Relationship between the TP53 SNP rs1800371 and Lung Cancer in African Americans

Noah Nichols\textsuperscript{1}, Khadijah Mitchell\textsuperscript{1}, Adriana Zingone\textsuperscript{1}, Claire Meaney\textsuperscript{1}, Elise D. Bowman\textsuperscript{1}, Bríd M. Ryan\textsuperscript{1}

\textsuperscript{1}Laboratory of Human Carcinogenesis, Centre for Cancer Research, NCI, Building 37, Room 3060C, Bethesda, MD, 20892,

*Corresponding author e-mail: Brid.Ryan@nih.gov

**ABSTRACT**

Germline and somatic mutations in TP53 have been investigated and intensely catalogued. One of these germline mutations, TP53 P47S SNP is only found in populations with African ancestry. The S47 variant was associated with an impaired ability to induce cell death following cisplatin treatment, and with a high rate of spontaneous cancer formation in mice engineered to carry S47. Lung cancer incidence is higher among men of African descent. We therefore tested the hypothesis that P47S was associated with increased risk of lung cancer. We included 926 controls and 425 cases, all of whom were African Americans. We genotyped rs1800371 using a pre-designed Taqman assay. No serine/serine genotypes were detected in the population. The proline/serine genotype was detected in 42 individuals (32/891 controls (3.5%) and 10/434 cases (2.3%). The serine allele was not associated with risk of lung cancer (OR: 0.65, 95\% C.I. 0.28-1.48). rs1800371 is not associated with risk of lung cancer among African Americans.

**KEYWORDS:** TP53, SNP, health disparities

INTRODUCTION

Lung cancer is the leading cause of cancer-related death in the United States and the second most common form of cancer in both men and women (Siegel et al., 2017). Several studies have demonstrated an increased lung cancer burden in African Americans (AA) (Siegel et al., 2016). Smoking is the major etiological cause of lung cancer (Doll and Peto, 1976; Peto, 2001); however, a combination of environmental and genetic factors could predispose an individual to lung cancer.

The p53 tumor suppressor (TP53) is widely studied and its importance in tumorigenesis is underpinned by the fact that it is the most frequently mutated gene in human cancers (Olivier et al., 2010). Somatic mutations in TP53 have been investigated and intensely catalogued. Additionally, individuals with Li-Fraumeni syndrome, who carry germline mutations in TP53, have an increased risk of developing cancer (Malkin, 1993). Acting as a transcription factor, p53 binds to regulatory elements in DNA and facilitates the transcription of genes that protect against tumor growth. Pathways activated by p53 include cell cycle arrest, apoptosis, and angiogenesis inhibition (Whibley et al., 2009).

Harris and colleagues first identified the TP53 P47S SNP (rs1800371) in 1992 (Felley-Bosco et al., 1993; Gerwin et al., 1992). This polymorphism, characterized by a C>T transition at position 1 of codon 47 (CCG-TCG), causes a non-synonymous amino acid change from proline to serine. Felley-Bosco et al., noted that the individuals in that study, who identified as Caucasian American, were homozygous for the proline allele. However, the CT genotype was observed in 3/37 African Americans. In a functional analysis of this polymorphism, the authors concluded that, in their model, the S47 allele functioned similar to wild-type p53.

Phosphorylation of the N-terminal domain of p53 facilitates its transactivation activity of pro-apoptotic genes (Kurihara et al., 2007). For example, p38 MAPK and HIPK2 transactivation by p53 occurs through proline-directed phosphorylation of codon 46 at the N-terminal domain. Therefore, it was hypothesized that loss of the adjacent proline residue at position 47 could impair apoptotic activation. Indeed, Li et al. demonstrated that the S47 variant is an inferior substrate for p38 MAPK and exhibited a decreased ability to induce apoptosis in vivo compared with wild-type p53 (Li et al., 2005). The S47 variant was also associated with an impaired ability to induce cell death following cisplatin treatment (Jennis et al., 2016). Furthermore, the serine variant was associated with decreased transcription of a subset of target genes involved in metabolism and pro-apoptotic signaling (Jennis et al., 2016). Mice engineered to carry S47 had a high rate of spontaneous cancer formation, as 80% of S47 homozygous mice developed hepatocellular carcinoma, B-cell lymphoma, and other tumor types (Jennis et al., 2016).

