Published in final edited form as:

Lung Cancer. 2018 September ; 123: 99–106. doi:10.1016/j.lungcan.2018.07.011.

Common Cancer-Driver Mutations and their Association with Abnormally Methylated Genes in Lung Adenocarcinoma from Never-Smokers

Mathewos Tessema^{*a}, Michael R. Rossi^b, Maria A. Picchi^a, Christin M. Yingling^a, Yong Lin^a, Suresh S. Ramalingam^c, and Steven A. Belinsky^a

^aLung Cancer Program, Lovelace Respiratory Research Institute, Albuquerque, NM, USA.

^bDepartments of Pathology and Laboratory Medicine, Radiation Oncology

^cHematology and Oncology, Emory University School of Medicine, Winship Cancer Institute, Atlanta, Georgia, USA.

Abstract

Objectives: Lung adenocarcinoma in never-smokers accounts for 15–20% of all lung cancer. Although targetable mutations are more prevalent in these tumors, the biological and clinical importance of coexisting and/or mutually exclusive abnormalities is just emerging. This study evaluates the relationships between common genetic and epigenetic aberrations in these tumors.

Materials and methods: Next-generation sequencing was employed to screen 20 commonly mutated cancer-driver genes in 112 lung adenocarcinomas from never-smokers. The relationship of these mutations with cancer-related methylation of 59 genes, and geographical/ethnic differences in the prevalence for mutations compared to multiple East Asian never-smoker lung adenocarcinoma cohorts was studied.

Results: The most common driver mutation detected in 40% (45/112) of the tumors was *EGFR*, followed by *TP53* (18%), *SETD2* (11%), and *SMARCA4* (11%). Over 72% (81/112) of the cases have mutation of at least one driver gene. While 30% (34/112) of the tumors have co-mutations of two or more genes, 42% (47/112) have only one driver gene mutation. Differences in the prevalence for some of these mutations were seen between adenocarcinomas in East Asian versus US (mainly Caucasian) never-smokers including a significantly lower rate of *EGFR* mutation among the US patients. Interestingly, aberrant methylation of multiple cancer-related genes was significantly associated with *EGFR* wildtype tumors. Among 15 differentially methylated genes

Corresponding Authors: Mathewos Tessema, Lovelace Respiratory Research Institute, 2425 Ridgecrest Dr. SE, Albuquerque, NM. 87108, U.S.A. Phone: 505-348-9123, Fax: 505-348-8567, mtessema@LRRI.org; and Steven A. Belinsky sbelinsk@LRRI.org.

Conflict of interest

The authors declare no conflict of interest.

Disclosure: This study was supported largely by grant U10 CA180950 and in part by R01CA183296, R01CA193532, and P30CA11800 from the National Institutes of Health.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

by *EGFR* mutation, 14 were more commonly methylated in *EGFR* wildtype compared to mutant tumors. These findings were independently validated using publicly available data.

Conclusion: Most lung adenocarcinomas from never-smokers harbor targetable mutation/comutations. In the absence of *EGFR* mutation that drives 40% of these tumors, *EGFR* wildtype tumors appear to develop by acquiring aberrant promoter methylation that silences tumor-suppressor genes.

Keywords

Lung adenocarcinoma; non-smokers; mutation; methylation

1. Introduction

Lung cancer is the leading cause of cancer morbidity and mortality worldwide [1, 2]. Although smoking remains the major cause of lung cancer, the percentage of never-smokers diagnosed with the disease is increasing. This together with the fact that never-smokers often develop lung adenocarcinoma has contributed to a shift in lung cancer histology. Lung adenocarcinoma has now surpassed squamous cell carcinoma, which is more commonly diagnosed among smokers, to become the most common type of all lung cancer subtypes. Currently, lung adenocarcinoma from never-smokers alone accounts for 15–20% of all lung cancer cases [3–5]. This is also clinically relevant because many of the currently targetable genetic abnormalities such as epidermal growth factor receptor (*EGFR*) mutations and translocation of the anaplastic lymphoma receptor kinase (*ALK*) are more prevalent among lung adenocarcinoma cases from never-smokers [6]. Due to the striking efficacy of tyrosine kinase inhibitors (TKIs) targeting the activated oncogenes, patients with these mutations are highly responsive to therapy and show better survival [7–11]. In contrast, the frequency of non-actionable mutations such as Kirsten rat sarcoma viral oncogene (*KRAS*) is higher in lung adenocarcinomas from smokers and co-mutation of *KRAS* together with a targetable mutation is often associated with resistance to the TKIs. As a result, never-smoker lung adenocarcinoma patients often benefit the most from molecular profiling and targeted therapy compared to other subgroups of nonsmall-cell lung cancer (NSCLC). In addition, the presence of targetable mutations in lung cancer and response to TKIs are also strongly associated with female sex and East Asian ancestry [3, 4].

The importance of mutation screening to identify molecular targets when selecting treatment regimens is now well established [11–13] and, tests for *EGFR* mutations [14, 15] and the *ALK* fusion gene [16, 17] are part of the standard clinical practice in many countries [18–20]. Thus, as drugs targeting several other cancer-driver genes [21] are entering the market, clearly understanding the presence and impact of driver mutations to the sensitivity or resistance of specific anticancer agents is critically important. As the number of targetable genetic abnormalities increases and the technology to detect them continues to advance, the number of patients identified with these molecular changes also continues to rise substantially [22]. Cancer-driver mutations are generally considered to be independent of one another and in NSCLC they are often mutually exclusive [23]. However, with the advent of next-generation sequencing (NGS)-based diagnostics, more recent studies have started to show that co-existence of driver mutations is not uncommon. A recent study demonstrated

approximately 12% of non-squamous NSCLC have concomitant driver mutations [24]. Coexistence of driver mutations is also clinically relevant because it provides targets for additional or combination therapy and may help to pre-identify potential sensitivity or resistance to a particular type of targeted therapy. For example, detection of *KRAS* mutation in an *EGFR* mutant lung cancer is a well-established indicator of resistant to EGFR TKIs. Similarly, *MET* activation in *EGFR* mutant lung cancer also indicates resistance to Gefitinib [25]. In contrast, coexistence *PIK3CA* with *EGFR* or *ALK* fusion indicates responsiveness to TKIs targeting EGFR as well as ALK [24]. Concomitant *EGFR* mutation and *ALK* fusion in a tumor was recently linked to long-lasting response both to EGFR inhibitor therapy and then ALK inhibitor therapy once the tumor progressed [26]. Furthermore, the discovery of TKIs such as Entrectinib that can effectively target multiple tyrosine kinases with proven tolerability and effectiveness in solid tumors harboring *NTRK1/2/3*, *ROS1*, or *ALK* gene fusions suggests that tumors with concomitant activation of multiple tyrosine kinases likely show increased sensitivity to such multi-targeting TKIs [27].

