The Oncologist, 2022, **27**, 476–486
https://doi.org/10.1093/oncolo/oyac035
Advance access publication 17 March 2022
Original Article

Targeted Next-Generation Sequencing Reveals Exceptionally High Rates of Molecular Driver Mutations in Never-Smokers With Lung Adenocarcinoma

Philip C. Mack[1,‡,*], Michael I. Klein[2,‡], Kristin L. Ayers[2], Xiang Zhou[2], Sunny Guin[2], Marc Fink[2], Michael Rossi[2], Hussam Al-Kateb[2], Timmy O'Connell[2], Feras M. Hantash[1,2], William K. Oh[1,2], Scott Newman[2], Eric E. Schadt[1,2,#], Rong Chen[1,2,#], Fred R. Hirsch[1,#]

[1]Center for Thoracic Oncology, Tisch Cancer Institute, Icahn School of Medicine, Mount Sinai, New York, NY, USA
[2]Sema4, Stamford, CT, USA

*Corresponding author: Philip C. Mack, Center for Thoracic Oncology, Tisch Cancer Institute, Mount Sinai Health System, One Gustave L. Levy Place, New York, NY 10029, USA. Tel: 212 241-0776; Email: philip.mack@mssm.edu
‡Co-first authors contributed equally to this work.
#Co-senior authors.

Abstract

Introduction: Historically, high rates of actionable driver mutations have been reported in never-smokers with lung adenocarcinoma (ADC). In the era of modern, comprehensive cancer mutation sequencing, this relationship necessitates a more detailed analysis.

Methods: All Mount Sinai patients between January 1, 2015, and June 1, 2020, with a diagnosis of ADC of any stage with known smoking status who received genomic testing were included. Most patients were analyzed using the Sema4 hotspot panel or the Oncomine Comprehensive Assay version 3 next-generation sequencing (NGS) panel conducted at Sema4. Patients were considered *fully genotyped* if they were comprehensively analyzed for alterations in *EGFR*, *KRAS*, *MET*, *ALK*, *RET*, *ROS1*, *BRAF*, *NTRK1-3*, and *ERBB2*, otherwise they were considered *partially genotyped*.

Results: Two hundred and thirty-six never-smokers and 671 smokers met the above criteria. Of the never-smokers, 201 (85%) had a driver mutation with 167 (71%) considered actionable (ie, those with US Food and Drug Administration-approved agents). Among smokers, 439 (65%) had an identified driver mutation with 258 (38%) actionable ($P < .0001$). When comprehensively sequenced, 95% (70/74) of never-smokers had a driver mutation with 78% (58/74) actionable; whereas, for smokers, 75% (135/180) had a driver with only 47% (74/180) actionable ($P < .0001$). Within mutations groups, EGFR G719X and KRAS G12Cs were more common to smokers. For stage IV patients harboring EGFR-mutant tumors treated with EGFR-directed therapies, never-smokers had significantly improved OS compared to smokers (hazard ratio = 2.71; $P = .025$). In multivariable analysis, Asian ancestry and female sex remained significant predictors of (1) OS in stage IV patients and (2) likelihood of harboring a receptor of fusion-based driver.

Conclusion: Comprehensive NGS revealed driver alterations in 95% of never-smokers, with the majority having an associated therapy available. All efforts should be exhausted to identify or rule out the presence of an actionable driver mutation in all metastatic lung ADC.

Key words: lung adenocarcinoma; lung cancer in never smokers; precision oncology; survival differences.

Implications for Practice

Driver mutations in lung cancer are often treatable with highly specific targeted therapies, typically associated with favorable clinical outcomes. Here, a modern, comprehensive assessment of lung adenocarcinoma tumor molecular profiles revealed that 95% of never-smokers had a detectable driver mutation, with 78% treatable with a US Food and Drug Administration-approved targeted therapy. Demographic factors, including age, ethnicity, and sex, were also evaluated by statistical modeling, demonstrating contributions to overall survival.

Introduction

A defining characteristic of lung adenocarcinoma (ADC) is the frequency and diversity of driver mutations that can be used to guide targeted therapy in metastatic disease.[1,2] In this context, a driver mutation can be described as a single oncogenic activating alteration that the tumor is exquisitely dependent on for its growth and survival. In ADC, driver mutations are currently limited to oncogenically activated signaling intermediaries and kinases, including receptor tyrosine kinases (RTKs) and novel fusion events.[3,4] A continuously expanding array of targeted agents are US Food and Drug Administration (FDA)-approved for the treatment of

Received: 24 November 2021; Accepted: 16 January 2022.

patients with driver-positive ADC, including inhibitors of aberrant *EGFR*, *ALK*, *ROS1*, *RET*, *BRAF*, *MET*, and *NTRK1*. In ADC, KRAS, and potentially other members of the MAPK pathway such as *NRAS*, *HRAS*, and *MEK1* activating mutations, are also classifiable as driver mutations, evidenced by in vitro studies and the observation that they tend to be mutually exclusive with other drivers.[3,4] *KRAS* mutations have been generally refractory to direct inhibition strategies, with the notable exception of the G12C amino acid substitution, associated with tobacco smoking, which has targeted agents recently FDA approved and in clinical trials.[5-8]

Within the ADC histology, patients with tumors harboring non-MAPK pathway driver mutations (RTK mutations and fusions) are typically associated with a set of clinical demographics consisting of light or never-smoking, female gender, Asian descent, and younger age.[9,10] Further, these tumors tend to have limited mutation burdens and are often less responsive to immune therapies. It is well understood that a history of smoking is a prognostic indicator of poor outcome, with the suggestion that non–small cell lung cancer (NSCLC) arising in never-smokers may be pathologically distinct from typical tumors associated with heavy smoking.[11-16] With the myriad of targeted treatments now available to non-squamous cases, outcomes of patients are increasingly dependent on their molecular eligibility for effective regimens. Notwithstanding the long-understood association between smoking and lung cancer, there remains a paucity of literature evaluating smoking history in the context of next-generation sequencing (NGS) of lung cancer-associated mutations and drivers. With the introduction of routine, comprehensive genomic sequencing conducted systematically across all patients with ADC, a more accurate depiction of the diversity of driver mutations, including rare variants and fusions, can provide an updated evaluation of the relationship between tumor molecular subtypes and patient smoking histories. Here, we investigated in detail the distinctions between never-smokers and smokers, focusing on the role and relationship between oncogenic drivers, demographics, and outcomes, using a well-annotated, real-world data set collected at the high-diversity Mount Sinai Health System (MSHS) in New York City, NY.

Patients and Methods

Patients

From a consecutive population of 2081 lung cancer patients treated at MSHS between January 1, 2015, and June 1, 2020, a cohort of 262 never-smokers and 764 smokers with a diagnosis of ADC of any stage was identified (Supplementary Fig. S1). Smoking status and race/ethnicity were self-reported. Patients were excluded if they met any of the following criteria: (1) patient had multiple cancer diagnoses, including multiple lung cancer diagnoses, dating back to 2010, (2) stage was not reported and could not be imputed from clinical or pathological TNM scores, and (3) smoking status could not be determined. Patients were defined as *never-smokers* if they reported as a never-smoker in at least one visit and did not ever report any history of smoking. Patients were categorized as current or former *smokers* if they reported any smoking history, including passive and light smoking, in any hospital visit. Approximately 4% of lung ADC patients did not have any smoking status reported and were excluded.

