Racial Diversity of Actionable Mutations in Non–Small Cell Lung Cancer

Aliccia Bollig-Fischer, PhD, Wei Chen, PhD, Shirish M. Gadgeel, MD, Angela S. Wenzlaff, MPH, Michele L. Cote, PhD, Ann G. Schwartz, PhD, MPH, and Gerold Bepler, MD, PhD

Introduction: Lung cancer is the leading cause of cancer-related deaths in the US. The reasons for higher incidence and poorer survival rates among black compared with white lung cancer patients have not been defined. We hypothesized that differential incidence of somatic cancer gene mutations may be a contributing factor. Previous genomic studies of non–small cell lung cancer (NSCLC) have not adequately represented black patients.

Methods: A matrix-assisted laser desorption/ionization and time-of-flight mass spectrometry approach was used to analyze tumor DNA for 214 coding mutations in 26 cancer genes previously identified in NSCLC. The samples included NSCLC from 335 white patients and 137 black patients. For 299 of these, normal matched DNA was available and analyzed.

Results: Epidermal growth factor receptor (EGFR) exon 19 deletions were only detected in women cases, with increased odds for black women compared with white women (odds ratio = 3.914, 95% confidence interval: 1.014–15.099, \( p = 0.048 \)). Beyond race, variations in mutation frequencies were seen by histology. DDR2 alterations, previously described as somatic mutations, were identified as constitutional variants.

Conclusions: This study is among the largest comparing somatic mutations in black and white patients. The results point to the molecular diversity of NSCLC and raise new questions as to the importance of inherited alleles. Genomic tumor testing will benefit both populations, although the mutation spectrum appears to vary by sex, race, and histology.

Key Words: Non–small cell lung cancer, African American, mutations, Oncogenes, Tumor suppressors.

(Lung Cancer. 2015;10: 250–255)

Lung cancers are a heterogeneous group of tumors traditionally categorized by histology, with the majority classified as non–small cell lung cancer (85%, NSCLC).1 NSCLC is further subdivided into adenocarcinomas (~45%), squamous cell carcinomas (~23%), and large cell carcinomas (~3%), with other subtypes representing the remaining approximately 28%.1 Large-scale gene mutation profiling is transforming our knowledge of the heterogeneity of NSCLC.2 Genomic research now links specific oncogenes and recurring mutations to the disease phenotype and provides a rationale for use of molecularly targeted treatments that are improving lung cancer patients’ outcomes.2

Black Americans are more likely to develop lung cancer and at an earlier age compared with white individuals despite lower rates of smoking in black adolescents.3–7 Moreover, blacks with lung cancer show worse outcomes, including shorter overall survival and increased mortality, which persist when correcting for socioeconomic factors and unequal access to care.7 These data suggest diversity in disease etiology across populations. The specifics of disease etiology govern the mutational processes giving rise to particular mutation patterns or signatures in cancer.4 Along these lines, study by Alexandrov et al.6 demonstrates that a specific mutation profile in lung cancers is attributable to smoking. The same study also reports distinct mutation signatures in lung cancers not yet linked to a cause. It is predicted that more causally linked cancer mutation signatures will emerge when both the numbers of specimens and diversity of patients under investigation increases. Thus, we hypothesized that the frequencies of specific mutations in NSCLC will vary between black and white populations. Previous studies have compared mutation frequencies according to race, but largely focused on a small subset of oncogenes—e.g., mutations in epidermal growth factor receptor (EGFR) and kirsten rat sarcoma viral oncogene homolog (KRAS)—and they yielded conflicting results.9–13 More comprehensive research is required to thoroughly investigate cancer gene mutation differences and to derive insight as to diagnostics and individualized treatment modalities that provide patient benefit. Here, we present findings from a large-scale genomic study, with a large proportion of black patients, to examine whether the landscape of cancer relevant mutation frequencies varies according to race.

