

Profiling the Mutational Landscape in Known Driver Genes and Novel Genes in African American Non-Small Cell Lung Cancer Patients



Christine M. Lusk^{1,2}, Donovan Watza^{1,2}, Greg Dyson^{1,2}, Douglas Craig^{1,2}, Valerie Ratliff^{1,2}, Angela S. Wenzlaff^{1,2}, Fulvio Lonardo^{1,3}, Aliccia Bollig-Fischer^{1,2}, Gerold Beppler^{1,2}, Kristen Purrington^{1,2}, Shirish Gadgeel^{1,2,4}, and Ann G. Schwartz^{1,2}

Abstract

Purpose: Identifying novel driver genes and mutations in African American non-small cell lung cancer (NSCLC) cases can inform targeted therapy and improve outcomes for this traditionally underrepresented population.

Experimental Design: Tumor DNA, RNA, and germline DNA were collected from African American NSCLC patients who participated in research conducted at the Karmanos Cancer Institute (KCI) in Detroit, Michigan. Known mutations were ascertained through the Sequenom LungCarta panel of 214 mutations in 26 genes, *RET/ROS1* fusions, amplification of *FGFR1*, and expression of ALK. Paired tumor and normal DNA was whole-exome sequenced for a subset of cases without known driver mutations.

Results: Of the 193 tumors tested, 77 known driver mutations were identified in 66 patients (34.2%). Sixty-seven of the 127 patients without a known driver mutation were sequenced. In 54 of these patients, 50 nonsynonymous mutations were predicted to have damaging effects among the 26 panel genes, 47 of which are not found in The Cancer Genome Atlas NSCLC white or African American samples. Analyzing the whole-exome sequence data using MutSig2CV identified a total of 88 genes significantly mutated at FDR $q < 0.1$. Only 5 of these genes were previously reported as oncogenic.

Conclusions: These findings suggest that broader mutation profiling including both known and novel driver genes in African Americans with NSCLC will identify additional mutations that may be useful in treatment decision-making.

Introduction

Lung cancer is the second leading form of cancer diagnosed in both men and women in the United States, affecting an estimated 234,030 persons in 2018, and is the leading cause of cancer-related death in the United States with an expected 154,050 deaths in 2018 (1). The highest incidence rates occur among African Americans (2) even though African Americans smoke fewer cigarettes per day than whites. Five-year survival for lung cancer is worse in African Americans (17.3%) than in whites (19.6%; ref. 2). Improvements in survival have been observed in patients with molecularly characterized tumors receiving targeted treatments (3). Non-small cell lung cancers (NSCLC) have been defined by molecular profile including genetic alterations in *EGFR*, *ALK*, *RET*, *ROS1*, and *BRAF* providing actionable treatment

targets with improved outcomes with treatment with targeted agents (4, 5).

Although the testing of a defined set of mutations in targeted genes forms the basis for therapeutic decision-making, the set of mutations tested and deemed actionable has been based on discovery in largely white populations. For example, although The Cancer Genome Atlas (TCGA) profiled over 1,000 NSCLC tumors, fewer than 10% originated in African Americans (TCGA data portal accessed on May 10, 2018). Similarly, the Lung Cancer Mutation Consortium (LCMC2) reported targeted mutation profiles in only 60 African Americans (6). Given that genetic profiles of lung tumors differ by smoking exposure, ethnicity, age, and sex, it is imperative to ask whether there are also differences by race. This study evaluated genetic profiling in African Americans with NSCLC with a focus on known driver mutations, as well as profiling novel alterations in both known oncogenic and novel genes through whole-exome sequencing, to provide a comprehensive approach through which to evaluate treatment decision-making, furthering options in personalized/precision medicine in African Americans.

¹Karmanos Cancer Institute, Detroit, Michigan. ²Department of Oncology, School of Medicine, Wayne State University, Detroit, Michigan. ³Department of Pathology, School of Medicine, Wayne State University, Detroit, Michigan. ⁴Currently at the Comprehensive Cancer Center, University of Michigan, Ann Arbor, Michigan.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

C.M. Lusk and D. Watza contributed equally to this article.

Corresponding Author: Ann G. Schwartz, Wayne State University, 4100 John R., Detroit, MI 48201. Phone: 313-578-4201; Fax: 313-578-4359; E-mail: schwarta@karmanos.org

Clin Cancer Res 2019;25:4300-8

doi: 10.1158/1078-0432.CCR-18-2439

©2019 American Association for Cancer Research.

Materials and Methods

Patient samples

A total of 193 African American NSCLC cases with both normal DNA, and tumor DNA and RNA were available for this study. Cases were selected from case-control studies conducted at Karmanos Cancer Institute (KCI; refs. 7, 8). Additional cases participated in a research registry and biorepository at the KCI, which has been enrolling lung cancer cases since 2010

Translational Relevance

African Americans continue to have poorer 5-year survival after a lung cancer diagnosis than whites for reasons that remain to be fully characterized. Targeted therapies based on specific molecular profiles have improved outcomes for some patients, although these profiles were primarily developed in patients of European descent. We evaluated the mutational landscape of both known driver and novel genes in a sample of 193 African American NSCLC cases. We found that 127 of 193 cases did not carry a known driver mutation. Whole-exome sequencing of paired tumor/normal DNA in 67 of these cases revealed 47 mutations with predicted damaging effects among 26 oncogenic driver genes that were not observed in TCGA samples. We also identified 88 genes, including 83 novel genes, which were significantly mutated indicating there are additional mutations in both known driver and novel genes that may be clinically relevant for African American NSCLC cases.

($n = 1,164$). Written informed consent was obtained for all participants. The Wayne State University (Detroit, MI) Institutional Review Board approved the procedures used in collecting and processing participant information, which were in accordance with the Declaration of Helsinki. Germline DNA was available from blood (61.0%), saliva (26.7%), or adjacent normal tissue identified by pathology review in formalin-fixed, paraffin-embedded (FFPE) blocks (10.8%). FFPE tumor tissue was the source of tumor DNA and RNA for all participants.

Archived FFPE tissues were reviewed by a pathologist and tumor-specific regions ($\sim 1 \text{ cm} \times \sim 1 \text{ cm}$) were marked. Marked regions containing tumor were macrodissected to create separate tumor sub-blocks. Five 10- μm curls were collected from tumor sub-blocks for DNA extraction. DNA was extracted from tissue using Qiagen's QIAamp DNA FFPE Kit. Normal DNA was isolated from whole blood and saliva using Qiagen Puregene reagents. Sample DNA quantity and quality were determined using a Nanodrop spectrophotometer and Quantifiler (Life Technologies) assays. Total RNA was isolated from FFPE tissue using the RNeasy FFPE Kit, with extended proteinase K and DNAase treatment (Life Technologies). Total purified RNA was quantified by UV spectrophotometry using the DropSense96 Microplate Spectrophotometer (Trinean) and A260/A280 and A260/A230 ratios.

