Comparison of Prevalence and Types of Mutations in Lung Cancers Among Black and White Populations

Joshua D. Campbell, PhD; Christopher Lathan, MD, MPH; Lynette Sholl, MD; Matthew Ducar; Mikenah Vega; Ashwini Sunkavalli; Ling Lin, PhD; Megan Hanna; Laura Schubert; Aaron Thorner, PhD; Nicholas Faris, MDiv; David R. Williams, PhD, MPH; Raymond U. Osarogiagbon, MBBS; Paul van Hummelen, PhD; Matthew Meyerson, MD, PhD; Laura MacConaill, PhD

IMPORTANCE Lung cancer is the leading cause of cancer death in the United States in all ethnic and racial groups. The overall death rate from lung cancer is higher in black patients than in white patients.

OBJECTIVE To compare the prevalence and types of somatic alterations between lung cancers from black patients and white patients. Differences in mutational frequencies could illuminate differences in prognosis and lead to the reduction of outcome disparities by more precisely targeting patients’ treatment.

DESIGN, SETTING, AND PARTICIPANTS Tumor specimens were collected from Baptist Cancer Center (Memphis, Tennessee) over the course of 9 years (January 2004-December 2012). Genomic analysis by massively parallel sequencing of 504 cancer genes was performed at Dana-Farber Cancer Institute (Boston, Massachusetts). Overall, 509 lung cancer tumors specimens (319 adenocarcinomas; 142 squamous cell carcinomas) were profiled from 245 black patients and 264 white patients.

MAIN OUTCOMES AND MEASURES The frequencies of genomic alterations were compared between tumors from black and white populations.

RESULTS Overall, 509 lung cancers were collected and analyzed (273 women [129 black patients; 144 white patients] and 236 men [116 black patients; 120 white patients]). Using 313 adenocarcinomas and 138 squamous cell carcinomas with genetically supported ancestry, overall mutational frequencies and copy number changes were not significantly different between black and white populations in either tumor type after correcting for multiple hypothesis testing. Furthermore, specific activating alterations in members of the receptor tyrosine kinase/Ras/Raf pathway including EGFR and KRAS were not significantly different between populations in lung adenocarcinoma.

CONCLUSIONS AND RELEVANCE These results demonstrate that lung cancers from black patients are similar to cancers from white patients with respect to clinically actionable genomic alterations and suggest that clinical trials of targeted therapies could significantly benefit patients in both groups.
lung cancer remains the leading cause of death from cancer in the United States. Black populations have the highest lung cancer mortality rate of any racial or ethnic group in the United States. The average annual age-adjusted incidence is 65.0 and 58.8 per 100,000 in black patients and white patients, respectively. Many factors may contribute to the disparities in both incidence and outcome including differences in access to health care, smoking behavior, and other socioeconomic variables. However, these known risk factors do not fully explain the differences in the incidence and outcome between these 2 populations.

Acquired alterations to the cancer genome—including somatic single-nucleotide variants, insertions and deletions (indels), copy number variations (CNVs), and structural rearrangements—contribute to tumorigenesis by activating pathways involved in cell growth and proliferation, resistance to apoptosis, and immune cell invasion. Alterations in receptor tyrosine kinases (RTKs) such as EGFR, ALK, RET, and ROS1 are clinically important as they confer sensitivity to kinase inhibitors in lung adenocarcinomas. Recent studies led by our group and The Cancer Genome Atlas (TCGA) have performed comprehensive molecular characterization of lung tumors to identify other potentially targetable alterations. We have shown that the frequency of acquired alterations in lung cancer driver genes are largely distinct between the 2 largest subclasses of non–small-cell lung cancer (NSCLC), lung adenocarcinoma, and lung squamous cell carcinoma.

In addition to histology, the frequency of acquired alterations in lung tumors can vary across sex, smoking status, and ancestral populations. For example, EGFR kinase domain mutations occur at a significantly higher frequency in lung adenocarcinomas from women compared with men, never-smokers compared with smokers, and East Asian populations compared with non-Asian populations. The higher rates of EGFR mutations in these populations have resulted in a corresponding higher rate of patients with clinical response to EGFR inhibitors. Studies examining mutational profiles of EGFR in tumors from patients with black ancestry have been less conclusive. Some studies have reported significantly lower frequencies of EGFR mutations in black patients compared with whites, while other studies did not observe any association of EGFR mutation status with ancestry or self-reported race. The interpretation of these findings has been limited by the presence of confounding factors such as sex and smoking status or by the fact that cohorts were pooled together from different clinical settings across different institutions. Moving beyond EGFR, a more recent study examining a panel of known driver mutations across several hundred NSCLC samples found a lower overall frequency of mutations in any gene in tumors from black patients compared with white patients. However, no studies have been performed that comprehensively characterize genomic alterations across hundreds of well-known cancer genes in a large number of tumors from both black and white populations.

