zhang, Y. Yan, H. Ma, J. Chen, M. Chu, F. Lu, Z. Zhang, F. Chen, X. Wang, L. Jin, J. Lu, B. Zhou, D. Lu, T. Wu, D. Lin and H. Shen (2011). "A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese." Nat Genet **43**(8): 792-796.

Hua, X., K. C. Ward, T. W. Gillespie, J. Lipscomb and M. Goodman (2014). "Non-small cell lung cancer treatment receipt and survival among African-Americans and whites in a rural area." J Community Health **39**(4): 696-705.

Huang, W., X. Cheng, J. Ji, J. Zhang and Q. Li (2016). "The Application Value of HNF-1beta Transcription Factor in the Diagnosis of Ovarian Clear Cell Carcinoma." Int J Gynecol Pathol **35**(1): 66-71.

Hung, R. J., J. D. McKay, V. Gaborieau, P. Boffetta, M. Hashibe, D. Zaridze, A. Mukeria, N. Szeszenia-Dabrowska, J. Lissowska, P. Rudnai, E. Fabianova, D. Mates, V. Bencko, L. Foretova, V. Janout, C. Chen, G. Goodman, J. K. Field, T. Liloglou, G. Xinarianos, A. Cassidy, J. McLaughlin, G. Liu, S. Narod, H. E. Krokan, F. Skorpen, M. B. Elvestad, K. Hveem, L. Vatten, J. Linseisen, F. Clavel-Chapelon, P. Vineis, H. B. Bueno-de-Mesquita, E. Lund, C. Martinez, S. Bingham, T. Rasmuson, P. Hainaut, E. Riboli, W. Ahrens, S. Benhamou, P. Lagiou, D. Trichopoulos, I. Holcatova, F. Merletti, K. Kjaerheim, A. Agudo, G. Macfarlane, R. Talamini, L. Simonato, R. Lowry, D. I. Conway, A. Znaor, C. Healy, D. Zelenika, A. Boland, M. Delepine, M. Foglio, D. Lechner, F. Matsuda, H. Blanche, I. Gut, S. Heath, M. Lathrop and P. Brennan (2008). "A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25." Nature **452**(7187): 633-637.

International HapMap Consortium (2003). "The International HapMap Project." Nature **426**(6968): 789-796.

Islami, F., E. M. Ward, E. J. Jacobs, J. Ma, A. Goding Sauer, J. Lortet-Tieulent and A. Jemal (2015). "Potentially preventable premature lung cancer deaths in the USA if overall population rates were reduced to those of educated whites in lower-risk states." Cancer Causes Control **26**(3): 409-418.

Jafri, M. A., S. A. Ansari, M. H. Alqahtani and J. W. Shay (2016). "Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies." Genome Med **8**(1): 69.

Jin, G., H. Ma, C. Wu, J. Dai, R. Zhang, Y. Shi, J. Lu, X. Miao, M. Wang, Y. Zhou, J. Chen, H. Li, S. Pan, M. Chu, F. Lu, D. Yu, Y. Jiang, J. Dong, L. Hu, Y. Chen, L. Xu, Y. Shu, S. Pan, W. Tan, B. Zhou, D. Lu, T. Wu, Z. Zhang, F. Chen, X. Wang, Z. Hu, D. Lin and H. Shen (2012). "Genetic variants at 6p21.1 and 7p15.3 are associated with risk of multiple cancers in Han Chinese." Am J Hum Genet **91**(5): 928-934.

Jones, C. C., W. S. Bush, D. C. Crawford, A. S. Wenzlaff, A. G. Schwartz, J. K. Wiencke, M. R. Wrensch, W. J. Blot, S. J. Chanock, E. L. Grogan and M. C. Aldrich (2017). "Germline genetic variants and lung cancer survival in African Americans." Cancer Epidemiol Biomarkers Prev.

Kamatani, Y., K. Matsuda, Y. Okada, M. Kubo, N. Hosono, Y. Daigo, Y. Nakamura and N. Kamatani (2010). "Genome-wide association study of hematological and biochemical traits in a Japanese population." Nat Genet **42**(3): 210-215.

Kaur, J., R. Adya, B. K. Tan, J. Chen and H. S. Randeva (2010). "Identification of chemerin receptor (ChemR23) in human endothelial cells: chemerin-induced endothelial angiogenesis." Biochem Biophys Res Commun **391**(4): 1762-1768.

Khutornenko, A. A., V. V. Roudko, B. V. Chernyak, A. B. Vartapetian, P. M. Chumakov and A. G. Evstafieva (2010). "Pyrimidine biosynthesis links mitochondrial respiration to the p53 pathway." Proc Natl Acad Sci U S A **107**(29): 12828-12833.

Kinnersley, B., M. Labussiere, A. Holroyd, A. L. Di Stefano, P. Broderick, J. Vijayakrishnan, K. Mokhtari, J. Y. Delattre, K. Gousias, J. Schramm, M. J. Schoemaker, S. J. Fleming, S. Herms, S. Heilmann, S. Schreiber, H. E. Wichmann, M. M. Nothen, A. Swerdlow, M. Lathrop, M. Simon, M. Bondy, M. Sanson and R. S. Houlston (2015). "Genome-wide association study identifies multiple susceptibility loci for glioma." Nat Commun **6**: 8559.

Kinoshita, F., T. Kondoh, K. Komori, T. Matsui, N. Harada, A. Yanai, M. Fukuda, K. Morifuji and T. Matsumoto (2011). "Miller syndrome with novel dihydroorotate dehydrogenase gene mutations." Pediatr Int **53**(4): 587-591.

Kumar, P., S. Henikoff and P. C. Ng (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nat Protoc **4**(7): 1073-1081.

Kumar, R., M. A. Seibold, M. C. Aldrich, L. K. Williams, A. P. Reiner, L. Colangelo, J. Galanter, C. Gignoux, D. Hu, S. Sen, S. Choudhry, E. L. Peterson, J. Rodriguez-Santana, W. Rodriguez-Cintron, M. A. Nalls, T. S. Leak, E. O'Meara, B. Meibohm, S. B. Kritchevsky, R. Li, T. B. Harris, D. A. Nickerson, M. Fornage, P. Enright, E. Ziv, L. J. Smith, K. Liu and E. G. Burchard (2010). "Genetic ancestry in lung-function predictions." N Engl J Med **363**(4): 321-330.

Lan, Q., C. A. Hsiung, K. Matsuo, Y. C. Hong, A. Seow, Z. Wang, H. D. Hosgood, 3rd, K. Chen, J. C. Wang, N. Chatterjee, W. Hu, M. P. Wong, W. Zheng, N. Caporaso, J. Y. Park, C. J. Chen, Y. H. Kim, Y. T. Kim, M. T. Landi, H. Shen, C. Lawrence, L. Burdett, M. Yeager, J. Yuenger, K. B. Jacobs, I. S. Chang, T. Mitsudomi, H. N. Kim, G. C. Chang, B. A. Bassig, M. Tucker, F. Wei, Z. Yin, C. Wu, S. J. An, B. Qian, V. H. Lee, D. Lu, J. Liu, H. S. Jeon, C. F. Hsiao, J. S. Sung, J. H. Kim, Y. T. Gao, Y. H. Tsai, Y. J. Jung, H. Guo, Z. Hu, A. Hutchinson, W. C. Wang, R. Klein, C. C. Chung, I. J. Oh, K. Y. Chen, S. I. Berndt, X. He, W. Wu, J. Chang, X. C. Zhang, M. S. Huang, H. Zheng, J. Wang, X. Zhao, Y. Li, J. E. Choi, W. C. Su, K. H. Park, S. W. Sung, X. O. Shu, Y. M. Chen, L. Liu, C. H. Kang, L. Hu, C. H. Chen, W. Pao, Y. C. Kim, T. Y. Yang, J. Xu, P. Guan, W. Tan, J. Su, C. L. Wang, H. Li, A. D. Sihoe, Z. Zhao, Y. Chen, Y. Y. Choi, J. Y. Hung, J. S. Kim, H. I. Yoon, Q. Cai, C. C. Lin, I. K. Park, P. Xu, J. Dong, C. Kim, Q. He, R. P. Perng, T. Kohno, S. S. Kweon, C. Y. Chen, R. Vermeulen, J. Wu, W. Y. Lim, K. C. Chen, W. H. Chow, B. T. Ji, J. K. Chan, M. Chu, Y. J. Li, J. Yokota, J. Li, H. Chen, Y. B. Xiang,

C. J. Yu, H. Kunitoh, G. Wu, L. Jin, Y. L. Lo, K. Shiraishi, Y. H. Chen, H. C. Lin, T. Wu, Y. L. Wu, P. C. Yang, B. Zhou, M. H. Shin, J. F. Fraumeni, Jr., D. Lin, S. J. Chanock and N. Rothman (2012). "Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia." Nat Genet **44**(12): 1330-1335.

Landi, M. T., N. Chatterjee, K. Yu, L. R. Goldin, A. M. Goldstein, M. Rotunno, L. Mirabello, K. Jacobs, W. Wheeler, M. Yeager, A. W. Bergen, Q. Li, D. Consonni, A. C. Pesatori, S. Wacholder, M. Thun, R. Diver, M. Oken, J. Virtamo, D. Albanes, Z. Wang, L. Burdette, K. F. Doheny, E. W. Pugh, C. Laurie, P. Brennan, R. Hung, V. Gaborieau, J. D. McKay, M. Lathrop, J. McLaughlin, Y. Wang, M. S. Tsao, M. R. Spitz, Y. Wang, H. Krokan, L. Vatten, F. Skorpen, E. Arnesen, S. Benhamou, C. Bouchard, A. Metspalu, T. Vooder, M. Nelis, K. Valk, J. K. Field, C. Chen, G. Goodman, P. Sulem, G. Thorleifsson, T. Rafnar, T. Eisen, W. Sauter, A. Rosenberger, H. Bickeboller, A. Risch, J. Chang-Claude, H. E. Wichmann, K. Stefansson, R. Houlston, C. I. Amos, J. F. Fraumeni, Jr., S. A. Savage, P. A. Bertazzi, M. A. Tucker, S. Chanock and N. E. Caporaso (2009). "A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma." Am J Hum Genet **85**(5): 679-691.