For comparison, clinical and genomic TCGA data for 1,144 samples were downloaded from NCI's Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov>, accessed on March 2018). Of the 1,144 samples, 933 self-reported white or African American (211 samples were either missing race or reported Asian or Native American ancestry). Hence, we analyzed 567 white or African American lung adenocarcinoma (517 white, 50 African American) and 366 white or African American lung squamous cell carcinoma (337 white, 29 African American) samples.

Detection of known lung cancer mutations

Both normal and tumor DNA were initially screened for a panel of 214 known tumor driver mutations in 26 genes using the Sequenom MassARRAY System—matrix-assisted laser desorption/ionization and time of flight (MALDI-TOF) mass spectrometry (MS)—and the Sequenom LungCarta panel (Supplementary

Table S1; ref. 9). The sensitivity of the technology allows for detection of a mutation that represents $\geq 8\%$ of the sample.

Detection of *ROS1* and *RET* fusion genes

Expression of known fusion genes characterized to be oncogenic was assessed using the Sequenom MassARRAY system, as described in Wijesinghe and colleagues (10). Sets of multiplexed assays tested for the presence of fusion genes in specimen-derived mRNA converted to double-stranded cDNA and amplified using 200 ng cDNA targets a short sequence (on average 100 base pairs) surrounding the fusion gene junction or adjacent wild-type exon junction. An extension reaction and MALDI-TOF MS analysis were used to detect the expected mass spectra of the wild-type or aberrant transcript of 15 *RET* or *ROS1* variants.

Detection of ALK

Tissue sections were deparaffinized. Antigen retrieval was performed using an EDTA buffer pH 8.0 in the deprotecting chamber (Biocare Medical). Endogenous peroxidase was quenched using 3% H_2O_2 . Nonspecific staining was blocked using CAS-block (Life Technologies). The sections were then incubated overnight with a 1/100 dilution of the primary antibody anti-ALK5F3 (Cell Signaling Technology). Visualization was achieved using the horseradish peroxidase-labeled polymer system Envision (Dako) and the chromogen DAB. The slides were counterstained using Mayer Hematoxylin and read by a single pathologist.

Detection of *FGFR1*

Copy number analysis for the human *FGFR1* gene was performed by real-time PCR using an independent TaqMan Assay (Life Technologies) specific for exon 15 (catalog no. Hs02702320; targeting Chr.8:38274932 on NCBI build 37, overlapping the intron 14–exon 15 boundary within the kinase domain region). Relative quantitation of gene copy number in cancer samples was done by the Livak ($2^{-\Delta\Delta C_t}$) method using CEPH 1347-02 control DNA and RPLPO as the reference gene (catalog no. 4326314E), both from Life Technologies (11). The average value across three replicates was calculated and used as the gene amplification level; a copy number threshold of 3.5 or higher constituted the threshold for positivity.

Whole-exome sequencing

Whole-exome sequencing was done on normal and tumor DNA from 67 samples, selected based on DNA quantity (dsDNA concentration assessed by PicoGreen fluorescence; Molecular Probes) and quality (peak fragment length assessed by Agilent Technologies 2200 TapeStation system), from the 127 tumors without known genetic alterations described above (Supplementary Fig. S1). Libraries were prepared per Illumina standard procedures. Pooled libraries were subjected to capture using the Illumina Nextera Rapid Capture Exome Enrichment Kit and sequenced on a HiSeq 2500 instrument.

Samples were run in two batches. In the first batch, 23 matched pairs were sequenced on eight lanes. To increase read depth, this experiment was run three times. In the second batch, 44 paired samples were sequenced on eight lanes and run twice, switching lane positions to mitigate any potential lane bias. Repeat runs for individual samples were combined at the ".bam" level and somatic variants were called on the basis of this combined file. Mean read depth for normal samples was 48 and the 90th percentile depth was 115, whereas mean read depth for tumor

samples was 36 and the 90th percentile depth was 90. Mutect2 was used after aligning with the BWA-MEM procedure to call somatic variants (12, 13). Mutect2 is designed to identify variant alleles in the tumor that differ from both the reference and the control (or "normal") genome. Because the reference genome is based primarily on non-African ancestry populations, somatic mutations could be missed if a site varies in the normal sample but matches the reference genome in the tumor (i.e., minor allele switched in African vs. European populations). As such, identification of potential somatic variants was performed two ways: (i) using the germline sequence as the "control" and (ii) using the tumor sequence as the "control." This allowed calling somatic variants where the reference allele is not appropriate for an African American population. Variants that had at least 10 reads in the normal sample and 10 reads in the tumor sample and at least a 1/6 difference in variant frequency between the tumor and normal sample were carried forward for further investigation. Identified mutations in known driver genes were classified using PolyPhen to determine potential functional consequences (14). PolyPhen predicts the functional significance of a nonsynonymous mutation by evaluating key features of the amino acid change, including (i) the functional activity of the site, (ii) spatial contact of the substitution with functional sites, (iii) compatibility of the substitution in the context of the family of homologous proteins, (iv) compatibility of the substitution in the secondary structure of the protein, and (v) compatibility of the substitution in the three-dimensional structure of the protein as it pertains to folding energy, surface area, and surface propensities.

To identify novel driver genes we utilized MutSig2CV, described in detail elsewhere (15). For this analysis, we considered the 60 cases with no known driver mutations. Briefly, MutSig2CV estimates background mutation rates (BMR) specific to particular genes, patients, and contexts (e.g., gene expression levels) and uses this BMR to evaluate the significance of a given gene's observed mutation rate. MutSig2CV performs three tests of significance for each gene, based on abundance (mutation rate), clustering (physical collocation of mutations), and evolutionary conservation. A Fisher combined probability test is used to estimate an overall *P* value and the FDR is used for multiple test correction. The MutSig2CV analysis was performed in MatLab. All other analyses were performed with R v3.4.3. Waterfall plots were created using the GenVisR package.

Survival analysis

Cox proportional hazards models were used to test associations between overall survival and particular mutation profiles. To determine which covariates to include in survival modeling, univariate association tests of each covariate with overall survival were performed. Covariates that were significant at $\alpha = 0.05$ were included in a multivariate Cox model, and those that remained significant in the multivariate model were retained for mutation profile survival modeling.

