BRIEF REPORT

Somatic Mutations and Ancestry Markers in Hispanic Lung Cancer Patients

Nicholas T. Gimbrone, BS,[a] Bhaswati Sarcar, PhD,[a] Edna R. Gordian, MA,[b] Jason I. Rivera, BS,[a] Christian Lopez, BS,[b] Sean J. Yoder, MS,[c] Jamie K. Teer, PhD,[d] Eric A. Welsh, PhD,[d] Alberto A. Chiappori, MD,[b,e] Matthew B. Schabath, PhD,[f] Gary W. Reuther, PhD,[a] Julie Dutil, PhD,[g] Miosotis Garcia, MD,[h] Ronald Ventosilla-Villanueva, MD,[i] Luis Vera-Valdivia, MD,[j] Alejandro Yabar-Berrocal, MD,[k] Rodrigo Motta-Guerrero, MD,[l] Pedro G. Santiago-Cardona, PhD,[g] Teresita Muñoz-Antonia, PhD,[b] W. Douglas Cress, PhD[a,*]

[a]*Molecular Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida*
[b]*Tumor Biology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida*
[c]*Molecular Genomics Core Facility, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida*
[d]*Biostatistics and Bioinformatics, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida*
[e]*Thoracic Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida*
[f]*Cancer Epidemiology Programs, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida*
[g]*Basic Sciences Department, Biochemistry and Cancer Biology Divisions, Ponce Health Sciences University, Ponce, Puerto Rico*
[h]*Endocrinology Department, Central University of the Caribbean, 9 San Juan Municipal Hospital, San Juan, Puerto Rico*
[i]*Peruvian Cayetano Heredia University, Daniel Alcides Carrión Hospital, Lima, Perú*
[j]*Edgardo Rebagliati Martins Essalud National Hospital, Lima, Perú*
[k]*Pathology Department, Rebagliati Martins Essalud National Hospital, Lima, Perú*
[l]*Peruvian Cayetano Heredia University, Cayetano Heredia National Hospital, Lima, Perú*

Received 26 May 2017; revised 17 August 2017; accepted 21 August 2017
Available online - 11 September 2017

ABSTRACT

Introduction: To address the lack of genomic data from Hispanic/Latino (H/L) patients with lung cancer, the Latino Lung Cancer Registry was established to collect patient data and biospecimens from H/L patients.

Methods: This retrospective observational study examined lung cancer tumor samples from 163 H/L patients, and tumor-derived DNA was subjected to targeted-exome sequencing (>1000 genes, including *EGFR, KRAS*, serine/threonine kinase 11 gene [*STK11*], and tumor protein p53 gene [*TP53*]) and ancestry analysis. Mutation frequencies in this H/L cohort were compared with those in a similar cohort of non-Hispanic white (NHW) patients and correlated with ancestry, sex, smoking status, and tumor histologic type.

Results: Of the adenocarcinomas in the H/L cohort (n = 120), 31% had *EGFR* mutations, versus 17% in the NHW control group ($p < 0.001$). *KRAS* (20% versus 38% [$p = 0.002$]) and *STK11* (8% versus 16% [$p = 0.065$]) mutations occurred at lower frequency, and mutations in *TP53* occurred at similar frequency (46% versus 40% [$p = 0.355$]) in H/L and NHW patients, respectively. Within the Hispanic cohort, ancestry influenced the rate of *TP53* mutations ($p = 0.009$) and may have influenced the rate of *EGFR, KRAS,* and *STK11* mutations.

Conclusions: Driver mutations in H/L patients with lung adenocarcinoma differ in frequency from those in NHW patients associated with their indigenous American ancestry. The spectrum of driver mutations needs to be further assessed in the H/L population.

Keywords: Lung cancer; Hispanic; *EGFR* mutations; Ancestry

*Corresponding author.

Disclosure: The authors declare no conflict of interest.