Epigenetic modification, particularly aberrant methylation of cancer-related genes, is one of the most common abnormalities that coexists with driver mutations in lung tumors and modifies the biology and sensitivity/resistance of tumors to various therapeutic agents. A combined analysis of RNA-seq, DNA-seq, and whole genome methylation of lung adenocarcinoma by the Cancer Genome Atlas (TCGA) project revealed association of some driver mutations to methylation of specific genes or a CpG island methylator phenotype (CIMP) that covers multiple genes [28]. For instance, tumors with *CDKN2A* methylation have overall fewer mutations but were significantly enriched for *SETD2* mutation, while *MYC* overexpression was significantly associated with CIMP-high phenotype. Interestingly, *SETD2* along with *SMARCA4* and *ARID1A* are some of the epigenetic regulators that are commonly mutated in lung adenocarcinoma and show cancer-driver properties [29]. Tiedemann et al. recently confirmed that *SETD2* mutations drive tumorigenesis by coordinated disruption of the epigenome and transcriptome [30]. Other studies compared the relationship between *EGFR* mutation in lung cancer with methylation of selected tumor-suppressor genes including *CHFR*, *CDKN2A*, *SPARC*, *RASSF1A* and *RNUNX3* and showed that methylation of these genes is inversely related to *EGFR* mutation [31–33]. Taken together, these findings suggest that epigenetic silencing of tumor-suppressor genes in lung cancer appears to function in supporting or as an alternative to oncogenic signaling mediated by some cancer-driver mutations. Thus, clearly understanding the relationship between these genetic and epigenetic abnormalities along with environmental and demographic factors that may modulate the interaction will help to optimize molecular screening and support selection of the most favorable and cost-effective therapeutic options for each patient. As recently demonstrated for AML, defining the genetic and epigenetic changes of a tumor allows selective combination of epigenetic therapy to reverse aberrantly silenced growth-regulatory genes and increase the efficacy of anti-cancer agents including TKIs [34].

In this study, a large number of well-annotated lung adenocarcinoma samples from never-smokers were screened using targeted NGS for 20 genes. These genes were selected because they are more commonly mutated in lung cancer and are recently nominated as drivers of lung cancer by two high-impact publications [28, 29], The majority of published data

regarding cancer-driver mutation in lung adenocarcinoma from never-smokers is based on patients of East Asian origin. Thus, our data, which is based on one of the largest number of lung adenocarcinoma cases from never-smokers in the US, was used to compare the frequencies of specific driver mutations between the two ethnic/geographic populations. Furthermore, the DNA methylation profile of CpG islands across the promoter regions of multiple cancer-related genes has been previously determined for the majority of tumors evaluated in this study [35–39]. The relationship between the 20 cancer-driver mutations and the methylation profile of 59 genes (Table S1), including some key genes in the TCGA CIMP analysis, was evaluated.

2. Materials and Methods

2.1. Study population

Fresh frozen lung tumors from curative intent resection of 112 never-smoker lung adenocarcinoma patients between 1992 and 2007 were obtained from the tumor banks at the University of New Mexico (UNM), Johns Hopkins, and the Mayo Clinic. Never-smokers were defined as patients who have never smoked or smoked less than 100 cigarettes in their life time. Distant normal lung tissue obtained from the most distant site of the resected lobe was available for a subset of these cases. Normal human bronchial epithelial cells (NHBE) collected from cancer-free smokers (n = 10) at the UNM through diagnostic bronchoscopy [40] and peripheral blood mononuclear cells (PBMC) obtained from healthy blood donors (free of clinically evident diseases, n = 10) were used as normal control. All samples were obtained with written informed consent from patients, and the study was approved by the institutes' Ethics Committees. Demographic and clinico-pathological characteristics of patients whose tumors were used for targeted sequencing in this study and for the DNA methylation analysis in the previous studies [35–39] are described in Table 1.

2.2. DNA extraction and Methylation analysis

DNA was extracted and quantified as described [36] and the methods used for DNA modification and methylation analysis are described [35–39]. Briefly, high molecular weight DNA was isolated using a standard phenol-chloroform extraction method and modified using the EZ DNA Methylation-Gold Kit™ (ZYMO Research) as described by the manufacturer. The semi-quantitative combined bisulfite modification and restriction analysis (CoBRA) and the qualitative methylation-specific PCR (MSP) assays were used to determine the methylation status of all 59 genes. The primers and PCR conditions used for each gene are described in detail [35–39]. The CoBRA primers were designed to amplify promoter CpG island regions with multiple BstU1 enzyme recognition sites (CGCG) regardless of the methylation status. The PCR products were then digested with BstU1 enzyme and methylation status was determined semi-quantitatively based on the level of digestion into smaller fragments, which indicates that methylation of the CGCG sites prevented modification into TGTG. The CoBRA results were then used to optimize the primers for the highly sensitivity MSP assay that amplifies only the methylated or unmethylated DNA based on the use of primers that are specific to methylated or unmethylated DNA, respectively. The methylation data generated from lung tumors using this highly sensitive MSP assay was used for comparisons with driver mutations.