The stage at diagnosis was directly reported in progress notes in 59% (730/1237) of single-cancer ADC patients. We imputed stage using clinical/pathological TNM values in an additional 25% (313/1237) of patients who did not have stage directly reported but did have either clinical or pathological TNM reported. The eighth edition of TNM mapping[17] was used for patients diagnosed after 2017, and the seventh edition of TNM mapping[18] was used for patients diagnosed before 2017. The remaining 194 patients for whom neither stage nor TNM values were reported were excluded. When possible, molecular profiling was performed in material immediately following a diagnosis of advanced disease.

All clinical and demographic data were obtained from MSHS Electronic Medical Record (*EMR*) *databases*. Original EMR data from the hospital data warehouse were processed using Sema4 Centrellis platform, consisting of natural language processing and machine learning-based automated abstraction engine, curation platform for manual review, patient journey, and cohort builder. Comprehensive patient journey was curated and integrated from both structured data and unstructured clinical notes using an automated abstraction engine and manual review.[19]

Study permission was granted by the Mount Sinai IRB. The procedures followed in this study were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration.

Genomic Testing

The vast majority of genomic testing was conducted using tissue-based Sema4 tests, with some (<10%) through external vendors using liquid biopsy assays such as Guardant Health 360 assay or Foundation One. External testing results were extracted from patient notes, whereas Sema4 test results were queried directly from the Sema4 genomic testing databases. From January 2015 until August 2018, the most commonly prescribed genetic test was the Ion AmpliSeq Cancer Hotspot Panel v2 with testing performed by Sema4. This panel evaluates 50 cancer genes for common "hotspot" mutations. The hotspot panel was routinely prescribed along with single-gene testing for *ALK* and *ROS1* fusions, since structural rearrangements could not be identified using the hotspot panel. From September 2018 onward, the Oncomine Comprehensive Assay version 3 (OCAv3) NGS panel (Oncomine, Thermo Fisher Scientific, Waltham, Massachusetts, USA) offered by Sema4 replaced the Sema4 hotspot panel as the routinely prescribed genetic test as it covers 161 cancer genes and is able to detect structural and copy number changes in addition to hotpot and other coding/splice mutations. This NGS assay is expected to capture all point mutations and small insertions and deletions in all the guideline lung cancer driver genes, with complete coverage of all key exons. Mutation burden was estimated in patients receiving the more comprehensive OCAv3 test by counting the number of variants discovered during testing that passed recommended quality control thresholds but were classified as variants of uncertain significance (VUS). Recent studies have found evidence that gene panels may be a reasonable proxy to mutational burden computed from whole-genome sequencing.[20]

Statistical Methodology

The baseline characteristics tables report the group-wise comparison *P*-values for the chi-square test for categorical variables, the *t*-test with equal variance for the continuous variables, and were created by the R package tableone

v0.12.0. The standardized mean difference is also reported. *P*-values and odds ratios (ORs) for contingency tables (2 × 2 tests for comparisons of percentages between the smoker and never-smoker group) were computed with Fisher's exact test. Logistic regression was applied to compute ORs for multivariable models. For the multivariable models, we considered smoking status, age, sex, and race. Subset analysis within both the fully and partially genotyped individuals was performed to confirm trends in the whole population were not driven by genetic testing biases.

For survival analysis, patients were followed longitudinally from their first diagnosis of Stage IV lung cancer until death or their last visit prior to the end of the study date (August 29, 2020) using a diagnosis cutoff of August 29, 2019, to ensure at least 1 year of potential follow-up and were required to contribute a minimum of 30 days of follow-up after diagnosis. The median follow-up time is reported as the known function time, that is, the time from diagnosis to the last visit date for the censored individuals, or the time from diagnosis to end of study date for those experiencing the event (death). The completion index, a measure of the completeness of the follow-up, is the ratio of the total observed time versus the potential time of follow-up over each individual in a study.[21] Kaplan-Meier curves were created in the R survival package (v3.2-7) and report the overall survival (OS) in months. Statistical comparisons between the never smoker and smoker groups were reported using the global log-rank test *P*-value with the hazard ratio (HR) estimated from Cox regression. Cox regression was performed for multi-variable models and the HR and Wald test *P*-value for each variable were reported. Fixed time-point survival rates, 95% confidence intervals (CIs), and log test *p*-values were computed in the R package ComparisonSurv v1.0.9.

Results

Patient Demographics and Characteristics

The distributions of gender, ethnicity, and age between the former/current smoker group and the never-smoker group were evaluated (Table 1). Within the never-smokers cohort, 75.9% of patients ($n = 262$) were female compared to 50.4% ($n = 385$) in smokers ($P < .001$). Never-smokers were enriched for patients of Asian or Hispanic/Latino descent; 5.2% of smokers were Asian, compared to 24.4% of never-smokers, and Hispanic/Latino represented 9.2% of the smokers and 14.9% of the never-smokers ($P < .001$). The median age at diagnosis was not statistically different between smokers and never-smokers, with medians of 68.5 and 68.3 years old,

Table 1. Characteristics of the full cohort ($n = 1026$) for the never-smoker versus the smoker population.

	Never smoker	Former or current smoker	*P*	SMD
n	262	764		
Age at diagnosis, years, median (IQR)	68.29 (59.74-75.95)	68.54 (61.37-75.15)	.286	0.072
Age categorical, *n* (%)			**.007**	0.204
<45 years	15 (5.7)	16 (2.1)		
45-75 years	183 (69.8)	580 (75.9)		
76+ years	64 (24.4)	168 (22.0)		
Gender, *n* (%)			**<.001**	0.55
Male	63 (24.0)	379 (49.6)		
Female	199 (75.9)	385 (50.4)		
Race/ethnicity, *n* (%)			**<.001**	0.631
Asian	64 (24.4)	40 (5.2)		
Black/African American	37 (14.1)	139 (18.2)		
Hispanic or Latino	39 (14.9)	70 (9.2)		
Other/unknown	25 (9.5)	111 (14.5)		
White	97 (37.0)	404 (52.9)		
Stage at diagnosis, *n* (%)			.597	0.099
1	123 (46.9)	364 (47.6)		
2	17 (6.5)	57 (7.5)		
3	23 (8.8)	83 (10.9)		
4	99 (37.8)	260 (34.0)		
Genetic testing received, *n* (%)			.251	0.118
Fully genotyped	74 (28.2)	180 (23.6)		
No genotyping	26 (9.9)	93 (12.2)		
Partially genotyped	162 (61.8)	491 (64.3)		
Year of diagnosis, *n* (%)			.014	0.212
2015-2016	90 (34.4)	341 (44.6)		
2017-2018	112 (42.7)	277 (36.3)		
2019-2020	60 (22.9)	146 (19.1)		

Bold values are less than .05.

respectively (Table 1). However, upon classifying patients as under 45 years old, 45-75 years old, and over 75 years old, we found an enrichment in the under 45 group in the never-smokers, 5.7% compared to 2.1% of smokers ($P = .007$). No statistical differences were observed in stage at diagnosis between smokers versus never-smokers.

Genomic Comparison of Never-Smokers Versus Smokers

Molecular testing was documented in 236/262 never-smokers and 671/764 smokers. Patients were considered *fully genotyped* if their genetic testing encompassed *KRAS*, *EGFR*, *ALK*, *ROS1*, *MET*, *BRAF*, *ERBB2*, *RET,* and *NTRK1*, and were considered *partially genotyped* if they were tested for just a subset of these alterations. No statistical difference was observed in the genetic testing rates between smokers and never-smokers (Table 1). Out of 907 patients who received genetic testing, 254 (28%) were fully genotyped, and the remaining 653 (72%) were partially genotyped.