Address for correspondence: W. Douglas Cress, PhD, MRC 4 East Room 4072G, 12902 Magnolia Dr., H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612. E-mail: douglas.cress@moffitt.org

ISSN: 1556-0864
http://dx.doi.org/10.1016/j.jtho.2017.08.019

Introduction

The U.S. Census Bureau projects that the U.S. Hispanic/Latino (H/L) population will reach 119 million by 2060.[1] However, H/L individuals represent only 3% of patients characterized in The Cancer Genome Atlas (TCGA) and indigenous Americans represent less than 0.5%.[2] Because precision medicine is driven by data that primarily represent non-Hispanic white (NHW) patients, there is a significant potential health disparity for underrepresented groups.[3–6] NSCLC is the leading cause of cancer death among H/L men and second only to breast cancer in H/L women.[3] In NHW individuals, half of NSCLC cases have been shown to possess at least one of several known driver mutations, including alterations in *KRAS, EGFR,* MNNG HOS Transforming gene (*MET*), erb-b2 receptor tyrosine kinase 2 gene (*HER2*), *BRAF*, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha gene (*PIK3CA*), AKT/serine threonine kinase 1 gene (*AKT1*), mitogen-activated protein kinase kinase 1 gene (*MAP2K1*), ALK receptor tyrosine kinase gene (*ALK*), and *MEK.*[7] African Americans demonstrate a similar frequency of these mutations.[8] The frequency of *EGFR* mutations has been extensively studied for some groups and determined to be approximately 30% for Asian, 15% for NHW, and 19% to 21% for African American populations.[5,9] However, a recent study of 241 African Caribbean patients found *EGFR* mutations to be common (37%).[10] Studies in H/L individuals have indicated an elevated frequency of *EGFR* mutations relative to that in NHW individuals.[11,12] Other common mutations in lung cancer among H/L individuals have not been sufficiently investigated.[4,13] To address the lack of molecular data for H/L individuals, our group established the Latino Lung Cancer Registry. We describe here our initial characterization of 163 H/L patients with lung cancer.

Methods

Patients and Tissues

Supplementary Table 1 highlights individual data for 163 patients, including molecular data, sex, smoking status, and tumor histologic type, as available. Samples designated *Moffitt* were from the Total Cancer Care (TCC) program and were collected between April 2006 and August 2010 under University of South Florida Protocols 00001222 and 00011723. Samples designated *Puerto Rico* were acquired through the Puerto Rico Bio-Bank under the Ponce Health Sciences University Protocol 080121-IF. Samples designated *Perú* were from patients with a diagnosis of lung adenocarcinoma who had an operation during the years 2013–2014 in Lima, Perú, where approval was obtained locally.

DNA Sequencing, Global Ancestry Estimations, and Comparison with NHW Patients

Detailed methods for DNA isolation, sequencing and ancestry-informative marker (AIM) analysis are described in the Supplementary Methods. Two data sets were combined to compare mutation frequencies between lung adenocarcinoma samples from H/L and NHW patients, including the TCGA cohort of 483 patients with lung adenocarcinoma and a cohort of 265 patients with adenocarcinoma obtained through the TCC.

Results

Patient Characteristics

Table 1 summarizes the demographics of the cohort overall and by site of collection. The most common histologic type in the cohort was adenocarcinoma (120 [74%]), followed by squamous cell carcinomas (27 [17%]). Known female patients (78) outnumbered known male patients (66), and never-smokers (45) represented 36% of those with known smoking status. We subjected the available samples to AIM analysis to define the relative contributions of European, African American, and indigenous American ancestries in the cohort and in each individual patient. Overall, the cohort was primarily European (67%), followed by indigenous American (21%) and then of African ancestry (12%). African ancestry was enriched in the Puerto Rican subcohort, and indigenous American ancestry was enriched in the group from Perú.

Prevalence of Somatic Mutations

Figure 1*A* presents the mutation frequency in the four most common drivers in lung adenocarcinoma in NHW and H/L patients. For H/L patients, within the adenocarcinoma histologic type, 37 (31%) had one or more *EGFR* variants and 17 (14%) had *KRAS* variants. This analysis also showed a trend toward reduced rates in serine/threonine kinase 11 gene (*STK11*) mutations ($p = 0.065$). The mutation rate of tumor protein p53 gene (*TP53*) was indistinguishable ($p = 0.717$) between H/L (49%) and NHW patients (46%). Supplementary Table 2 lists other genes that may be mutated at a different frequency in H/L patients compared with in NHW patients.