Lane, R. S., S. E. Frost, G. R. Howe and L. B. Zablotska (2010). "Mortality (1950-1999) and cancer incidence (1969-1999) in the cohort of Eldorado uranium workers." Radiat Res **174**(6): 773-785.

Lathan, C. S., C. Okechukwu, B. F. Drake and G. G. Bennett (2010). "Racial differences in the perception of lung cancer: the 2005 Health Information National Trends Survey." Cancer **116**(8): 1981-1986.

Le Gallo, M., A. J. O'Hara, M. L. Rudd, M. E. Urick, N. F. Hansen, N. J. O'Neil, J. C. Price, S. Zhang, B. M. England, A. K. Godwin, D. C. Sgroi, N. I. H. I. S. C. C. S. Program, P. Hieter, J. C. Mullikin, M. J. Merino and D. W. Bell (2012). "Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes." Nat Genet **44**(12): 1310-1315.

Lee, Y., K. A. Yoon, J. Joo, D. Lee, K. Bae, J. Y. Han and J. S. Lee (2013). "Prognostic implications of genetic variants in advanced non-small cell lung cancer: a genome-wide association study." Carcinogenesis **34**(2): 307-313.

Levy, M. A., C. M. Lovly and W. Pao (2012). "Translating genomic information into clinical medicine: lung cancer as a paradigm." Genome Res **22**(11): 2101-2108.

Li, B. and S. M. Leal (2008). "Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data." Am J Hum Genet **83**(3): 311-321.

Li, C., Z. Yin, W. Wu, X. Li, Y. Ren and B. Zhou (2013). "Genetic variations in TERT-CLPTM1L genes and risk of lung cancer in Chinese women nonsmokers." PLoS One **8**(5): e64988.

Lingner, J., T. R. Hughes, A. Shevchenko, M. Mann, V. Lundblad and T. R. Cech (1997). "Reverse transcriptase motifs in the catalytic subunit of telomerase." Science **276**(5312): 561-567.

Lips, E. H., V. Gaborieau, J. D. McKay, A. Chabrier, R. J. Hung, P. Boffetta, M. Hashibe, D. Zaridze, N. Szeszenia-Dabrowska, J. Lissowska, P. Rudnai, E. Fabianova, D. Mates, V. Bencko, L. Foretova, V. Janout, J. K. Field, T. Liloglou, G. Xinarianos, J. McLaughlin, G. Liu, F. Skorpen, M. B. Elvestad, K. Hveem, L. Vatten, E. Study, S. Benhamou, P. Lagiou, I. Holcatova, F. Merletti, K. Kjaerheim, A. Agudo, X. Castellsague, T. V. Macfarlane, L. Barzan, C. Canova, R. Lowry, D. I. Conway, A. Znaor, C. Healy, M. P. Curado, S. Koifman, J. Eluf-Neto, E. Matos, A. Menezes, L. Fernandez, A. Metspalu, S. Heath, M. Lathrop and P. Brennan (2010). "Association between a 15q25 gene variant, smoking quantity and tobacco-related cancers among 17 000 individuals." Int J Epidemiol **39**(2): 563-577.

Liu, Y., L. Cao, Z. Li, D. Zhou, W. Liu, Q. Shen, Y. Wu, D. Zhang, X. Hu, T. Wang, J. Ye, X. Weng, H. Zhang, D. Zhang, Z. Zhang, F. Liu, L. He and Y. Shi (2014). "A genome-wide association study identifies a locus on TERT for mean telomere length in Han Chinese." PLoS One **9**(1): e85043.

Lohinai, Z., T. Klikovits, J. Moldvay, G. Ostoros, E. Raso, J. Timar, K. Fabian, I. Kovalszky, I. Kenessey, C. Aigner, F. Renyi-Vamos, W. Klepetko, B. Dome and B. Hegedus (2017). "KRAS-mutation incidence and prognostic value are metastatic site-specific in lung adenocarcinoma: poor prognosis in patients with KRAS mutation and bone metastasis." Sci Rep **7**: 39721.

Loomis, D., J. Dement, D. Richardson and S. Wolf (2010). "Asbestos fibre dimensions and lung cancer mortality among workers exposed to chrysotile." Occup Environ Med **67**(9): 580-584.

Loomis, D., J. M. Dement, L. Elliott, D. Richardson, E. D. Kuempel and L. Stayner (2012). "Increased lung cancer mortality among chrysotile asbestos textile workers is more strongly associated with exposure to long thin fibres." Occup Environ Med **69**(8): 564-568.

Loomis, D., J. M. Dement, S. H. Wolf and D. B. Richardson (2009). "Lung cancer mortality and fibre exposures among North Carolina asbestos textile workers." Occup Environ Med **66**(8): 535-542.

Lorenzo Bermejo, J. and K. Hemminki (2005). "Familial lung cancer and aggregation of smoking habits: a simulation of the effect of shared environmental factors on the familial risk of cancer." Cancer Epidemiol Biomarkers Prev **14**(7): 1738-1740.

Lovly, C. M., N. T. McDonald, H. Chen, S. Ortiz-Cuaran, L. C. Heukamp, Y. Yan, A. Florin, L. Ozretic, D. Lim, L. Wang, Z. Chen, X. Chen, P. Lu, P. K. Paik, R. Shen, H. Jin, R. Buettner, S. Ansen, S. Perner, M. Brockmann, M. Bos, J. Wolf, M. Gardizi, G. M. Wright, B. Solomon, P. A. Russell, T. M. Rogers, Y. Suehara, M. Red-Brewer, R. Tieu, E. de Stanchina, Q. Wang, Z. Zhao, D. H. Johnson, L. Horn, K. K. Wong, R. K. Thomas, M. Ladanyi and W. Pao (2014). "Rationale for co-targeting IGF-1R and ALK in ALK fusion-positive lung cancer." Nat Med **20**(9): 1027-1034.

Ludovini, V., F. Bianconi, L. Pistola, R. Chiari, V. Minotti, R. Colella, D. Giuffrida, F. R. Tofanetti, A. Siggillino, A. Flacco, E. Baldelli, D. Iacono, M. G. Mameli, A. Cavaliere and L. Crino (2011). "Phosphoinositide-3-kinase catalytic alpha and KRAS mutations are important predictors of resistance to therapy with epidermal growth factor receptor tyrosine kinase inhibitors in patients with advanced non-small cell lung cancer." J Thorac Oncol **6**(4): 707-715.

Lynch, T. J., D. W. Bell, R. Sordella, S. Gurubhagavatula, R. A. Okimoto, B. W. Brannigan, P. L. Harris, S. M. Haserlat, J. G. Supko, F. G. Haluska, D. N. Louis, D. C. Christiani, J. Settleman and D. A. Haber (2004). "Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib." N Engl J Med **350**(21): 2129-2139.

Machiela, M. J. and S. J. Chanock (2015). "LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants." Bioinformatics **31**(21): 3555-3557.

Manrai, A. K., B. H. Funke, H. L. Rehm, M. S. Olesen, B. A. Maron, P. Szolovits, D. M. Margulies, J. Loscalzo and I. S. Kohane (2016). "Genetic Misdiagnoses and the Potential for Health Disparities." N Engl J Med **375**(7): 655-665.

Maples, B. K., S. Gravel, E. E. Kenny and C. D. Bustamante (2013). "RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference." Am J Hum Genet **93**(2): 278-288.

Marchetti, A., L. Felicioni, S. Malatesta, M. Grazia Sciarrotta, L. Guetti, A. Chella, P. Viola, C. Pullara, F. Mucilli and F. Buttitta (2011). "Clinical features and outcome of patients with non-small-cell lung cancer harboring BRAF mutations." J Clin Oncol **29**(26): 3574-3579.

Marchini, J., B. Howie, S. Myers, G. McVean and P. Donnelly (2007). "A new multipoint method for genome-wide association studies by imputation of genotypes." Nat Genet **39**(7): 906-913.

Margolis, M. L., J. D. Christie, G. A. Silvestri, L. Kaiser, S. Santiago and J. Hansen-Flaschen (2003). "Racial differences pertaining to a belief about lung cancer surgery: results of a multicenter survey." Ann Intern Med **139**(7): 558-563.

Martin, A. R., C. R. Gignoux, R. K. Walters, G. L. Wojcik, B. M. Neale, S. Gravel, M. J. Daly, C. D. Bustamante and E. E. Kenny (2017). "Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations." Am J Hum Genet **100**(4): 635-649.

Matakidou, A., T. Eisen and R. S. Houlston (2005). "Systematic review of the relationship between family history and lung cancer risk." Br J Cancer **93**(7): 825-833.

McKay, J. D., R. J. Hung, V. Gaborieau, P. Boffetta, A. Chabrier, G. Byrnes, D. Zaridze, A. Mukeria, N. Szeszenia-Dabrowska, J. Lissowska, P. Rudnai, E. Fabianova, D. Mates, V. Bencko, L. Foretova, V. Janout, J. McLaughlin, F. Shepherd, A. Montpetit, S. Narod, H. E. Krokan, F. Skorpen, M. B. Elvestad, L. Vatten, I. Njolstad, T. Axelsson, C. Chen, G. Goodman, M. Barnett, M. M. Loomis, J. Lubinski, J. Matyjasik, M. Lener, D. Oszutowska, J. Field, T. Liloglou, G. Xinarianos, A. Cassidy, E. Study, P. Vineis, F. Clavel-Chapelon, D. Palli, R. Tumino, V. Krogh, S. Panico, C. A. Gonzalez, J. Ramon Quiros, C. Martinez, C. Navarro, E. Ardanaz, N. Larranaga, K. T. Kham, T. Key, H. B. Bueno-de-Mesquita, P. H. Peeters, A. Trichopoulou, J. Linseisen, H. Boeing, G. Hallmans, K. Overvad, A. Tjonneland, M. Kumle, E. Riboli, D. Zelenika, A. Boland, M. Delepine, M. Foglio, D. Lechner, F. Matsuda, H. Blanche, I. Gut, S. Heath, M. Lathrop and P. Brennan (2008). "Lung cancer susceptibility locus at 5p15.33." Nat Genet **40**(12): 1404-1406.