Driver abnormalities were categorized into **Tiers** according to the availability of targeted therapeutics (Table 2). **Tier 1** consists of driver mutations with a matched, FDA-approved targeted treatment in lung cancer. **Tier 2** consists of drivers targetable by agents in advanced phase III clinical investigation with potential for near-future approval in lung cancer. **Tier 3** consists of drivers that are not clinically targetable with current treatments, but with agents in development. In our cohort, **Tier 3** drivers consisted of activating mutations in *KRAS* non-G12C, *NRAS, HRAS*, and *BRAF* non-V600E.

Frequencies of specific driver mutations were compared relative to smoking status (Fig. 1). The distribution of *KRAS*, *EGFR*, and *ALK* abnormalities were statistically different based on smoking status, with *KRAS* mutations more common to smokers, and *EGFR* and *ALK* more common to never-smokers (Fig. 1A). Tier 1 *EGFR* mutations were found in 55.5% of never-smokers versus 17.0% of smokers (P < .001). ALK fusions were found in 6.4% ($n = 15/236$) of never smokers versus 1.6% ($n = 11/671$) of smokers ($P < .001$). Abnormalities in *ROS1*, *MET,* and *RET* were numerically more common in never-smokers but not individually significant. In contrast, abnormalities in *BRAF* and *NRAS* were numerically more common to smokers. No *NTRK1* fusions were detected in this population.

In Fig. 1B, differences in driver distribution by smoking status are shown, subdivided by Tier category. Overall, 85% of never-smokers harbored a driver mutation (Tiers 1-3), compared to 65% of smokers (OR = 3.0, $P < .0001$, Table 3). Using multivariable logistic regression, we found that never-smokers were still more likely to harbor a driver mutation after the adjustment for potential confounders of ethnicity and sex (OR = 2.4, $P < .0001$). Tier 1 alterations (mutations associated with an FDA-approved drug) were detected in 71% (167/236) of never-smokers compared to 38% (258/671) of smokers (OR = 3.9, $P < .0001$; multi-variable OR = 3.2, *P*-value < .0001). While smoking remained highly significant, both females (OR = 1.44, $P = .0036$) and Asians (versus Caucasians, OR = 2.49, $P = .00060$) were still significantly enriched in the *EGFR*-mutant group after adjusting for smoking status.

Unsurprisingly, the percentage of driver-positive patients was higher in the fully-genotyped cohort (Fig. 1C) (Supplementary Fig. S2). Strikingly, 70/74 (95%) of fully-genotyped never-smokers had an oncogenic driver, and 58/74 (78%) had a Tier 1 driver (Fig. 1C, Table 3). *MET* exon 14 skipping was observed in 5.4% of fully genotyped never-smokers and 5% of fully genotyped smokers. This alteration was almost never observed in partially genotyped patients (Supplementary Fig. S2), since the vast majority were genotyped prior to August of 2018 via the Sema4 hotspot panel which did not test for *MET* exon 14 skipping. These results imply that Fig. 1A and B underestimate the true rates of *MET* and other driver alterations in both smokers and never-smokers.

The MAPK pathway mutations *KRAS* G12C and *BRAF* V600E currently represent the only Tier 1 driver alterations that do not directly involve activation of RTKs. RTK-based drivers (EGFR, HER2, MET, ALK, RET, and ROS1) were significantly less frequent in smokers (OR 0.13, $P < .00001$; multivariable OR 0.16, $P < .0001$, Fig. 2). A multivariable model was generated to look at the relationship between RTK-based drivers and demographics (Table 4). In the single-variable model, the presence of an RTK driver was significantly associated with never-smoking, female gender, and Asian or Latino ancestry. Smoking and Asian ancestry remained significant in multivariable analysis. In contrast, driver alterations in the MAPK pathway (RAS and RAF) were significantly more frequent in smokers compared to never-smokers, with 41% (275/761) of smokers harboring a RAS or RAF alteration, compared to 14% (33/236) of never-smokers (OR = 4.3, $P < .0001$; multivariable OR = 4.3, $P < .0001$). In the multivariable model, the presence of a MAPK pathway driver was significantly associated with a history of smoking, female, and Caucasian ancestry (Table 4).

Next, we examined risk factors for non-EGFR-RTK-positive tumor (tumors including Tier 1 variant genes *MET*, *HER2*, *ALK*, *RET*, *ROS1*). These mutations were also more common in the never-smoker group (OR 2.54, $P = .00056$). In a multivariable model including sex and race, smoking status remained highly significant, but no associations with race or sex were observed (data not shown).

In fully-genotyped patients, mutations not classified as drivers (including variants of unknown significance) were

Table 2. Categorization of driver mutations by Tiers based on treatment availability.

Tier 1: Known oncogene with FDA approved targeted drug in lung cancer	Tier 2: Under active investigation/nearing approval	Tier 3: Known oncogene without any available targeted drug
1. EGFR exon 19 deletion 2. EGFR L858R 3. EGFR rare activating mutations 4. EGFR exon 20 insertion 5. ALK fusion 6. ROS1 fusion 7. MET exon 14 skipping 8. BRAF V600E 9. RET fusion 10. NTRK Fusion 11. KRAS G12C	1. ERBB2 Exon 20 insertion Point mutations Amplification 2. NRG2 Fusions 3. Met amplification	1. KRAS non-G12C G12x Q61x G13x KRAS_nos* 2. KRAS + BRAF (KRAS hotspot with a BRAF nonv600e) 3. HRAS hotspot 4. NRAS hotspot 5. BRAF non V600E

Note: In several cases, the notes only stated patient EGFR positive and the patient received EGFR TKI.
Abbreviations: FDA, US Food and Drug Administration; nos, not otherwise specified.