Results

Cohort characteristics and panel-based driver gene mutation status

The characteristics of the study population are presented in Table 1. African American cases were 59.6% female, 6.7% were never smokers, the mean age of diagnosis was 61.4 years, and 66.3% were adenocarcinomas. Of the 193 cases in this study,

Table 1. Description of African American NSCLC cases

Variable	N = 193
Age, mean (SD)	61.4 (10.9)
Gender, no. (%)	
Male	78 (40.4)
Female	115 (59.6)
Smoking status, no. (%)	
Never	13 (6.7)
Ever	180 (93.3)
Pack-years, mean (SD)	40.2 (26.8)
Family history of lung cancer, no. (%)	
No	149 (77.2)
Yes	44 (22.8)
History of any COPD, no. (%)	
No	120 (64.9)
Yes	65 (35.1)
Missing	8
Stage, no. (%)	
I	113 (58.6)
II	21 (10.9)
III	35 (18.1)
IV	24 (12.4)
Histology, no. (%)	
Adenocarcinoma	128 (66.3)
Squamous cell	49 (25.4)
Other NSCLC ^a	16 (8.3)

^aOther NSCLC includes large cell, adenosquamous, and not otherwise specified.

77 known pathogenic mutations were identified in 66 cases (34.2%) via the LungCarta panel or through assays detecting *RET/ROS* fusions, *FGFR1* amplification, or *ALK* expression. The most frequent panel mutations were in *KRAS* (11.4%), *TP53* (7.3%), and *EGFR* (5.7%; Table 2). One case had a *RET/ROS* fusion event, and 6 (3.1%) *FGFR1* amplifications and 7 (3.6%) *ALK*-positive tumors were identified. There were 11 cases with cooccurring somatic mutations (6%), 4 of whom had both a *TP53* and a *KRAS* mutation. Females were 2.4 times more likely than males to have any one of these specific mutations ($P = 0.008$), explained by a 12-fold higher proportion of *EGFR* mutations in females versus males ($P = 0.016$). Age, smoking status, pack-years, family history of lung cancer, history of chronic obstructive pulmonary disease (COPD), stage at diagnosis, and histology were not associated with a positive tumor mutation profile for this limited set of genes ($P > 0.05$).

We found gender, histology (adenocarcinoma/squamous cell/other NSCLC), and stage were each associated with overall survival in univariate Cox modeling (Supplementary Table S2). In a multivariate model including all three covariates, histology and stage remained significant (age P -value = 0.1902); therefore, mutation carrier status analyses were adjusted for histology and stage. There was no association between mutation carrier status and overall survival adjusted for stage and histology [HR 1.11; 95% confidence interval (CI), 0.74–1.65; $P = 0.623$; Fig. 1].

When stratified by histology, the subtype-specific profile in African Americans is comparable with that seen in other populations for several driver genes; genetic alterations in *ALK*, *EGFR*, and *KRAS* occurred most often in the 128 adenocarcinomas, whereas *FGFR1* amplification was more common in the 49 squamous cell carcinomas (Table 2). Overall, adenocarcinomas were more likely to harbor driver mutations than squamous cell carcinomas (37% vs. 24% of samples carrying at least one panel mutation, respectively). Among the 16 samples with other NSCLC histologies (e.g., large cell, adenosquamous), *KRAS* mutations

Table 2. List of known somatic mutations in African American NSCLC cases, by histology

Gene	Total (N = 193)	Adenocarcinoma (N = 128)		Squamous cell (N = 49)		Other NSCLC (N = 16)	
	No. (%)	No. (%)	Mutation(s)	No. (%)	Mutation(s)	No. (%)	Mutation(s)
ALK	7 (3.6)	7 (5.5)	ALK-EML4 fusion	0		0	
EGFR	11 (5.7)	10 (7.8)	E746-A750del, E746-T751>S, G719C, L858R	0		1 (6.3)	D761N
EPHA3	1 (0.5)	1 (0.8)	A435S	0		0	
ERBB2	1 (0.5)	0		1 (2.0)	M774-A775insAYVM	0	
FGFR1	6 (3.1)	2 (1.6)	3.5-fold amplification	4 (8.2)		0	
KRAS	22 (11.4)	18 (14.1)	G12C, G12D, G12V, G13C, Q61H	0		4 (25.0)	G12C, G12D, G12V, G13C
MET	1 (0.5)	0		0		1 (6.3)	N375S
NOTCH1	2 (1.0)	1 (0.8)	V1671I	1 (2.0)	V1671I	0	
NRAS	3 (1.6)	3 (2.3)	Q61L	0		0	
NRF2	2 (1.0)	0		2 (4.1)	D29H, E79Q	0	
NTRK2	1 (0.5)	0		1 (2.0)	G261R	0	
PIK3CA	2 (1.0)	1 (0.8)	E545K	1 (2.0)	E545K	0	
PTEN	2 (1.0)	0		2 (4.1)	R223*	0	
RET	1 (0.5)	1 (0.8)	Exon 12 fusion with exon 15 of <i>KIF5B</i>	0		0	
STK11	1 (0.5)	1 (0.8)	Q37*L	0		0	
TP53	14 (7.3)	11 (8.6)	V157F, R158P, R248L, R273C, R282W	3 (6.1)	G245C, R158P, R273H	0	