Previous work has established the increased frequency of the S47 allele in African Americans and other populations with West African ancestry. Despite convincing evidence from mouse models and cell culture experiments suggesting that the S47 variant is functionally significant and mechanistically differs from the proline variant, several studies have failed to show an association with cancer risk or survival (Alawadi et al., 2011; Almeida et al., 2009; Jaiswal et al., 2011; Kaur et al., 2014; Mostaid et al., 2014; Sameer et al., 2010;
Santos et al., 2011; Singamsetty et al., 2014), though many of these studies included small sample sizes. Herein, we genotyped 425 lung cancer cases and 926 controls. The objective of this study is to determine whether rs1800371 is associated with lung cancer risk and survival in African Americans and admixed European Americans.

METHODS AND MATERIALS

Patient Population

Participants were selected from the ongoing National Cancer Institute—Maryland Case-Control Study. We included 926 controls and 425 cases for whom DNA was available (all African American). The NCI-Maryland Lung Cancer Case-Control study is an ongoing study that started in 1998 and includes participants recruited from the greater Baltimore and Eastern Shore regions of Maryland (Mechanic et al., 2007): Cases, recruited from seven hospitals in the greater Baltimore metropolitan area, were diagnosed with histologically confirmed NSCLC (of any stage). Population controls, frequency matched to cases on age, race, and sex, were recruited from Maryland Department of Motor Vehicles records.

General inclusion criteria included having been born in the United States, an ability to speak English well enough to be interviewed, being in good health and not residing in an institution. Trained interviewers administered a standardized questionnaire to capture information regarding demographics, socioeconomic characteristics, tobacco smoking history (current, former [those who reported to have quit smoking one year prior to the interview], or never [those who smoked less than 100 cigarettes over their lifetime]), as well as pack-years of smoked cigarettes, alcohol history, medical history, family cancer history, reproductive history, and occupational history. Body mass index (BMI) was determined using reported measures of height and weight by the participant.

<table>
<thead>
<tr>
<th>Table 1. Demographic characteristics of the cases and controls.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>African Americans</strong></td>
</tr>
<tr>
<td><strong>Controls</strong> N=926</td>
</tr>
<tr>
<td>Age (mean, sd)</td>
</tr>
<tr>
<td>Gender (N, %)</td>
</tr>
<tr>
<td>Males</td>
</tr>
<tr>
<td>Females</td>
</tr>
<tr>
<td>Smoking (N, %)</td>
</tr>
<tr>
<td>Never</td>
</tr>
<tr>
<td>Former</td>
</tr>
<tr>
<td>Current</td>
</tr>
<tr>
<td>Former smoker: Quit within 1 year of interview</td>
</tr>
<tr>
<td>Pack-years of smoking (Median, IQR)</td>
</tr>
<tr>
<td>Family History of Lung Cancer (N, %)</td>
</tr>
</tbody>
</table>
Genotyping

Genomic DNA was isolated from cheek cells or buffy coat samples containing white blood cells from patients in the case–control study using Flexigene DNA Kit (Qiagen), according to the manufacturer’s instructions. Case, control, and duplicate samples (10%) were randomized and blinded for processing. Genotypes were obtained using a standard TaqMan SNP Genotyping Assay for rs1800391 and followed the manufacturer’s protocol. Genotyping failed for 4 individuals. Cases and controls were randomized across all the plates. Duplicate concordance was 100%. We also genotyped a series of European Americans (n=1349). All were homozygous wild-type for the SNP (CC).

Global Genetic Ancestry

Global genetic ancestry analysis was performed as previously described (Al-Alem et al., 2014). Briefly, DNA from blood was genotyped for 100 ancestry informative markers (AIMs) using the Sequenom MassARRAY iPLEX platform. The AIMs panel consisted of carefully selected autosomal markers that were previously identified and validated for estimating continental ancestry information in admixed populations (Kosoy et al., 2009; Nassir et al., 2009). Individual SNP genotype calls were generated using Sequenom TYPER software. A genotype concordance rate of 99.5% was observed for all markers. Genotyping call rates exceeded 97% for all individuals included in the analyses. Individual admixture estimates for each study participant were calculated using a model-based clustering method as implemented in the program STRUCTURE v2.3 (Falush et al., 2003). STRUCTURE 2.3 was run using parental population genotypes from West Africans, Europeans, and Native Americans (Kosoy et al., 2009) under the admixture model using the Bayesian Markov chain Monte Carlo method (K=3, assuming three founding populations) and a burn-in length of 30,000 for 70,000 repetitions.

