The Association between Polluted Neighborhoods and *TP53*-Mutated Non–Small Cell Lung Cancer

Loretta Erhunmwunsee[1,2], Sam E. Wing[2], Jenny Shen[2], Hengrui Hu[2], Ernesto Sosa[2], Lisa N. Lopez[2], Catherine Raquel[2], Melissa Sur[2], Pilar Ibarra-Noriega[2], Madeline Currey[2], Janet Lee[3], Jae Y. Kim[1], Dan J. Raz[1], Arya Amini[4], Sagus Sampath[4], Marianna Koczywas[5], Erminia Massarelli[5], Howard L. West[5], Karen L. Reckamp[6], Rick A. Kittles[2], Ravi Salgia[5], Victoria L. Seewaldt[2], Susan L. Neuhausen[2], and Stacy W. Gray[2,5]

ABSTRACT

Background: Poor patients often reside in neighborhoods of lower socioeconomic status (SES) with high levels of airborne pollutants. They also have higher mortality from non–small cell lung cancer (NSCLC) than those living in wealthier communities. We investigated whether living in polluted neighborhoods is associated with somatic mutations linked with lower survival rates, i.e., *TP53* mutations.

Methods: In a retrospective cohort of 478 patients with NSCLC treated at a comprehensive cancer center between 2015 and 2018, we used logistic regression to assess associations between individual demographic and clinical characteristics, including somatic *TP53* mutation status and environmental risk factors of annual average particulate matter ($PM_{2.5}$) levels, and neighborhood SES.

Results: 277 patients (58%) had somatic *TP53* mutations. Of those, 45% lived in neighborhoods with "moderate" Environmental Protection Agency–defined $PM_{2.5}$ exposure, compared with 39% of patients without *TP53* mutations. We found significant associations between living in neighborhoods with "moderate" versus "good" $PM_{2.5}$ concentrations and minority population percentage [OR, 1.06; 95% confidence interval (CI), 1.04–1.08]. There was a significant association between presence of *TP53* mutations and $PM_{2.5}$ exposure (moderate versus good: OR, 1.66; 95% CI, 1.02–2.72) after adjusting for patient characteristics, other environmental factors, and neighborhood-level SES.

Conclusions: When controlling for individual- and neighborhood-level confounders, we find that the odds of having a *TP53*-mutated NSCLC are increased in areas with higher $PM_{2.5}$ exposure.

Impact: The link between pollution and aggressive biology may contribute to the increased burden of adverse NSCLC outcomes in individuals living in lower SES neighborhoods.

Introduction

Lung cancer is the leading cause of cancer-related death in the United States. Racial/ethnic minorities and those of lower socioeconomic status (SES) have higher rates of non–small cell lung cancer (NSCLC) incidence and mortality (1, 2). It has been assumed that the disparities in outcomes are caused primarily by higher rates of tobacco use and lower access to care in underserved populations; however, studies have shown that disparities persist after controlling for smoking and access to care (3–6). What has not yet been thoroughly evaluated is whether exposure to adverse social conditions – including increased air pollution – promote aggressive tumor biology, thereby increasing NSCLC outcome disparities.

Studies have reported that air pollution is associated with increased lung cancer risk and mortality, independent of cigarette smoking (7–9). While national average air pollution concentrations had been declining through 2016, levels have risen modestly since (10). Despite these overall declines, air pollution continues to be a major factor in increasing nonsmoking lung cancer incidence (11). The added threats of climate change and wildfires may make this problem worse, especially for vulnerable populations (12). Despite producing less air pollution, low SES individuals frequently live near toxin-producing industries and highways and have jobs with greater exposure to disease-associated chemicals (13, 14). Thus, lower SES communities in the United States and around the globe (15, 16) are exposed to higher concentrations of particulate matter (PM) in air pollution, which are associated with higher rates of NSCLC incidence and mortality, heart disease, and pediatric asthma (17–22).

PM in air pollution, commonly measured as the concentration of PM with an aerodynamic diameter of less than 2.5 μm ($PM_{2.5}$), represents a heterogeneous mixture of inorganic, organic, and biological compounds. $PM_{2.5}$ is classified by the International Agency for Research on Cancer (IARC) as a Group 1 (most severe) carcinogen. The associations of elevated $PM_{2.5}$ exposure with lower SES (13–16) and with worse mortality from NSCLC (19) are well established. However, little is known about the biological mechanism(s) underlying the relationship between air pollution and NSCLC mortality. It is essential to understand the impact of environmental and social conditions on tumor biology to understand causes of health inequity. Studies have shown that PM activates cell-cycle arrest and induces somatic mutations in cell-cycle checkpoint genes (23). PM also alters epigenetic modifications, including DNA methylation and expression of miRNAs. However, these results were largely ascertained without

[1]Department of Surgery, City of Hope Comprehensive Cancer Center, Duarte, California. [2]Department of Population Sciences, City of Hope Comprehensive Cancer Center, Duarte, California. [3]Vital Research, Los Angeles, California. [4]Department of Radiation Oncology, City of Hope Comprehensive Cancer Center, Duarte, California. [5]Department of Medical Oncology, City of Hope Comprehensive Cancer Center, Duarte, California. [6]Department of Medical Oncology, Cedars-Sinai Medical Center, Los Angeles, California.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Corresponding Author: Loretta Erhunmwunsee, City of Hope Comprehensive Cancer Center, 1500 East Duarte Road, Duarte, CA 91010. Phone: 626-218-7287; Fax: 626-218-7215; E-mail: LorettaE@coh.org

Cancer Epidemiol Biomarkers Prev 2021;30:1498–505

doi: 10.1158/1055-9965.EPI-20-1555

taking SES and neighborhood context into account. Therefore, these findings are limited in their ability to explain the population health impact of $PM_{2.5}$ on biology or oncologic disparities.