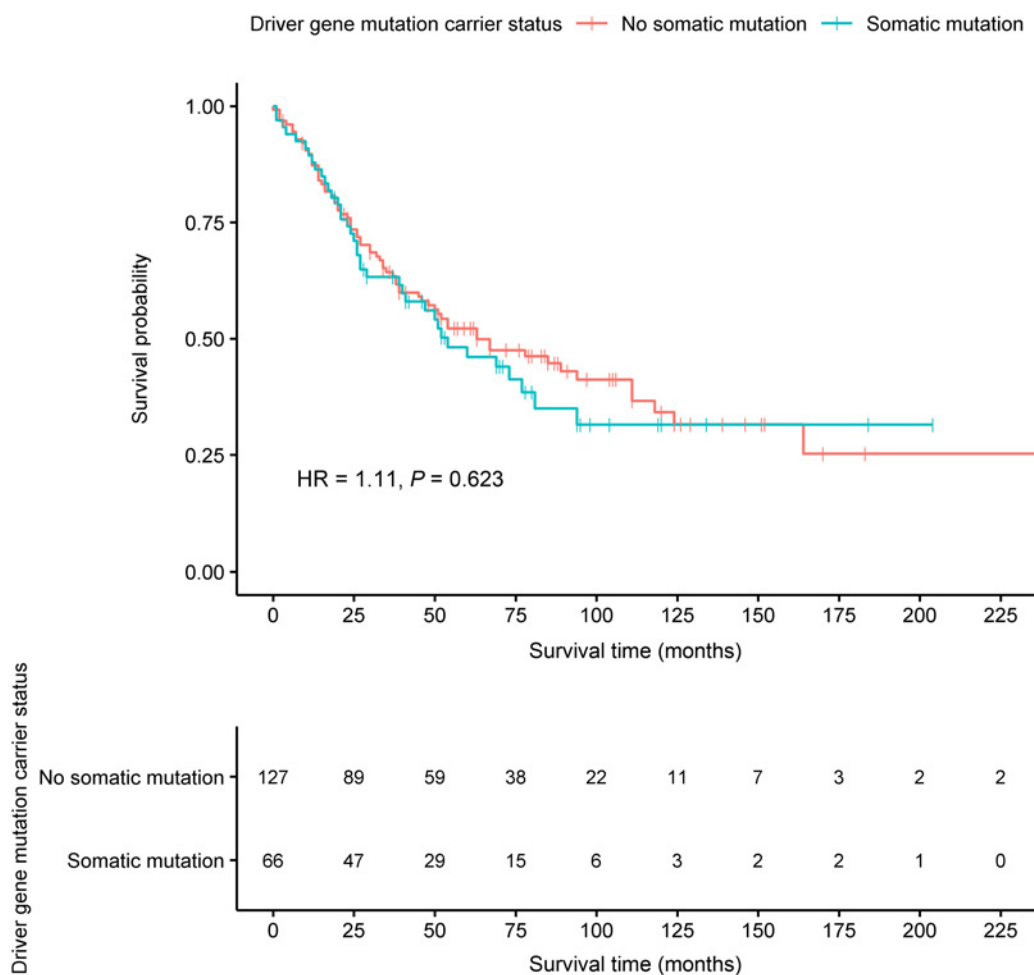


Figure 1. Driver gene mutation carrier status and overall survival for 193 African American NSCLC cases. Cox proportional hazards modeling was used to estimate differences in mortality risk by mutation carrier status, adjusted for stage and histology. Kaplan-Meier estimates were used to plot the survival curves for cases with and without mutations.

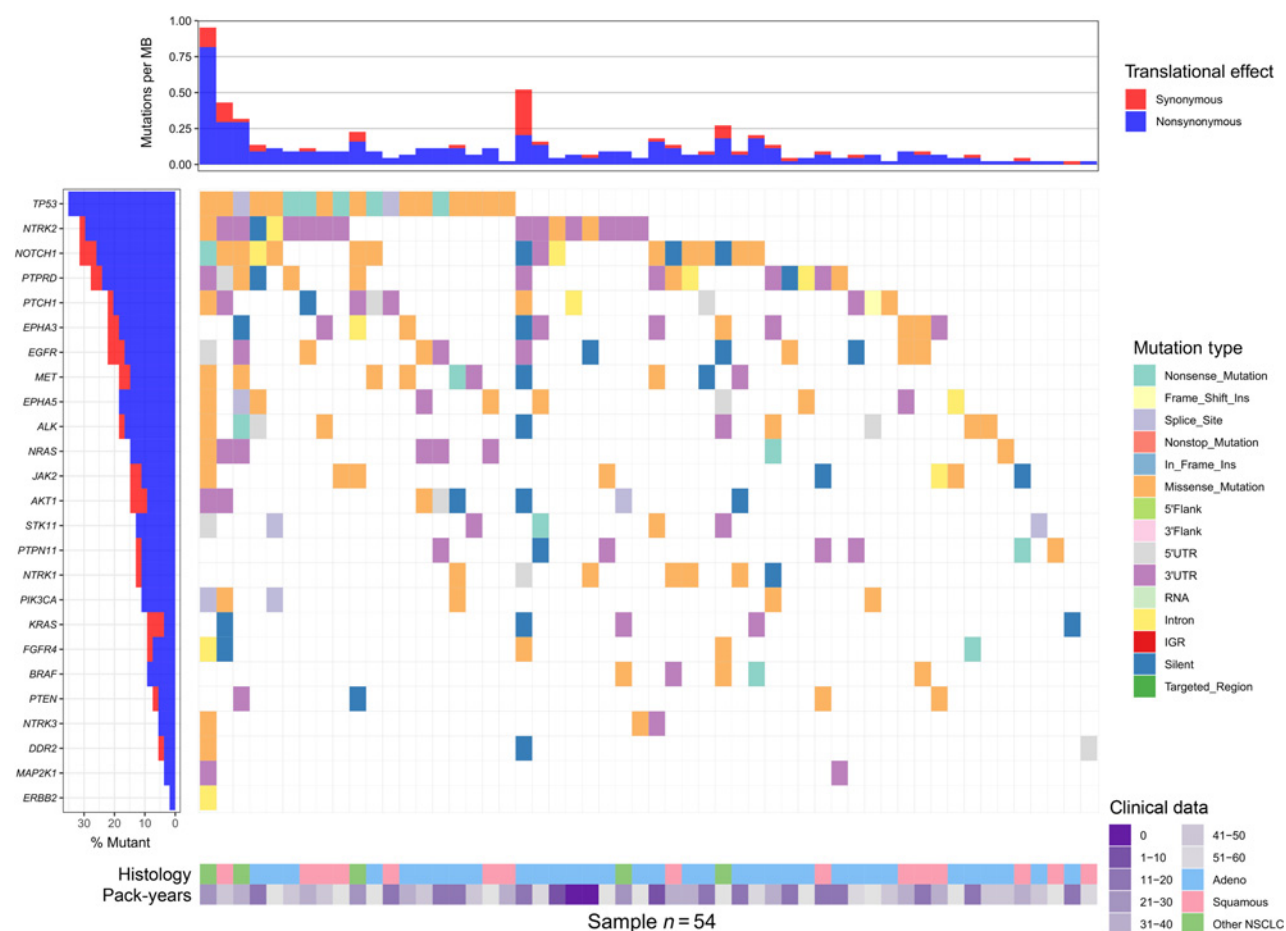


Figure 2.

Mutational landscape of 26 oncogenic driver genes among African American NSCLC cases with no previously identified driver mutations. Percent mutated samples for each gene are presented in the left panel; rate of mutations per sample is presented in the top panel. One LungCarta panel gene, *NFE2L2*, did not have additional mutations in this sample and thus is not represented in the figure; similarly, samples with no mutations in these genes are not represented in the figure ($N = 6$).

were most frequent, occurring in 25% of tumors. There were no statistically significant associations between the occurrence of known driver mutations (yes/no) and covariates, including gender, within histologic groups.

Exome sequencing and additional mutations in driver genes

From the 127 cases not carrying one of the tested genetic alterations, 67 of the highest quality samples were selected for whole-exome sequencing of both germline and tumor DNA. Among the 26 panel genes, seven of the 67 tumor/normal pairs sequenced were found to carry a driver mutation not recognized in the Sequenom assay (i.e., false negative on the panel, $n = 7$, 10.5%). Of the remaining 60 sequenced tumors, 54 (80.6%) were found to carry somatic exonic alterations in at least one of the known driver genes, including 34 adenocarcinomas, 15 squamous cell carcinomas, and 5 other NSCLC tumors (Supplementary Fig. S1). A total of 260 mutations were identified, of which 110 (42%) were nonsynonymous (Fig. 2). The majority of these tumors (29 of 54, 54%) carried more than one nonsynonymous somatic mutation in these genes, some with multiple mutations in the same gene, with no association between number of muta-

tions and pack-years of smoking. To estimate each mutation's likelihood of biological impact, we assessed the context of these mutations within their respective genes using PolyPhen (14). Of the 110 unique nonsynonymous mutations, 50 (45%) were predicted to be "probably" damaging by PolyPhen (Supplementary Table S3). In 34 adenocarcinomas, "probably" damaging mutations were identified in *AKT1*, *EPHA5*, *FGFR4*, *JAK2*, *MET*, *NOTCH1*, *NTRK1*, *NTRK2*, *PIK3CA*, *PTPRD*, *STK11*, and *TP53* (Supplementary Table S4). In 15 squamous cell carcinomas, "probably" damaging mutations were seen in *EPHA3*, *EPHA5*, *NOTCH1*, *NTRK1*, *PTEN*, *PTPRD*, and *TP53* (Supplementary Table S5).