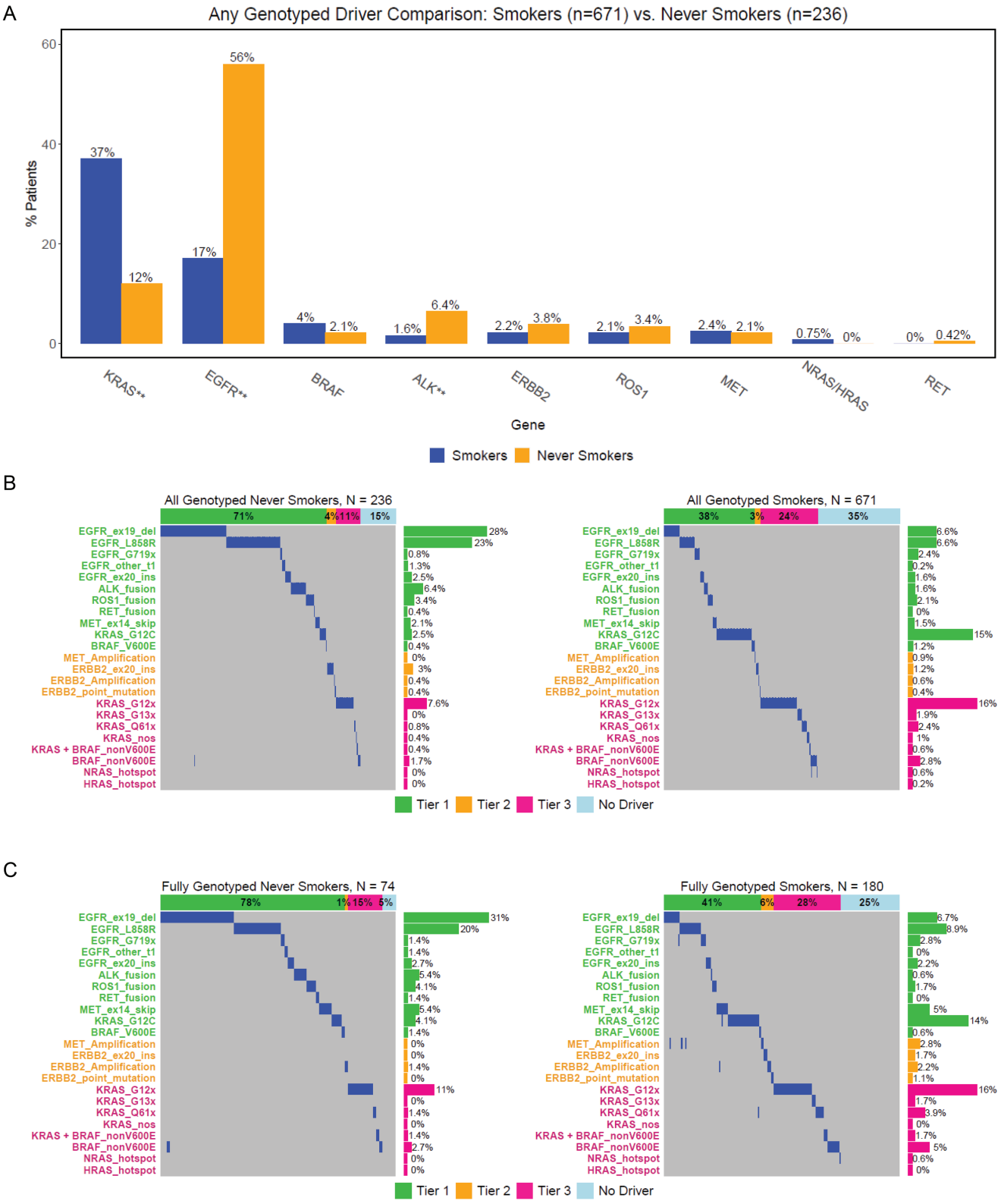

Figure 1. (**A**) Comparison of driver genes between the smoker and never-smoker populations. Patients having any driver with any amount of genetic testing were included. Any patients with a Tier 1-3 driver mutation were considered to have a driver mutation in that gene. Only genes present in both groups are presented. (**B**) Rates of mutation type in never smokers (left) versus smokers (right) organized by tier and frequency (bars on right side of the plot). All genotyped individuals are included. The overall frequency by tier is plotted above the oncoprint with light blue representing the proportion with no driver found. All drivers on the left-hand side of the plot represent a single variant except KRAS + BRAF non-V600E, which emerged as common tandem and has been reported in the literature. (**C**) Only fully genotyped individuals are included (a subset of **B**).

Table 3. Frequencies of mutations within Tier groups, subdivided by smoking status and extent of genotyping.

	N	Has T1 driver	Has T1-T2 driver	Has T1-T3 driver	No driver found
All tested never-smokers	236	71%	75%	85%	15%
All tested smokers	671	38%	41%	65%	35%
Fully genotyped never-smokers	74	78%	80%	95%	5%
Fully genotyped smokers	180	41%	47%	75%	25%
Partially genotyped never-smokers	162	67%	72%	81%	19%
Partially genotyped smokers	491	37%	39%	62%	38%

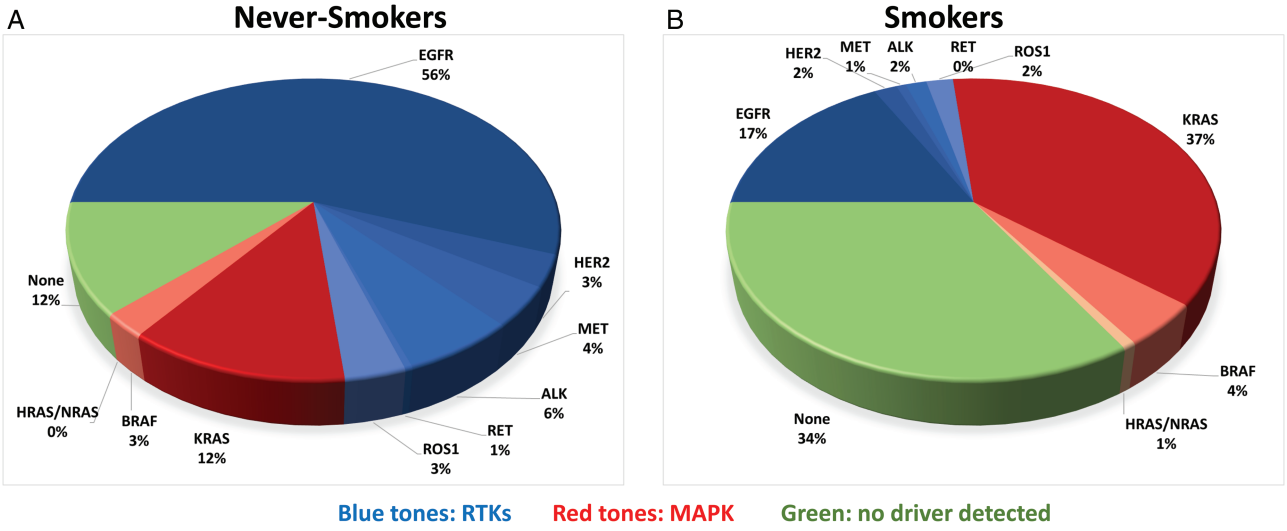

Figure 2. Pie charts depicting the distribution of RTK-mediated driver mutations (Blue tones) and MAPK signalling pathway drivers (red tones) in (**A**) never-smokers and (**B**) smokers. RTK driver abnormalities include *EGFR, MET, HER2, ALK, RET,* and *ROS1*; MAPK abnormalities include *KRAS, BRAF, NRAS,* and *HRAS.*

more commonly observed in smokers compared to never-smokers (Supplementary Fig. S3). The mean number of non-clinically significant alterations detected by OCAv3 and reported as VUS (indicated in Supplementary Fig. S3 with pink triangles) was 1.09 (range 0 to 7) for never-smokers and 2.18 (range 0-13) for smokers (*t*-test $P < .0001$).

Differences Detected in the Proportions of EGFR and KRAS Subtypes

Differences in the proportions of specific oncogenic alterations within *EGFR* and *KRAS* are noted in Fig. 3. There was a scarcity of *EGFR* G719X in never-smokers (Fig. 3A). These substitutions were only detected in 1.5% (2/131) of never-smokers with an *EGFR* driver, whereas they were observed in 14% (16/114) of smokers who harbored an *EGFR* driver ($P = .00028$). We observed enrichment in EGFR Exon20ins within the smoking group (OR = 2.2, not significant; never-smokers 4.6% vs. Smokers: 9.6%). *KRAS* G12C substitutions, the most frequent *KRAS* mutation in NSCLC, accounted for 44% (102/232) of *KRAS* mutations in smokers, compared to 24% (6/25) in never-smokers ($P = 0.058$, Fig. 3B).

Survival Differences Based on Smoking History

OS was assessed in patients with stage IV ADC who received genetic testing. The median survival time for the stage IV cohort was 27.9 months (95% CI, 21.9-36.0). The median follow-up time for the cohort was 32.6 months, and the dropout rate was 34% with a completion index of 76.4. We did not observe a significant difference between the dropout rate and follow-up times between the never-smoker and current/former smoker groups (Supplementary Table S1).