NSCLC comprises multiple genomic subtypes; some subtypes are more aggressive than others. For instance, *EGFR* mutations are associated with better overall survival in patients with NSCLC (24–27). *TP53* mutations, in contrast, are associated with shorter survival in patients with NSCLC (27, 28) and contribute to more aggressive NSCLC biology. Some *TP53* mutation hotspots are associated with cigarette smoking via DNA adduct formation, while others are not clearly associated with cigarette smoking (29, 30). Moreover, there is a strong association between aggressive NSCLC biology and cigarette smoking (31, 32), but whether mutations associated with worse outcomes are also linked to adverse social and environmental conditions remains unclear.

In a Chinese study, patients with NSCLC living in areas with higher levels of pollution had a higher frequency of *TP53* mutations (33), but no such study has been conducted in the United States. To further understand the relationships between adverse social conditions—such as living in an area with elevated $PM_{2.5}$ exposure—and aggressive somatic NSCLC biology, we conducted an exploratory cross-sectional study within a retrospective cohort at a comprehensive cancer center. We hypothesized that individuals with NSCLC who had greater exposure to air pollution [i.e., moderate vs. good $PM_{2.5}$ levels, as defined by the US Environmental Protection Agency (EPA)] would have higher odds of somatic *TP53* mutations. An understanding of this relationship could improve our ability to adequately stratify individuals who may be at high risk of developing aggressive disease, thus enabling more accurate screening of high-risk patients.

Materials and Methods

Study subjects

We included all patients with a primary NSCLC diagnosis who were treated at City of Hope Comprehensive Cancer Center (COH) from 2015 through 2018 and had documentation of somatic tumor sequencing in the electronic health record (EHR). Exclusion criteria were: (i) diagnosis of small cell lung cancer, carcinoid tumors, or sarcomas; (ii) *in situ* lung cancer; (iii) <18 years of age; and/or (iv) multiple primary NSCLCs with different somatic phenotypes. This study was approved by the COH Institutional Review Board. Patients included in the study provided written consent, and the study was conducted in accordance with the International Ethical Guidelines for Biomedical Research Involving Human Subjects (CIOMS).

Measures

Patient data

We obtained individual demographic data (i.e., age, sex, and race/ethnicity), disease characteristics (i.e., stage and histology), smoking history, and home address from the EHR and the accredited COH hospital-based cancer registry. Smoking history information is assessed by clinicians to aid in clinical decision making and was collected either in clinic notes or patient surveys. We obtained somatic test results by reviewing testing reports in the EHR. Our institutional policy mandates that all third-party laboratory reports be uploaded into the EHR. We included test results from multiple samples for each patient. For patients with discrepant genetic reports, study staff prioritized findings from tissue over blood-based assays. The laboratories that conducted testing are listed in Supplementary Table S1. All patients with somatic *TP53* mutations were included in our main analysis. We utilized the OncoKB knowledge base (34) to better characterize the spectrum of *TP53* variants within our cohort and evaluate the proportion of loss-of-function variants to those of unknown significance (Supplementary Table S2).

Neighborhood-level air pollution and socioeconomic data

Residential addresses of patients at the time of diagnosis were geocoded into corresponding coordinates using the Spatialitics Health Geocoder (Spatialitics LLC). Environmental risk factors including $PM_{2.5}$, ozone, National Air Toxics Assessment (NATA) air toxics cancer risk, and traffic proximity, as well as neighborhood (census block group)-level demographic indicators (percent minority and percent less than high school education) were extracted from the EPA's Environmental Justice Screening and Mapping Tool (EJSCREEN). EJSCREEN was developed by the EPA to track environmental and health risks across the country. EJSCREEN leverages data from the US Census, the EPA, and the US Department of Transportation to generate these estimates (35). We assigned the annual average exposure to $PM_{2.5}$ and ozone in the year or two prior to diagnosis. EJSCREEN has only made the following vintages of $PM_{2.5}$ data available: 2011, 2012, 2013, 2014, and 2016. Therefore, patients diagnosed in 2015 and 2016 were assigned 2014 values and patients diagnosed in 2017 and 2018 were assigned 2016 values. NATA air toxics cancer risk is an estimation of lifetime cancer risk due to hazardous air pollution inhalation. These risks are based on modeled estimates of air toxic concentrations within a given area and combined with epidemiologic evidence of the pollutant's carcinogenicity to generate an estimate of overall cancer risk. Traffic proximity is a function of traffic counts on nearby major roads inversely weighted by the distance from the roadway. NATA, $PM_{2.5}$ and ozone estimates are generated at the census tract-level and EJSCREEN assigns all block groups within a given census tract the same exposure to these pollution indices.

Percent minority is defined as the percent of individuals within a block group who list their racial status as a race other than white alone. Percent less than high school education is defined as the percent of individuals within a block group age 25 and over with less than a high school degree. The block group in which the patient's home address was located was used to assign exposures. Average annual $PM_{2.5}$ levels for patients in the study ranged from 5.34 $\mu g/m^3$ to 16.3 $\mu g/m^3$, with approximately 84% of patients having a $PM_{2.5}$ level between 10.0 to 14.0 $\mu g/m^3$. Using EPA standards as a guide, we dichotomized average annual $PM_{2.5}$ and ozone concentrations as "good" ($PM_{2.5}$: 0–12.0 $\mu g/m^3$; ozone: 0–54 ppb) and "moderate" ($PM_{2.5}$: 12.1–35.4 $\mu g/m^3$; ozone 55–70 ppb). Grouping exposure levels according to EPA standards provided an interpretation of results that follows the conventions for air quality. There are several concentration levels beyond moderate, including unhealthy to hazardous. However, within the United States, it is rare for an average annual $PM_{2.5}$ level to be within these concentrations. NATA cancer risk and traffic proximity were dichotomized by a median split.