Comparison with TCGA exome sequence data

These findings were compared with publicly available TCGA data, excluding samples that harbored one of the previously identified driver mutations. Of the 933 white or African American TCGA samples with sequence data, 38 (48%) of 79 African Americans (19 adenocarcinomas, 19 squamous cell carcinomas) and 427 (50%) of 854 whites (240 adenocarcinomas, 187 squamous cell carcinomas) did not have a previously identified driver

Table 3. Predicted damaging somatic mutations in common to both sequenced African American NSCLC cases and TCGA lung adenocarcinomas (LUAD) and lung squamous cell (LUSC) samples with no previously identified driver mutations in these genes

Gene	ID	Consequence	Detroit African Americans		TCGA		P
			Sample(s)	Frequency	Sample(s)	Frequency	
<i>MET</i>	chr7:g.116771969A>T	Y1003F	Adeno	1/37	LUAD whites	1/240	0.2543
<i>TP53</i>	chr17:g.7674241G>A	S241F	Adeno	1/37	LUAD whites	1/240	0.2543
<i>TP53</i>	chr17:g.7675151C>A	G154V	Adeno	1/37	LUAD, LUSC whites	1/240, 2/187	0.2543, 0.4244

mutation, ALK-positive expression, *FGFR1* amplification, or a *RET/ROS* fusion event. The relative frequency of mutations in 3 panel genes (*EPHA3*, *MET*, and *NOTCH1*) differed significantly between our adenocarcinoma sample and the TCGA adenocarcinoma sample (Supplementary Table S4); no significant differences were found in squamous cell samples (Supplementary Table S5).

Only 3 of 50 predicted damaging mutations within the 26 driver genes were also reported in TCGA samples in whites, and none were reported in TCGA African American samples (Table 3). The 47 novel "probably" damaging mutations unique to our African American cases are presented in Supplementary Table S6 and Supplementary Fig. S2. One of these mutations occurred more than once: *NTRK1* R692L. Most driver genes yielded too few nonsynonymous mutations to analyze individually with respect to outcomes. Among our 60 African American cases without a panel mutation, 29 carried at least one predicted damaging driver gene mutation. Survival was no different among these 29 cases when compared with the remaining 31 cases after adjusting for stage and histology (HR 1.66; 95% CI, 0.81–3.41; $P = 0.167$; Supplementary Fig. S3).

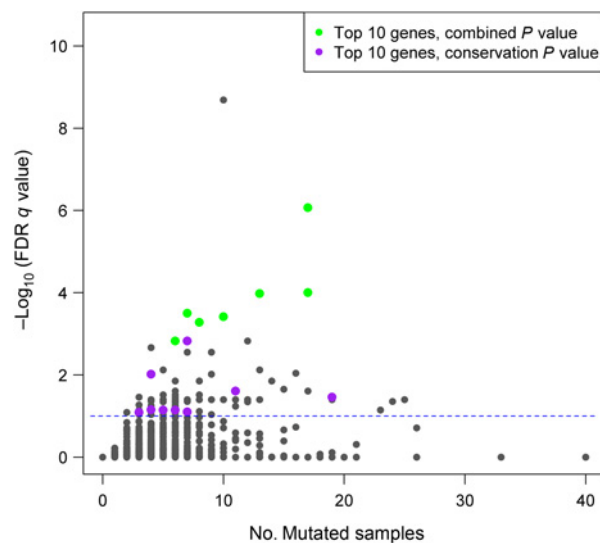


Figure 3. Analysis of significantly mutated genes by MutSig2CV among 60 African American NSCLC cases with no known driver mutations. Results for each of 13,079 genes are presented as significance according to MutSig2CV algorithm (FDR q value) relative to the number of samples carrying at least one mutation in the gene. Dashed line indicates FDR significance threshold ($q < 0.1$); 88 genes were significant after multiple test correction. The top 10 significantly mutated genes based on combined P value (abundance, clustering, and evolutionary conservation) and the top 10 significantly mutated genes based specifically on conservation are highlighted in green and purple, respectively.

Exome sequencing and novel genetic variation

To extend our investigation beyond known driver genes, we utilized MutSig2CV in the whole-exome sequence data to identify potential novel drivers in African Americans. Of the 13,079 genes analyzed by the algorithm, 88 genes were significantly mutated after multiple test correction (FDR q -value < 0.1 , Fig. 3). The significant genes included 5 known drivers reported in the COSMIC Cancer Gene Consensus (CGC; *EGFR*, *KEAP1*, *PABPC1*, *PMS2*, and *TP53*), 3 of which were reported in NSCLC tumors (*EGFR*, *KEAP1*, and *TP53*) and 2 of which were not (*PABPC1* and *PMS2*). When we compared our results with the TCGA pan-lung cancer analysis (16), only 4 genes were in common: *EGFR*, *KEAP1*, and *TP53*, known NSCLC drivers also reported in CGC, and the lysine methyltransferase gene *MLL3* (aka *KMT2C*).

Among the top 10 significantly mutated genes (based on combined P value, see Supplementary Table S7), 5 of these genes are either known drivers (*TP53*) or not functionally relevant to tumorigenesis (*TBC1D29*, *OR11H12*, *TCP10*, *OR52I2*, *VSIG10*). Of the 4 remaining genes [*CDC27*, *OXCT2*, *GSTM1*, *NHEDC1* (a.k.a *SLC9B1*)], there are a total of 57 unique nonsynonymous mutations across 30 cases, 10 of which are predicted damaging mutations (1 in each of 10 cases). These predicted damaging mutations characterize 16.7% of cases in our sample. Alternatively, the top 10 genes based on frequency of mutations in evolutionary conserved sites (functional significance P value) are *RHPN2*, *CRYGD*, *LLPH*, *PRIM1*, *RSPH4A*, *ZC3HC1*, *TP53*, *MLL3*, *CR2*, and *LRMP*. Excluding known drivers (*TP53*) and genes with unknown or biologically irrelevant protein products (*CRYGD*, *LLPH*), there are 7 remaining genes with 63 unique nonsynonymous mutations in 29 cases, of which 20 are predicted damaging mutations across 17 cases (3 cases carry 2 damaging mutations each, 14 cases carry 1 mutation each), or 28.3% of our sample.