The median OS in the smoking cohort ($n = 189$) was 23.2 months (95% CI, 14.8-31.0), significantly shorter compared to never-smokers ($n = 79$) where the median OS was not reached in the study time frame (HR = 2.2 [95% CI, 1.4 to 3.5], $P = .00036$) (Fig. 4A). The 2-year survival rate was 66% (95% CI, 56-79) and 48% (95% CI, 41-57) for the never-smoker group and smoker groups, respectively ($P = .015$).

Baseline characteristics of stage IV patients are reported in Supplementary Table S2 and both single-variable and multivariable analyses including smoking status, age, sex, race/ethnicity, and *EGFR* mutation status are presented in Supplementary Table S3. Variables with *P*-values < .10 for any factor were incorporated in the multivariable model. The effect of smoking became nonsignificant when we accounted for the effects of sex, race/ethnicity, and *EGFR* mutation status (HR = 1.13, $P = .66$). Comparing patients with tumors harboring *EGFR* mutations (Fig. 4B), the median OS for smokers was 34 months in contrast to never-smokers where the median was 52 months (HR = 2.71

Table 4. Logistic regression for Tier 1-3 RTK and MAPK driver mutations.

		RTK (*n* = 335 carriers)				MAPK (*n* = 308 carriers)			
		Single variable		Multivariable		Single variable		Multivariable	
Variable	*n*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*
Age of diagnosis	907	1.00 (0.98-1.01)	.60			1.01 (0.99-1.02)	.36		
Age of diagnosis category									
45-75 years	679	Ref		Ref		Ref		Ref	
<45 years	31	2.95 (1.41-6.19)	**.0041**	2.10 (0.91-4.87)	.082	0.36 (0.14-0.95)	**.038**	0.54 (0.20-1.48)	.23
76+ years	197	1.25 (0.90-1.73)	.18	1.27 (0.88-1.83)	.20	0.94 (0.67-1.31)	.72	0.92 (0.65-1.31)	.65
Smoking status									
Never smoker	236	Ref		Ref		Ref		Ref	
Smoker	671	0.13 (0.09-0.18)	**<.00001**	0.16 (0.11-0.22)	**<.00001**	4.27 (2.87-6.36)	**<.00001**	4.33 (2.83-6.63)	**<.00001**
Gender									
Female	519	Ref		Ref		Ref		Ref	
Male	388	0.62 (0.47-0.82)	**.00075**	0.89 (0.65-1.22)	.47	0.79 (0.60-1.04)	.096	0.61 (0.45-0.82)	**.0013**
Race/ethnicity									
White/Caucasian	443	Ref		Ref		Ref		Ref	
Asian	94	4.13 (2.58-6.60)	**<.00001**	2.28 (1.34-3.89)	**.0024**	0.29 (0.16-0.52)	**2.00E-05**	0.52 (0.29-0.96)	**.036**
Black or African American	157	1.20 (0.82-1.77)	.34	1.21 (0.80-1.83)	.38	0.57 (0.38-0.85)	**.0053**	0.56 (0.37-0.84)	**.0053**
Hispanic or Latino	97	2.01 (1.29-3.15)	**.0021**	1.55 (0.94-2.55)	.085	0.44 (0.27-0.73)	**.0015**	0.56 (0.33-0.95)	**.031**
Other/unknown	116	1.01 (0.65-1.56)	.98	0.99 (0.61-1.60)	.97	0.78 (0.51-1.19)	.24	0.83 (0.53-1.28)	0.40

Note: Logistic regression models for smokers and never smokers who were given genetic testing for those carrying either a Tier 1-2 RTK or MAPK driver mutation. The odds ratio, 95% CI, and *P*-value are reported for both the univariate and multivariable models. Variables with *P*-values < .10 for any category were retained in the multivariable model. Bold values are less than .05.
Abbreviations: CI, confidence interval; OR, odds ratio.

[95% CI, 1.13-6.51]; *P* = .026). Only smoking status and race/ethnicity were modestly associated with OS (*P* < .10) and were brought forth for further analysis. The race/ethnicity distribution for patients with *EGFR*-mutant ADC is shown in Supplementary Table S4. Neither reached statistical significance in the multivariable models, but smoking trended with poorer OS, and Asians trended toward better OS (Supplementary Table S5).

Discussion

It has long been appreciated that lung cancer in patients without a significant exposure to tobacco smoke is associated with both different demographics and tumor molecular characteristics compared to patients with tobacco smoke-associated cancer; nevertheless, there remains a paucity of recent literature directly and systematically defining these differences using modern molecular diagnostics.[22] Here, we investigated the relationship between smoking, patient demographics, tumor oncogenetics, and outcome in ADC patients enrolled at the MSHS between January 1, 2015, and June 1, 2020.

Although tobacco smoking remains the most significant risk factor for the development of lung cancer, several studies noted a positive predictive association with the outcome of Immune checkpoint inhibitors (ICIs).[23-25] A meta-analysis of randomized trials in advanced NSCLC found that, while ICI was of significant benefit to smokers, chemotherapy outperformed ICI in never-smokers.[26] In a second meta-analysis comparing ICI monotherapy to chemotherapy in NSCLC, smokers had superior PFS and OS from ICIs compared to never-smokers.[27] The relationship between smoking and outcome of patients with NSCLC is undoubtedly multi-dimensional with smoking contributing to secondary comorbidities, increased rate and type of co-mutations, overall genomic complexity, and chromosomal instability, along with lifestyle and demographic associations. Although a link between ancestral genetics and the likelihood of an *EGFR*-mutation driven tumor has yet to be revealed, recent studies have given support for this hypothesis.[28] Smoking status is clearly and inextricably linked to rates of driver mutations and availability of therapies.

In our study, never-smokers comprised 26% of the population and were significantly more likely to harbor oncogenic driver mutations, be eligible for FDA-approved therapies, have improved outcomes in stage IV disease, be of Asian or Hispanic/Latino ancestry, have driver abnormalities in RTKs, and have improved outcomes in *EGFR*-targeted treatment. In contrast, smoking-associated tumors were more likely to have MAPK-associated driver abnormalities. The subtypes of *EGFR* mutations were divergent by smoking status, with Exon19del significantly more frequent in the never-smokers cohort and G719X mutations significantly more frequent in smokers. Perhaps most notably, in the subset of patients who received comprehensive NGS of all driver mutations, the never-smoker subgroup was driver-positive in 95% of cases.