MATERIALS AND METHODS

Patients and Tissues

Biospecimen collection and outcomes data compilation were done according to the Helsinki Declaration and approved by the Wayne State University School of Medicine.
Institutional Review Board. Fresh-frozen or formalin-fixed paraffin-embedded specimens were collected from patients who underwent a surgical resection for diagnosed or suspected lung cancer. Frozen specimen procurement procedures were implemented to reduce the resection to freezing time interval to less than 30 minutes. The overall procurement period ranged from 1985 to 2012 from five different case series. Frozen and archived FFPE specimens were reviewed to verify NSCLC diagnosis and to determine that tumor cell content was equal or greater than 70%. Only tumors pathologically confirmed to be NSCLC were included in the analysis. DNA from adjacent normal lung tissue was available and analyzed from 299 of the 472 cases. Demographic and clinical outcomes data collected included: the dates of birth, diagnosis, and last follow-up or death; sex; race; tumor histology; pathological, and clinical tumor stage (AJC staging manual, version 6.0) and self-reported smoking history (defined as life-time never smoker for those who had smoked <100 cigarettes, former smoker for those who had quit cigarette smoking for more than 1 year, and smoker for all others).

Genetic Analysis

Frozen NSCLC and normal tissue specimens (~100 mg) were pulverized using sterilized and frozen mortar and pestles and DNA was isolated using resin or phenol-based extraction methods. DNA was isolated from paraffin-embedded tissue using EZ1 Advanced magnetic bead technology, the EZ1 DNA de-paraffinization method and the EZ1 DNA Tissue kit (Qiagen, Valencia, CA). All sample DNA quantity and quality was assessed using a Nanodrop spectrophotometer and Quantifiler assay (Life Technologies, Carlsbad, CA), a real-time polymerase chain reaction-based approach for estimation of amplifiable DNA. A standard curve of known DNA concentrations was also analyzed for quantitation of each DNA sample.

Mutations were analyzed using the Sequenom MassARRAY System employing matrix-assisted laser desorption/ionization and time of flight mass spectrometry and the Sequenom LungCarta panel (Sequenom, San Diego, CA). The panel targets 214 sequence mutations in 26 oncogenes and tumor suppressors, which were previously identified in NSCLC. The genes include: AKT1, ALK, BRAF, DDR2, EGFR, EPHA3, EPHA5, ERBB2, FGFR4, JAK2, KRAS, MAP2K1, STK11, MET, NOTCH1, NRAS, NRP2, NTRK1, NTRK2, NTRK3, PIK3CA, PTCH1, PTEN, PTPN11, PTPRD, and TP53. The coding mutation types include synonymous and non-synonymous nonsense and missense point mutations, transversions and transitions, and short insertions and deletions (Supplementary Table 1, Supplementary Digital Content 2, http://links.lww.com/JTO/A748). The methodology starts with polymerease chain reaction of small DNA segments (<100 base pairs) encompassing the DNA mutation sites in a multiplex reaction. A follow-up extension reaction adds a single base to extension primers, the nature of the base pair added to a primer affects its mass and time-of-flight; thus, a wild-type and mutation sequence can be differentiated and quantified. The sensitivity of the approach allows for detection of a mutation that represents ≥10% of the sample. The assay precision and accuracy for both FFPE and frozen specimen DNA were validated for as little as 60 ng by us and others however, typically 480 ng of sample DNA was analyzed for this study.

Statistical Analysis

Descriptive statistics were provided for patients’ demographic and clinical characteristics. Multivariable logistic regression was used to estimate the odds of having the genetic alteration, adjusted for potential confounders such as race (Black, reference: White), age (as continuous variable), sex (women, reference: men), tumor stage (II, III, or IV, reference: I), histology (squamous, other NSCLC, reference: adenocarcinoma), and smoking status (never, unknown, reference: ever). All statistical tests were traditional two sided at a significance level of 0.05. Given the nature of this exploratory study, p values were not adjusted for multiple testing. Single nucleotide polymorphism (SNP) allele frequency within each subgroup of race was compared with the National Institutes of Health Heart, Lung and Blood Institute Grand Opportunity Exome Sequencing 6500 Project database† using Fisher’s exact test. The statistical software R 3.0.0 (The R foundation for statistical computing; http://www.r-project.org/foundation/) was used for all analyses.

RESULTS

Patient Characteristics

Descriptive statistics for the 472 NSCLC patients—137 black individuals and 335 white individuals—who contributed specimens to this study are listed in Table 1. The proportions of histology, sex, stage, and smoking history are given stratified by race; median age and pack years smoked by race are also indicated.