Statistical analysis

Using logistic regression, we assessed the associations of various patient characteristics, environmental risk factors, and neighborhood-level demographics with somatic *TP53* mutations. In adjusted models, we controlled for covariates in three modeling stages. The first model included demographic and clinical variables (i.e., age, sex, smoking status, race/ethnicity, stage, histology), the second added neighborhood-level environmental air pollution exposures (i.e., $PM_{2.5}$, ozone, NATA cancer risk, and traffic proximity), and the third added metrics of neighborhood-level sociodemographics (i.e., percent minority

population and educational attainment less than high school education). We also conducted a sensitivity analysis comparing only those patients who had a *TP53* variant classified as loss-of-function or likely-loss-of-function in OncoKB (Supplementary Table S2) to those who tested *TP53*-mutation negative. Furthermore, we explored different operationalizations of $PM_{2.5}$, examining linear and threshold relationships with the presence of a *TP53* mutation (Supplementary Table S3). Subgroup analysis were conducted on patients diagnosed only in 2017 and 2015 to assess the impact of using only patients for whom a one-year lag was available, as opposed to one- and two-year lags with the whole sample. ORs and 95% confidence intervals (CI) were calculated for all models. All tests were two-sided, and statistical significance was defined as $P < 0.05$. To explore possible heterogeneity of effects between subgroups, we included an interaction term between $PM_{2.5}$ exposure and covariates of interest, including smoking status, sex, and histology. All statistical analyses were performed using SAS 9.4 (SAS Institute).

Results

Of the 694 patients identified in our initial sampling frame, 505 underwent testing for somatic *TP53* mutations. Twenty-seven of these patients were missing covariate data and were excluded, leaving 478 patients included in the final analysis (Supplementary Fig. S1). Ninety-three percent of patients resided in California, with 64% of these individuals reporting a home address in Los Angeles county. Those patients not residing in California lived in Arizona, Georgia, Maryland, Nevada, and Ohio. Males (Fisher exact $P = 0.03$), those with late stage (Fisher exact $P = 0.005$), and with adenocarcinomas (Fisher exact $P < 0.001$) were more likely to have received somatic *TP53* sequencing (Supplementary Table S4). Tumor samples from 62% of patients were tested within 90 days of diagnosis. We identified 344 *TP53* variants in 277 patients. We found that 23% (79/344) of these mutations were located in pathogenic *TP53* hotspot residues as defined in previous literature (29, 30, 32), with the highest mutation frequency of 7% (24/344) at amino acid site R273. In addition, 40% (139/344) of the variants identified in our cohort were classified by OncoKB as loss-of-function or likely loss-of-function mutations (Supplementary Table S2). The remaining variants in our cohort were not classified by OncoKB and were excluded from this Supplementary table but retained in the main analysis. Patient demographics and disease characteristics are summarized in **Table 1**. The mean age of patients at diagnosis was 67 years. Fifty percent of the cohort was female, 55% non-Hispanic White, and 37% reported never smoking cigarettes. Most patients were diagnosed with Stage IV NSCLC (70%) and had adenocarcinoma (85%). Forty-three percent of the participants lived in block groups with concentrations of $PM_{2.5}$ defined by the EPA as moderate and 277 (58%) had an identified *TP53* somatic alteration. We used the sociodemographic variables of percent less than high school education and percent minority population at the block group level. We excluded neighborhood-level income as a covariate because of its high degree of collinearity with other SES variables.

Neighborhood conditions and $PM_{2.5}$ levels

Measures of neighborhood sociodemographic variables were associated with concentrations of $PM_{2.5}$. In crude logistic regression analyses, a 1% increase in a census block group's minority population was associated with an OR of 1.06 (95% CI, 1.05–1.07) of being in the moderate versus good $PM_{2.5}$ category. When adjusting for patient demographics and disease characteristics, this estimate remained statistically significant (OR 1.06, 95% CI, 1.04–1.08) (Supplementary Table S5). Neighborhood percent minority and percent less than a high school education were highly correlated (Pearson correlation coefficient $r = 0.73$), but no statistically significant relationship between neighborhood educational attainment and $PM_{2.5}$ was detected in the fully adjusted model. In contrast, when omitting percent minority from the adjusted model, percent less than high school education was significantly associated with elevated $PM_{2.5}$ exposure (OR, 1.04; 95% CI, 1.02–1.07), indicating collinearity between the neighborhood-level education and minority variables. In addition, in the fully adjusted model, living in a neighborhood above versus below the median NATA cancer risk was strongly associated with elevated $PM_{2.5}$ exposure (OR, 4.13, 95% CI 2.43–7.03).

TP53

Patients with *TP53* mutations were more likely to be exposed to higher levels of air pollution, with 45% of patients with *TP53* mutations living in neighborhoods with moderate $PM_{2.5}$ exposure compared to 39% of patients without a *TP53* mutation. Individuals with *TP53* mutations had three additional pack-years of smoking on average compared to those without. The covariate-adjusted logistic regression analyses on the odds of having a *TP53* mutation are displayed in **Table 2**. We found that *TP53* mutations were significantly associated with higher $PM_{2.5}$ exposure (moderate vs. good: OR, 1.68; 95% CI, 1.06–2.66) after adjusting for patient characteristics (including smoking status) and other environmental factors (Model 2). After adding neighborhood-level sociodemographic factors in Model 3, this association with $PM_{2.5}$ remained significant (OR, 1.66; 95% CI, 1.02–2.72). Adjusting for covariates, cigarette smoking was also positively associated with *TP53* mutations (current vs. never smokers, OR, 2.43; 95% CI, 1.29–4.60 and former vs. never smokers, OR, 1.65; 95% CI, 1.05–2.59). In addition, we did not detect any statistically significant interactions between $PM_{2.5}$ exposure and sex, smoking, or histology on either the multiplicative or additive scales. Men were less likely to be in the higher $PM_{2.5}$ category (OR, 0.71; 95% CI, 0.52–0.96) and more likely to be former or current smokers (Fisher exact $P < 0.001$) relative to women. We did not detect a relationship between $PM_{2.5}$ and *TP53* when $PM_{2.5}$ was treated either as a linear variable or in quantiles (e.g., tertiles or quartiles) (Supplementary Table S3). When restricting to only *TP53* variants classified as loss-of-function or likely-loss-of-function in OncoKB, the association with $PM_{2.5}$ remained consistent (OR, 1.65; 95% CI, 1.01–2.70). Among patients for whom one-year lag exposure data were available (i.e., patients diagnosed in 2015 and 2017), we found n strengthened adjusted association relative to the main result (OR, 2.98; 95% CI, 1.38–6.43).