2.3. Genetic analysis

Targeted NGS was applied to screen a panel of 20 genes selected based on two recent publications that 1) identified frequent mutation of these genes in lung adenocarcinoma and 2) nominated the genes as drivers of lung adenocarcinoma because of the cancer-driver properties of the mutations [29, 41]. Mutation analysis was performed at the Emory Integrated Genomics Core and Biostatistics and Bioinformatics Shared resource of Winship Cancer Institute of Emory University using a combination of multiple mutation assays each as described in detail [42, 43]. These assays include the SNaPshot (Life Technologies, Grand Island, NY) multiplex PCR with single base primer extension, [44] the TruSight (Illumina Inc, San Diego, CA) tumor target amplicon-based library preparation kit, the HaloPlex Cancer NGS (Agilent Technologies, Santa Clara, CA), and the Fluidigm 48.48 Access Array custom designed for lung adenocarcinoma related mutations as described [43]. DNA samples from lung adenocarcinoma tumor-normal pairs were used for the different sample preparations (SNaPshot, TruSight tumor, Access Array, and HaloPlex probe capture libraries) and NGS was completed using the Illumina MiSeq as described [42, 43]. Sequence alignment, index and primer trimming, and variant calls were performed using onboard MiSeq software. FASTQ files were analyzed and variants were detected using the various manufacturers' recommended software as described in detail [43].

2.4. Statistical analysis

Gene mutation, methylation, and patient characteristics including age, race, sex, stage of lung cancer, survival, and status (alive or dead) were summarized with mean and standard deviation for continuous variables and proportions for categorical variables. The association between mutation and methylation as well as mutation and patient characteristics was assessed using two-sided Fisher's exact test. Proportional hazards models were used to examine each gene with adjustment for patient characteristics. Missing data were censored for all comparative analysis. Survival time was calculated from time of diagnosis until death or last follow-up. The effects of variables including the presence of 0, 1, or 2 mutant genes on survival were assessed with Kaplan-Meier estimator and the log-rank test. All analyses were conducted in SAS 9.4.

3. Results

3.1. Cancer-driver mutations in lung adenocarcinoma from never-smokers

Targeted NGS was applied to screen a panel of 20 genes using DNA from 112 fresh frozen lung adenocarcinoma samples obtained from never-smokers. Although only the nonsynonymous mutations were included in the analysis for this study, all mutations detected across the 11 oncogenes and 9 tumor suppressor genes (TSGs) are shown in Table S2. *EGFR* mutation was detected in 40.2% (45/112) of the cases representing by far the most prevalent genetic abnormality (Fig. 1A). Nearly half of the *EGFR* mutant tumors had missense (22/45) or inframe (22/45) mutations while one tumor had a truncating mutation. *TP53* showed the second most frequent mutation at 17.9% (20/112) followed by *SETD2* and *SMARCA4* (both 10.7%), *MET* (6.3%), *KRAS* and *NFI* (both 5.4%). The co-mutation plot also revealed that over half of the *EGFR* mutant tumors (23/45) had at least one additional driver mutation (Fig. 1A). The most common co-mutations found in the *EGFR* mutant

tumors (n = 45) were *TP53* (10, 22.2%), *SETD2* (7, 15.6%), and *SMARCA4* (6, 13.3%), while *MET* (3, 6.7%), *NFI*, and *STK11* (both 2, 4.4%) were also co-mutated in few of the *EGFR* mutant tumors. Mutation of 2 or more genes was found in 30.4% (34/112) tumors, while 42.0% (47/112) tumors had only one cancer-driver mutation. Thirty-one tumors (27.7%) had normal genotype (wildtype) for all 20 genes (Fig. 1B). The median overall survival of patients with 0, 1, and 2 mutant genes was 54, 48, and 39 months from time of diagnosis, respectively, and was not significantly different between the three groups or from the 48 months median survival for all 112 patients. Similarly, none of the individual mutant genes significantly changed the overall survival. However, one tumor, (Fig. 1A, the 3rd sample), had by far the most complex genetic abnormality compared to the remaining samples with 8 of the 20 driver genes including *CDKN2A*, *EGFR*, *ERBB2*, *NFI*, *PTEN*, *SETD2*, *STK11*, and *TP53*, found to be mutated. The survival of this patient, who died 28 months after surgery, was 20 and 11 months shorter than the median survival for all patients and those with co-mutations of 2 or more genes, respectively.

The potential differences in the mutation and co-mutation patterns of the 20 cancer driver genes based on their oncogenic or tumor suppresser functions was also investigated by evaluating the oncogenes and TSGs separately. Considering only the 11 oncogenes, 50 (44.6%), 54 (48.2%), and 8 (7.1%) tumors had mutation of 0, 1, or 2 genes, respectively (Table 2). Similarly, evaluation of the 9 TSGs alone showed that 65 (58.0%), 38 (33.9%), and 9 (8.0%) tumors had mutation of 0, 1, or 2 genes, respectively. Interestingly, among the 34 tumors that contain > 2 mutant genes out of 20, 28 (82.4%) had co-mutations of at least one oncogene and one TSG (O/T). In contrast, 8 (23.5%) and 9 (26.5%) tumors had co-mutations of 2 oncogenes (O/O) or 2 TSGs (T/T), respectively, for a total of 17 tumors (50%) with either O/O or T/T co-mutation. This indicates a significantly higher ($p < 0.01$) co-mutation between an oncogene and a TSG and could be biologically relevant as it may indicate that the cooperation between oncogenic and TSG abnormalities are more common and/or more successful in driving tumorigenesis than co-mutations of oncogenes or TSGs (Table 2).