Frequency of Cancer Gene Mutations According to Race

An initial overview of the data from mass spectrometry analysis showed that for 180 of the 472 NSCLC specimens (38%) a mutation was detected (characteristics and mutations identified for each specimen are provided in the online Supplementary Data file, Supplementary Digital Content 2, http://links.lww.com/JTO/A749). Comparing frequencies for individual mutations across race showed no significant differences for point mutations; however, there was a higher frequency for the EGFR E746-T751>S deletion in exon 19 in tumor specimens from black patients compared with white patients (p = 0.076; Table 2). When the combined frequency of all variations of deletions detected at E746 (i.e., E746-T751 and E746-T750) were considered, we observed that these mutations occurred exclusively in women and more often in black women than in white women (odds ratio [OR]: 3.914; 95% confidence interval [CI]: 1.014–15.099; p = 0.048), after adjusting for age, stage, histology, race, and smoking status (Table 3). Across all mutation types, tumors from women were 80% more likely to carry at least one mutation than those from men after adjustment for smoking, age, race, stage, and histologic type (OR: 1.815; 95% CI: 1.200–2.747; p = 0.005)
fully reported; however, mapping of co-occurring mutations and multiple mutations, correlations cannot be tested for or meaning-
black and white patients. Given the small numbers of mul-
and multiple mutations observed in NSCLC specimens from
com/JTO/A748). Figure 1 summarizes the number of single
Table 2, Supplementary Digital Content 1, http://links.lww.
= 0.025) (Supplementary
0.582; 95% CI: 0.363–0.933;
p
having a mutation was decreased for the black patients (OR:
(0.2%). Also of note, two
DDR2
SNPs were detected in matched
tissue from a single white patient in our cohort of 472 patients
this study, we identified it in normal and corresponding malignant
given that might be inherited. Our efforts identified seven constitutional
mutations in c-met (MET), notch 1 (NOTCH1), and serine/threonine kinase 11 (STK11). The
differentially detected SNPs included NOTCH p.V1671I and
STK11 p.Y272Y (both increased in blacks; p < 0.0001), and
MET p.N375S (increased in whites; p = 0.0117). Using the
National Institutes of Health Heart, Lung and Blood Institute
Grand Opportunity Exome Sequencing 6500 Project data-
based,17 we observed that all three SNPs demonstrated simi-
lar differences in comparing black and white frequencies in
the general US population. However, NOTCH1 p.V1671I
(p = 0.01), and MET p.N375S (p = 0.076) appear to be more
frequent in the cancer patient population (Supplementary
Table 3, Supplementary Digital Content 1, http://links.lww.
com/JTO/A748). All of these SNPs are in the National Center
for Biotechnology Information SNP database although they
have previously been highlighted as mutations in the literature
or COSMIC database.13,19–22

We proceeded to analyze matched normal lung DNA from
299 cases to confirm these SNPs and to identify other variants
that might be inherited. Our efforts identified seven constitutional
polymorphisms in the dataset, described in Table 4. Included
among them was EGFR p.R776H, an activating mutation only
recently identified as inherited in a single case study of a non-
smoking mother–daughter pair diagnosed with lung cancer.20
In this study, we identified it in normal and corresponding malignant
tissue from a single white patient in our cohort of 472 patients
(0.2%). Also of note, two DDR2 SNPs were detected in matched
normal and tumor specimens: DDR2 p.L63V in a white patient
displays a greater diversity of known mutations in tumors
from white patients, which is consistent with overall signi-
ificantly more known mutations having been identified in
NSCLC specimens from white patients.

**Distinguishing Constitutional Polymorphisms from Somatic Mutations**

In an overview of the data from mass spectrometry analysis, we identified that there were significant differences across race for genetic alterations in tumors that are also present in germline DNA, representing previously described single nucleotide polymorphisms in c-met (MET), notch 1 (NOTCH1), and serine/threonine kinase 11 (STK11). The differentially detected SNPs included NOTCH p.V1671I and STK11 p.Y272Y (both increased in blacks; p < 0.0001), and MET p.N375S (increased in whites; p = 0.0117). Using the National Institutes of Health Heart, Lung and Blood Institute Grand Opportunity Exome Sequencing 6500 Project database,13,19–22 we observed that all three SNPs demonstrated similar differences in comparing black and white frequencies in the general US population. However, NOTCH1 p.V1671I (p = 0.01), and MET p.N375S (p = 0.076) appear to be more frequent in the cancer patient population (Supplementary Table 3, Supplementary Digital Content 1, http://links.lww.com/JTO/A748). All of these SNPs are in the National Center for Biotechnology Information SNP database although they have previously been highlighted as mutations in the literature or COSMIC database.13,19–22