Discussion

In this study, we investigated the associations of $PM_{2.5}$ exposure with both neighborhood social conditions and somatic *TP53* mutations. We found that $PM_{2.5}$ exposure in our population varied and was associated with demographic and neighborhood social characteristics. We also found an association between air pollution before the time of diagnosis and *TP53*-mutated NSCLC. A previous study (33) reported a link between highly polluted regions and specific somatic NSCLC mutations in a cohort of patients living in China. The study performed whole genome sequencing of NSCLCs in "highly polluted regions" and found that patients in these regions had 3 times as many mutated genes, including *TP53*, as those in control regions. However, the study did not consider the impact of specific pollutants at the home address, but instead compared NSCLC patients living in two municipalities with historically high rates of smoky coal use for indoor cooking and

Table 1. Demographic and exposure characteristics for patients who had somatic *TP53* testing.

Characteristic	Value[a]	
	***TP53* Mutation (*n* = 277)**	**No *TP53* mutation (*n* = 201)**
Age at diagnosis, Mean (SD)	66 (12)	68 (12)
Sex		
Female	143 (52%)	95 (47%)
Male	134 (48%)	106 (53%)
Race/Ethnicity		
Asian	78 (28%)	69 (34%)
Black	13 (5%)	9 (4%)
Non-Hispanic White	157 (57%)	107 (53%)
Hispanic White	29 (10%)	16 (8%)
Stage		
I–II	33 (12%)	31 (15%)
III	50 (18%)	29 (14%)
IV	194 (70%)	141 (70%)
Histology		
Adenocarcinoma	230 (83%)	178 (89%)
Squamous	26 (9%)	12 (6%)
Other	21 (8%)	11 (5%)
Smoking status		
Current	50 (18%)	23 (11%)
Former	136 (49%)	91 (45%)
Never	91 (33%)	87 (43%)
Pack-years, mean (SD)[b]	21 (25)	18 (26)
Particulate matter exposure		
Good (0–12.0 $\mu g/m^3$)	152 (55%)	123 (61%)
Moderate (12.1–35.4 $\mu g/m^3$)	125 (45%)	78 (39%)
Ozone		
Good (0–54 ppb)	142 (59%)	102 (51%)
Moderate (55–70 ppb)	135 (41%)	99 (49%)
NATA Cancer risk		
$<50^{th}$ Percentile	138 (50%)	98 (49%)
$\geq 50^{th}$ Percentile	139 (50%)	103 (51%)
Traffic proximity		
$<50^{th}$ Percentile	134 (48%)	100 (50%)
$\geq 50^{th}$ Percentile	143 (52%)	101 (50%)
Percent minority population, mean (SD)	60% (0.23)	60% (0.23)
Percent less than high school education, mean (SD)	14% (0.10)	15% (0.11)

[a]Data are presented as number (percentage) of patients unless otherwise indicated.
[b]Average pack-year was calculated among 473 patients due to missing data.

heating (highly polluted region) to those patients living in areas with air pollution levels more representative of the rest of the country (control region). In addition, they did not account for social determinant data and their sample size was smaller than our study ($N = 164$). Our study is the first U.S.-based cohort to evaluate both key neighborhood social determinant data and individual patient EHR data, including detailed smoking data and report their impact on NSCLC biology.

$PM_{2.5}$ is a widely studied air pollutant whose concentrations have been consistently linked with SES. Lower SES communities in the United States and globally (15, 16) are exposed to higher concentrations of $PM_{2.5}$. These higher levels are tied to worse disease rates and mortality. Furthermore, $PM_{2.5}$ is also classified as a Group 1 human carcinogen by the IARC and is associated with high rates of NSCLC incidence and mortality.

Ambient air pollution is a global concern, with an estimated 12.8% of lung cancer deaths attributed to fine PM exposure (36). These effects on mortality are particularly notable in rapidly growing countries like India and China (37, 38). Vehicular travel and industrial pollution, as well as adverse neighborhood social conditions, are steadily increasing globally (39–41). The impact of these trends also have the potential to be magnified by climate change and the resultant increase in wildfire activity (12). Thus, the association between elevated $PM_{2.5}$ levels and aggressive biology has impactful public health implications. This relationship may help explain the increased burden of adverse health outcomes in NSCLC that is observed in neighborhoods comprised of lower SES populations (42–45).

The biological mechanism underpinning if and how exposure to air pollution alters NSCLC biology is not yet clear. Prior studies have demonstrated an association between exposure to air pollutants and *TP53* mutations in mouse models and human lung cell lines (46, 47). Specifically, *TP53* mutations have been observed in mouse cell lines that were experimentally exposed to different environmental toxins, including benzo(a)pyrene (BaP) and 3-nitrobenzanthrone (3-NBA). Exposure to 3-NBA, which is linked with diesel exhaust, was associated with mutations similar to those seen in tobacco smokers with NSCLC,

Table 2. Adjusted ORs (95% CIs) of *TP53* mutations.