Discussion

Outcomes after a lung cancer diagnosis have been uniformly poor until the identification of targetable driver mutations in selected patients, and more recently with immunotherapy. Systematic genomic testing for known driver mutations in *EGFR*, *KRAS*, *BRAF*, *PIK3CA*, *HER2*, and *ALK* in NSCLC is now standard practice and drives treatment decision-making. However, our knowledge of the frequency and therapeutic impact of test results is based on tumor and patient characteristics developed in primarily white populations. In this study, we show that African American patients with NSCLC are less likely than white patients to carry a known driver mutation based on testing using a defined panel, but upon sequencing, many more alterations can be identified in these genes and in novel genes that are predicted to be damaging.

Profiling of genetic alterations in adenocarcinomas and squamous cell carcinomas of the lung in TCGA highlights the genetic complexity of lung cancers compared with most other cancers (17–22), however few African Americans have been included

in these efforts (23). Of the 1,144 adenocarcinomas and squamous cell carcinomas profiled in TCGA, fewer than 10% of tumors originated in African Americans (TCGA data portal accessed on May 10, 2018). In a comprehensive review of the mutational landscape in 14 genes in 1,367 adenocarcinomas of the lung from the LCMC2, a driver oncogenic alteration was reported in 60% of these tumors (5); however, the publication did not include the racial breakdown of the sample.

Of those studies that included African Americans, most sample sizes are small, were limited to a defined set of driver mutations, and results show that African Americans are less likely to carry a known driver mutation (24–28). In an early study in 335 white and 137 African American patients with NSCLC, we reported that only 32% of African Americans carried a known driver mutation included on a lung cancer-specific panel compared with 41% of whites (24). We also participated in a study pooling African American samples from five centers that evaluated classic driver mutations in *EGFR*, *KRAS*, *NRAS*, *AKT1*, *PIK3CA*, *ERBB2*, and *MEK1* in 260 African American patients, as well as *ALK* translocations in nonsquamous tumors (27). Only 23.5% of tumors carried a known driver mutation, with most alterations occurring in *KRAS* and *EGFR*. Steuer and colleagues (6) evaluated a series of targeted mutations in 10 genes in 60 African Americans with metastatic adenocarcinoma in the LCMC2. Fifty-three percent of patients had mutations in at least one of the following genes: *KRAS*, *EGFR*, *ALK*, *ERBB2*, *BRAF*, *PIK3CA*, *MET*, *NRAS*, *MEK1*, and *AKT1*. In this series, *KRAS* and *EGFR* mutations were most frequent. Although this percentage is higher than what we found, the case series studied was restricted to late-stage adenocarcinomas; we did not restrict our sample in the same way. We report similar distributions of known driver mutations by histologic type in African Americans, with *EGFR*, *KRAS*, and *ALK* alterations seen primarily in adenocarcinomas, whereas *FGFR1* amplification was seen primarily, but not exclusively, in squamous cell carcinomas.

In an extended evaluation of 81 genes in 99 NSCLCs in African Americans, only 24 (24%) cases carried a classic driver mutation (27), with an additional 227 nonsilent variants identified in these 81 genes. These results are similar to our results suggesting that previously identified mutations in driver genes, characterized in mostly white patient populations and assessed as part of panels, are less frequent in African American patients. However, additional sequencing of these genes showed another 50% of samples carried a predicted damaging mutation that was not included on the panel.

Within the set of panel genes included in this study, we identified three nonsynonymous, "probably damaging" mutations in African American NSCLC cases that had also been reported only in whites in TCGA. Two were *TP53* mutations in adenocarcinomas occurring in 5.4% of African Americans in our study versus 0.8% of adenocarcinomas and 1.1% of squamous cell carcinomas in whites in TCGA. *TP53* is frequently mutated in lung cancer, yet it is not associated with a potential targeted treatment. A *MET* mutation that occurred in 1 (2.7%) African American with adenocarcinoma also was reported in 1 (0.3%) white patient with adenocarcinoma in TCGA. The usefulness of these data for treatment decision making is unknown and needs confirmation.

Several of the previously unreported 47 nonsynonymous, "probably" damaging mutations in panel genes we identified in African Americans might serve as targets for treatment. The two novel *ALK* mutations occurred in nonsquamous, non-

adenocarcinomas, as did a *BRAF* mutation and 2 of 4 *MET* mutations. Six (12.8%) of the 47 mutations were in *NTRK* genes and 9 (19.1%) were in *TP53*, findings that were not limited to any one histologic type. These findings suggest that using histology as a decision point for targeted sequencing or using limited mutation panels in African Americans might miss potentially damaging genetic alterations and lead to suboptimal treatment decisions.

We note that the LungCarta panel was developed based on significantly mutated genes identified by Ding and colleagues (9) and therefore is not a comprehensive profile of NSCLC driver mutations. The low rate of *TP53* mutations in our sample (~7%) is evidence of the limited scope of this panel. However, we also found a relatively lower rate of *TP53* somatic mutations in the subset of sequenced cases (28%) compared with TCGA samples without panel mutations (80%–90%). This is consistent with the observation that the genetic profile in African American cases differs from other published data. Although the tumor samples that were sequenced were of higher DNA quality and quantity compared with those that were not sequenced, we did not find any significant differences between exome-sequenced ($N = 67$) and nonsequenced ($N = 60$) cases with respect to stage ($P = 0.267$), histology ($P = 0.675$), date of diagnosis ($P = 0.362$), or overall survival (HR 1.02; 95% CI, 0.61–1.57; $P = 0.925$).

In addition to novel mutations in known driver genes, we also identified 88 genes that were significantly mutated in African American NSCLC cases with no known driver mutations, 83 of which were novel. Of the 5 known drivers found in COSMIC CGC, *EGFR*, *KEAP1*, and *TP53* have been previously associated with NSCLC (29). Those not previously associated with lung cancer, *PABPC1* and *PMS2*, are potential candidates for further study. Mutations in the poly(A) binding protein gene *PABPC1* have been reported in head and neck squamous cell carcinoma (HNSCC) and biliary tract carcinoma (29). Germline variants in *PMS2*, which encodes a protein involved in DNA mismatch repair, have been associated with multiple cancer types including colorectal, ovarian, endometrial, and glioma (29). Somatic mutations in *PMS2* may increase sensitivity to DNA-damaging agents and are currently under investigation in combination with PARP inhibitors (<https://clinicaltrials.gov>). Further investigation of these genes in lung cancer is warranted.