We proceeded to analyze matched normal lung DNA from
299 cases to confirm these SNPs and to identify other variants
that might be inherited. Our efforts identified seven constitutional
polymorphisms in the dataset, described in Table 4. Included
among them was EGFR p.R776H, an activating mutation only
recently identified as inherited in a single case study of a non-
smoking mother–daughter pair diagnosed with lung cancer.20
In this study, we identified it in normal and corresponding malignant
tissue from a single white patient in our cohort of 472 patients
(0.2%). Also of note, two DDR2 SNPs were detected in matched
normal and tumor specimens: DDR2 p.L63V in a white patient

**TABLE 1. Patient Characteristics**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Black (n = 137)</th>
<th>White (n = 335)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age Median</td>
<td>60.8</td>
<td>66.1</td>
</tr>
<tr>
<td>Range (min, max)</td>
<td>(29.3, 84.7)</td>
<td>(25.8, 83.7)</td>
</tr>
<tr>
<td>Gender Women</td>
<td>80 (58%)</td>
<td>116 (35%)</td>
</tr>
<tr>
<td>Men</td>
<td>57 (42%)</td>
<td>219 (65%)</td>
</tr>
<tr>
<td>Histology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>92 (67%)</td>
<td>162 (48%)</td>
</tr>
<tr>
<td>Squamous cell</td>
<td>29 (21%)</td>
<td>133 (40%)</td>
</tr>
<tr>
<td>NSCLC</td>
<td>16 (12%)</td>
<td>40 (12%)</td>
</tr>
<tr>
<td>Stage I</td>
<td>75 (55%)</td>
<td>219 (65%)</td>
</tr>
<tr>
<td>II</td>
<td>17 (12%)</td>
<td>70 (21%)</td>
</tr>
<tr>
<td>III or IV</td>
<td>45 (33%)</td>
<td>46 (14%)</td>
</tr>
<tr>
<td>Smoking history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever</td>
<td>123 (90%)</td>
<td>273 (81%)</td>
</tr>
<tr>
<td>Never</td>
<td>13 (9%)</td>
<td>22 (7%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (1%)</td>
<td>40 (12%)</td>
</tr>
<tr>
<td>Pack years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>38</td>
<td>50</td>
</tr>
<tr>
<td>Range (min, max)</td>
<td>(1, 130)</td>
<td>(1, 300)</td>
</tr>
</tbody>
</table>

<sup>NSCLC not sub-categorized in pathology review.</sup>
<sup>NSCLC, non–small cell lung cancer.</sup>

**TABLE 2. Frequency and Significance<sup>a</sup> for Specific Mutations Identified in Tumor Tissues among 472 Patients by Race**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Variant</th>
<th>Black</th>
<th>White</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFR</td>
<td>delE746-T751&gt;S</td>
<td>2%</td>
<td>0%</td>
<td>0.076</td>
</tr>
<tr>
<td>EGFR</td>
<td>G719C</td>
<td>1%</td>
<td>0%</td>
<td>0.084</td>
</tr>
<tr>
<td>EGFR</td>
<td>delE746-A750</td>
<td>4%</td>
<td>2%</td>
<td>0.116</td>
</tr>
<tr>
<td>TP53</td>
<td>R158L</td>
<td>1%</td>
<td>3%</td>
<td>0.189</td>
</tr>
</tbody>
</table>

<sup>Top genes shown, ranked according to smallest p value of significance.</sup>

**TABLE 3. Logistic Regression for Odds Ratio of having EGFR Mutations delE746-T751>S or delE746-A750 Discovered Only in Women Samples**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds Ratio</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race: Black</td>
<td>3.914</td>
<td>1.014</td>
<td>15.099</td>
<td>0.048</td>
</tr>
<tr>
<td>Age</td>
<td>1.056</td>
<td>0.993</td>
<td>1.132</td>
<td>0.081</td>
</tr>
<tr>
<td>Stage II</td>
<td>1.666</td>
<td>0.367</td>
<td>7.560</td>
<td>0.508</td>
</tr>
<tr>
<td>Stages III, IV</td>
<td>0.572</td>
<td>0.125</td>
<td>2.611</td>
<td>0.471</td>
</tr>
<tr>
<td>Histology NSCLC</td>
<td>0.000</td>
<td>0.000</td>
<td>Inf</td>
<td>0.991</td>
</tr>
<tr>
<td>Histology squamous</td>
<td>0.618</td>
<td>0.106</td>
<td>3.605</td>
<td>0.593</td>
</tr>
<tr>
<td>Smoke never</td>
<td>7.070</td>
<td>2.016</td>
<td>24.790</td>
<td>0.002</td>
</tr>
<tr>
<td>Smoke unknown</td>
<td>9.281</td>
<td>1.286</td>
<td>66.973</td>
<td>0.027</td>
</tr>
</tbody>
</table>