Variable	Model 1	Model 2	Model 3
Age at diagnosis (per year)	0.99 (0.97–1.00)	0.99 (0.97–1.00)	0.99 (0.97–1.00)
Cigarette smoking			
Never smoker	Ref	Ref	Ref
Former smoker	1.61 (1.03–2.51)	1.65 (1.05–2.59)	1.65 (1.05–2.59)
Current smoker	2.28 (1.22–4.26)	2.30 (1.23–4.32)	2.43 (1.29–4.60)
Sex			
Male	Ref	Ref	Ref
Female	1.35 (0.92–2.00)	1.37 (0.92–2.02)	1.35 (0.91–2.01)
Race/Ethnicity			
Non-Hispanic White	Ref	Ref	Ref
Asian	0.93 (0.60–1.44)	0.82 (0.51–1.32)	0.79 (0.48–1.33)
Black	0.90 (0.37–2.23)	0.94 (0.38–2.35)	0.95 (0.37–2.44)
Hispanic White	1.28 (0.65–2.52)	1.15 (0.57–2.31)	1.29 (0.62–2.70)
Stage			
I–II	0.77 (0.44–1.35)	0.78 (0.44–1.36)	0.77 (0.44–1.36)
III	1.11 (0.66–1.87)	1.14 (0.67–1.93)	1.13 (0.67–1.93)
IV	Ref	Ref	Ref
Histology			
Squamous	Ref	Ref	Ref
Adeno	0.57 (0.27–1.18)	0.53 (0.25–1.11)	0.50 (0.24–1.06)
Other	0.79 (0.29–2.19)	0.76 (0.27–2.13)	0.73 (0.26–2.04)
$PM_{2.5}$			
Good	-	Ref	Ref
Moderate	-	1.68 (1.06–2.66)	1.66 (1.02–2.72)
Ozone			
Good	-	Ref	Ref
Moderate	-	1.01 (0.68–1.50)	1.00 (0.67–1.50)
NATA Cancer risk			
$<50^{th}$ Percentile	-	Ref	Ref
$\geq 50^{th}$ Percentile	-	0.80 (0.51–1.27)	0.83 (0.52–1.34)
Traffic proximity			
$<50^{th}$ Percentile	-	Ref	Ref
$\geq 50^{th}$ Percentile	-	1.12 (0.75–1.67)	1.18 (0.78–1.78)
Percent minority population	-	-	1.01 (0.99–1.02)
Percent less than high school education	-	-	0.98 (0.95–1.01)

such as G>T transversions (47). In that study, 38% of cell lines exposed to 3-NBA possessed a G>T transversion in *TP53*. The exact mechanistic link between $PM_{2.5}$ and *TP53* mutations is unclear. In human lung cells, $PM_{2.5}$ has been found to disrupt the *TP53*–retinoblastoma protein signaling pathway, resulting in the proliferation of tumor cells (46). An *in vitro* study of human epithelial cells found that $PM_{2.5}$ exposure resulted in hypomethylation of the *TP53* promoter region and inhibition of *TP53* expression (48).

Moderate levels of $PM_{2.5}$ contain higher concentrations of polycyclic aromatic hydrocarbons (PAH), including benzo[a]pyrene (BaP) and polar compounds, which are highly mutagenic (33). Higher concentrations of BaP have been associated with several mutations, including those in *TP53*. Yu and colleagues (33) suggested that the mutations are induced by DNA adducts that are formed by the release of reactive intermediates when BaP and other PAHs are metabolized. Other studies have found that inhalation of $PM_{2.5}$ particles may attract lymphocytes to tissues, resulting in angiogenesis and inflammation that could promote tumor growth (36, 49).

Our findings suggest that living in a polluted neighborhood with moderate $PM_{2.5}$ levels (12.1–35.4 $\mu g/m^3$), which are considered "acceptable" by the EPA, is associated with higher odds of *TP53* mutations. This link between exposure to air pollution and aggressive NSCLC phenotypes may contribute to the poor outcomes observed in NSCLC patients from underserved populations. However, other explanations for the link between adverse social conditions and *TP53* mutations may include secondhand smoking exposure and additional, unmeasured environmental exposures. Our findings also suggest that exposure to moderate levels of $PM_{2.5}$ may not, in fact, be "acceptable" and may warrant reconsideration of current $PM_{2.5}$ classifications to reflect the potential health implications of chronic exposure.

Our study also found an association of elevated $PM_{2.5}$ levels with neighborhood minority population and percent less than high school education. These findings reflect the high correlation between social conditions and pollution that other studies have shown (13–16). Despite finding an association between $PM_{2.5}$ and both neighborhood composition and *TP53* mutations, our results did not reveal a direct correlation between the neighborhood social conditions and these mutations. There is a complicated relationship between neighborhood conditions, lung cancer outcomes, and biology. We have shown that vulnerable communities have elevated $PM_{2.5}$ levels and that higher $PM_{2.5}$ levels are associated with aggressive NSCLC biology. Our findings suggest that $PM_{2.5}$ may contribute to the link between lower SES neighborhoods and NSCLC mutations that are associated with poor outcomes. Nevertheless, despite finding an association between $PM_{2.5}$ and both neighborhood composition and aggressive biology, we

did not find that $PM_{2.5}$ was a true mediator between these two variables. Instead, $PM_{2.5}$ is likely one of many intermediaries that facilitates the complicated relationship between neighborhood conditions and NSCLC biology.

Limitations and strengths

The study had several limitations. Patients' addresses were collected at the time of diagnosis at COH and we were unable to assess patients' historical exposure to carcinogens, making it difficult to establish temporality. In addition, exposure misclassification is possible, as people may move frequently throughout their lifetime. Nevertheless, this misclassification is likely to be independent and non-differential with respect to outcome status. Therefore, the bias due to this exposure misclassification is likely toward the null, leading to a conservative estimation of the true effect size (50). Furthermore, the EJSCREEN tool pulls data from census block group information. Because this information is an average of factors within an area, it does not reflect individual addresses and is therefore subject to measurement error, as the assumption is that exposures at the census block group level accurately reflect a patient's individual exposure profile. However, it does address our research question, as we investigated neighborhood effects independent of individual-level characteristics, mirroring previous research (51–55). In addition, EJSCREEN did not offer $PM_{2.5}$ data for all of the recent years of interest. We found that restricting analyses to patients for whom one-year lag exposure data was available generated a stronger measure of association, indicating that increased lag time for some patients may have biased our overall association to the null. The nested nature of patients within block groups was given consideration, but with nearly all census block groups having fewer than five patients, multilevel modeling was not possible. Another limitation is that neighborhood-level education is only one component of neighborhood SES, and excluding neighborhood income from our analysis due to collinearity may restrict our ability to comprehensively understand how neighborhood SES impacts biology. We also acknowledge that our cohort is unique in that a third of the patients were nonsmoking Asians, and only a few Hispanics and African Americans were included. This racial composition may impact our ability to generalize these results to populations with different racial/ethnic compositions. Finally, this is a single institution study with moderate sample size, limiting the power to detect differences between subgroups and possible interactions. For example, a previous study of over 400,0000 participants found a synergistic interaction between smoking and exposure to air pollution in lung cancer mortality (56) which we did not observe. Our study sample size was sufficient to detect a significant association between $PM_{2.5}$ and *TP53* mutations.