Meza, R., C. Meernik, J. Jeon and M. L. Cote (2015). "Lung cancer incidence trends by gender, race and histology in the United States, 1973-2010." PLoS One **10**(3): e0121323.

Mocellin, S., D. Verdi, K. A. Pooley, M. T. Landi, K. M. Egan, D. M. Baird, J. Prescott, I. De Vivo and D. Nitti (2012). "Telomerase reverse transcriptase locus polymorphisms and cancer risk: a field synopsis and meta-analysis." J Natl Cancer Inst **104**(11): 840-854.

Mok, T. S., Y. L. Wu, S. Thongprasert, C. H. Yang, D. T. Chu, N. Saijo, P. Sunpaweravong, B. Han, B. Margono, Y. Ichinose, Y. Nishiwaki, Y. Ohe, J. J. Yang, B. Chewaskulyong, H. Jiang, E. L. Duffield, C. L. Watkins, A. A. Armour and M. Fukuoka (2009). "Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma." N Engl J Med **361**(10): 947-957.

Molineros, J. E., A. K. Maiti, C. Sun, L. L. Looger, S. Han, X. Kim-Howard, S. Glenn, A. Adler, J. A. Kelly, T. B. Niewold, G. S. Gilkeson, E. E. Brown, G. S. Alarcon, J. C. Edberg, M. Petri,

R. Ramsey-Goldman, J. D. Reveille, L. M. Vila, B. I. Freedman, B. P. Tsao, L. A. Criswell, C. O. Jacob, J. H. Moore, T. J. Vyse, C. L. Langefeld, J. M. Guthridge, P. M. Gaffney, K. L. Moser, R. H. Scofield, M. E. Alarcon-Riquelme, B. Network, S. M. Williams, J. T. Merrill, J. A. James, K. M. Kaufman, R. P. Kimberly, J. B. Harley and S. K. Nath (2013). "Admixture mapping in lupus identifies multiple functional variants within IFIH1 associated with apoptosis, inflammation, and autoantibody production." PLoS Genet **9**(2): e1003222.

Moolgavkar, S. H., T. R. Holford, D. T. Levy, C. Y. Kong, M. Foy, L. Clarke, J. Jeon, W. D. Hazelton, R. Meza, F. Schultz, W. McCarthy, R. Boer, O. Gorlova, G. S. Gazelle, M. Kimmel, P. M. McMahon, H. J. de Koning and E. J. Feuer (2012). "Impact of reduced tobacco smoking on lung cancer mortality in the United States during 1975-2000." J Natl Cancer Inst **104**(7): 541-548.

Moreno-Estrada, A., C. R. Gignoux, J. C. Fernandez-Lopez, F. Zakharia, M. Sikora, A. V. Contreras, V. Acuna-Alonzo, K. Sandoval, C. Eng, S. Romero-Hidalgo, P. Ortiz-Tello, V. Robles, E. E. Kenny, I. Nuno-Arana, R. Barquera-Lozano, G. Macin-Perez, J. Granados-Arriola, S. Huntsman, J. M. Galanter, M. Via, J. G. Ford, R. Chapela, W. Rodriguez-Cintron, J. R. Rodriguez-Santana, I. Romieu, J. J. Sienra-Monge, B. del Rio Navarro, S. J. London, A. Ruiz-Linares, R. Garcia-Herrera, K. Estrada, A. Hidalgo-Miranda, G. Jimenez-Sanchez, A. Carnevale, X. Soberon, S. Canizales-Quinteros, H. Rangel-Villalobos, I. Silva-Zolezzi, E. G. Burchard and C. D. Bustamante (2014). "Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits." Science **344**(6189): 1280-1285.

Murabito, J. M., C. L. Rosenberg, D. Finger, B. E. Kreger, D. Levy, G. L. Splansky, K. Antman and S. J. Hwang (2007). "A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study." BMC Med Genet **8 Suppl 1**: S6.

National Comprehensive Cancer Network (2017). "National Comprehensive Cancer Network (NCCN) clinical practice guidelines in oncology: non-small cell lung cancer." Version 6.2017. from https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf.

National Research Council Committee on Health Risks of Exposure to Radon (1999). Health Effects of Exposure to Radon: BEIR VI. Washington (DC), National Academies Press.

National Toxicology Program "2016. Report on Carcinogens, Fourteenth Edition.; Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service. http://ntp.niehs.nih.gov/go/roc14/."

Ng, S. B., K. J. Buckingham, C. Lee, A. W. Bigham, H. K. Tabor, K. M. Dent, C. D. Huff, P. T. Shannon, E. W. Jabs, D. A. Nickerson, J. Shendure and M. J. Bamshad (2010). "Exome sequencing identifies the cause of a mendelian disorder." Nat Genet **42**(1): 30-35.

Nitadori, J., M. Inoue, M. Iwasaki, T. Otani, S. Sasazuki, K. Nagai and S. Tsugane (2006). "Association between lung cancer incidence and family history of lung cancer: data from a large-scale population-based cohort study, the JPHC study." Chest **130**(4): 968-975.

Niu, N., D. J. Schaid, R. P. Abo, K. Kalari, B. L. Fridley, Q. Feng, G. Jenkins, A. Batzler, A. G. Brisbin, J. M. Cunningham, L. Li, Z. Sun, P. Yang and L. Wang (2012). "Genetic association with overall survival of taxane-treated lung cancer patients - a genome-wide association study in human lymphoblastoid cell lines followed by a clinical association study." BMC Cancer **12**: 422.

Ooi, W. L., R. C. Elston, V. W. Chen, J. E. Bailey-Wilson and H. Rothschild (1986). "Increased familial risk for lung cancer." J Natl Cancer Inst **76**(2): 217-222.

Paez, J. G., P. A. Janne, J. C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F. J. Kaye, N. Lindeman, T. J. Boggon, K. Naoki, H. Sasaki, Y. Fujii, M. J. Eck, W. R. Sellers, B. E. Johnson and M. Meyerson (2004). "EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy." Science **304**(5676): 1497-1500.

Panagiotou, O. A., R. C. Travis, D. Campa, S. I. Berndt, S. Lindstrom, P. Kraft, F. R. Schumacher, A. Siddiq, S. I. Papatheodorou, J. L. Stanford, D. Albanes, J. Virtamo, S. J. Weinstein, W. R. Diver, S. M. Gapstur, V. L. Stevens, H. Boeing, H. B. Bueno-de-Mesquita, A. B. Gurrea, R. Kaaks, K. T. Khaw, V. Krogh, K. Overvad, E. Riboli, D. Trichopoulos, E. Giovannucci, M. Stampfer, C. Haiman, B. Henderson, L. Le Marchand, J. M. Gaziano, D. J. Hunter, S. Koutros, M. Yeager, R. N. Hoover, P. C. The, S. J. Chanock, S. Wacholder, T. J. Key and K. K. Tsilidis (2014). "A Genome-wide Pleiotropy Scan for Prostate Cancer Risk." Eur Urol.

Pao, W. and J. Chmielecki (2010). "Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer." Nat Rev Cancer **10**(11): 760-774.

Pao, W., V. Miller, M. Zakowski, J. Doherty, K. Politi, I. Sarkaria, B. Singh, R. Heelan, V. Rusch, L. Fulton, E. Mardis, D. Kupfer, R. Wilson, M. Kris and H. Varmus (2004). "EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib." Proc Natl Acad Sci U S A **101**(36): 13306-13311.

Pao, W., V. A. Miller, K. A. Politi, G. J. Riely, R. Somwar, M. F. Zakowski, M. G. Kris and H. Varmus (2005). "Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain." PLoS Med **2**(3): e73.

Pao, W., T. Y. Wang, G. J. Riely, V. A. Miller, Q. Pan, M. Ladanyi, M. F. Zakowski, R. T. Heelan, M. G. Kris and H. E. Varmus (2005). "KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib." PLoS Med **2**(1): e17.

Papi, A., G. Casoni, G. Caramori, I. Guzzinati, P. Boschetto, F. Ravenna, N. Calia, S. Petruzzelli, L. Corbetta, G. Cavallesco, E. Forini, M. Saetta, A. Ciaccia and L. M. Fabbri (2004). "COPD increases the risk of squamous histological subtype in smokers who develop non-small cell lung carcinoma." Thorax **59**(8): 679-681.

Park, S. L., S. G. Carmella, M. Chen, Y. Patel, D. O. Stram, C. A. Haiman, L. Le Marchand and S. S. Hecht (2015). "Mercapturic Acids Derived from the Toxicants Acrolein and Crotonaldehyde in the Urine of Cigarette Smokers from Five Ethnic Groups with Differing Risks for Lung Cancer." PLoS One **10**(6): e0124841.

Park, S. L., M. D. Fesinmeyer, M. Timofeeva, C. P. Caberto, J. M. Kocarnik, Y. Han, S. A. Love, A. Young, L. Dumitrescu, Y. Lin, R. Goodloe, L. R. Wilkens, L. Hindorff, J. H. Fowke, C. Carty, S. Buyske, F. R. Schumacher, A. Butler, H. Dilks, E. Deelman, M. L. Cote, W. Chen, M. Pande, D. C. Christiani, J. K. Field, H. Bickebller, A. Risch, J. Heinrich, P. Brennan, Y. Wang, T. Eisen, R. S. Houlston, M. Thun, D. Albanes, N. Caporaso, U. Peters, K. E. North, G. Heiss, D. C. Crawford, W. S. Bush, C. A. Haiman, M. T. Landi, R. J. Hung, C. Kooperberg, C. I. Amos, L. Le Marchand and I. Cheng (2014). "Pleiotropic associations of risk variants identified for other cancers with lung cancer risk: the PAGE and TRICL consortia." J Natl Cancer Inst **106**(4): dju061.

Pasaniuc, B., S. Sankararaman, G. Kimmel and E. Halperin (2009). "Inference of locus-specific ancestry in closely related populations." Bioinformatics **25**(12): i213-221.