Mutations associated with an FDA-approved therapy were significantly more frequent in the never-smoker population (71%) compared to those with a history of smoking (38%). Oncogenic drivers of any Tier were observed in 85% of never-smokers compared to 65% of smokers. Activating mutations within the MAPK signal transduction pathway (*KRAS*, *BRAF*, *NRAS*) were more commonly observed in smokers, representing 41% of cases compared to never-smokers at

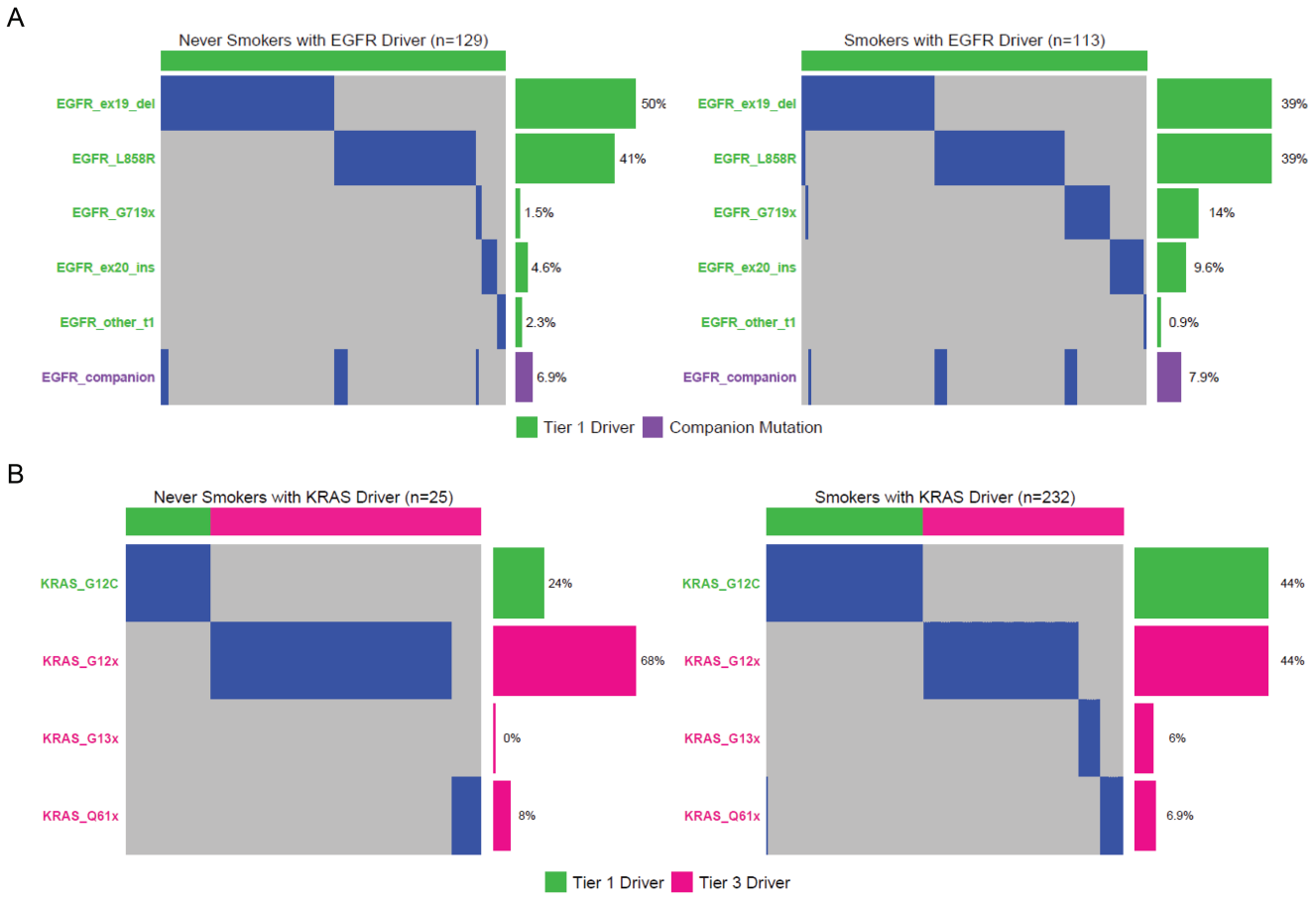

Figure 3. (**A**) Rates of *EGFR* mutation type in (left) never-smokers versus (right) smokers organized by Tier and frequency (bars on right side of the plot). Purple represents the proportion of tandem (companion) *EGFR* mutations. (**B**) Rates of KRAS mutation type in (left) never smokers versus (right) smokers organized by Tier and frequency (bars on right side of the plot). All genotyped individuals are included.

14%. In contrast, activating mutations in receptors and oncogenic fusions were commonplace in never-smokers at 68% compared to smokers at just 23%.

When subdivided by the presence of an "actionable" oncogenic driver, defined here as a Tier 1 or Tier 2 (ie, those with FDA-approved targeted therapies available or in advanced phase clinical trials), 75% of never-smokers were positive. In contrast, actionable oncogenic drivers were observed in only 41% of fully-genotyped smokers (n = 671). Of note, *KRAS* G12C mutations are considered here as an actionable (Tier 1) alteration due to the development of effective mutation-specific inhibitors that have been FDA approved.[5-8] In our study, *KRAS* G12C mutations comprised 37.1% of "actionable" cases in smokers versus just 3.4% of never-smokers.

In contrast to other *KRAS*-driven tumor types, the G12C variant is common in ADC, likely due to the smoking-associated carcinogenic process which stimulates the generation of nucleotide transversions (particular G → T).[29-31] Not surprisingly, we observed this alteration far more frequently in smokers compared to never-smokers (15% versus 2.5%). A similar process may be playing a role in the increased rate of *EGFR* G719X cases observed in the smoking cohort. Several of these point mutations are also the product of transversions, including the G719A and G719C variants, which both result from a G → T transversion, and G719S, which results from a GG → TC substitution.

In stage IV patients, never-smokers significantly and substantially outperformed smokers in OS. The presence of an actionable (Tier 1) driver was a principal predictor of survival, with female sex and Asian ancestry remaining significant. While this is largely due to the increased availability targeted therapy options, smoking may contribute to this phenomenon through additional mechanisms, including increased co-morbidities and a difference in the overall oncogenetic composition of the tumors. Among stage IV patients harboring *EGFR*-mutant tumors, smoking and Asian descent were significant in the single-variable model, but lost significance in the multivariable model, possible due to the small sample size.

Consistent with previous reports, we found that patients with ADC who self-reported as never-smokers were significantly more likely to be female, and of Asian descent.[32-34] We did not observe a significant difference in median age but noted a small but significant increase in the number of cases younger than 45 in the never-smoking cohort. We also observed an increased proportion of Hispanic/Latino patients who were never-smokers (14% in never-smokers versus 9% amongst smokers, P = .014); an observation that was reported in a large, prospective study of lung cancer care and outcomes in the US.[35] We observed that, in addition to a history of never smoking, female sex, Asian and Latino/Hispanic descent remained significant predictors of the presence of RTK-based driver abnormalities (*EGFR*, *MET*, *HER2*, *ALK*,