The study had several strengths. The use of data from patients' EHR, including tumor sequencing results and detailed smoking data, is more extensive than what is readily available from the typical cancer registry. Also, our institution had previously enumerated the distribution of various NSCLC mutations, as well as the survival of patients who harbor them (27); thus, we have established expertise in their evaluation. Moreover, we combined both individual- and neighborhood-level variables, allowing for the integration of multilevel risk factors. Finally, EJSCREEN offers high spatial resolution of both neighborhood exposures at the level of census block groups, which are approximately one-third the size of census tracts. Our findings add evidence of the mutagenic role of air pollution in lung cancer, providing a platform for future studies, including the exploration of multilevel models.

Future research should include prospective studies to validate the etiologic role of environmental and social determinants of health in the development of aggressive NSCLC biology. Inclusion of multilevel factors and various biological profiles will allow for a more comprehensive understanding of this complex social, environmental, and biological relationship. Additionally, given the lack of literature on the association between NSCLC outcomes and specific *TP53* variants, future research should focus on variant-level differences that are driving adverse outcomes.

Conclusion

Our study reports an association between $PM_{2.5}$ and aggressive NSCLC biology. The integration of socioeconomic, environmental, biological, and clinical factors is paramount as we aim to eliminate health inequities. In the future, this comprehensive information may improve our ability to adequately stratify individuals who may be at high risk of developing aggressive disease, allowing for more accurate screening of high-risk patients.

Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH https://doi.org/10.13039/100000002.

Authors' Disclosures

L. Erhunmwunsee reports grants from NIH K12 award during the conduct of the study, as well as grants from AstraZeneca Pharmaceuticals outside the submitted work. D.J. Raz reports personal fees from Roche and AstraZeneca outside the submitted work. E. Massarelli reports personal fees from AstraZeneca and Merck outside the submitted work. K.L. Reckamp reports personal fees and non-financial support as a consultant (honoraria to self) from Calithera, Euclises, Guardant, Precision Health, Amgen, AstraZeneca, Blueprint, Boehringer Ingelheim, Daiichi Sankyo, EMD Serono, Genentech, Janssen, Lilly, Merck KGA, Seattle Genetics, Takeda, and Tesaro and grant/research support to institution from AbbVie, Acea, Adaptimmune, Guardant, Molecular Partners, Seattle Genetics, Boehringer Ingelheim, Bristol Myers Squibb, Genentech, GlaxoSmithKline, Janssen, Loxo Oncology, Spectrum, Takeda, Xcovery, Zeno, Calithera, Daiichi Sankyo, and Elevation Oncology outside the submitted work. S.W. Gray reports personal fees from TripTych Health Partners outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

L. Erhunmwunsee: Conceptualization, resources, data curation, formal analysis, supervision, funding acquisition, visualization, methodology, writing–original draft, project administration, writing–review and editing. **S.E. Wing:** Data curation, formal analysis, supervision, writing–original draft, project administration, writing–review and editing. **J. Shen:** Conceptualization, methodology, writing–original draft, project administration, writing–review and editing. **H. Hu:** Data curation, formal analysis, validation, visualization. **E. Sosa:** Writing–original draft, writing–review and editing. **L.N. Lopez:** Data curation, writing–original draft, writing–review and editing. **C. Raquel:** Data curation, writing–original draft. **M. Sur:** Conceptualization, methodology, writing–review and editing. **P. Ibarra-Noriega:** Data curation, formal analysis, writing–original draft. **M. Currey:** Conceptualization, resources, data curation, formal analysis, supervision, funding acquisition, visualization, methodology, writing–original draft, project administration, writing–review and editing. **J. Lee:** Conceptualization, resources, investigation, methodology, writing–review and editing. **J.Y. Kim:** Conceptualization, resources, methodology, writing–review and editing. **D.J. Raz:** Conceptualization, resources, supervision, methodology, writing–original draft, project administration, writing–review and editing. **A. Amini:** Conceptualization, resources, data curation, formal analysis, methodology, writing–review and editing. **S. Sampath:** Conceptualization, resources, methodology, writing–review and editing. **M. Koczywas:** Conceptualization, resources, methodology, writing–review and editing. **E. Massarelli:** Conceptualization, resources, methodology, writing–review and editing. **H.L. West:** Conceptualization, resources, methodology, writing–review and editing. **K.L. Reckamp:** Conceptualization, resources, writing–review and editing. **R.A. Kittles:** Writing–review and editing. **R. Salgia:** Conceptualization, resources, methodology, writing–review and editing. **V.L. Seewaldt:** Conceptualization, resources, methodology, writing–review and editing. **S.L. Neuhausen:** Conceptualization, resources, methodology, writing–review and editing. **S.W. Gray:** Conceptualization,

supervision, methodology, writing–original draft, project administration, writing–review and editing.

Acknowledgments

We thank Kerin Higa, PhD, for reviewing and editing the manuscript. Research reported in this publication was supported by the NCI of the NIH under award number K17CA001727. One hundred percent of the total costs for this publication were by financed a nongovernmental source, for a total of $30,000.00.

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

Received October 30, 2020; revised January 13, 2021; accepted May 19, 2021; published first June 4, 2021.