Patel, Y. M., S. L. Park, Y. Han, L. R. Wilkens, H. Bickeboller, A. Rosenberger, N. Caporaso, M. T. Landi, I. Bruske, A. Risch, Y. Wei, D. C. Christiani, P. Brennan, R. Houlston, J. McKay, J. McLaughlin, R. Hung, S. Murphy, D. O. Stram, C. Amos and L. Le Marchand (2016). "Novel Association of Genetic Markers Affecting CYP2A6 Activity and Lung Cancer Risk." Cancer Res **76**(19): 5768-5776.

Peralta, C. A., N. Risch, F. Lin, M. G. Shlipak, A. Reiner, E. Ziv, H. Tang, D. Siscovick and K. Bibbins-Domingo (2010). "The Association of African Ancestry and elevated creatinine in the Coronary Artery Risk Development in Young Adults (CARDIA) Study." Am J Nephrol **31**(3): 202-208.

Polo, S. E., A. Kaidi, L. Baskcomb, Y. Galanty and S. P. Jackson (2010). "Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4." EMBO J **29**(18): 3130-3139.

Popejoy, A. B. and S. M. Fullerton (2016). "Genomics is failing on diversity." Nature **538**(7624): 161-164.

Prevention, C. f. D. C. a. (2005). "Annual smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 1997-2001." MMWR Morb Mortal Wkly Rep **54**(25): 625-628.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick and D. Reich (2006). "Principal components analysis corrects for stratification in genome-wide association studies." Nat Genet **38**(8): 904-909.

Pritchard, J. K., M. Stephens and P. Donnelly (2000). "Inference of population structure using multilocus genotype data." Genetics **155**(2): 945-959.

Pronk, A., J. Coble and P. A. Stewart (2009). "Occupational exposure to diesel engine exhaust: a literature review." J Expo Sci Environ Epidemiol **19**(5): 443-457.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly and P. C. Sham (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." Am J Hum Genet **81**(3): 559-575.

Rainger, J., H. Bengani, L. Campbell, E. Anderson, K. Sokhi, W. Lam, A. Riess, M. Ansari, S. Smithson, M. Lees, C. Mercer, K. McKenzie, T. Lengfeld, B. Gener Querol, P. Branney, S. McKay, H. Morrison, B. Medina, M. Robertson, J. Kohlhase, C. Gordon, J. Kirk, D. Wieczorek and D. R. Fitzpatrick (2012). "Miller (Genee-Wiedemann) syndrome represents a clinically and biochemically distinct subgroup of postaxial acrofacial dysostosis associated with partial deficiency of DHODH." Hum Mol Genet **21**(18): 3969-3983.

Rama, D., G. Esendagli and D. Guc (2011). "Expression of chemokine-like receptor 1 (CMKLR1) on J744A.1 macrophages co-cultured with fibroblast and/or tumor cells: modeling the influence of microenvironment." Cell Immunol **271**(1): 134-140.

Rapley, E. A., C. Turnbull, A. A. Al Olama, E. T. Dermitzakis, R. Linger, R. A. Huddart, A. Renwick, D. Hughes, S. Hines, S. Seal, J. Morrison, J. Nsengimana, P. Deloukas, U. K. T. C. Collaboration, N. Rahman, D. T. Bishop, D. F. Easton and M. R. Stratton (2009). "A genome-wide association study of testicular germ cell tumor." Nat Genet **41**(7): 807-810.

Roscoe, R. J., J. A. Deddens, A. Salvan and T. M. Schnorr (1995). "Mortality among Navajo uranium miners." Am J Public Health **85**(4): 535-540.

Roscoe, R. J., K. Steenland, W. E. Halperin, J. J. Beaumont and R. J. Waxweiler (1989). "Lung cancer mortality among nonsmoking uranium miners exposed to radon daughters." JAMA **262**(5): 629-633.

Ruark, E., S. Seal, H. McDonald, F. Zhang, A. Elliot, K. Lau, E. Perdeaux, E. Rapley, R. Eeles, J. Peto, Z. Kote-Jarai, K. Muir, J. Nsengimana, J. Shipley, U. K. T. C. Collaboration, D. T. Bishop, M. R. Stratton, D. F. Easton, R. A. Huddart, N. Rahman and C. Turnbull (2013). "Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14." Nat Genet **45**(6): 686-689.

Rubin, D. B. (2009). Multiple Imputation for Nonresponse in Surveys. Hoboken, NJ, USA, John Wiley & Sons.

Ruiz-Narvaez, E. A., L. Sucheston-Campbell, J. T. Bensen, S. Yao, S. Haddad, C. A. Haiman, E. V. Bandera, E. M. John, L. Bernstein, J. J. Hu, R. G. Ziegler, S. L. Deming, A. F. Olshan, C. B. Ambrosone, J. R. Palmer and K. L. Lunetta (2016). "Admixture Mapping of African-American Women in the AMBER Consortium Identifies New Loci for Breast Cancer and Estrogen-Receptor Subtypes." Front Genet **7**: 170.

Saccone, N. L., R. C. Culverhouse, T. H. Schwantes-An, D. S. Cannon, X. Chen, S. Cichon, I. Giegling, S. Han, Y. Han, K. Keskitalo-Vuokko, X. Kong, M. T. Landi, J. Z. Ma, S. E. Short, S. H. Stephens, V. L. Stevens, L. Sun, Y. Wang, A. S. Wenzlaff, S. H. Aggen, N. Breslau, P. Broderick, N. Chatterjee, J. Chen, A. C. Heath, M. Heliovaara, N. R. Hoft, D. J. Hunter, M. K. Jensen, N. G. Martin, G. W. Montgomery, T. Niu, T. J. Payne, L. Peltonen, M. L. Pergadia, J. P. Rice, R. Sherva, M. R. Spitz, J. Sun, J. C. Wang, R. B. Weiss, W. Wheeler, S. H. Witt, B. Z. Yang, N. E. Caporaso, M. A. Ehringer, T. Eisen, S. M. Gapstur, J. Gelernter, R. Houlston, J. Kaprio, K. S. Kendler, P. Kraft, M. F. Leppert, M. D. Li, P. A. Madden, M. M. Nothen, S. Pillai, M. Rietschel, D. Rujescu, A. Schwartz, C. I. Amos and L. J. Bierut (2010). "Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD." PLoS Genet **6**(8).

Salari, K., S. Choudhry, H. Tang, M. Naqvi, D. Lind, P. C. Avila, N. E. Coyle, N. Ung, S. Nazario, J. Casal, A. Torres-Palacios, S. Clark, A. Phong, I. Gomez, H. Matallana, E. J. Perez-Stable, M. D. Shriver, P. Y. Kwok, D. Sheppard, W. Rodriguez-Cintron, N. J. Risch, E. G. Burchard and E. Ziv (2005). "Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics." Genet Epidemiol **29**(1): 76-86.

Samet, J. M., E. Avila-Tang, P. Boffetta, L. M. Hannan, S. Olivo-Marston, M. J. Thun and C. M. Rudin (2009). "Lung cancer in never smokers: clinical epidemiology and environmental risk factors." Clin Cancer Res **15**(18): 5626-5645.

Samet, J. M., D. M. Kutvirt, R. J. Waxweiler and C. R. Key (1984). "Uranium mining and lung cancer in Navajo men." N Engl J Med **310**(23): 1481-1484.