Statistical Analysis

Chi-square tests were used to compare categorical variables between cases and controls. A logistic regression model was used to estimate odds ratios and 95% confidence interval (CI), before and after adjustment for known and potential confounders, including age, gender, smoking status, pack-years of smoking and BMI. Cox proportional hazard models were used to estimate hazard ratios (HR) and 95% CI for disease-free survival after adjustment for age, gender, smoking status, pack-years, tumor stage and histological subtype. Survival was calculated using the difference in time.
between the date of surgery or diagnosis and the date of last known follow-up or death from lung cancer. All statistical tests were two-sided, and a p-value of 0.05 was used to assess significance. We used Stata version 14 to perform all analyses.

RESULTS

Frequency of rs1800371 genotypes
The Proline/Serine genotype was detected in 42 individuals (3.1%), giving an allele frequency of 1.5%. Of these, 32 (3.5%) were controls and 10 (2.3%) were cases (Table 2). No Serine/Serine genotypes were reported. Also, the Proline/Serine genotype was only detected in African Americans, which is consistent with the ancestral origin of the serine allele.

<table>
<thead>
<tr>
<th>rs1800371</th>
<th>European American</th>
<th>African American</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Case</td>
</tr>
<tr>
<td>Proline/Proline</td>
<td>717 (100%)</td>
<td>632 (100%)</td>
</tr>
<tr>
<td>Proline/Serine</td>
<td>32 (3.5%)</td>
<td></td>
</tr>
<tr>
<td>Serine/Serine</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relationship between rs1900371 with lung cancer risk
As the serine allele was only detected in African Americans, we restricted our analysis to this subgroup of the population. We observed a nonsignificant association between the SNP and lung cancer risk both before (OR: 0.64, 95% C.I. 0.31-1.32) and after adjustment for potential confounders (OR: 0.55, 95% C.I. 0.28-1.48) (Table 3). As African Americans are an admixed population, we also adjusted our analysis for global genetic ancestry, however, this adjustment did not alter the coefficients (Table 3).

<table>
<thead>
<tr>
<th>rs1800371</th>
<th>OR (95% C.I.)</th>
<th>P-value</th>
<th>OR (95% C.I.)</th>
<th>P-value</th>
<th>OR (95% C.I.)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proline/Proline</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Proline/Serine</td>
<td>0.64 (0.31-1.32)</td>
<td>0.23</td>
<td>0.65 (0.28-1.48)</td>
<td>0.30</td>
<td>0.65 (0.28-1.50)</td>
<td>0.32</td>
</tr>
<tr>
<td>Serine/Serine</td>
<td>NA</td>
<td></td>
<td>NA</td>
<td></td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

* Unadjusted
b Adjusted for age at diagnosis, gender, smoking status, pack-years of smoking, bmi
* Adjusted for age at diagnosis, gender, smoking status, pack-years of smoking, bmi, West African ancestry, European ancestry and Native American ancestry
Relationship between rs1800371 with lung cancer survival

We tested the relationship between rs1800371 with lung cancer specific survival among the African American cases for whom we had survival data (n=439). We observed a non-significant relationship between the serine allele with prognosis (HR: 0.82, 95% C.I. 0.34-2.00).

<table>
<thead>
<tr>
<th>rs1800371</th>
<th>HR (95% C.I.) a N=439</th>
<th>P-value</th>
<th>HR (95% C.I.) b N=360</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proline/Proline</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Proline/Serine</td>
<td>0.82 (0.34-2.00)</td>
<td>0.67</td>
<td>0.56 (0.23-1.38)</td>
<td>0.21</td>
</tr>
<tr>
<td>Serine/Serine</td>
<td>NA</td>
<td></td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

DISCUSSION

Animal and experimental studies using human cells have suggested that the serine allele of rs1800371 predisposes individuals to risk of cancer (Basu et al., 2016; Jennis et al., 2016). In this study, we examined whether carriers of the serine allele had an increased risk of lung cancer. However, we did not find evidence for an association, neither did we find an association with survival.

Mostaid et al. did not find an association between S47 and lung cancer risk in a Bangladeshi population. (Mostaid et al., 2014). Similarly, no risk associations were observed in esophageal, breast, colorectal and bladder cancer studies (Alawad et al., 2011; Jaiswal et al., 2011; Kaur et al., 2014; Sameer et al., 2010; Santos et al., 2011). In contrast, Singamsetty et al., focusing on a south Indian population, observed a significant difference in the proportion of S47 alleles in cases which had a protective effect on CRC risk (Singamsetty et al., 2014). Furthermore, one study has demonstrated an association with increased risk and the S47 variant. In 2013, an extra-axial brain tumor study revealed a small increase in risk for tumor development in a Brazilian population (Almeida et al., 2009). In breast cancer, particularly among pre-menopausal women, the serine allele is associated with increased risk among African Americans (Murphy et al., 2017).