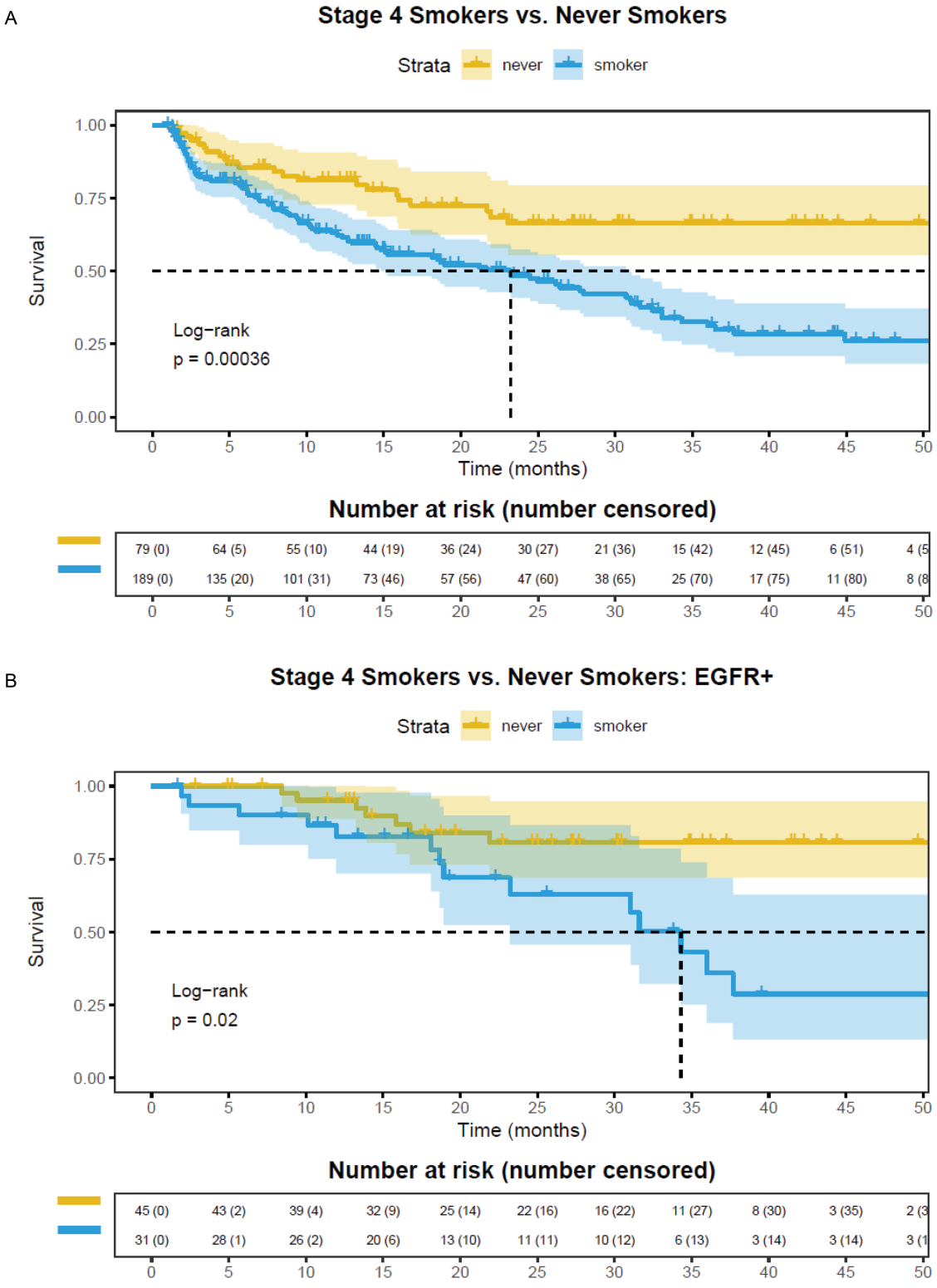

Figure 4. Survival curves for those diagnosed with stage IV lung cancer for never-smokers versus smokers with 95% CIs. (**A**) Full stage IV cohort. (**B**) Restricted to individuals with *EGFR* Tier 1 driver mutations.

RET, ROS1). MAPK mutations (RAS, RAF) were associated with female sex as well, but in contrast to RTK, were associated with smoking and Caucasian ancestry.

A limitation of this study is that not all cases had complete molecular genotyping. In contrast to the patients who received comprehensive OCAv3 genotyping, most of the patients diagnosed before August 2018 received an older multi-gene hotspot panel together with single-gene analysis of *ALK* and *ROS1*. These patients were generally not profiled for *RET* fusions, *NTRK* fusions nor *MET* exon 14 skipping. The analysis presented here includes all positive findings of a driver mutation, but this approach may skew results toward targets such as *EGFR*, *KRAS*, *ALK*, and *ROS1*, as these genes were included in all the genetic tests. The NGS approach

used for most patients in this study is designed to be as inclusive as possible of fusion forms, but likely misses some percentage of complex structural variants or rare partner genes. Additionally, the strict definition of *never-smoker* used in this study means that the *smoking* cohort includes a proportion of light smokers who may have ADC tumors more similar to never-smokers.

These results compare favorably to a similar study conducted by Memorial Sloan Kettering using the MSK Impact NGS panel. They observed that 68% (204/301) of never-smokers were female compared to smokers at 57%. Our results showed a similar distribution with 74% female in never-smokers versus 50% in smokers.[36] Both the MSK and our dataset had higher components of never-smokers (MSK:32%; Mount Sinai 26%), with concomitantly elevated *EGFR* mutation-positive subsets (MSK:31%; Mount Sinai 27%), in comparison to the 2014 The Cancer Genome Atlas study of comprehensive profiling in lung ADC, which estimated a never-smoking component of 14%, with 11% *EGFR* mutant-positive.[3]

In summary, we observed that the vast majority (85%-95%) of never-smokers harbored a recognized driver mutation, with 75%-80% positive for an actionable driver mutation associated with an available FDA-approved targeted therapy or nearing potential approval. This stands in contrast to smokers, where 65%-75% were found to harbor driver mutations with only 41%-47% eligible for a currently FDA-approved targeted therapy. Nevertheless, the detection of an additional 15% of smokers with *KRAS* G12C mutations highlights the requirement of molecular testing *in all stage IV ADC cases*, regardless of smoking or other clinical features. In all cases, and particularly in never-smokers, all efforts should be exhausted to identify an actionable driver mutation, including the use of plasma NGS approaches where tissue is insufficient.

Funding

This work was supported by Sema4 and the Center for Thoracic Oncology Seed funding.

Conflict of Interest

Michael I. Klein, Kristin L. Ayers, Xiang Zhou, Sunny Guin, Marc Fink, Michael Rossi, Hussam Al-Kateb, Feras M. Hantash, Scott Newman, Eric E. Schadt, Rong Chen: Sema4 (E, OI); **Timmy O'Connell, William Oh:** Janssen, Merck, Pfizer (C/A); Sema4 (E, OI). **Philip C. Mack**: Guardant Health, Amgen (H); **Fred R. Hirsch:** Bristol-Myers Squibb, Merck, Novartis, Genentech, AstraZeneca/Daiichi, Sanofi/Regeneron, OncoCyte, Amgen (C/A), Patent through the University of Colorado. (no royalties) "EGFR copy number and Protein Expression as Predictive Biomarker for EGFR directed Therapy" (IP). The other authors indicated no financial relationships.

(C/A) Consulting/advisory relationship; (RF) Research funding; (E) Employment; (ET) Expert testimony; (H) Honoraria received; (OI) Ownership interests; (IP) Intellectual property rights/inventor/patent holder; (SAB) Scientific advisory board

Author Contributions

Conception/design: P.C.M., M.I.K., K.L.A., F.M.H., S.N., R.C., F.R.H. Provision of study material/patients: M.I.K., K.L.A., X.Z., S.G., M.F., M.R., H.A.-K., F.M.H., W.K.O., S.N., E.E.S., R.C. Collection and/or assembly of data: P.C.M., M.I.K., K.L.A., X.Z., S.G., M.F., M.R., H.A.-K., T.O'C., F.M.H., W.K.O., S.N., E.E.S., R.C., F.R.H. Data analysis and interpretation: P.C.M., M.I.K., K.L.A., X.Z., S.G., M.F., M.R., H.A.-K., F.M.H., W.K.O., S.N., E.E.S., R.C., F.R.H. Manuscript writing: P.C.M., M.I.K., K.L.A., F.M.H., W.K.O., S.N., E.E.S., R.C., F.R.H. Final approval of manuscript: All authors.