3.2. Association of mutations with clinico-pathological features, demography, and smoking

Some clinically relevant cancer-driver mutations in lung cancer show strong association with specific clinico-pathological features of the tumor as well as the demography and/or smoking history of the patient. For example, *EGFR* mutation is significantly more prevalent in female never-smoker adenocarcinoma patients of East Asian ancestry [3, 4]. However, there is a strong association between these variables as female patients are more frequently never-smokers than male patients while never-smokers more often develop adenocarcinoma compared to smokers. Thus, to better understand these associations, the relationship between the variables and driver mutations within our study that exclusively used lung adenocarcinomas from never-smokers, as well as our data with previous related publications were compared. Evaluation of *EGFR* mutation based on the sex of never-smoker lung adenocarcinoma patients revealed that the prevalence for *EGFR* mutation between male (7/19, 37%) and female (38/93, 41%) patients was similar ($p = 0.745$). The frequency of mutations in smoker and never-smoker lung adenocarcinoma cases among US based patients

were also compared using our results and data from the TCGA and Imielinski et al. studies [29, 41]. Although the two studies evaluated lung adenocarcinoma from a large number of smokers (n = 181 and 135), the number of tumors from never-smokers (n = 32 and 27, respectively) was quite small. As shown in Table S3 for the TCGA data, the prevalence for *ALK* fusion and mutations of *BRAF*, *EGFR*, *KRAS*, *NF1*, and *TP53* were significantly different between lung tumors from smokers and never-smokers. While *BRAF*, *KRAS*, *NF1*, and *TP53* mutations were significantly more prevalent in smokers compared to never-smokers, *ALK* fusion and *EGFR* mutation were more common in lung tumors from never-smokers than smokers ($p < 0.05$). With the exception of *ALK* fusion (not performed in our study) and *ERBB2*, all other significant (or borderline for *KRAS*) differences in the TCGA data were confirmed in the larger sample size in our study (Table S3). The study by Imielinski et al. also showed similar results for the 11 mutations that are common with our and TCGA studies. Because of the similar findings in the three studies, a combined analysis was performed for the 11 genes commonly evaluated. The results revealed that *BRAF*, *KEAP1*, *KRAS*, *NF1*, *SKT11*, and *TP53* mutations were significantly more prevalent in lung adenocarcinomas from smokers (n = 316) than never-smokers (n = 171) whereas *EGFR* mutation was significantly more prevalent in never-smokers than smokers (Table 3). The differences in the prevalence for *PIK3CA*, *PTEN*, *SETD2*, and *SMARCA4* mutations between lung adenocarcinomas from smokers and never-smokers were not statistically significant.

3.3. Mutations and promoter methylation of cancer-related genes in lung adenocarcinoma

The methylation profile of CpG islands across the promoter regions of multiple cancer-related genes [35–39] has been defined for the majority (75/112) of lung adenocarcinomas evaluated by targeted sequencing. Of these, 59 genes whose promoter CpG islands were not methylated in normal lung or peripheral blood mononuclear cells (PBMCs) but aberrantly methylated in lung tumors at frequencies between 20 and 80% were selected and their relationship with the driver mutations was compared. The most striking association between aberrant methylation and driver mutation that follows a similar pattern across multiple genes was seen for *EGFR* mutation. Out of 15 differentially methylated genes that showed discernable methylation differences based on *EGFR* mutation status, 14 were more commonly methylated in *EGFR* wildtype compared to mutant tumors (Table 4). The CpG islands across the promoters of 7 of these genes showed significantly more prevalent methylation in the *EGFR* wildtype tumors than *EGFR* mutant tumors ($p < 0.05$) while 7 additional genes were more frequently methylated in the *EGFR* wildtype than mutant tumors at a statistically borderline significance ($p < 0.062$). Only one gene (*TOX3*) showed significantly more frequent promoter methylation in *EGFR* mutant compared to wildtype tumors (Table 4). For the remaining 19 sequenced genes, the number of mutant tumors with methylation data was 12 (*TP53*), 11 (*SMARCA4*), 9 (*SETD2*), or lower (all others), that was too small to rationally compare the relationship between mutation and methylation.

The publicly available quantitative genome-wide methylation data generated by The Cancer Genome Atlas (TOGA) project using the HumanMethylation450 beadchip and mutation data generated using exome sequencing were used for validation of our findings. Out of the lung

adenocarcinoma samples evaluated both for methylation and mutation in the TOGA database, 53 samples were from never smokers and include 38 EGFR wildtype and 15 EGFR mutant samples. For comparisons of the quantitative methylation data from TOGA, only probes interrogating the methylation status of CpGs within the promoter CpG islands of the candidate genes were used. The summary statistics comparing the quantitative methylation data of the 15 differentially methylated genes identified above (Table 4) with EGFR mutation status and the number of probes used for the analysis of each gene are shown in Table S4. The HumanMethylation450 beadchip measures methylation levels in beta-values (β -values) that ranges from 0 (0% methylation) to 1 (100% methylation). Despite the small number of EGFR mutant samples ($n = 15$), 6 of the 15 genes showed significantly higher methylation among the EGFR wildtype tumors (Table S4). This shows clear and independent validation of our findings shown in Table 4.

4. Discussion

This study investigated the genetic and epigenetic abnormalities of large number of cancer-related genes in lung tumors from never-smokers. Targeted sequencing was used to screen mutation of 20 clinically relevant lung cancer-driver genes using one of the largest well-annotated sets of lung adenocarcinoma samples from never-smokers in the US. Evaluation of the relationships between the most frequent mutations and aberrant promoter methylation of 59 genes in these tumors identified a novel epigenetic pattern based on *EGFR* mutation. *EGFR* mutations, primarily inframe and missense type, were by far the most prevalent nonsynonymous genetic abnormality detected in over 40% of lung adenocarcinoma from never-smokers. Contrary to the common perception that tumors with one cancer-driver mutation rarely have another driver mutation [23], more than half of the *EGFR* mutant tumors (23/45) also have co-mutation of at least one other driver gene. *TP53*, *SETD2*, and *SMARCA4* genes were more frequently co-mutated with *EGFR* at 22%, 16%, and 13% of the tumors, respectively. However, in agreement with previous observations of mutual exclusion, only 1/6 (*KRAS*), 1/4 (*ERBB2*), 2/6 (*NF1*) and 0/2 *BRAF* mutations were found in the *EGFR* mutant tumors. Interestingly, *EGFR* wildtype tumors show significantly more frequent aberrant promoter methylation of multiple cancer-related genes compared to the *EGFR* mutant tumors. This suggests that in the absence of the major cancer driving oncogenic signaling of mutant *EGFR*, lung tumors in never-smokers develop through epigenetic silencing of multiple important tumor-suppressor genes. As the presence or absence of additional genetic and/or epigenetic abnormalities in *EGFR* mutant tumors modulate the biology and/or sensitivity to anti-EGFR therapy, clearly defining the relationship between these abnormalities will contribute to further optimization of targeted therapy. The high prevalence of co-mutations and the unique epigenetic profile based on *EGFR* mutation shown in this study indicates that an improved understanding of the relationship could benefit a large number of lung adenocarcinoma cases from never-smokers than previously estimated.