In further evaluation of the top 10 most significant genes based on the combined P value, we excluded from further consideration *TP53*, and those genes with little biological relevance to cancer, including the pseudogene *TBC1D29*, *OR11H12*, *TCP10*, *OR52I2*, and *VSIG10*, leaving 4 candidate driver genes for consideration. The *CDC27* gene product is a cell-cycle regulator that interacts with mitotic checkpoint proteins. This gene has been implicated in lung adenocarcinoma and gastric cancer (30, 31). Knockdown and overexpression studies show that this gene likely enhances tumor cell migration and invasion (31). Somatic mutations were reported in two lung adenocarcinomas in an exome sequencing study of 16 never-smoking patients with lung adenocarcinoma, although no mutations were detected in an additional set of 54 patients (32). The CoA-transferase enzyme (*OXC2*) is involved in the breakdown of fatty acids to ketone bodies, with no characterized role in lung cancer. *GSTM1* has been studied as a lung cancer susceptibility gene due to its involvement in the detoxification of tobacco smoke. Germline variants have been associated with lung cancer risk in numerous populations (33). Molecular studies also report enhanced platinum sensitivity in lung cancer

cell lines when targeting *GSTM1* with analogue inhibitors suggesting a role in multidrug resistance pathways (34). Finally, *SLC9B1* knockdown studies suggest the cation/proton antiporter family plays a pivotal role in ion homeostasis with respect to lung cancer, many different solute carrier members have been targets of drug therapies (35, 36).

We also examined the top 10 most significant genes based on frequency of mutations in evolutionary conserved sites (functional *P* value). There were seven potential novel driver genes after excluding *TP53*, *CRYGD*, and *LLPH* from further consideration. The *RHPN2* gene product raphophilin 2 can activate RhoA, which when overexpressed has been shown to drive mesenchymal transformation in malignant cell lines (37), and the gene was significantly mutated in a sample of 335 Chinese lung adenocarcinoma tumors (38). The DNA primase gene *PRIM1* encodes a protein involved in DNA replication and repair, with reported differential expression in lung adenocarcinoma tissue versus adjacent normal tissue (39). Molecular studies demonstrated that depletion of *PRIM1* in combination with *ATR* deficiency resulted in S-phase stasis and synthetic lethality in cancer cell lines (40). The radial spoke head component protein encoded by *RSPH4A* is involved in cilia motility and mucociliary clearance and is highly expressed in lung tissue (41), although no studies have previously linked mutations in this gene to lung cancer. The zinc finger gene product encoded by *ZC3HC1* is involved in cell division with no studies investigating associations between the gene and lung cancer. The methyltransferase gene *MLL3* is an interesting candidate driver gene; it was also significantly mutated in TCGA pan-lung cancer samples (16) and has been implicated in other cancers such as leukemia and gastric cancer. The *MLL* family of genes are the target of phase I trials in hematologic malignancies (<https://clinicaltrials.gov>). The complement receptor *CR2* is highly expressed in lymph node tissue and is involved in B-cell activation. Although this gene has not been previously implicated in lung tumorigenesis, the correlation between inflammatory processes and lung cancer is evidenced by the increased risk of lung cancer conferred by COPD. A nonsynonymous germline variant (V141L, rs7969931) in the lymphoid restricted membrane protein (*LRMP*) gene was associated with survival in a cohort of 361 patients with lung adenocarcinoma from Italy (42). Other treatment relevant mutations identified include those within *KEAP1* and *AKAP4*. *KEAP1* is the target of a phase II trial in advanced stage lung cancer with a novel mTOR inhibitor (<https://clinicaltrials.gov>). *AKAP4* gene silencing in tumor cell lines showed markedly reduced cell proliferation and invasion *in vitro* and *in vivo* when implanted in mice (43). *AKAP4* mutations

are being utilized as part of a neo-antigen panel in a dendritic cell therapy in early-stage trials for solid tumors (<https://clinicaltrials.gov>).

In conclusion, the frequency of actionable mutations in a limited, established panel of driver genes is lower in African Americans compared with whites. NSCLC in African Americans patients may contain alternate pathogenic (damaging) mutations in driver genes that may be effective targets for therapeutic interventions. Furthermore, we identified previously uncharacterized genes that are candidate driver genes in African Americans. The significance of these alternative mutations and approaches to personalized treatment planning in diverse populations warrants further investigation.

Disclosure of Potential Conflicts of Interest

S. Gadgeel is a consultant/advisory board member for Genentech/Roche, Takeda, AstraZeneca, Pfizer, Abbvie, Xcovery, Boehringer-Ingelheim, and Bristol-Myers Squibb. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: C.M. Lusk, D. Watzka, S. Gadgeel, A.G. Schwartz

Development of methodology: C.M. Lusk, D. Watzka, V. Ratliff, A. Bollig-Fischer, G. Bepler, A.G. Schwartz

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S. Gadgeel, A.G. Schwartz

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): C.M. Lusk, D. Watzka, G. Dyson, D. Craig, F. Lonardo, G. Bepler, K. Purrington, S. Gadgeel, A.G. Schwartz

Writing, review, and/or revision of the manuscript: C.M. Lusk, D. Watzka, G. Dyson, D. Craig, F. Lonardo, A. Bollig-Fischer, G. Bepler, K. Purrington, S. Gadgeel, A.G. Schwartz

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): C.M. Lusk, D. Craig, A.S. Wenzlaff, A.G. Schwartz

Study supervision: A.G. Schwartz

Acknowledgments

This research was supported by the NIH grants/contracts (R21CA184778, R01CA141769, P30CA022453, HHSN261201300011L, T32CA009531) and the Herrick Foundation.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 27, 2018; revised December 7, 2018; accepted April 5, 2019; published first April 12, 2019.

References

- American Cancer Society. Cancer Facts & Figures 2018. Atlanta, GA: American Cancer Society; 2018.
- SEER Cancer Stat Facts [cited 2018 Mar 7]. Available from: <http://seer.cancer.gov/statfacts/html>.
- Raparia K, Villa C, DeCamp MM, Patel JD, Mehta MP. Molecular profiling in non-small cell lung cancer: a step toward personalized medicine. *Arch Pathol Lab Med* 2013;137:481–91.
- Shames DS, Wistuba II. The evolving genomic classification of lung cancer. *J Pathol* 2014;232:121–33.
- Aisner DL, Sholl LM, Berry LD, Rossi MR, Chen H, Fujimoto J, et al. The impact of smoking and TP53 mutations in lung adenocarcinoma patients with targetable mutations—the lung cancer mutation consortium (LCMC2). *Clin Cancer Res* 2018;24:1038–47.
- Steuer CE, Behera M, Berry L, Kim S, Rossi M, Sica G, et al. Role of race in oncogenic driver prevalence and outcomes in lung adenocarcinoma: results from the lung cancer mutation consortium. *Cancer* 2016;122:766–72.
- Schwartz AG, Cote ML, Wenzlaff AS, Land S, Amos CI. Racial differences in the association between SNPs on 15q25.1, smoking behavior, and risk of non-small cell lung cancer. *J Thorac Oncol* 2009;4:1195–201.
- Schwartz AG, Lusk CM, Wenzlaff AS, Watzka D, Pandolfi S, Mantha L, et al. Risk of lung cancer associated with COPD phenotype based on quantitative image analysis. *Cancer Epidemiol Biomarkers Prev* 2016;25:1341–7.
- Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 2008;455:1069–75.