Sampson, J. N., W. A. Wheeler, M. Yeager, O. Panagiotou, Z. Wang, S. I. Berndt, Q. Lan, C. C. Abnet, L. T. Amundadottir, J. D. Figueroa, M. T. Landi, L. Mirabello, S. A. Savage, P. R. Taylor, I. De Vivo, K. A. McGlynn, M. P. Purdue, P. Rajaraman, H. O. Adami, A. Ahlbom, D. Albanes, M. F. Amary, S. J. An, U. Andersson, G. Andriole, Jr., I. L. Andrulis, E. Angelucci, S. M. Ansell, C. Arici, B. K. Armstrong, A. A. Arslan, M. A. Austin, D. Baris, D. A. Barkauskas, B. A. Bassig, N. Becker, Y. Benavente, S. Benhamou, C. Berg, D. Van Den Berg, L. Bernstein, K. A. Bertrand, B. M. Birmann, A. Black, H. Boeing, P. Boffetta, M. C. Boutron-Ruault, P. M. Bracci, L. Brinton, A. R. Brooks-Wilson, H. B. Bueno-de-Mesquita, L. Burdett, J. Buring, M. A. Butler, Q. Cai, G. Cancel-Tassin, F. Canzian, A. Carrato, T. Carreon, A. Carta, J. K. Chan, E. T. Chang, G. C. Chang, I. S. Chang, J. Chang, J. Chang-Claude, C. J. Chen, C. Y. Chen, C. Chen, C. H. Chen, C. Chen, H. Chen, K. Chen, K. Y. Chen, K. C. Chen, Y. Chen, Y. H. Chen, Y. S. Chen, Y. M. Chen, L. H. Chien, M. D. Chirlaque, J. E. Choi, Y. Y. Choi, W. H. Chow, C. C. Chung, J. Clavel, F. Clavel-Chapelon, P. Cocco, J. S. Colt, E. Comperat, L. Conde, J. M. Connors, D. Conti, V. K. Cortessis, M. Cotterchio, W. Cozen, S. Crouch, M. Crous-Bou, O. Cussenot, F. G. Davis, T. Ding, W. R. Diver, M. Dorronsoro, L. Dossus, E. J. Duell, M. G. Ennas, R. L. Erickson, M. Feychting, A. M. Flanagan, L. Foretova, J. F. Fraumeni, Jr., N. D. Freedman, L. E. Beane Freeman, C. Fuchs, M. Gago-Dominguez, S. Gallinger, Y. T. Gao, S. M. Gapstur, M. Garcia-Closas, R. Garcia-Closas, R. D. Gascoyne, J. Gastier-Foster, M. M. Gaudet, J. M. Gaziano, C. Giffen, G. G. Giles, E. Giovannucci, B. Glimelius, M. Goggins, N. Gokgoz, A. M. Goldstein, R. Gorlick, M. Gross, R. Grubb, 3rd, J. Gu, P. Guan, M. Gunter, H. Guo, T. M. Habermann, C. A. Haiman, D. Halai, G. Hallmans, M. Hassan, C. Hattinger, Q. He, X. He, K. Helzlsouer, B. Henderson, R. Henriksson, H. Hjalgrim, J. Hoffman-Bolton, C. Hohensee, T. R. Holford, E. A. Holly, Y. C. Hong, R. N. Hoover, P. L. Horn-Ross, G. M. Hosain, H. D. Hosgood, 3rd, C. F. Hsiao, N. Hu, W. Hu, Z. Hu, M. S. Huang, J. M. Huerta, J. Y. Hung, A. Hutchinson, P. D. Inskip, R. D. Jackson, E. J. Jacobs, M. Jenab, H. S. Jeon, B. T. Ji, G. Jin, L. Jin, C. Johansen, A. Johnson, Y. J. Jung, R. Kaaks, A. Kamineni, E. Kane, C. H. Kang, M. R. Karagas, R. S. Kelly, K. T. Khaw, C. Kim, H. N. Kim, J. H. Kim, J. S. Kim, Y. H. Kim, Y. T. Kim, Y. C. Kim, C. M. Kitahara, A. P. Klein, R. J. Klein, M. Kogevinas, T. Kohno, L. N. Kolonel, C. Kooperberg, A. Kricker, V. Krogh, H. Kunitoh, R. C. Kurtz, S. S. Kweon, A. LaCroix, C. Lawrence, F. Lecanda, V. H. Lee, D. Li, H. Li, J. Li, Y. J. Li, Y. Li, L. M. Liao, M. Liebow, T. Lightfoot, W. Y. Lim, C. C. Lin, D. Lin, S. Lindstrom, M. S. Linet, B. K. Link, C. Liu, J. Liu, L. Liu, B. Ljungberg, J. Lloreta, S. Di Lollo, D. Lu, E. Lund, N. Malats, S. Mannisto, L. Le Marchand, N. Marina, G. Masala, G. Mastrangelo, K. Matsuo, M. Maynadie, J. McKay, R. McKean-Cowdin, M. Melbye, B. S. Melin, D. S. Michaud, T. Mitsudomi, A. Monnereau, R. Montalvan, L. E. Moore, L. M. Mortensen, A. Nieters, K. E. North, A. J. Novak, A. L. Oberg, K. Offit, I. J. Oh, S. H. Olson, D. Palli, W. Pao, I. K. Park, J. Y. Park, K. H. Park, A. Patino-Garcia, S. Pavanello, P. H. Peeters, R. P. Perng, U. Peters, G. M. Petersen, P. Picci, M. C. Pike, S. Porru, J. Prescott, L. Prokunina-Olsson, B. Qian, Y. L. Qiao, M. Rais, E. Riboli, J. Riby, H. A. Risch, C. Rizzato, R. Rodabough, E. Roman, M. Roupret, A. M. Ruder, S. Sanjose, G. Scelo, A. Schned, F. Schumacher, K. Schwartz, M. Schwenn, K. Scotlandi, A. Seow, C. Serra, M. Serra, H. D. Sesso, V. W. Setiawan, G. Severi, R. K. Severson, T. D. Shanafelt, H. Shen, W. Shen, M. H. Shin, K. Shiraishi, X. O. Shu, A. Siddiq, L. Sierrasesumaga, A. D. Sihoe, C. F. Skibola, A.

Smith, M. T. Smith, M. C. Southey, J. J. Spinelli, A. Staines, M. Stampfer, M. C. Stern, V. L. Stevens, R. S. Stolzenberg-Solomon, J. Su, W. C. Su, M. Sund, J. S. Sung, S. W. Sung, W. Tan, W. Tang, A. Tardon, D. Thomas, C. A. Thompson, L. F. Tinker, R. Tirabosco, A. Tjonneland, R. C. Travis, D. Trichopoulos, F. Y. Tsai, Y. H. Tsai, M. Tucker, J. Turner, C. M. Vajdic, R. C. Vermeulen, D. J. Villano, P. Vineis, J. Virtamo, K. Visvanathan, J. Wactawski-Wende, C. Wang, C. L. Wang, J. C. Wang, J. Wang, F. Wei, E. Weiderpass, G. J. Weiner, S. Weinstein, N. Wentzensen, E. White, T. E. Witzig, B. M. Wolpin, M. P. Wong, C. Wu, G. Wu, J. Wu, T. Wu, W. Wu, X. Wu, Y. L. Wu, J. S. Wunder, Y. B. Xiang, J. Xu, P. Xu, P. C. Yang, T. Y. Yang, Y. Ye, Z. Yin, J. Yokota, H. I. Yoon, C. J. Yu, H. Yu, K. Yu, J. M. Yuan, A. Zelenetz, A. Zeleniuch-Jacquotte, X. C. Zhang, Y. Zhang, X. Zhao, Z. Zhao, H. Zheng, T. Zheng, W. Zheng, B. Zhou, M. Zhu, M. Zucca, S. M. Boca, J. R. Cerhan, G. M. Ferri, P. Hartge, C. A. Hsiung, C. Magnani, L. Miligi, L. M. Morton, K. E. Smedby, L. R. Teras, J. Vijai, S. S. Wang, P. Brennan, N. E. Caporaso, D. J. Hunter, P. Kraft, N. Rothman, D. T. Silverman, S. L. Slager, S. J. Chanock and N. Chatterjee (2015). "Analysis of Heritability and Shared Heritability Based on Genome-Wide Association Studies for Thirteen Cancer Types." J Natl Cancer Inst **107**(12): djv279.

Sasaki, T., E. C. Gan, A. Wakeham, S. Kornbluth, T. W. Mak and H. Okada (2007). "HLA-B-associated transcript 3 (Bat3)/Scythe is essential for p300-mediated acetylation of p53." Genes Dev **21**(7): 848-861.

Sato, Y., N. Yamamoto, H. Kunitoh, Y. Ohe, H. Minami, N. M. Laird, N. Katori, Y. Saito, S. Ohnami, H. Sakamoto, J. Sawada, N. Saijo, T. Yoshida and T. Tamura (2011). "Genome-wide association study on overall survival of advanced non-small cell lung cancer patients treated with carboplatin and paclitaxel." J Thorac Oncol **6**(1): 132-138.

Schubauer-Berigan, M. K., R. D. Daniels and L. E. Pinkerton (2009). "Radon exposure and mortality among white and American Indian uranium miners: an update of the Colorado Plateau cohort." Am J Epidemiol **169**(6): 718-730.

Schwartz, A. G., M. L. Cote, A. S. Wenzlaff, S. Land and C. I. Amos (2009). "Racial differences in the association between SNPs on 15q25.1, smoking behavior, and risk of non-small cell lung cancer." J Thorac Oncol **4**(10): 1195-1201.

Schwartz, A. G., A. S. Wenzlaff, C. H. Bock, J. J. Ruterbusch, W. Chen, M. L. Cote, A. S. Artis, A. L. Van Dyke, S. J. Land, C. C. Harris, S. R. Pine, M. R. Spitz, C. I. Amos, A. M. Levin and P. M. McKeigue (2011). "Admixture mapping of lung cancer in 1812 African-Americans." Carcinogenesis **32**(3): 312-317.

Schwartz, A. G., P. Yang and G. M. Swanson (1996). "Familial risk of lung cancer among nonsmokers and their relatives." Am J Epidemiol **144**(6): 554-562.

Selikoff, I. J., E. C. Hammond and J. Churg (1968). "Asbestos exposure, smoking, and neoplasia." JAMA **204**(2): 106-112.

Shay, J. W. and W. E. Wright (2011). "Role of telomeres and telomerase in cancer." Semin Cancer Biol **21**(6): 349-353.

Sherr, C. J. and F. McCormick (2002). "The RB and p53 pathways in cancer." Cancer Cell **2**(2): 103-112.

Shopland, D. R. (1995). "Tobacco use and its contribution to early cancer mortality with a special emphasis on cigarette smoking." Environ Health Perspect **103 Suppl 8**: 131-142.

Shriner, D., F. Tekola-Ayele, A. Adeyemo and C. N. Rotimi (2014). "Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry." Sci Rep **4**: 6055.

Siegel, R. L., K. D. Miller and A. Jemal (2017). "Cancer Statistics, 2017." CA Cancer J Clin **67**(1): 7-30.

Signorello, L. B., M. K. Hargreaves and W. J. Blot (2010). "The Southern Community Cohort Study: investigating health disparities." J Health Care Poor Underserved **21**(1 Suppl): 26-37.

Signorello, L. B., M. K. Hargreaves, M. D. Steinwandel, W. Zheng, Q. Cai, D. G. Schlundt, M. S. Buchowski, C. W. Arnold, J. K. McLaughlin and W. J. Blot (2005). "Southern community cohort study: establishing a cohort to investigate health disparities." J Natl Med Assoc **97**(7): 972-979.

Signorello, L. B., S. M. Williams, W. Zheng, J. R. Smith, J. Long, Q. Cai, M. K. Hargreaves, B. W. Hollis and W. J. Blot (2010). "Blood vitamin d levels in relation to genetic estimation of African ancestry." Cancer Epidemiol Biomarkers Prev **19**(9): 2325-2331.

Silverman, D. T. (2017). "Diesel exhaust causes lung cancer: now what?" Occup Environ Med **74**(4): 233-234.

Silverman, D. T., C. M. Samanic, J. H. Lubin, A. E. Blair, P. A. Stewart, R. Vermeulen, J. B. Coble, N. Rothman, P. L. Schleiff, W. D. Travis, R. G. Ziegler, S. Wacholder and M. D. Attfield (2012). "The Diesel Exhaust in Miners study: a nested case-control study of lung cancer and diesel exhaust." J Natl Cancer Inst **104**(11): 855-868.

Sisti, J. and P. Boffetta (2012). "What proportion of lung cancer in never-smokers can be attributed to known risk factors?" Int J Cancer **131**(2): 265-275.