The model was built in the 196 women-only group. Reference groups: stage I, adenocarcinoma, or smoker. CI, confidence interval; NSCLC, non–small cell lung cancer.
and DDR2 p.G505S in a black patient. Both variants were previously characterized as activating mutations in NSCLC.\textsuperscript{15}

**Mutations Associated with Histologic Subtype**

According to our data, mutation frequencies were lower in squamous cell carcinoma than in adenocarcinoma (OR: 0.583; 95% CI: 0.374–0.909; \( p = 0.017 \)). Further investigation of the frequency of somatic mutations in our 472 patient cohort according to histology highlighted potential differences specifically for PIK3CA p.E545K (increased rate in squamous cell carcinoma, \( p = 0.029 \)) and KRAS p.G12C/V (increased rate in adenocarcinoma, \( p < 0.001 \) and \( p = 0.016 \)) (Supplementary Table 4, Supplementary Digital Content 1, http://links.lww.com/JTO/A748). In general, adenocarcinomas exhibited greater numbers and higher diversity of mutations compared with squamous cell carcinomas (Fig. 2). It is noted that the level of diversity for co-occurring mutations discovered in squamous cell carcinoma appears to resemble the relative level of diversity observed for black patients. Any observation of similarity is potentially misleading and not consistent with a histology bias in the study population, where to the contrary, squamous cell cancer is significantly less represented in our black patient cohort compared with adenocarcinomas (21% versus 67%, \( p < 0.001 \)).

**DISCUSSION**

Certain epidemiological data establishing disparities in NSCLC occurrence in black versus white patients are also suggestive of differences in disease etiology,\textsuperscript{3–7} and provide

---

**TABLE 4.** Description of Constitutional Polymorphisms and Frequencies in NSCLC Specimens

<table>
<thead>
<tr>
<th>Gene Variant</th>
<th>dbSNP# or Germline Reference</th>
<th>Cosmic# or Mutation Reference</th>
<th>Black (Frequency)</th>
<th>White (Frequency)</th>
<th>( p ) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDR2 (G505S)</td>
<td>rs115169993</td>
<td>Hammerman 2011 (activating)</td>
<td>1 (1%; CI: 0%, 4%)</td>
<td>0 (0%; CI: 0%, 1%)</td>
<td>0.2903</td>
</tr>
<tr>
<td>DDR2 (L63V)</td>
<td>rs144594252</td>
<td>Hammerman 2011 (activating)</td>
<td>0 (0%; CI: 0%, 3%)</td>
<td>1 (0%; CI: 0%, 2%)</td>
<td>1</td>
</tr>
<tr>
<td>EGFR (R776H)</td>
<td>van Noesel 2013</td>
<td>COSM22940</td>
<td>0 (0%; CI: 0%, 3%)</td>
<td>1 (0%; CI: 0%, 2%)</td>
<td>1</td>
</tr>
<tr>
<td>MET (N375S)</td>
<td>rs33917957</td>
<td>COSM28925; Krishnaswamy 2009 (inhibitor resistant)</td>
<td>1 (1%; CI: 0%, 4%)</td>
<td>19 (6%; CI: 4%, 9%)</td>
<td>0.0117</td>
</tr>
<tr>
<td>NOTCH1 (V1671I)</td>
<td>rs2229968</td>
<td>COSM33750</td>
<td>12 (9%; CI: 5%, 15%)</td>
<td>0 (0%; CI: 0%, 1%)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>STK11 (F354L)</td>
<td>Launonen 2000</td>
<td>COSM1360</td>
<td>0 (0%; CI: 0%, 3%)</td>
<td>5 (1%; CI: 1%, 3%)</td>
<td>0.3278</td>
</tr>
<tr>
<td>STK11 (Y272Y)</td>
<td>rs9282859</td>
<td>COSM29005</td>
<td>21 (15%; CI: 10%, 22%)</td>
<td>1 (0%; CI: 0%, 2%)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