To understand the relationship between tumoral genomic alterations and black ancestry, we profiled the exome sequences of 504 cancer genes in 509 lung cancers from black patients and white patients. These tumors were collected from patients in the same clinical setting and were well matched for clinical covariates. We did not identify enrichment of mutations or copy number changes in either population. These results suggest that targeted therapies and biomarker-driven clinical trials should significantly benefit patients in both groups.

### Methods

#### Sample Collection

Formalin-fixed paraffin-embedded (FFPE) lung tumor specimens were collected from black and white patients from Baptist Cancer Center in Memphis, Tennessee. All specimens were obtained from patients with approval from the relevant institutional review boards. Lung cancer cases from January 2004 to December 2012 were identified from a search of pathology department records across 3 health care systems (Baptist Memorial Health Care, Methodist-LeBonheur Healthcare, and St Francis Healthcare) in metropolitan Memphis. Race was categorized by patient self-identification. From each institution, specimens from black patients were paired with specimens from white patients by matching for histology, sex, and age.

#### Sequencing and Analysis

In total, 607 tumors underwent DNA isolation and sequencing of 504 cancer-related genes using the OncoPanel platform (eMethods in Supplement 1). Overall, 550 libraries with sufficient material were submitted for sequencing. Thirty tumors were excluded for failing sequencing quality control (eMethods in Supplement 1), and an additional 11 tumors were excluded due to insufficient clinical annotation, leaving a total of 509 tumors for analysis. Data were aligned and processed as previously described (eMethods in Supplement 1). Principal component analysis (PCA) and other statistical analyses are described in eMethods in Supplement 1. Because we did not have paired normal tissue for these patients, we excluded variants found in exomes of noncancer tissue using the Exome Aggregation Consortium database (eFigure 1 in Supplement 1).

### Results

#### Study Cohort

In total, 509 tumor specimens from 245 black patients and 264 white patients were analyzed (Table) (eTable 2 in Supplement 1).
This included 319 adenocarcinomas, 142 squamous cell carcinomas, 47 NSCLC of other histology, and 1 small-cell carcinoma. Twenty-nine samples were obtained from a biopsy of a nonlung site for a putative lung cancer metastasis. The majority of tumors were stage I (n = 305) or stage II (n = 98). We performed secondary pathological review for 380 lung cancers (eMethods in Supplement 1) and confirmed histological classification in 223 of 247 lung adenocarcinomas (90%) and 80 of 97 squamous cell carcinomas (82%).

### Association of Self-reported Race With Germline Variation

Previous studies have noted that self-reported clinical variables such as race and sex are not always congruent with underlying biological ancestry or sex as defined by genetic variation and structure. To examine the relationship between self-reported race and biological ancestry, we performed PCA on single-nucleotide variants (SNVs) likely to be inherited and not somatically acquired (eMethods and eFigure 1 in Supplement 1). Principal component 1 (PC1) was strongly associated with self-reported race \( (P = 7.6 \times 10^{-129}) \) (Figure 1). Two SNV profiles from reported white patients clustered with the samples from black patients, and 6 SNV profiles from reported black patients clustered with the white patient samples. Additionally, 3 samples were outliers along PC2. These II samples were excluded from all further analyses examining the association between ancestry and genomic alterations.

### Mutational Signatures

Various exogenous exposures or endogenous biological processes can contribute to the overall mutational load observed in lung tumors. We identified 4 mutational signatures present in the cohort \( ^{37} \) (eMethods in Supplement 1). These signatures were strongly correlated with previously characterized signatures of smoking (designated as SI4), aberrant APOBEC (apolipoprotein B messenger RNA editing enzyme, catalytic polypeptide-like) activity (SI2 and SI13), mismatch repair (MMR) and aging (SI6/SI15 and SI1/SI5), and ultraviolet (UV) radiation exposure (SI7) (Figure 2) (eFigure 2 in Supplement 1). The smoking-related signature SI4 was enriched for cytosine (C) > adenine (A) transversions and strongly associated with smoking status in lung adenocarcinoma \( (P = 9.1 \times 10^{-13}) \) (Figure 2) and to a lesser extent in lung squamous cell carcinoma \( (P = 0.08) \). No signatures were associated with ancestry, suggesting that the mutagenic processes are similar in lung tumors from black and white populations in lung adenocarcinoma.