Data Availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Supplementary Material

Supplementary material is available at *The Oncologist* online.

References

1. Gandara DR, Riess JW, Kelly K, Li T, Mack PC, Lara PN Jr. Evolution and increasing complexity of the therapeutic landscape in advanced non-small-cell lung cancer. *Clin Lung Cancer.* 2017;18(1):1-4.
2. Hirsch FR, Scagliotti GV, Mulshine JL, et al. Lung cancer: current therapies and new targeted treatments. *Lancet.* 2017;389(10066):299-311.
3. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma, *Nature*. 2014;511(7511):543-50.
4. Mack PC, Banks KC, Espenschied CR, et al. Spectrum of driver mutations and clinical impact of circulating tumor DNA analysis in non-small cell lung cancer: analysis of over 8000 cases. *Cancer.* 2020;126(14):3219-3228.
5. Lanman BA, Allen JR, Allen JG, et al. Discovery of a covalent inhibitor of KRASG12C (AMG 510) for the treatment of solid tumors. *J Med Chem*. 2020;63(1):52-65.
6. Hong DS, Fakih MG, Strickler JH, et al. KRASG12C inhibition with sotorasib in advanced solid tumors. *N Engl J Med.* 2020;383(13):1207-1217.
7. Hallin J, Engstrom LD, Hargis L, et al. The KRASG12C inhibitor MRTX849 provides insight toward therapeutic susceptibility of KRAS-mutant cancers in mouse models and patients. *Cancer Discov.* 2020;10(1):54-71.
8. Goebel L, Müller MP, Goody RS, Rauh D. KRasG12C inhibitors in clinical trials: a short historical perspective. *RSC Med Chem.* 2020;11(7):760-770.
9. Scagliotti GV, Longo M, Novello S. Nonsmall cell lung cancer in never smokers. *Curr Opin Oncol.* 2009;21(2):99-104.
10. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers—a different disease. *Nat Rev Cancer.* 2007;7(10):778-790.
11. Janjigian YY, McDonnell K, Kris MG, et al. Pack-years of cigarette smoking as a prognostic factor in patients with stage IIIB/IV nonsmall cell lung cancer. *Cancer.* 2010;116(3):670-675.
12. Tammemagi CM, Neslund-Dudas C, Simoff M, Kvale P. Smoking and lung cancer survival: the role of comorbidity and treatment. *Chest*. 2004;125(1):27-37.
13. Nordquist LT, Simon GR, Cantor A, Alberts WM, Bepler G. Improved survival in never-smokers vs current smokers with primary adenocarcinoma of the lung. *Chest*. 2004;126(2):347-351.
14. Kawaguchi T, Takada M, Kubo A, et al. Gender, histology, and time of diagnosis are important factors for prognosis: analysis of 1499 never-smokers with advanced non-small cell lung cancer in Japan. *J Thorac Oncol.* 2010;5(7):1011-1017.
15. Kogure Y, Ando M, Saka H, et al. Histology and smoking status predict survival of patients with advanced non-small-cell lung

cancer. Results of West Japan Oncology Group (WJOG) Study 3906L. *J Thorac Oncol*. 2013;8(6):753-758.
16. Toh CK, Gao F, Lim WT, et al. Never-smokers with lung cancer: epidemiologic evidence of a distinct disease entity. *J Clin Oncol*. 2006;24(15):2245-2251.
17. Goldstraw P, Chansky K, Crowley J, et al. International Association for the Study of Lung Cancer Staging and Prognostic Factors Committee, Advisory Boards, and Participating Institutions; International Association for the Study of Lung Cancer Staging and Prognostic Factors Committee Advisory Boards and Participating Institutions. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2016;11(1):39-51.
18. Sobin LH, Compton CC. TNM seventh edition: what's new, what's changed: communication from the International Union Against Cancer and the American Joint Committee on Cancer. *Cancer*. 2010;116(22):5336-5339.
19. Ayers KL, Ma M, Debussche G, et al. A composite biomarker of neutrophil-lymphocyte ratio and hemoglobin level correlates with clinical response to PD-1 and PD-L1 inhibitors in advanced non-small cell lung cancers. *BMC Cancer*. 2021;21(1):441.
20. Fancello L, Gandini S, Pelicci PG, Mazzarella L. Tumor mutational burden quantification from targeted gene panels: major advancements and challenges. *J Immunother Cancer*. 2019;7(1):183.
21. Clark TG, Altman DG, De Stavola BL. Quantification of the completeness of follow-up. *Lancet*. 2002;359(9314):1309-1310.
22. Ferketich AK, Niland JC, Mamet R, et al. Smoking status and survival in the national comprehensive cancer network non-small cell lung cancer cohort. *Cancer*. 2013;119(4):847-853.
23. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med*. 2015;373(17):1627-1639.
24. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med*. 2015;373(2):123-135.
25. Reck M, Rodríguez-Abreu D, Robinson AG, et al. KEYNOTE-024 Investigators. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med*. 2016;375(19):1823-1833.
26. Kim JH, Kim HS, Kim BJ. Prognostic value of smoking status in non-small-cell lung cancer patients treated with immune checkpoint inhibitors: a meta-analysis. *Oncotarget*. 2017;8(54):93149-93155.
27. Li B, Huang X, Fu L. Impact of smoking on efficacy of PD-1/PD-L1 inhibitors in non-small cell lung cancer patients: a meta-analysis. *Onco Targets Ther*. 2018;11:3691-3696.
28. Carrot-Zhang J, Soca-Chafre G, Patterson N, et al. Genetic ancestry contributes to somatic mutations in lung cancers from admixed Latin American populations. *Cancer Discov*. 2021;11(3):591-598.
29. Kim HR, Shim HS, Chung JH, et al. Distinct clinical features and outcomes in never-smokers with nonsmall cell lung cancer who harbor EGFR or KRAS mutations or ALK rearrangement. *Cancer*. 2012;118(3):729-739.
30. Eisenstadt E, Warren AJ, Porter J, Atkins D, Miller JH. Carcinogenic epoxides of benzo[a]pyrene and cyclopenta[cd]pyrene induce base substitutions via specific transversions. *Proc Natl Acad Sci USA*. 1982;79(6):1945-1949.
31. Hainaut P, Pfeifer GP. Patterns of p53 G–>T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. *Carcinogenesis*. 2001;22(3):367-374.
32. Matsuo K, Ito H, Yatabe Y, et al. Risk factors differ for non-small-cell lung cancers with and without EGFR mutation: assessment of smoking and sex by a case-control study in Japanese. *Cancer Sci*. 2007;98(1):96-101.
33. Shi Y, Au JS, Thongprasert S, et al. A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER). *J Thorac Oncol*. 2014;9(2):154-162.
34. Midha A, Dearden S, McCormack R. EGFR mutation incidence in non-small-cell lung cancer of adenocarcinoma histology: a systematic review and global map by ethnicity (mutMapII). *Am J Cancer Res*. 2015;5(9):2892-2911.
35. Clément-Duchêne C, Stock S, Xu X, et al. Survival among never-smokers with lung cancer in the Cancer Care Outcomes Research and Surveillance study. *Ann Am Thorac Soc*. 2016;13(1):58-66.
36. Jordan EJ, Kim HR, Arcila ME, et al. Prospective comprehensive molecular characterization of lung adenocarcinomas for efficient patient matching to approved and emerging therapies. *Cancer Discov*. 2017;7(6):596-609.