The CT genotype, conferring one copy of the serine allele and one copy of the proline allele, was observed in 42/1435 (2.9%) of African Americans. This frequency is somewhat lower than the occurrence of the CT genotype in African populations (~6-8%) previously reported, but the lower frequency in our cohort could reflect genetic admixture. Of note, the minor allele frequency of the T allele in this study (1.5%) is similar to that reported in a recent report of African American women (~1.4%) (Murphy et al., 2017).

P53 mediates its tumor suppressive function by inhibiting cell growth, inducing senescence and modulating cell metabolism. It accomplishes these functions via transcriptional activation of target genes, some of which is driven by post-translational modifications such as phosphorylation. For example, phosphorylation of serine 46, the amino acid next to proline 47 is required for efficient p53-mediated cell death in several cell line systems and in mice (Bulavin et al., 1999; Jennis et al., 2016). Phosphorylation of serine 46 requires the presence of a proline at 47, thus
serine 47 significantly impairs the induction of cell death in cell line models (Li et al., 2005). However, the impact of this variant on the p53 signaling pathway and cancer risk in an intact animal has never been assessed. In a follow up study, Jennis and colleagues found that the S47 variant is modestly impaired for most p53 functions, and that it is significantly impaired for the ability to transactivate a subset of p53 target genes to induce cell death following cisplatin or ferroptosis in a mouse model (Jennis et al., 2016).

We had hypothesized that rs1800371 would be associated with increased risk of lung cancer. This hypothesis was primarily based on the observation that 80% of mice carrying two copies of the serine allele developed spontaneous cancers. There are several possible reasons why an observation with cancer risk was not observed in our study. First, it is possible that this SNP is not associated with lung cancer. In the mouse study, the main tumor type observed was hepatocellular carcinoma. This is possibly related to the fact that one of the main gene targets disrupted by serine 47 is GLS2 (Jennis et al., 2016). While some penetrance of spontaneous tumors was observed among mice carrying one serine allele, the frequency was lower. Also, our study population did not identify any individuals carrying two serine alleles and it is possible that both alleles are needed for higher disease penetrance. Moreover, smoking is the main etiological cause of lung cancer in our population and it is possible that the mechanism of smoking-induced lung carcinogenesis does not involve ferroptosis. TP53 has many splice variants, and one key difference between the human and mouse is the presence and absence of the dominant negative delta133 isoform in human and mice, respectively. However, given that this variant occurs in codon 47, and delta133 is missing the first 133 codons of the full-length gene, it seems unlikely that these differences in isoform expression could impact function. Finally, as shown in previous studies, TP53-related variants do not always function in isolation. For example, the two promoter variants SNP309 and SNP285 of the MDM2 gene, negative regulator of p53, functionally interact; SNP285 acts as an antagonist to SNP309 by overriding the effect of SNP309 on SP1-mediated transcription (Knappskog et al., 2011; Ryan et al., 2012). Therefore, it is possible that the tumor suppressive function of serine 47 interacts with another human polymorphism in the TP53 pathway.

Our study suggests that rs1800371 is not associated with increased lung cancer risk among African Americans, but we cannot rule out the possibility that carriers of two serine alleles would have increased susceptibility. However, given the penetrance of hepatocellular carcinoma in mice carrying the serine allele, and disparities in liver cancer incidence in the US (Islami et al., 2017), case control studies of this cancer type and rs1800371 should be considered.

Acknowledgements
Community Health Representative(s), Community Partners, and community members who participated in the study. This work was supported by the intramural research program of the Center for Cancer Research, National Cancer Institute.

Conflict of interest statement
The author has declared that no competing or conflict of interests exists. The funders had no role in study design, writing of the manuscript and decision to publish.
Authors’ contributions
Conceptualization: BMR, Formal analysis: NN, AZ, BMR, KM, CM Methodology: AZ, NN, EBSupervision: AZ, BMR Writing ± original draft: NN, BMR Writing ± review & editing: NN, KM, BMR, AZ, EB, CM,

REFERENCES


www.companyofscientists.com/index.php/chd e8 Cancer Health Disparities
marker set for determining continental origin: validation and extension using human genome diversity panels. BMC Genet 10, 39.