References

1. Erhunmwunsee L, Joshi MB, Conlon DH, Harpole DH Jr. Neighborhood-level socioeconomic determinants impact outcomes in nonsmall cell lung cancer patients in the Southeastern United States. Cancer 2012;118:5117–23.
2. Lynge E. Unemployment and cancer: a literature review. IARC Sci Publ 1997; 343–51.
3. Abdelsattar ZM, Hendren S, Wong SL. The impact of health insurance on cancer care in disadvantaged communities. Cancer 2017;123:1219–27.
4. Forrest LF, Adams J, Wareham H, Rubin G, White M. Socioeconomic inequalities in lung cancer treatment: systematic review and meta-analysis. PLoS Med 2013;10:e1001376.
5. Johnson AM, Hines RB, Johnson JA III, Bayakly AR. Treatment and survival disparities in lung cancer: the effect of social environment and place of residence. Lung Cancer 2014;83:401–7.
6. Everist MM, Howard LE, Aronson WJ, Kane CJ, Amling CL, Cooperberg MR, et al. Socioeconomic status, race, and long-term outcomes after radical prostatectomy in an equal access health system: results from the SEARCH database. Urol Oncol 2019;37:289.e211–7.
7. Pope Iii CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 2002;287:1132–41.
8. Raaschou-Nielsen O, Andersen ZJ, Beelen R, Samoli E, Stafoggia M, Weinmayr G, et al. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). Lancet Oncol 2013;14:813–22.
9. Turner MC, Krewski D, Pope CA, Chen Y, Gapstur SM, Thun MJ. Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers. Am J Respir Crit Care Med 2011;184:1374–81.
10. United States Environmental Protection Agency. Particulate matter (PM2.5) trends; 2020. Available from: https://www.epa.gov/air-trends/particulate-matter-pm25-trends#pmnat.
11. Turner MC, Andersen ZJ, Baccarelli A, Diver WR, Gapstur SM, Pope CA, et al. Outdoor air pollution and cancer: an overview of the current evidence and public health recommendations. CA Cancer J Clin 2020;70:460–79.
12. Liu JC, Mickley LJ, Sulprizio MP, Dominici F, Yue X, Ebisu K, et al. Particulate air pollution from wildfires in the Western US under climate change. Clim Change 2016;138:655–66.
13. Bowe B, Xie Y, Yan Y, Al-Aly Z. Burden of cause-specific mortality associated with PM2.5 air pollution in the United States. JAMA Netw Open 2019;2: e1915834.
14. Tessum CW, Apte JS, Goodkind AL, Muller NZ, Mullins KA, Paolella DA, et al. Inequity in consumption of goods and services adds to racial-ethnic disparities in air pollution exposure. Proc Natl Acad Sci U S A 2019;116: 6001–6.
15. Hajat A, Hsia C, O'Neill MS. Socioeconomic disparities and air pollution exposure: a global review. Curr Environ Health Rep 2015;2:440–50.
16. Naess O, Piro FN, Nafstad P, Smith GD, Leyland AH. Air pollution, social deprivation, and mortality: a multilevel cohort study. Epidemiology 2007;18: 686–94.
17. Garcia E, Berhane KT, Islam T, McConnell R, Urman R, Chen Z, et al. Association of changes in air quality with incident asthma in children in California, 1993–2014. JAMA 2019;321:1906–15.
18. Hayes RB, Lim C, Zhang Y, Cromar K, Shao Y, Reynolds HR, et al. PM2.5 air pollution and cause-specific cardiovascular disease mortality. Int J Epidemiol 2020;49:25–35.
19. Huang F, Pan B, Wu J, Chen E, Chen L. Relationship between exposure to PM2.5 and lung cancer incidence and mortality: a meta-analysis. Oncotarget 2017;8: 43322–31.
20. Kaufman JD, Adar SD, Barr RG, Budoff M, Burke GL, Curl CL, et al. Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): a longitudinal cohort study. Lancet 2016;388:696–704.
21. Pun VC, Kazemiparkouhi F, Manjourides J, Suh HH. Long-term PM2.5 exposure and respiratory, cancer, and cardiovascular mortality in older US adults. Am J Epidemiol 2017;186:961–9.
22. Thurston GD, Burnett RT, Turner MC, Shi Y, Krewski D, Lall R, et al. Ischemic heart disease mortality and long-term exposure to source-related components of U.S. fine particle air pollution. Environ Health Perspect 2016;124:785–94.
23. Longhin E, Holme JA, Gutzkow KB, Arlt VM, Kucab JE, Camatini M, et al. Cell cycle alterations induced by urban PM2.5 in bronchial epithelial cells: characterization of the process and possible mechanisms involved. Part Fibre Toxicol 2013;10:63–63.
24. Li W-Y, Zhao T-T, Xu H-M, Wang Z-N, Xu Y-Y, Han Y, et al. The role of EGFR mutation as a prognostic factor in survival after diagnosis of brain metastasis in non-small cell lung cancer: a systematic review and meta-analysis. BMC Cancer 2019;19:145.
25. Kosaka T, Yatabe Y, Endoh H, Kuwano H, Takahashi T, Mitsudomi T. Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. Cancer Res 2004;64:8919–23.
26. Sonobe M, Manabe T, Wada H, Tanaka F. Mutations in the epidermal growth factor receptor gene are linked to smoking-independent, lung adenocarcinoma. Br J Cancer 2005;93:355–63.
27. Mambetsariev I, Wang Y, Chen C, Nadaf S, Pharaon R, Fricke J, et al. Precision medicine and actionable alterations in lung cancer: a single institution experience. PLoS One 2020;15:e0228188.
28. Gu J, Zhou Y, Huang L, Ou W, Wu J, Li S, et al. TP53 mutation is associated with a poor clinical outcome for non-small cell lung cancer: Evidence from a meta-analysis. Mol Clin Oncol 2016;5:705–13.
29. Barta JA, McMahon SB. Lung-enriched mutations in the p53 tumor suppressor: a paradigm for tissue-specific gain of oncogenic function. Mol Cancer Res 2019;17:3.
30. Shajani-Yi Z, de Abreu FB, Peterson JD, Tsongalis GJ. Frequency of somatic TP53 mutations in combination with known pathogenic mutations in colon adenocarcinoma, non-small cell lung carcinoma, and gliomas as identified by next-generation sequencing. Neoplasia 2018;20:256–62.
31. Gibelin C, Couraud S. Somatic alterations in lung cancer: do environmental factors matter? Lung Cancer 2016;100:45–52.
32. Mogi A, Kuwano H. TP53 mutations in nonsmall cell lung cancer. J Biomed Biotechnol 2011;2011:583929.
33. Yu X-J, Yang M-J, Zhou B, Wang G-Z, Huang Y-C, Wu L-C, et al. Characterization of somatic mutations in air pollution-related lung cancer. EBioMedicine 2015;2:583–90.
34. Chakravarty D, Gao J, Phillips S, Kundra R, Zhang H, Wang J, et al. OncoKB: a precision oncology knowledge base. JCO Precis Oncol 2017:1–16.
35. U.S. Environmental Protection Agency (EPA). EJSCREEN Technical Documentation; 2019. Accessed from: https://www.epa.gov/ejscreen/technical-documentation-ejscreen.
36. Evans J, van Donkelaar A, Martin RV, Burnett R, Rainham DG, Birkett NJ, et al. Estimates of global mortality attributable to particulate air pollution using satellite imagery. Environ Res 2013;120:33–42.
37. Fajersztajn L, Veras M, Barrozo LV, Saldiva P. Air pollution: a potentially modifiable risk factor for lung cancer. Nat Rev Cancer 2013;13:674–8.
38. Apte JS, Marshall JD, Cohen AJ, Brauer M. Addressing global mortality from ambient PM2.5. Environ Sci Technol 2015;49:8057–66.
39. White AJ, Kresovich JK, Keller JP, Xu Z, Kaufman JD, Weinberg CR, et al. Air pollution, particulate matter composition and methylation-based biologic age. Environ Int 2019;132:105071.
40. Landrigan PJ, Fuller R, Acosta NJR, Adeyi O, Arnold R, Basu NN, et al. The Lancet Commission on pollution and health. Lancet 2018;391:462–512.