Cancer-driver mutations are generally considered mutually exclusive and tumors with one driver mutation rarely have a second driver mutation. The fact that some well-known mutations such as *EGFR* and *KRAS* are often mutually exclusive and the logical explanation that tumors with a major driver mutation may gain little growth advantage by having another

driver mutation likely led to the wider acceptance of this school of thought. However, in part due to the increase in the number of mutations that are considered drivers and expansion of targeted and genome/exome/transcriptome-wide sequencing of primary tumors, co-existence of driver mutations is increasingly reported. Yet, whether these mutations co-exist within the same cancer cells or are present in different cellular clones within a tumor is not clearly defined. Nevertheless, the potential interaction between pathways regulated by these mutant genes and the observations that some co-mutations modify the sensitivity or resistance of drugs targeting one of the mutations indicates their clinical relevance. The discovery of co-mutations in more than 30% of the tumors in our study further indicates that such genetic abnormalities are more frequent than previously expected. Thus, careful evaluation of their associations and impact on tumor biology and targeted therapy could benefit a large number of lung cancer patients.

Previous studies that screened cancer-driver mutations in lung adenocarcinomas from never-smokers documented that *EGFR* mutation is the most prevalent genetic abnormality present in 50 to 65% of the cases [45–47]. However, these reports are largely based on East Asian patients. Thus, the prevalence for *EGFR* and other mutations that are more common among never-smokers and may differ based on race or geographical location of these patients is not well defined. Some recent publications evaluated driver mutations in lung adenocarcinomas using US based patients and reported a slightly lower prevalence for *EGFR* mutation (~40%) [28, 29]. However, because of the few number of never-smokers evaluated in these studies (~30 cases per study), the currently available data is not sufficient to make conclusive determination regarding the racial/geographical differences in lung cancer-driver mutations. Thus, this study screened the most common and clinically relevant cancer-driver mutations using one of the largest number lung adenocarcinoma cases from never-smokers in the US. Our mutation data largely validated the previous reports and the combined analysis of the studies confirmed that *EGFR* mutation in US never-smoker lung adenocarcinoma patients is less frequent than similar patients of East Asian origin.

Aberrant promoter CpG island methylation of tumor-suppressor genes is one of the most common abnormalities in cancer. Our group has published multiple papers describing promoter hypermethylation mediated silencing of multiple genes in lung cancer from smokers and never-smokers [35–39]. In this study, the relationship between the most common driver mutations and methylation of 59 genes that were methylated in 20 to 80% of lung adenocarcinomas from never-smokers were compared. The most striking relationship between a driver mutation and epigenetic silencing of these genes was found for *EGFR* mutation. Fourteen of the 15 differentially methylated genes based on *EGFR* mutation were more commonly methylated in the *EGFR* wildtype tumors. The significantly higher methylation of six of these genes in the *EGFR* wildtype tumors was also independently validated using the publicly available quantitative methylation and exome sequencing data from TCGA. This suggests that in the absence of the *EGFR* activating mutations that drive over 40% of lung adenocarcinomas in never-smokers, *EGFR* wildtype tumors may develop through accumulation of epigenetic silencing of multiple tumor-suppressor genes. Complete loss of function (LOF) of critical tumor suppresser genes could result in failure to control cell proliferation or inability to induce cell death pathways and serve as an additional or alternative cancer promoting mechanism. Inactivation of both alleles of tumor suppresser

genes leading to such total LOF is often described as the “Knudson’s two-hit model” and the two alleles could be both hit (inactivated) by genetic, epigenetic or the combination of the two abnormalities inactivating one allele each [48, 49],

Our observation that *EGFR* wildtype tumors may develop through epigenetic repression of tumor-suppressor genes is also supported by multiple previous publications. Koga et al. reported that *EGFR* mutation and *CHFR* hypermethylation in lung adenocarcinoma are mutually exclusive [31]. Similarly, Suzuki et al. reported that lung tumors with wildtype *EGFR* show more prevalent hypermethylation of *p16* (34% vs. 16%) and *SPARC* (60% vs. 37%) genes compared to *EGFR* mutant tumors [32]. Yanagawa et al. also compared the relationship between methylation of 10 cancer-related genes and mutations of *EGFR*, *KRAS*, and *TP53* in 62 lung adenocarcinoma cases from smokers and never-smokers and found that the prevalence for methylation of *FHIT*, *RASSF1A*, and *RUNX3* genes in *EGFR* mutant tumors was significantly lower compared to *EGFR* wild-type tumors. In contrast, the significantly higher rate of *TOX3* methylation in *EGFR* mutant than wildtype tumors demonstrates that aberrant methylation is not always inversely associated to *EGFR* mutation. Rather, though not as common as *EGFR* wildtype tumors, methylation of some genes is strongly associated and may contribute to the development of *EGFR* mutant tumors. Comparison of *EGFR* mutation and methylation of 30 genes by Pesek et al using 121 lung tumors similarly found that the methylation rates of *APC*, *CDKN2B*, *ESR1*, and *VHL* was significantly higher in *EGFR* mutant than wildtype tumors [50]. Whereas, methylation of four genes (*CDKN2A*, *CHFR*, *MGMT*, and *RASSF1*) that were commonly evaluated in Pesek et al. and our studies similarly had no significant association with *EGFR* mutation. Although 75/121 tumors evaluated by Pesek et al. were adenocarcinomas, only 22 were from non-smokers. Thus, to control the impact of smoking and tumor histology that are strongly associated with *EGFR* mutation, the higher methylation rate of the four genes in *EGFR* mutant tumors needs to be validated exclusively in adenocarcinomas from never-smokers. Taken together, these findings support the premise that tumors that lack the cancer-driver role of *EGFR* mutations develop through epigenetic silencing of multiple cancer-related genes and could be more sensitive to epigenetic therapy. Despite previous publication reporting similar trends using methylation of different genes, additional studies are required to independently validate and clearly define the inverse relationship between methylation of tumor-suppressor genes and *EGFR* mutation in lung adenocarcinoma.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Data generated by The Cancer Genome Atlas (TCGA) pilot project established by the NCI and NHGRI was used to validate part of our findings. The dbGaP accession number for TCGA data is phs000178.v8.p7. Information about TCGA and the investigators and institutions that constitute the TCGA research network can be found at <http://cancergenome.nih.gov/>.