### Table. Cohort Demographics by Self-reported Race

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Black (n = 245)</th>
<th>White (n = 264)</th>
<th>All (n = 509)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>116</td>
<td>120</td>
<td>236</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>129</td>
<td>144</td>
<td>273</td>
<td></td>
</tr>
<tr>
<td>Pack-years</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Median (range)</td>
<td>30 (0-120)</td>
<td>40 (0-120)</td>
<td>38 (0-120)</td>
<td></td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>26</td>
<td>23</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Former</td>
<td></td>
<td></td>
<td></td>
<td>.33</td>
</tr>
<tr>
<td>Heavy</td>
<td>92</td>
<td>116</td>
<td>208</td>
<td></td>
</tr>
<tr>
<td>Light</td>
<td>18</td>
<td>12</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heavy</td>
<td>86</td>
<td>88</td>
<td>174</td>
<td></td>
</tr>
<tr>
<td>Light</td>
<td>19</td>
<td>14</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>4</td>
<td>11</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Histology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSCLC</td>
<td></td>
<td></td>
<td></td>
<td>.94</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>150</td>
<td>169</td>
<td>319</td>
<td></td>
</tr>
<tr>
<td>Squamous</td>
<td>69</td>
<td>73</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>Large-cell undifferentiated</td>
<td>8</td>
<td>6</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Adenosquamous</td>
<td>6</td>
<td>7</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Large-cell neuroendocrine</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Undefined</td>
<td>6</td>
<td>4</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Small-cell carcinoma</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td>.60</td>
</tr>
<tr>
<td>I</td>
<td>142</td>
<td>163</td>
<td>305</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>53</td>
<td>45</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>28</td>
<td>32</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>10</td>
<td>9</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>12</td>
<td>15</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: NA, not applicable or no data available; NSCLC, non–small-cell lung cancer.

*Unless otherwise indicated, data are reported as number of patients.*
noma or lung squamous cell carcinoma ($P > .05$ for all) (eFigure 3 in Supplement 1).

**Lack of Association Between Genomic Alterations and Ancestry**

To assess the relationship between ancestry and mutation status, we examined the mutational frequencies in 313 lung adenocarcinomas and 138 lung squamous cell carcinomas. This set excluded 6 adenocarcinomas and 4 squamous cell carcinomas from patients whose ancestry was not confirmed on PC1 or were outliers on PC2 in the PCA analysis (Figure 1). Clinical variables were not significantly associated with ancestry in either set of tumors with the exception of smoking pack-years ($P < .05$ for all) (eTables 3 and 4 in Supplement 2). The total number of putative somatic mutations, the percentage of genes amplified, or the percentage of genes deleted per tumor were not associated with ancestry (eFigure 4 in Supplement 1). Similarly, other quality control metrics were not associated with ancestry (eFigure 4 in Supplement 1).

When examining previously characterized cancer genes in lung adenocarcinoma, the most commonly mutated oncoproteins included $KRAS$ (34%), $EGFR$ (14%), and $BRAF$ (7%), while the most commonly mutated tumor suppressors included $TP53$ (60%), $STK11$ (14%), and $ARID1A$ (9%) (Figure 3). The most commonly mutated lung squamous cell carcinoma oncogenes included $NFE2L2$ (12%), and $PIK3CA$ (9%) while tumor suppressors included $TP53$ (88%), $CDKN2A$ (17%), and $PTEN$ (14%) (eFigure 5 in Supplement 1). Using the Fisher exact test, we did not observe any statistical association between ancestry and mutation frequencies for any gene after correction for multiple hypothesis testing in either tumor type ($FDR Q > .05$) (eTables 5 and 6 in Supplement 2). We also compared the mutational frequencies in this cohort with the mutational frequencies found in tumors from TCGA and related studies. In lung adenocarcinoma, $TP53$ had a higher mutation frequency in tumors from black patients ($P = .02$), $NF1$ had a lower mutation frequency in tumors from white patients ($P = .045$), and $KEAP1$ had lower mutational frequencies in tumors from both populations ($P < .001$) in this cohort compared with tumors from TCGA ($n = 660$) (Figure 3). No significant differences were observed between the lung squamous cell carcinomas from this cohort and those from TCGA ($n = 484$) using the Fisher exact test ($P > .05$ for all) (eFigure 5 in Supplement 1).