Smeenk, G., W. W. Wiegant, H. Vrolijk, A. P. Solari, A. Pastink and H. van Attikum (2010). "The NuRD chromatin-remodeling complex regulates signaling and repair of DNA damage." J Cell Biol **190**(5): 741-749.

Solovieff, N., C. Cotsapas, P. H. Lee, S. M. Purcell and J. W. Smoller (2013). "Pleiotropy in complex traits: challenges and strategies." Nat Rev Genet **14**(7): 483-495.

Spitz, M. R., C. I. Amos, Q. Dong, J. Lin and X. Wu (2008). "The CHRNA5-A3 region on chromosome 15q24-25.1 is a risk factor both for nicotine dependence and for lung cancer." J Natl Cancer Inst **100**(21): 1552-1556.

Spitz, M. R., C. I. Amos, S. Land, X. Wu, Q. Dong, A. S. Wenzlaff and A. G. Schwartz (2013). "Role of selected genetic variants in lung cancer risk in African Americans." J Thorac Oncol **8**(4): 391-397.

Stewart, B. W., C. Wild, International Agency for Research on Cancer and World Health Organization (2014). World cancer report 2014. Lyon, France, Geneva, Switzerland, International Agency for Research on Cancer WHO Press.

Su, Z., D. Dias-Santagata, M. Duke, K. Hutchinson, Y. L. Lin, D. R. Borger, C. H. Chung, P. P. Massion, C. L. Vnencak-Jones, A. J. Iafrate and W. Pao (2011). "A platform for rapid detection of multiple oncogenic mutations with relevance to targeted therapy in non-small-cell lung cancer." J Mol Diagn **13**(1): 74-84.

Sudmant, P. H., T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, J. Huddleston, Y. Zhang, K. Ye, G. Jun, M. Hsi-Yang Fritz, M. K. Konkel, A. Malhotra, A. M. Stutz, X. Shi, F. Paolo Casale, J. Chen, F. Hormozdiari, G. Dayama, K. Chen, M. Malig, M. J. Chaisson, K. Walter, S. Meiers, S. Kashin, E. Garrison, A. Auton, H. Y. Lam, X. Jasmine Mu, C. Alkan, D. Antaki, T. Bae, E. Cerveira, P. Chines, Z. Chong, L. Clarke, E. Dal, L. Ding, S. Emery, X. Fan, M. Gujral, F. Kahveci, J. M. Kidd, Y. Kong, E. W. Lameijer, S. McCarthy, P. Flicek, R. A. Gibbs, G. Marth, C. E. Mason, A. Menelaou, D. M. Muzny, B. J. Nelson, A. Noor, N. F. Parrish, M. Pendleton, A. Quitadamo, B. Raeder, E. E. Schadt, M. Romanovitch, A. Schlattl, R. Sebra, A. A. Shabalin, A. Untergasser, J. A. Walker, M. Wang, F. Yu, C. Zhang, J. Zhang, X. Zheng-Bradley, W. Zhou, T. Zichner, J. Sebat, M. A. Batzer, S. A. McCarroll, C. Genomes Project, R. E. Mills, M. B. Gerstein, A. Bashir, O. Stegle, S. E. Devine, C. Lee, E. E. Eichler and J. O. Korbel (2015). "An integrated map of structural variation in 2,504 human genomes." Nature **526**(7571): 75-81.

Tan, X. L., A. M. Moyer, B. L. Fridley, D. J. Schaid, N. Niu, A. J. Batzler, G. D. Jenkins, R. P. Abo, L. Li, J. M. Cunningham, Z. Sun, P. Yang and L. Wang (2011). "Genetic variation predicting cisplatin cytotoxicity associated with overall survival in lung cancer patients receiving platinum-based chemotherapy." Clin Cancer Res **17**(17): 5801-5811.

Tang, S., Y. Pan, Y. Wang, L. Hu, S. Cao, M. Chu, J. Dai, Y. Shu, L. Xu, J. Chen, G. Jin, Z. Hu, H. Ma and H. Shen (2015). "Genome-wide association study of survival in early-stage non-small cell lung cancer." Ann Surg Oncol **22**(2): 630-635.

Thorgeirsson, T. E., F. Geller, P. Sulem, T. Rafnar, A. Wiste, K. P. Magnusson, A. Manolescu, G. Thorleifsson, H. Stefansson, A. Ingason, S. N. Stacey, J. T. Bergthorsson, S. Thorlacius, J. Gudmundsson, T. Jonsson, M. Jakobsdottir, J. Saemundsdottir, O. Olafsdottir, L. J. Gudmundsson, G. Bjornsdottir, K. Kristjansson, H. Skuladottir, H. J. Isaksson, T. Gudbjartsson, G. T. Jones, T. Mueller, A. Gottsater, A. Flex, K. K. Aben, F. de Vegt, P. F. Mulders, D. Isla, M. J. Vidal, L. Asin, B. Saez, L. Murillo, T. Blondal, H. Kolbeinsson, J. G. Stefansson, I. Hansdottir, V. Runarsdottir, R. Pola, B. Lindblad, A. M. van Rij, B. Dieplinger, M. Haltmayer, J. I. Mayordomo, L. A. Kiemeney, S. E. Matthiasson, H. Oskarsson, T. Tyrfingsson, D. F. Gudbjartsson, J. R. Gulcher, S. Jonsson, U. Thorsteinsdottir, A. Kong and K. Stefansson (2008). "A variant associated with nicotine dependence, lung cancer and peripheral arterial disease." Nature **452**(7187): 638-642.

Thun, M. J., S. J. Henley, D. Burns, A. Jemal, T. G. Shanks and E. E. Calle (2006). "Lung cancer death rates in lifelong nonsmokers." J Natl Cancer Inst **98**(10): 691-699.

Timofeeva, M. N., R. J. Hung, T. Rafnar, D. C. Christiani, J. K. Field, H. Bickeboller, A. Risch, J. D. McKay, Y. Wang, J. Dai, V. Gaborieau, J. McLaughlin, D. Brenner, S. A. Narod, N. E. Caporaso, D. Albanes, M. Thun, T. Eisen, H. E. Wichmann, A. Rosenberger, Y. Han, W. Chen, D. Zhu, M. Spitz, X. Wu, M. Pande, Y. Zhao, D. Zaridze, N. Szeszenia-Dabrowska, J. Lissowska, P. Rudnai, E. Fabianova, D. Mates, V. Bencko, L. Foretova, V. Janout, H. E. Krokan, M. E. Gabrielsen, F. Skorpen, L. Vatten, I. Njolstad, C. Chen, G. Goodman, M. Lathrop, S. Benhamou, T. Vooder, K. Valk, M. Nelis, A. Metspalu, O. Raji, Y. Chen, J. Gosney, T. Liloglou, T. Muley, H. Dienemann, G. Thorleifsson, H. Shen, K. Stefansson, P. Brennan, C. I. Amos, R. Houlston, M. T. Landi and T. Transdisciplinary Research in Cancer of the Lung Research (2012). "Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls." Hum Mol Genet **21**(22): 4980-4995.

Tishkoff, S. A., F. A. Reed, F. R. Friedlaender, C. Ehret, A. Ranciaro, A. Froment, J. B. Hirbo, A. A. Awomoyi, J. M. Bodo, O. Doumbo, M. Ibrahim, A. T. Juma, M. J. Kotze, G. Lema, J. H. Moore, H. Mortensen, T. B. Nyambo, S. A. Omar, K. Powell, G. S. Pretorius, M. W. Smith, M. A. Thera, C. Wambebe, J. L. Weber and S. M. Williams (2009). "The genetic structure and history of Africans and African Americans." Science **324**(5930): 1035-1044.

Tokuhata, G. K. and A. M. Lilienfeld (1963). "Familial aggregation of lung cancer among hospital patients." Public Health Rep **78**: 277-283.

Tompkins, J. D., X. Wu, Y. L. Chu and C. Her (2009). "Evidence for a direct involvement of hMSH5 in promoting ionizing radiation induced apoptosis." Exp Cell Res **315**(14): 2420-2432.

Tong, J. K., C. A. Hassig, G. R. Schnitzler, R. E. Kingston and S. L. Schreiber (1998). "Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex." Nature **395**(6705): 917-921.

Truong, T., R. J. Hung, C. I. Amos, X. Wu, H. Bickeboller, A. Rosenberger, W. Sauter, T. Illig, H. E. Wichmann, A. Risch, H. Dienemann, R. Kaaks, P. Yang, R. Jiang, J. K. Wiencke, M. Wrensch, H. Hansen, K. T. Kelsey, K. Matsuo, K. Tajima, A. G. Schwartz, A. Wenzlaff, A. Seow, C. Ying, A. Staratschek-Jox, P. Nurnberg, E. Stoelben, J. Wolf, P. Lazarus, J. E. Muscat, C. J. Gallagher, S. Zienolddiny, A. Haugen, H. F. van der Heijden, L. A. Kiemeney, D. Isla, J. I. Mayordomo, T. Rafnar, K. Stefansson, Z. F. Zhang, S. C. Chang, J. H. Kim, Y. C. Hong, E. J. Duell, A. S. Andrew, F. Lejbkowicz, G. Rennert, H. Muller, H. Brenner, L. Le Marchand, S. Benhamou, C. Bouchardy, M. D. Teare, X. Xue, J. McLaughlin, G. Liu, J. D. McKay, P. Brennan and M. R. Spitz (2010). "Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium." J Natl Cancer Inst **102**(13): 959-971.

Turnbull, C., E. R. Perdeaux, D. Pernet, A. Naranjo, A. Renwick, S. Seal, R. M. Munoz-Xicola, S. Hanks, I. Slade, A. Zachariou, M. Warren-Perry, E. Ruark, M. Gerrard, J. Hale, M. Hewitt, J. Kohler, S. Lane, G. Levitt, M. Madi, B. Morland, V. Neefjes, J. Nicholson, S. Picton, B. Pizer, M. Ronghe, M. Stevens, H. Traunecker, C. A. Stiller, K. Pritchard-Jones, J. Dome, P. Grundy and N. Rahman (2012). "A genome-wide association study identifies susceptibility loci for Wilms tumor." Nat Genet **44**(6): 681-684.