10. Wijesinghe P, Beppler G, Bollig-Fischer A. A mass spectrometry assay to simultaneously analyze ROS1 and RET fusion gene expression in non-small-cell lung cancer. *J Thorac Oncol* 2015;10:381–6.
11. Heist RS, Mino-Kenudson M, Sequist LV, Tammireddy S, Morrissey L, Christiani DC, et al. FGFR1 amplification in squamous cell carcinoma of the lung. *J Thorac Oncol* 2012;7:1775–80.
12. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013;31:213–9.
13. Cai L, Yuan W, Zhang Z, He L, Chou KC. In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. *Sci Rep* 2016;6:36540.
14. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013;7:Unit7 20.
15. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013;499:214–8.
16. Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Pedamallu CS, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet* 2016;48:607–16.
17. Pan Y, Wang R, Ye T, Li C, Hu H, Yu Y, et al. Comprehensive analysis of oncogenic mutations in lung squamous cell carcinoma with minor glandular component. *Chest* 2014;145:473–9.
18. Wang R, Pan Y, Li C, Zhang H, Garfield D, Li Y, et al. Analysis of major known driver mutations and prognosis in resected adenosquamous lung carcinomas. *J Thorac Oncol* 2014;9:760–8.
19. Wright GM, Do H, Weiss J, Alam NZ, Rathi V, Walkiewicz M, et al. Mapping of actionable mutations to histological subtype domains in lung adenocarcinoma: implications for precision medicine. *Oncotarget* 2014;5:2107–15.
20. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012;489:519–25.
21. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 2012;150:1107–20.
22. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014;511:543–50.
23. Spratt DE, Chan T, Waldron L, Speers C, Feng FY, Ogunwobi OO, et al. Racial/ethnic disparities in genomic sequencing. *JAMA Oncol* 2016;2:1070–4.
24. Bollig-Fischer A, Chen W, Gadgeel SM, Wenzlaff AS, Cote ML, Schwartz AG, et al. Racial diversity of actionable mutations in non-small cell lung cancer. *J Thorac Oncol* 2015;10:250–5.
25. Heon S, Johnson BE. Adjuvant chemotherapy for surgically resected non-small cell lung cancer. *J Thorac Cardiovasc Surg* 2012;144:S39–42.
26. Reinersman JM, Johnson ML, Riely GJ, Chitale DA, Nicastrì AD, Soff GA, et al. Frequency of EGFR and KRAS mutations in lung adenocarcinomas in African Americans. *J Thorac Oncol* 2011;6:28–31.
27. Araujo LH, Lammers PE, Matthews-Smith V, Eisenberg R, Gonzalez A, Schwartz AG, et al. Somatic mutation spectrum of non-small-cell lung cancer in African Americans: a pooled analysis. *J Thorac Oncol* 2015;10:1430–6.
28. Bauml J, Mick R, Zhang Y, Watt CD, Vachani A, Aggarwal C, et al. Frequency of EGFR and KRAS mutations in patients with non-small cell lung cancer by racial background: do disparities exist? *Lung Cancer* 2013;81:347–53.
29. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer* 2018;18:696–705.
30. Bidkhori G, Narimani Z, Hosseini Ashtiani S, Moeini A, Nowzari-Dalini A, Masoudi-Nejad A. Reconstruction of an integrated genome-scale co-expression network reveals key modules involved in lung adenocarcinoma. *PLoS One* 2013;8:e67552.
31. Xin Y, Ning S, Zhang LCui M. CDC27 facilitates gastric cancer cell proliferation, invasion and metastasis via twist-induced epithelial-mesenchymal transition. *Cell Physiol Biochem* 2018;50:501–11.
32. Ahn JW, Kim HS, Yoon JK, Jang H, Han SM, Eun S, et al. Identification of somatic mutations in EGFR/KRAS/ALK-negative lung adenocarcinoma in never-smokers. *Genome Med* 2014;6:18.
33. Yu P, Kusuma JD, Suarez MAR, Pamela Koong Shiao SY. Lung cancer susceptibility from GSTM1 deletion and air pollution with smoking status: a meta-prediction of worldwide populations. *Oncotarget* 2018;9:31120–32.
34. Wang CH, Wu HT, Cheng HM, Yen TJ, Lu IH, Chang HC, et al. Inhibition of glutathione S-transferase M1 by new gabosine analogues is essential for overcoming cisplatin resistance in lung cancer cells. *J Med Chem* 2011;54:8574–81.
35. Chintapalli VR, Kato A, Henderson L, Hirata T, Woods DJ, Overend G, et al. Transport proteins NHA1 and NHA2 are essential for survival but have distinct transport modalities. *Proc Natl Acad Sci USA* 2015;112:11720–5.
36. Lin L, Yee SW, Kim RB, Giacomini KM. SLC transporters as therapeutic targets: emerging opportunities. *Nat Rev Drug Discov* 2015;14:543–60.
37. Danussi C, Akavia UID, Niola F, Jovic A, Lasorella A, Pe'er D, et al. RHPN2 drives mesenchymal transformation in malignant glioma by triggering RhoA activation. *Cancer Res* 2013;73:5140–50.
38. Wu K, Zhang X, Li F, Xiao D, Hou Y, Zhu S, et al. Frequent alterations in cytoskeleton remodelling genes in primary and metastatic lung adenocarcinomas. *Nat Commun* 2015;6:10131.
39. Xu H, Ma J, Wu J, Chen L, Sun F, Qu C, et al. Gene expression profiling analysis of lung adenocarcinoma. *Braz J Med Biol Res* 2016;49:e4861.
40. Job A, Schmitt LM, von Wenserski L, Lankat-Buttgereit B, Gress TM, Buchholz M, et al. Inactivation of PRIM1 function sensitizes cancer cells to ATR and CHK1 inhibitors. *Neoplasia* 2018;20:1135–43.
41. Frommer A, Hjeij R, Loges NT, Edelbusch C, Jahnke C, Raidt J, et al. Immunofluorescence analysis and diagnosis of primary ciliary dyskinesia with radial spoke defects. *Am J Respir Cell Mol Biol* 2015;53:563–73.
42. Manenti G, Galbiati F, Pettinicchio A, Spinola M, Piconese S, Leoni VP, et al. A V141L polymorphism of the human LRMP gene is associated with survival of lung cancer patients. *Carcinogenesis* 2006;27:1386–90.
43. Saini S, Agarwal S, Sinha A, Verma A, Parashar D, Gupta N, et al. Gene silencing of A-kinase anchor protein 4 inhibits cervical cancer growth *in vitro* and *in vivo*. *Cancer Gene Ther* 2013;20:413–20.