Funding

This study was supported largely by grant U10 CA180950 and in part by R01CA183296, R01CA193532, and P30CA11800 from the National Institutes of Health.

References

- [1]. Siegel RL, Miller KD and Jemal A, Cancer Statistics, 2017, CA Cancer J Clin 67(1) (2017) 7–30. [PubMed: 28055103]
- [2]. Torre LA, Bray F, Siegel RL, et al., Global cancer statistics, 2012, CA Cancer J Clin 65(2) (2015) 87–108. [PubMed: 25651787]
- [3]. Subramanian J and Govindan R, Lung cancer in never smokers: a review, J Clin Oncol 25(5) (2007) 561–70. [PubMed: 17290066]
- [4]. Sun S, Schiller JH and Gazdar AF, Lung cancer in never smokers--a different disease, Nat Rev Cancer 7(10) (2007) 778–90. [PubMed: 17882278]
- [5]. Wakelee HA, Chang ET, Gomez SL, et al., Lung cancer incidence in never smokers, J Clin Oncol 25(5) (2007) 472–8. [PubMed: 17290054]
- [6]. Shaw AT, Yeap BY, Mino-Kenudson M, et al., Clinical features and outcome of patients with non-small-cell lung cancer who harbor EML4-ALK, J Clin Oncol 27(26) (2009) 4247–53. [PubMed: 19667264]
- [7]. Solomon BJ, Mok T, Kim DW, et al., First-line crizotinib versus chemotherapy in ALK-positive lung cancer, N Engl J Med 371(23) (2014) 2167–77. [PubMed: 25470694]
- [8]. Shaw AT, Kim DW, Nakagawa K, et al., Crizotinib versus chemotherapy in advanced ALK-positive lung cancer, N Engl J Med 368(25) (2013) 2385–94. [PubMed: 23724913]
- [9]. Kwak EL, Bang YJ, Camidge DR, et al., Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer, N Engl J Med 363(18) (2010) 1693–703. [PubMed: 20979469]
- [10]. Sequist LV, Yang JC, Yamamoto N, et al., Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations, J Clin Oncol 31(27) (2013) 3327–34. [PubMed: 23816960]
- [11]. Maemondo M, Inoue A, Kobayashi K, et al., Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR, N Engl J Med 362(25) (2010) 2380–8. [PubMed: 20573926]
- [12]. Kobayashi K and Hagiwara K, Epidermal growth factor receptor (EGFR) mutation and personalized therapy in advanced nonsmall cell lung cancer (NSCLC), Target Oncol 8(1) (2013) 27–33. [PubMed: 23361373]
- [13]. Mitsudomi T, Morita S, Yatabe Y, et al., Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial, Lancet Oncol 11(2) (2010) 121–8. [PubMed: 20022809]
- [14]. Lynch TJ, Bell DW, Sordella R, et al., Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib, N Engl J Med 350(21) (2004) 2129–39. [PubMed: 15118073]
- [15]. Paez JG, Janne PA, Lee JC, et al., EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy, Science 304(5676) (2004) 1497–500. [PubMed: 15118125]
- [16]. Choi YL, Takeuchi K, Soda M, et al., Identification of novel isoforms of the EML4-ALK transforming gene in non-small cell lung cancer, Cancer Res 68(13) (2008) 4971–6. [PubMed: 18593892]
- [17]. Soda M, Choi YL, Enomoto M, et al., Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer, Nature 448(7153) (2007) 561–6. [PubMed: 17625570]
- [18]. Hagiwara K and Kobayashi K, Importance of the cytological samples for the epidermal growth factor receptor gene mutation test for non-small cell lung cancer, Cancer Sci 104(3) (2013) 291–7. [PubMed: 23240603]
- [19]. Pirker R, Herth FJ, Kerr KM, et al., Consensus for EGFR mutation testing in non-small cell lung cancer: results from a European workshop, J Thorac Oncol 5(10) (2010) 1706–13. [PubMed: 20871269]
- [20]. Yatabe Y, Kerr KM, Utomo A, et al., EGFR mutation testing practices within the Asia Pacific region: results of a multicenter diagnostic survey, J Thorac Oncol 10(3) (2015) 438–45. [PubMed: 25376513]