VanderWeele, T. J., K. Asomaning, E. J. Tchetgen Tchetgen, Y. Han, M. R. Spitz, S. Shete, X. Wu, V. Gaborieau, Y. Wang, J. McLaughlin, R. J. Hung, P. Brennan, C. I. Amos, D. C. Christiani and X. Lin (2012). "Genetic variants on 15q25.1, smoking, and lung cancer: an assessment of mediation and interaction." Am J Epidemiol **175**(10): 1013-1020.

Velez Edwards, D. R., K. S. Tsosie, S. M. Williams, T. L. Edwards and S. B. Russell (2014). "Admixture mapping identifies a locus at 15q21.2-22.3 associated with keloid formation in African Americans." Hum Genet **133**(12): 1513-1523.

Walsh, K. M., C. I. Amos, A. S. Wenzlaff, I. P. Gorlov, J. D. Sison, X. Wu, M. R. Spitz, H. M. Hansen, E. Y. Lu, C. Wei, H. Zhang, W. Chen, S. M. Lloyd, M. L. Frazier, P. M. Bracci, M. F. Seldin, M. R. Wrensch, A. G. Schwartz and J. K. Wiencke (2012). "Association study of

nicotinic acetylcholine receptor genes identifies a novel lung cancer susceptibility locus near CHRNA1 in African-Americans." Oncotarget **3**(11): 1428-1438.

Walsh, K. M., I. P. Gorlov, H. M. Hansen, X. Wu, M. R. Spitz, H. Zhang, E. Y. Lu, A. S. Wenzlaff, J. D. Sison, C. Wei, S. M. Lloyd, W. Chen, M. L. Frazier, M. F. Seldin, L. J. Bierut, P. M. Bracci, M. R. Wrensch, A. G. Schwartz, J. K. Wiencke and C. I. Amos (2013). "Fine-mapping of the 5p15.33, 6p22.1-p21.31, and 15q25.1 regions identifies functional and histology-specific lung cancer susceptibility loci in African-Americans." Cancer Epidemiol Biomarkers Prev **22**(2): 251-260.

Wang, J., M. R. Spitz, C. I. Amos, A. V. Wilkinson, X. Wu and S. Shete (2010). "Mediating effects of smoking and chronic obstructive pulmonary disease on the relation between the CHRNA5-A3 genetic locus and lung cancer risk." Cancer **116**(14): 3458-3462.

Wang, Z., W. J. Seow, K. Shiraishi, C. A. Hsiung, K. Matsuo, J. Liu, K. Chen, T. Yamji, Y. Yang, I. S. Chang, C. Wu, Y. C. Hong, L. Burdett, K. Wyatt, C. C. Chung, S. A. Li, M. Yeager, A. Hutchinson, W. Hu, N. Caporaso, M. T. Landi, N. Chatterjee, M. Song, J. F. Fraumeni, Jr., T. Kohno, J. Yokota, H. Kunitoh, K. Ashikawa, Y. Momozawa, Y. Daigo, T. Mitsudomi, Y. Yatabe, T. Hida, Z. Hu, J. Dai, H. Ma, G. Jin, B. Song, Z. Wang, S. Cheng, Z. Yin, X. Li, Y. Ren, P. Guan, J. Chang, W. Tan, C. J. Chen, G. C. Chang, Y. H. Tsai, W. C. Su, K. Y. Chen, M. S. Huang, Y. M. Chen, H. Zheng, H. Li, P. Cui, H. Guo, P. Xu, L. Liu, M. Iwasaki, T. Shimazu, S. Tsugane, J. Zhu, G. Jiang, K. Fei, J. Y. Park, Y. H. Kim, J. S. Sung, K. H. Park, Y. T. Kim, Y. J. Jung, C. H. Kang, I. K. Park, H. N. Kim, H. S. Jeon, J. E. Choi, Y. Y. Choi, J. H. Kim, I. J. Oh, Y. C. Kim, S. W. Sung, J. S. Kim, H. I. Yoon, S. S. Kweon, M. H. Shin, A. Seow, Y. Chen, W. Y. Lim, J. Liu, M. P. Wong, V. H. Lee, B. A. Bassig, M. Tucker, S. I. Berndt, W. H. Chow, B. T. Ji, J. Wang, J. Xu, A. D. Sihoe, J. C. Ho, J. K. Chan, J. C. Wang, D. Lu, X. Zhao, Z. Zhao, J. Wu, H. Chen, L. Jin, F. Wei, G. Wu, S. J. An, X. C. Zhang, J. Su, Y. L. Wu, Y. T. Gao, Y. B. Xiang, X. He, J. Li, W. Zheng, X. O. Shu, Q. Cai, R. Klein, W. Pao, C. Lawrence, H. D. Hosgood, 3rd, C. F. Hsiao, L. H. Chien, Y. H. Chen, C. H. Chen, W. C. Wang, C. Y. Chen, C. L. Wang, C. J. Yu, H. L. Chen, Y. C. Su, F. Y. Tsai, Y. S. Chen, Y. J. Li, T. Y. Yang, C. C. Lin, P. C. Yang, T. Wu, D. Lin, B. Zhou, J. Yu, H. Shen, M. Kubo, S. J. Chanock, N. Rothman and Q. Lan (2016). "Meta-analysis of genome-wide association studies identifies multiple lung cancer susceptibility loci in never-smoking Asian women." Hum Mol Genet **25**(3): 620-629.

Warner, K. E., D. W. Sexton, B. W. Gillespie, D. T. Levy and F. J. Chaloupka (2014). "Impact of tobacco control on adult per capita cigarette consumption in the United States." Am J Public Health **104**(1): 83-89.

Wasswa-Kintu, S., W. Q. Gan, S. F. Man, P. D. Pare and D. D. Sin (2005). "Relationship between reduced forced expiratory volume in one second and the risk of lung cancer: a systematic review and meta-analysis." Thorax **60**(7): 570-575.

Wellmann, K. F. (1964). "[Smoking and Health. On the Report of the Advisory Committee to the Surgeon General of the Public Health Service]." Dtsch Med Wochenschr **89**: 1085-1086.

White, R. M., J. Cech, S. Ratanasirintrawoot, C. Y. Lin, P. B. Rahl, C. J. Burke, E. Langdon, M. L. Tomlinson, J. Mosher, C. Kaufman, F. Chen, H. K. Long, M. Kramer, S. Datta, D. Neuberg, S. Granter, R. A. Young, S. Morrison, G. N. Wheeler and L. I. Zon (2011). "DHODH modulates transcriptional elongation in the neural crest and melanoma." Nature **471**(7339): 518-522.

Willer, C. J., Y. Li and G. R. Abecasis (2010). "METAL: fast and efficient meta-analysis of genomewide association scans." Bioinformatics **26**(17): 2190-2191.

William D. Travis, E. B., H. Konrad Muller-Hermelink, Curtis C. Harris (Eds.) (2004). World Health Organization Classification of Tumors. Pathology and Genetics of Tumors of the Lung, Pleura, Thymus and Heart. Lyon, IARC Pres.

Winkler, C. A., G. W. Nelson and M. W. Smith (2010). "Admixture mapping comes of age." Annu Rev Genomics Hum Genet **11**: 65-89.

Wolpin, B. M., C. Rizzato, P. Kraft, C. Kooperberg, G. M. Petersen, Z. Wang, A. A. Arslan, L. Beane-Freeman, P. M. Bracci, J. Buring, F. Canzian, E. J. Duell, S. Gallinger, G. G. Giles, G. E. Goodman, P. J. Goodman, E. J. Jacobs, A. Kamineni, A. P. Klein, L. N. Kolonel, M. H. Kulke, D. Li, N. Malats, S. H. Olson, H. A. Risch, H. D. Sesso, K. Visvanathan, E. White, W. Zheng, C. C. Abnet, D. Albanes, G. Andreotti, M. A. Austin, R. Barfield, D. Basso, S. I. Berndt, M. C. Boutron-Ruault, M. Brotzman, M. W. Buchler, H. B. Bueno-de-Mesquita, P. Bugert, L. Burdette, D. Campa, N. E. Caporaso, G. Capurso, C. Chung, M. Cotterchio, E. Costello, J. Elena, N. Funel, J. M. Gaziano, N. A. Giese, E. L. Giovannucci, M. Goggins, M. J. Gorman, M. Gross, C. A. Haiman, M. Hassan, K. J. Helzlsouer, B. E. Henderson, E. A. Holly, N. Hu, D. J. Hunter, F. Innocenti, M. Jenab, R. Kaaks, T. J. Key, K. T. Khaw, E. A. Klein, M. Kogevinas, V. Krogh, J. Kupcinskas, R. C. Kurtz, A. LaCroix, M. T. Landi, S. Landi, L. Le Marchand, A. Mambrini, S. Mannisto, R. L. Milne, Y. Nakamura, A. L. Oberg, K. Owzar, A. V. Patel, P. H. Peeters, U. Peters, R. Pezzilli, A. Piepoli, M. Porta, F. X. Real, E. Riboli, N. Rothman, A. Scarpa, X. O. Shu, D. T. Silverman, P. Soucek, M. Sund, R. Talar-Wojnarowska, P. R. Taylor, G. E. Theodoropoulos, M. Thornquist, A. Tjonneland, G. S. Tobias, D. Trichopoulos, P. Vodicka, J. Wactawski-Wende, N. Wentzensen, C. Wu, H. Yu, K. Yu, A. Zeleniuch-Jacquotte, R. Hoover, P. Hartge, C. Fuchs, S. J. Chanock, R. S. Stolzenberg-Solomon and L. T. Amundadottir (2014). "Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer." Nat Genet **46**(9): 994-1000.

Wrensch, M. R., R. Miike, J. D. Sison, K. T. Kelsey, M. Liu, A. McMillan, C. Quesenberry and J. K. Wiencke (2005). "CYP1A1 variants and smoking-related lung cancer in San Francisco Bay area Latinos and African Americans." Int J Cancer **113**(1): 141-147.