*Among 472 patients (137 African American, 335 European ancestry) with Wilson’s 95% CI.
dbSNP, data base single nucleotide polymorphism; CI, confidence interval.
reasonable support for the notion that mutation profiles from the two populations will be different. Our study analyzed a broad variety of sequence mutations in NSCLC-linked genes and showed that the frequencies of coding mutations were within previously reported ranges for NSCLC (adenocarcinoma and squamous cell combined: KRAS 11%, EGFR 8%, and PIK3CA 2%); and their frequencies were similar for black and white groups. The exception to this rule is the frequency of small deletions in \textit{EGFR} exon 19 at p.E746. Results showed the deletion was only detected in lung tumors from women, and with increased frequency in lung tumors from black women. The data suggest that women and black women more specifically may benefit from testing for this lesion, which is a predictive biomarker of response to anti-\textit{EGFR} therapy. The clinical importance of a small deletion in \textit{EGFR} exon 19 is further underscored by early evidence linking the aberration to longer progression-free survival relative to other \textit{EGFR} mutations in metastatic cancer treated with first-line \textit{EGFR} tyrosine kinase inhibitors.\textsuperscript{23}

The types and or signatures of mutations that define cancer subtypes are attributable to causative mutational processes and evidence surrounding small deletions like p.E746-751 or p.E746-750 identified in \textit{EGFR} exon 19 points specifically to defects in the DNA mismatch repair process.\textsuperscript{24} Furthermore, significant differences in the numbers of specimens for black patients showing no coding mutation, 68% versus 59% for whites, indicates that further genomic analyses should be done to more comprehensively characterize the range of mutation types occurring in NSCLC of black patients. For example, analysis of a larger gene set or copy number variation analysis may increase the percentage of tumors where an oncogene mutation is discovered. Whole genome sequencing of tumors from black patients may also identify novel genetic alterations in this population.

Our study is only the second to report a rare \textit{EGFR} single nucleotide germ line variant previously reported in a single case study to be oncogenic.\textsuperscript{20} Also, ours is the first report of the discovery of \textit{DDR2} variants in both normal and NSCLC tissue from individual patients. Specifically, the \textit{DDR2} p.G505S and \textit{DDR2} p.L63V variants are reported to be activating mutations in the cancer literature—mutations that supposedly render a cancer cell sensitive to the antigrowth effects of the targeted drug dasatinib.\textsuperscript{15} It cannot be determined from this study what bearing these inherited \textit{DDR2} variants may have had on cancer initiation and outcomes. Their discovery brings to mind ongoing efforts to understand to what degree inherited \textit{EGFR} p.T790M mutations lead to an increased risk of developing lung cancer\textsuperscript{25-27}; the same consideration must be given to \textit{DDR2} variants.

Finally, our analyses also honed in on histology-associated differences for individual mutation frequencies, adding to the growing body of evidence and emerging understanding that histologic subtypes are distinguished by the mutations they harbor.\textsuperscript{2,28,29} It was observed that adenocarcinomas exhibited greater numbers and greater diversity of mutations, and more often co-occurring mutations compared with squamous cell carcinomas (Fig. 2). It should not be interpreted that squamous cell carcinoma is less apt to harbor driver mutations; on the contrary, the spectrum of cancer driver gene mutations associated with squamous cell histology is anticipated to be
different. Thus, pointing to a need for future work to engage expanded genomic analyses to identify the full spectrum of cancer gene lesions in lung cancer subtypes. The same notions apply to our analyses across race, where fewer mutations were identified in black patient samples.

In summary, this study designed to test for differences in known somatic coding mutation frequencies across populations, presents a number of novel and unexpected findings. The important significant findings associated with race support the need for further study to molecularly define NSCLC in diverse patient populations. The particular discovery that DDR2 variants (DDR2 p.G505S and DDR2 p.L63V) classify as inherited alleles forces the need for further evaluation of their occurrence and oncogenic role in NSCLC.

**ACKNOWLEDGMENTS**

The authors thank the patients who consented to provide tumor material that was used in this study.

**REFERENCES**