- [21]. Saito M, Shiraishi K, Kunitoh H, et al., Gene aberrations for precision medicine against lung adenocarcinoma, *Cancer Sci* 107(6) (2016) 713–20. [PubMed: 27027665]
- [22]. Drilon A, Wang L, Arcila ME, et al., Broad, Hybrid Capture-Based Next-Generation Sequencing Identifies Actionable Genomic Alterations in Lung Adenocarcinomas Otherwise Negative for Such Alterations by Other Genomic Testing Approaches, *Clin Cancer Res* 21(16) (2015) 3631–9. [PubMed: 25567908]
- [23]. Pao W and Hutchinson KE, Chipping away at the lung cancer genome, *Nat Med* 18(3) (2012) 349–51. [PubMed: 22395697]
- [24]. Martorell PM, Huerta M, Quilis AC, et al., Coexistence of EGFR, KRAS, BRAF, and PIK3CA Mutations and ALK Rearrangement in a Comprehensive Cohort of 326 Consecutive Spanish Nonsquamous NSCLC Patients, *Clin Lung Cancer* (2017).
- [25]. Engelman JA, Zejnullahu K, Mitsudomi T, et al., MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling, *Science* 316(5827) (2007) 1039–43. [PubMed: 17463250]
- [26]. Baldi L, Mengoli MC, Bisagni A, et al., Concomitant EGFR mutation and ALK rearrangement in lung adenocarcinoma is more frequent than expected: report of a case and review of the literature with demonstration of genes alteration into the same tumor cells, *Lung Cancer* 86(2) (2014) 291–5. [PubMed: 25312989]
- [27]. Drilon A, Siena S, Ou SI, et al., Safety and Antitumor Activity of the Multitargeted Pan-TRK, ROS1, and ALK Inhibitor Entrectinib: Combined Results from Two Phase I Trials (ALKA-372–001 and STARTRK-1), *Cancer Discov* 7(4) (2017) 400–409. [PubMed: 28183697]
- [28]. Cancer N Genome Atlas Research, Comprehensive molecular profiling of lung adenocarcinoma, *Nature* 511(7511) (2014) 543–50. [PubMed: 25079552]
- [29]. Imielinski M, Berger AH, Hammerman PS, et al., Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing, *Cell* 150(6) (2012) 1107–20. [PubMed: 22980975]
- [30]. Tiedemann RL, Hlady RA, Hanavan PD, et al., Dynamic reprogramming of DNA methylation in SETD2-deregulated renal cell carcinoma, *Oncotarget* 7(2) (2016) 1927–46. [PubMed: 26646321]
- [31]. Koga T, Takeshita M, Yano T, et al., CHFR hypermethylation and EGFR mutation are mutually exclusive and exhibit contrastive clinical backgrounds and outcomes in non-small cell lung cancer, *Int J Cancer* 128(5) (2011) 1009–17. [PubMed: 20473935]
- [32]. Suzuki M, Shigematsu H, Iizasa T, et al., Exclusive mutation in epidermal growth factor receptor gene, HER-2, and KRAS, and synchronous methylation of nonsmall cell lung cancer, *Cancer* 106(10) (2006) 2200–7. [PubMed: 16598760]
- [33]. Yanagawa N, Tamura G, Oizumi H, et al., Inverse correlation between EGFR mutation and FHIT, RASSF1A and RUNX3 methylation in lung adenocarcinoma: relation with smoking status, *Anticancer Res* 31(4) (2011) 1211–4. [PubMed: 21508367]
- [34]. Shih AH, Meydan C, Shank K, et al., Combination Targeted Therapy to Disrupt Aberrant Oncogenic Signaling and Reverse Epigenetic Dysfunction in IDH2-and TET2-Mutant Acute Myeloid Leukemia, *Cancer Discov* 7(5) (2017) 494–505. [PubMed: 28193779]
- [35]. Tessema M, Klinge DM, Yingling CM, et al., Re-expression of CXCL14, a common target for epigenetic silencing in lung cancer, induces tumor necrosis, *Oncogene* 29(37) (2010) 5159–70. [PubMed: 20562917]
- [36]. Tessema M, Willink R, Do K, et al., Promoter methylation of genes in and around the candidate lung cancer susceptibility locus 6q23–25, *Cancer Res* 68(6) (2008) 1707–14. [PubMed: 18339850]
- [37]. Tessema M, Yingling CM, Grimes MJ, et al., Differential epigenetic regulation of TOX subfamily high mobility group box genes in lung and breast cancers, *PLoS One* 7(4) (2012) e34850. [PubMed: 22496870]
- [38]. Tessema M, Yingling CM, Liu Y, et al., Genome-wide unmasking of epigenetically silenced genes in lung adenocarcinoma from smokers and never smokers, *Carcinogenesis* 35(6) (2014) 1248–57. [PubMed: 24398667]

- [39]. Tessema M, Yu YY, Stidley CA, et al., Concomitant promoter methylation of multiple genes in lung adenocarcinomas from current, former and never smokers, *Carcinogenesis* 30(7) (2009) 1132–8. [PubMed: 19435948]
- [40]. Belinsky SA, Palmisano WA, Gilliland FD, et al., Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers, *Cancer Res* 62(8) (2002) 2370–7. [PubMed: 11956099]
- [41]. TCGA TCGARN, Comprehensive genomic characterization of squamous cell lung cancers, *Nature* 489(7417) (2012) 519–25. [PubMed: 22960745]
- [42]. Cassidy RJ, Zhang X, Patel PR, et al., Next-generation sequencing and clinical outcomes of patients with lung adenocarcinoma treated with stereotactic body radiotherapy, *Cancer* (2017).
- [43]. Fisher KE, Zhang L, Wang J, et al., Clinical Validation and Implementation of a Targeted Next-Generation Sequencing Assay to Detect Somatic Variants in Non-Small Cell Lung, Melanoma, and Gastrointestinal Malignancies, *J Mol Diagn* 18(2) (2016) 299–315. [PubMed: 26801070]
- [44]. Su Z, Dias-Santagata D, Duke M, et al., 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) (2011) 74–84. [PubMed: 21227397]
- [45]. Dong YU, Ren W, Qi J, et al., EGFR, ALK, RET, KRAS and BRAF alterations in never-smokers with non-small cell lung cancer, *Oncol Lett* 11(4) (2016) 2371–2378. [PubMed: 27073482]
- [46]. Ha SY, Choi SJ, Cho JH, et al., Lung cancer in never-smoker Asian females is driven by oncogenic mutations, most often involving EGFR, *Oncotarget* 6(7) (2015) 5465–74. [PubMed: 25760072]
- [47]. Serizawa M, Koh Y, Kenmotsu H, et al., Assessment of mutational profile of Japanese lung adenocarcinoma patients by multitarget assays: a prospective, single-institute study, *Cancer* 120(10) (2014) 1471–81. [PubMed: 24700479]
- [48]. Knudson AG, Jr., Mutation and cancer: statistical study of retinoblastoma, *Proc Natl Acad Sci U S A* 68(4) (1971) 820–3. [PubMed: 5279523]
- [49]. Jones PA and Laird PW, Cancer epigenetics comes of age, *Nat Genet* 21(2) (1999) 163–7. [PubMed: 9988266]
- [50]. Pesek M, Kopeckova M, Benesova L, et al., Clinical significance of hypermethylation status in NSCLC: evaluation of a 30-gene panel in patients with advanced disease, *Anticancer Res* 31(12) (2011) 4647–52. [PubMed: 22199344]

Highlights

- Lung adenocarcinoma in never smokers is a growing health concern in many countries.
- Some targetable cancer driver mutations like EGFR are more common in these tumors.
- Mutation of 20 cancer driver genes was screened in 112 of these tumors.
- The association between mutations and 59 aberrantly methylated genes was studied.
- EGFR mutation is the most prevalent found in 40% of these tumors.
- Methylation of many genes is significantly more common in EGFR wild type tumors.