We also determined if the number of copy number gains or losses for individual genes was significantly different between populations in either tumor type. Similar to the mutational profiles, we did not observe significant differences after correction for multiple hypothesis testing for the majority of genes (eFigures 6 and 7 in Supplement 1; eTables 7-10 in Supplement 2). The only exception is $SULT1A1$, which displayed a modestly significantly higher number of gains in tumors from black patients ($FDR Q = .06$). Finally, while the num-

---

**Figure 1. Association of Germline Variation With Self-reported Race Using PCA**

![PCA plot showing self-reported race distribution](image-url)
In lung adenocarcinoma, alterations in members of the Ras/Raf/RTK signaling pathways are of particular interest because they confer sensitivity to clinically approved targeted therapeutics.39 We first examined the mutational profile of EGFR and observed 46 mutations across 41 tumors (Figure 4A). Five tumors had more than one EGFR mutation including p.V769M/p.ELREA746del, p.S768I/p.V769L, p.V769M/p.L858R, p.T790M/p.L858R, and p.G719A/p.L861Q. In-frame indels in exon 19 were observed in both men and women and did not show enrichment in black vs white women as has been previously reported23 (Fisher exact $P = .80$). The overall frequency of activating EGFR mutations was not associated with...
ancestry ($P = .51$) (Figure 4B). Using lung adenocarcinomas from patients with sufficient clinical annotation ($n = 303$), we performed logistic regression examining EGFR mutation status as a function of ancestry, sex, and smoking status. A significant increase was observed in never-smokers compared with ever-smokers ($P = 2.85 \times 10^{-6}$), and a modest increase in females compared with males ($P = .08$). However, no significant association was observed between EGFR mutation status and ancestry ($P = .35$) (eFigure 8 in Supplement 1 and eTable 12 in Supplement 2).

To determine if the overall frequency of mutations from any gene in the pathway was associated with race, we designated lung adenocarcinomas that had a previously characterized activating somatic single-nucleotide variants, indel, amplification, or gene fusion in a known RTK/Ras/Raf driver as “oncogene positive” ($n = 128$) while the remaining lung adenocarcinomas were considered “oncogene negative” ($n = 185$). Only fusions with canonical breakpoints were considered. MET and ERBB2 were considered amplified if the tumor had a log2 copy number ratio greater than 1. $^{15}$ Ancestry was not significantly associated with oncogene negative status using a Fisher exact test ($P = .36$) (Figure 4B). This lack of association was also observed in a logistic regression model controlling for smoking status and sex ($P = .36$) (eFigure 8 in Supplement 1 and eTable 13 in Supplement 2). These results indicate that the proportions of Ras/Raf/RTK pathway alterations are similar be-
Lung Cancer Mutations Among Black and White Populations

Figure 4. Known Activating Ras/Raf/RTK Pathway Alterations in Tumors From Black and White Populations

We examined previously characterized activating mutations in genes that are members of the Ras/Raf/RTK pathway. A total of 23 activating single-nucleotide variations (SNVs) and 23 in-frame indels were found in EGFR across 313 lung adenocarcinomas. These mutations are grouped by smoking status, sex, and ancestry. One EGFR p.L858R mutation in tumor from a white female is not shown in this figure due to lack of smoking annotation. B, We did not observe a significant enrichment or depletion of known activating mutations in EGFR (left) or all activating mutations in the pathway (right) between tumors from black and white populations (P > .05 for all).

Discussion

The primary goal of this study was to determine if the frequencies of putative acquired alterations in lung tumors are different between black and white ancestries. We examined variants in 504 genes from 313 lung adenocarcinomas and 138 lung squamous cell carcinomas and did not observe associations between ancestry and somatic genomic alterations within each histological subtype. As the targeted sequencing panel used in this study was limited to cancer genes, we did not have the ability to perform comprehensive ancestral inference and classification. However, we were able to capture SNVs in the cancer genes likely to be inherited due to their relatively high frequency in nondisease tissue from the Exome Aggregation Consortium database. We observed that the variation among these markers was strongly associated with self-reported race and largely confirmed the status of the self-reported race annotation in the majority of cases. The misclustering of tumors from 8 patients could be owing to errors in clinical data collection or differences in the patient’s self-identified race from their biological ancestry.

We observed 4 mutational signatures that have been previously characterized in cancer including those related to smoke exposure, aberrant APOBEC activity, MMR deficiency or aging, and UV exposure. Similar to 3 tumors from TCGA, we observed a single lung squamous cell carcinoma from a white male patient that had a pattern of UV-associated mutations often observed in skin cancers and may represent a metastasis to the lung.