41. Smith KR, Ezzati M. How environmental health risks change with development: the epidemiologic and environmental risk transitions revisited. Annu Rev Environ Resour 2005;30:291–333.
42. Eisner MD, Blanc PD, Omachi TA, Yelin EH, Sidney S, Katz PP, et al. Socioeconomic status, race and COPD health outcomes. J Epidemiol Community Health 2011;65:26–34.
43. Bradley CJ, Given CW, Roberts C. Race, socioeconomic status, and breast cancer treatment and survival. J Natl Cancer Inst 2002;94:490–6.
44. Ward E, Jemal A, Cokkinides V, Singh GK, Cardinez C, Ghafoor A, et al. Cancer disparities by race/ethnicity and socioeconomic status. CA Cancer J Clin 2004; 54:78–93.
45. Clegg LX, Reichman ME, Miller BA, Hankey BF, Singh GK, Lin YD, et al. Impact of socioeconomic status on cancer incidence and stage at diagnosis: selected findings from the surveillance, epidemiology, and end results: National Longitudinal Mortality Study. Cancer Causes Control 2009;20: 417–35.
46. Abbas I, Verdin A, Escande F, Saint-Georges F, Cazier F, Mulliez P, et al. In vitro short-term exposure to air pollution PM2.5–0.3 induced cell cycle alterations and genetic instability in a human lung cell coculture model. Environ Res 2016;147: 146–58.
47. Kucab JE, Phillips DH, Arlt VM. Linking environmental carcinogen exposure to TP53 mutations in human tumours using the human TP53 knock-in (Hupki) mouse model. FEBS J 2010;277:2567–83.
48. Zhou W, Tian D, He J, Wang Y, Zhang L, Cui L, et al. Repeated PM2.5 exposure inhibits BEAS-2B cell P53 expression through ROS-Akt-DNMT3B pathway-mediated promoter hypermethylation. Oncotarget 2016;7:20691–703.
49. Li R, Zhou R, Zhang J. Function of PM2.5 in the pathogenesis of lung cancer and chronic airway inflammatory diseases. Oncol Lett 2018;15:7506–14.
50. Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. Int J Epidemiol 2005;34:680–7.
51. Lim JW, Ashing-Giwa KT. Examining the effect of minority status and neighborhood characteristics on cervical cancer survival outcomes. Gynecol Oncol 2011;121:87–93.
52. Lee CC, Chien SH, Hung SK, Yang WZ, Su YC. Effect of individual and neighborhood socioeconomic status on oral cancer survival. Oral Oncol 2012;48:253–61.
53. Gomez SL, Glaser SL, McClure LA, Shema SJ, Kealey M, Keegan THM, et al. The California Neighborhoods Data System: a new resource for examining the impact of neighborhood characteristics on cancer incidence and outcomes in populations. Cancer Causes Control 2011;22:631–47.
54. Cheng I, Shariff-Marco S, Koo J, Monroe KR, Yang J, John EM, et al. Contribution of the neighborhood environment and obesity to breast cancer survival: the California Breast Cancer Survivorship Consortium. Cancer Epidemiol Biomarkers Prev 2015;24:1282–90.
55. DeRouen MC, Schupp CW, Koo J, Yang J, Hertz A, Shariff-Marco S, et al. Impact of individual and neighborhood factors on disparities in prostate cancer survival. Cancer Epidemiol 2018;53:1–11.
56. Turner MC, Cohen A, Jerrett M, Gapstur SM, Diver WR, Pope CA, et al. Interactions between cigarette smoking and fine particulate matter in the Risk of Lung Cancer Mortality in Cancer Prevention Study II. Am J Epidemiol 2014;180: 1145–9.