Figure 1: Common somatic mutations in lung adenocarcinoma from never-smokers.

A) DNA from fresh frozen lung adenocarcinoma samples of 112 never-smokers (arranged 1 to 112 left to right) was evaluated using targeted DNA sequencing of 20 cancer-driver genes. The presence or absence and type of nonsynonymous mutation detected in 19 of the genes are displayed, no nonsynonymous mutation was found for ALK. The prevalence for mutation of each gene is shown on the right side of the plot. **B)** Pie chart demonstrating the

most common mutations with co-mutations shown only for the gene with the most frequent and clinically relevant mutation.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:

Relationship between patient characteristics and mutations of 20 cancer-driver genes

Characteristics	Sequenced (N = 112), n (%) *	Mutant genes, n (%)				Methylation [†] evaluated (N = 75), n (%)
		0	1	2	EGFR	
Age: Median, (range)	69, (24–91)					70, (24–91)
Gender:						
Female	93 (83)	25 (27)	40 (43)	28 (30)	38 (41)	63 (84)
Male	19 (17)	6 (32)	7(37)	6 (32)	7(37)	12 (16)
Race:						
White	97 (87)	28 (29)	37 (38)	32 (33)	37 (38)	65 (87)
Black	3(3)	1 (33)	1 (33)	1 (33)	1 (33)	3(4)
Asian/Pacific Islander	2(2)	0(0)	2 (100)	0(0)	2 (100)	2(3)
Others	5(4)	1 (20)	3 (60)	1 (20)	2 (40)	3(4)
Not Available	5(4)	1 (20)	4 (80)	0(0)	3 (60)	2(3)
Stages:						
Stage I	79 (71)	21 (27)	31 (39)	27 (34)	32 (41)	51 (68)
Stage II	6(5)	2 (33)	1 (17)	3 (50)	2 (33)	6(8)
Stage III	21 (19)	7(33)	11 (52)	3 (14)	9 (43)	14 (19)
Stage IV	5(4)	1 (20)	3 (60)	1 (20)	2 (40)	4(5)
Not Available	1 (1)	0(0)	1 (100)	0(0)	0(0)	-

* 'N' and 'n' represent the total number of samples evaluated and positive samples, respectively.

[†] A large subset (75/112) of the sequenced samples with sufficient tissue for multiple assays was used for methylation analysis of the 59 genes. The patient characteristics of these samples were similar to the total samples evaluated.

Table 2:

Mutation patterns of oncogenes and tumor suppressor genes (TSGs)

Roles of the mutant genes in cancer development	Number of Mutated genes,n (%)		
	0	1	2
All genes (n = 20)	31 (28%)	47 (42%)	34 (30%)
Oncogenes (n = 11)	50 (45%)	54 (48%)	8 (7%)
TSGs (n = 9)	65 (58%)	38 (34%)	9 (8%)
Co-mutations (34 tumors)	2 mutant genes n (%)	P-values(Fisher exact test)	
Oncogene/Oncogene + TSG/TSG	(8 + 9), 17 (50%)		0.0095
Oncogene/TSG	28 (82%)		

Table 3:

Combined analysis of cancer driver mutations in lung adenocarcinoma from smokers and never smokers

Genes	Prevalence for mutation		P-values [†] (Fisher exact test)
	Smokers (n = 316)	Never smokers (n =171)	
BRAF	32 (10.1%)	2 (1.2%)	0.0003 [*]
EGFR	41 (13.0%)	65 (38.0%)	< 0.0001 [*]
KEAP1	54 (17.1%)	7 (4.1%)	< 0.0001
KRAS	105 (33.2%)	14 (8.2%)	< 0.0001 [*]
NF1	41 (13.0%)	7 (4.1%)	0.0021 [*]
PIK3CA	16 (5.1%)	3 (1.8%)	0.0872
PTEN	4 (1.3%)	4 (2.3%)	0.4600
SETD2	26 (8.2%)	15 (8.8%)	0.8649
SMARCA4	27 (8.5%)	12 (7.0%)	0.6039
STK11	58 (18.4%)	9 (5.3%)	< 0.0001
TP53	163 (51.6%)	33 (19.3%)	< 0.0001 [*]

[†] P-values were determined using 2-tailed Fisher exact test.^{*} Significantly different mutations validated in our samples.

Table 4:

Differentially methylated genes in lung tumors with or without EGFR mutation

Gene	<u>Prevalence for promoter methylation</u>		P-value (Fisher exact test)
	EGFR wildtype (n = 46)	EGFR mutant (n = 29)	
ACP5	41%	10%	0.004 *
ICAM5	70%	41%	0.018 *
CNRIP1	78%	52%	0.023 *
SFRP2	76%	48%	0.024 *
PCDH10	74%	48%	0.029 *
ZNF365	37%	14%	0.036 *
TOX3	54%	79%	0.047 **
TOX2	33%	10%	0.049 *
HES5	30%	10%	0.051
JPH3	85%	64%	0.051
ANK1	48%	25%	0.052
SYNE1	63%	38%	0.056
LOX	58%	34%	0.060
GJB2	61%	38%	0.062
TSLC1 (CADM1)	61%	38%	0.062

* Statistically significant difference ($p < 0.05$)

** The only gene with significantly higher methylation in EGFR mutant than wild-type tumors.