Several previous studies have compared mutational frequencies between lung tumors from black and white patients and have reported conflicting results. Most of these earlier studies have focused on EGFR owing to its importance as a clinically actionable target and its higher prevalence in East Asian populations. Studies by Leidner et al and Yang et al reported statistically significant lower frequencies of EGFR mutations in black patients compared with white patients; Bauml et al reported lower frequencies in black female smokers compared with white female smokers; and Bollig-Fischer et al found higher rates of EGFR exon 19 indels in black females compared with white females. In contrast, Cote et al, Reinersman et al, Yamaguchi et al, and Araujo et al did not observe any association of EGFR mutations with race. Of the studies that did not reach statistical significance, Reinersman et al found a higher frequency of EGFR mutations in blacks, while the others found lower frequencies. Over-all, the frequency of EGFR mutations in black patients across all of these studies range from 2% to 19%. Several confounding factors may contribute to these discrepancies. First, EGFR mutation frequency varies by smoking history and by sex. While some studies were balanced for clinical covariates between populations or used logistic regression to control for unbalanced variables, other studies did not take this into consideration. Furthermore, some studies used lung tumors from black patients that were collected in different clinical settings or were profiled with a different genomic technology than the tumors collected from white patients. Comparisons of populations collected in different clinical settings may be influenced by factors such as sample quality, inclusion criteria, regional smoking behavior, and accessibility to health care, which are difficult to quantify and control in statistical models.
More recent studies have moved beyond EGFR and KRAS to profiling mutations across panels of genes. Bollig-Fischer et al\(^2\) examined 214 coding mutations in 26 cancer genes and observed that EGFR exon 19 indels were enriched in black women (\(P = .048\)). Araujo et al\(^3\) examined 38 SNVs across 8 genes as well as EGFR and ERBB2 indels and ALK translocations. They did not find any significant differences in individual genes, but did observe an overall reduction in the frequency of tumors from black patients that had any driver alteration (\(P = .04\)). Due to the exploratory nature of these studies, no correction for multiple hypothesis testing was performed. However, foregoing this correction increases the risk of observing false positives due to chance and reduces the likelihood of replication in future studies. This may be a contributing factor in our inability to replicate the findings from these studies.

Our cohort of lung cancers was collected from institutions in the same region and clinical setting, profiled with the same DNA sequencing assay and was well matched for clinical variables with the exception of pack-years of smoking. Although the value of pack-years was associated with ancestry, we did not observe significant differences in the smoking-associated signature (S14) suggesting that differences between the populations in smoke exposure are minimal in our cohort. While this study profiled the largest number of genes compared with any of the previous studies, we cannot rule out the possibility that another gene outside of our assayed regions will have substantially different mutational frequencies between black and white populations. However, since the previously characterized recurrently altered genes in both tumor types\(^4\) did not display significant differences, we believe that the possibility of another gene displaying strong differences between populations is unlikely. We did observe a modest increase in the proportion of tumors with copy number gains in SULT1A1 from black patients. However, the variability in this region is likely germline as previously reported.\(^4\)

### Limitations

The mutational frequencies in this cohort were not significantly different from those observed in TCGA cohorts for most genes suggesting that the tumors in our cohort are not substantially different from those in TCGA. The only exceptions were in lung adenocarcinoma and included the tumor suppressors TP53, NFI1, and KEAP1. Because we did not have access to paired normal tissue, we limited the missense and splice site mutations in our analysis to those previously observed in human cancers as annotated in the COSMIC (Catalogue Of Somatic Mutations In Cancer) database. This may have led to the exclusion of some true novel somatic mutations and contributed to the lower mutational frequencies observed in these genes.

### Conclusions

In addition to the mutational frequencies of individual genes, we examined the differences in the frequency of previously curated Ras/Raf/RTK pathway alterations known to drive cellular transformation in lung adenocarcinoma. We again did not observe significant associations between ancestry and the mutational frequencies of the pathway members included in this panel. Therefore, we believe that clinical sequencing coupled with targeted therapies against these alterations will benefit patients in both populations. Overall, these results suggest that differences in exomic mutations do not contribute to the observed racial disparities in incidence and mortality between black and white populations. Further investigation into other genomic factors such as mutations in noncoding regions, epigenetic alterations, and gene expression changes or other socioeconomic variables beyond smoking behavior and access to health care may be required to fully explain these disparities.
Lung Cancer Mutations Among Black and White Populations

Original Investigation Research