Wu, C., B. Xu, P. Yuan, X. Miao, Y. Liu, Y. Guan, D. Yu, J. Xu, T. Zhang, H. Shen, T. Wu and D. Lin (2010). "Genome-wide interrogation identifies YAP1 variants associated with survival of small-cell lung cancer patients." Cancer Res **70**(23): 9721-9729.

Wu, M. C., S. Lee, T. Cai, Y. Li, M. Boehnke and X. Lin (2011). "Rare-variant association testing for sequencing data with the sequence kernel association test." Am J Hum Genet **89**(1): 82-93.

Wu, X., L. Wang, Y. Ye, J. A. Aakre, X. Pu, G. C. Chang, P. C. Yang, J. A. Roth, R. S. Marks, S. M. Lippman, J. Y. Chang, C. Lu, C. Deschamps, W. C. Su, W. C. Wang, M. S. Huang, D. W. Chang, Y. Li, V. S. Pankratz, J. D. Minna, W. K. Hong, M. A. Hildebrandt, C. A. Hsiung and P. Yang (2013). "Genome-wide association study of genetic predictors of overall survival for non-small cell lung cancer in never smokers." Cancer Res **73**(13): 4028-4038.

Wu, X., Y. Xu, K. Feng, J. D. Tompkins and C. Her (2013). "MutS homologue hMSH5: recombinational DSB repair and non-synonymous polymorphic variants." PLoS One **8**(9): e73284.

Wu, X., Y. Ye, R. Rosell, C. I. Amos, D. J. Stewart, M. A. Hildebrandt, J. A. Roth, J. D. Minna, J. Gu, J. Lin, S. C. Buch, T. Nukui, J. L. Ramirez Serrano, M. Taron, A. Cassidy, C. Lu, J. Y. Chang, S. M. Lippman, W. K. Hong, M. R. Spitz, M. Romkes and P. Yang (2011). "Genome-wide association study of survival in non-small cell lung cancer patients receiving platinum-based chemotherapy." J Natl Cancer Inst **103**(10): 817-825.

Wynder, E. L. and E. A. Graham (1950). "Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma; a study of 684 proved cases." J Am Med Assoc **143**(4): 329-336.

Xue, Y., J. Wong, G. T. Moreno, M. K. Young, J. Cote and W. Wang (1998). "NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities." Mol Cell **2**(6): 851-861.

Xun, W. W., P. Brennan, A. Tjonneland, U. Vogel, K. Overvad, R. Kaaks, F. Canzian, H. Boeing, A. Trichopoulou, E. Oustoglou, Z. Giotaki, M. Johansson, D. Palli, C. Agnoli, R. Tumino, C. Sacerdote, S. Panico, H. B. Bueno-de-Mesquita, P. H. Peeters, E. Lund, M. Kumle, L. Rodriguez, A. Agudo, M. J. Sanchez, L. Arriola, M. D. Chirlaque, A. Barricarte, G. Hallmans, T. Rasmuson, K. T. Khaw, N. Wareham, T. Key, E. Riboli and P. Vineis (2011). "Single-nucleotide polymorphisms (5p15.33, 15q25.1, 6p22.1, 6q27 and 7p15.3) and lung cancer survival in the European Prospective Investigation into Cancer and Nutrition (EPIC)." Mutagenesis **26**(5): 657-666.

Yamada, M., N. Sato, S. Ikeda, T. Arai, M. Sawabe, S. Mori, Y. Yamada, M. Muramatsu and M. Tanaka (2015). "Association of the chromodomain helicase DNA-binding protein 4 (CHD4) missense variation p.D140E with cancer: potential interaction with smoking." Genes Chromosomes Cancer **54**(2): 122-128.

Yang, J. J., C. Cheng, M. Devidas, X. Cao, Y. Fan, D. Campana, W. Yang, G. Neale, N. J. Cox, P. Scheet, M. J. Borowitz, N. J. Winick, P. L. Martin, C. L. Willman, W. P. Bowman, B. M. Camitta, A. Carroll, G. H. Reaman, W. L. Carroll, M. Loh, S. P. Hunger, C. H. Pui, W. E. Evans and M. V. Relling (2011). "Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia." Nat Genet **43**(3): 237-241.

Yang, P., Y. Li, R. Jiang, J. M. Cunningham, Y. Li, F. Zhang and M. de Andrade (2010). "A rigorous and comprehensive validation: common genetic variations and lung cancer." Cancer Epidemiol Biomarkers Prev **19**(1): 240-244.

Yoon, K. A., J. H. Park, J. Han, S. Park, G. K. Lee, J. Y. Han, J. I. Zo, J. Kim, J. E. Lee, A. Takahashi, M. Kubo, Y. Nakamura and J. S. Lee (2010). "A genome-wide association study reveals susceptibility variants for non-small cell lung cancer in the Korean population." Hum Mol Genet **19**(24): 4948-4954.

You, M., D. Wang, P. Liu, H. Vikis, M. James, Y. Lu, Y. Wang, M. Wang, Q. Chen, D. Jia, Y. Liu, W. Wen, P. Yang, Z. Sun, S. M. Pinney, W. Zheng, X. O. Shu, J. Long, Y. T. Gao, Y. B. Xiang, W. H. Chow, N. Rothman, G. M. Petersen, M. de Andrade, Y. Wu, J. M. Cunningham, J. S. Wiest, P. R. Fain, A. G. Schwartz, L. Girard, A. Gazdar, C. Gaba, H. Rothschild, D. Mandal, T. Coons, J. Lee, E. Kupert, D. Seminara, J. Minna, J. E. Bailey-Wilson, C. I. Amos and M. W. Anderson (2009). "Fine mapping of chromosome 6q23-25 region in familial lung cancer families reveals RGS17 as a likely candidate gene." Clin Cancer Res **15**(8): 2666-2674.

Young, R. P., F. Duan, C. Chiles, R. J. Hopkins, G. D. Gamble, E. M. Greco, C. Gatsonis and D. Aberle (2015). "Airflow Limitation and Histology Shift in the National Lung Screening Trial. The NLST-ACRIN Cohort Substudy." Am J Respir Crit Care Med **192**(9): 1060-1067.

Young, R. P., R. J. Hopkins, T. Christmas, P. N. Black, P. Metcalf and G. D. Gamble (2009). "COPD prevalence is increased in lung cancer, independent of age, sex and smoking history." Eur Respir J **34**(2): 380-386.

Yu, D. D., Y. Y. Jing, S. W. Guo, F. Ye, W. Lu, Q. Li, Y. L. Dong, L. Gao, Y. T. Yang, Y. Yang, M. C. Wu and L. X. Wei (2015). "Overexpression Of Hepatocyte Nuclear Factor-1beta Predicting Poor Prognosis Is Associated With Biliary Phenotype In Patients With Hepatocellular Carcinoma." Sci Rep **5**: 13319.

Zanetti, K. A., Z. Wang, M. Aldrich, C. I. Amos, W. J. Blot, E. D. Bowman, L. Burdette, Q. Cai, N. Caporaso, C. C. Chung, E. M. Gillanders, C. A. Haiman, H. M. Hansen, B. E. Henderson, L. N. Kolonel, L. L. Marchand, S. Li, L. H. McNeill, B. M. Ryan, A. G. Schwartz, J. D. Sison, M. R. Spitz, M. Tucker, A. S. Wenzlaff, J. K. Wiencke, L. Wilkens, M. R. Wrensch, X. Wu, W. Zheng, W. Zhou, D. Christiani, J. R. Palmer, T. M. Penning, A. G. Rieber, L. Rosenberg, E. A. Ruiz-Narvaez, L. Su, A. Vachani, Y. Wei, A. S. Whitehead, S. J. Chanock and C. C. Harris (2016). "Genome-wide association study confirms lung cancer susceptibility loci on chromosomes 5p15 and 15q25 in an African-American population." Lung Cancer **98**: 33-42.

Zeng, C., W. Wen, A. K. Morgans, W. Pao, X. Shu and W. Zheng (2015). "Disparities by race, age, and sex in the improvement of survival for major cancers: Results from the national cancer institute surveillance, epidemiology, and end results (seer) program in the united states, 1990 to 2010." JAMA Oncology **1**(1): 88-96.

Zhang, M., L. Hu, H. Shen, J. Dong, Y. Shu, L. Xu, G. Jin, T. Tian, Z. Hu and H. Shen (2010). "Candidate variants at 6p21.33 and 6p22.1 and risk of non-small cell lung cancer in a Chinese population." Int J Mol Epidemiol Genet **1**(1): 11-18.

Zhao, M. M., Y. Zhang, L. Shen, Y. W. Ren, X. L. Li, Z. H. Yin and B. S. Zhou (2014). "Genetic variations in TERT-CLPTM1L genes and risk of lung cancer in a Chinese population." Asian Pac J Cancer Prev **15**(6): 2809-2813.

Zhao, S., M. Choi, J. D. Overton, S. Bellone, D. M. Roque, E. Cocco, F. Guzzo, D. P. English, J. Varughese, S. Gasparrini, I. Bortolomai, N. Buza, P. Hui, M. Abu-Khalaf, A. Ravaggi, E. Bignotti, E. Bandiera, C. Romani, P. Todeschini, R. Tassi, L. Zanotti, L. Carrara, S. Pecorelli, D. A. Silasi, E. Ratner, M. Azodi, P. E. Schwartz, T. J. Rutherford, A. L. Stiegler, S. Mane, T. J. Boggon, J. Schlessinger, R. P. Lifton and A. D. Santin (2013). "Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma." Proc Natl Acad Sci U S A **110**(8): 2916-2921.

Zheng, L., L. Enewold, S. H. Zahm, C. D. Shriver, J. Zhou, A. Marrogi, K. A. McGlynn and K. Zhu (2012). "Lung cancer survival among black and white patients in an equal access health system." Cancer Epidemiol Biomarkers Prev **21**(10): 1841-1847.

Zhu, S., X. Yan, Z. Xiang, H. F. Ding and H. Cui (2013). "Leflunomide reduces proliferation and induces apoptosis in neuroblastoma cells in vitro and in vivo." PLoS One **8**(8): e71555.