The relationship between driver mutations and patient characteristics was also explored. Figure 1*B* presents the mutation rates of *EGFR*, *KRAS*, *STK11,* and *TP53* in male and female patients in each group. Notably, the rate of *EGFR* mutations was nearly 50% in H/L female patients. Mutations in *KRAS* and *STK11* were reduced in female H/L patients and to a lesser extent in male H/L patients. The rate of *TP53* mutations was similar between male and female patients (see

Table 1. Demographics of the Cohorts, Combined and Individual Results

Demographic Characteristic	Total Patients (N = 163)		Moffitt Cohort (n = 93)		PR Cohort (n = 46)		Perú Cohort (n = 24)	
	No.	%	No.	%	No.	%	No.	%
Histologic type								
Adenocarcinoma	120	74	65	70	31	67	24	100
Squamous cell carcinoma	27	17	15	16	12	26	0	0
Other[a]	16	10	13	14	4	7	0	0
Sex								
Male	66	46	35	38	20	63	11	58
Female	78	54	58	62	12	38	8	42
Unknown	19	NA	0	NA	14	NA	5	NA
Smoking								
Ever	80	64	60	66	17	81	3	23
Never	45	36	31	34	4	19	10	77
Unknown	38	NA	2	NA	25	NA	11	NA
Ancestry								
% European		67		79		64		63
% African		12		10		22		6
% Indigenous American		21		11		14		31

[a]The Moffitt cohort includes seven carcinoids; three non-small cell lung carcinomas, not otherwise specified; two large cell carcinomas; and one adenosquamous carcinoma. The Puerto Rico cohort contains three carcinomas, not otherwise specified, and one NSCLC, not otherwise specified.
PR, Puerto Rico; NA, not available.

Supplementary Table 3 for *p* values). Figure 1*C* presents a comparison between ever-smokers and never-smokers among H/L and NHW patients. Mutation rates in *KRAS* and *STK11* were much higher in smokers for both the H/L and NHW cohorts. In fact, no *STK11* mutations were observed in the 45 H/L never-smokers. In NHW patients, *EGFR* mutations were four times more common in never-smokers than in smokers. In the H/L cohort, this effect was much smaller. Mutations in *TP53* were not strongly associated with smoking in either cohort. Finally, mutations were examined as a function of smoking by sex. Although the numbers are too small for statistical significance (see Supplementary Table 5), Figure 1*D* reveals that *EGFR* mutations in male patients in the H/L cohort were reduced in smokers but there was no difference detected in female smokers versus in nonsmokers. In contrast, a similar analysis in NHW patients (Fig. 1*E*) shows a clear increase in *EGFR* mutations in both male and female nonsmokers.

In general, H/L individuals are characterized by a wide mixture of ancestries. Therefore, to explore the role that ancestry may play in mutation frequency in the H/L cohort, the Ward linkage method and Euclidean distance metric were used to separate the cohort into three ancestral clusters (Fig. 2*A*). The mutation frequency of various genes was then compared between the three clusters. Figure 2*B* reveals that *EGFR* mutations were more common in the indigenous American cluster but appeared identical in the European and African clusters (see Supplementary Table 7 for *p* values). In contrast, *TP53* mutations were much less common in the indigenous American cluster, whereas the frequency of *TP53* mutations appeared identical in the European and African clusters. *KRAS* mutations were shown to be rare in the African cluster, and *STK11* mutations were rare in both the African and indigenous American clusters relative to in the European cluster.

Discussion

In previous reports, *EGFR* exhibited a higher rate of mutation and *KRAS* was suggested to have a lower mutation rate in H/L patients than in NHW patients.[3,4,11,14] The data presented here confirm these observations. Of the 120 H/L adenocarcinomas that we analyzed, 37 (31%) had *EGFR* mutations compared with 17% in our TCC NHW cohort, and 17 H/L patients (14%) had *KRAS* mutations. In a published NHW cohort, 11% of lung adenocarcinomas had *EGFR* mutations and 35% had *KRAS* mutations.[15] In addition, our data suggest that the increase in *EGFR* mutations within our H/L cohort is driven by females, with 48% having *EGFR* mutations (Fig. 1*B*). Our data also suggest altered mutation rates for other genes commonly mutated in NSCLC. Among patients with adenocarcinoma, the frequencies of *KRAS* and *STK11* mutations were reduced in H/L patients relative to those in NHW patients.

Smoking is the primary driver of mutations in lung cancer. The overall percentage of nonsmokers in the H/L cohort was 36%. In comparison, an NHW cohort showed 9% nonsmokers[15] and the TCGA lung adenocarcinoma cohort showed 7% nonsmokers. Nonetheless, H/L individuals appeared to have a pattern similar to that

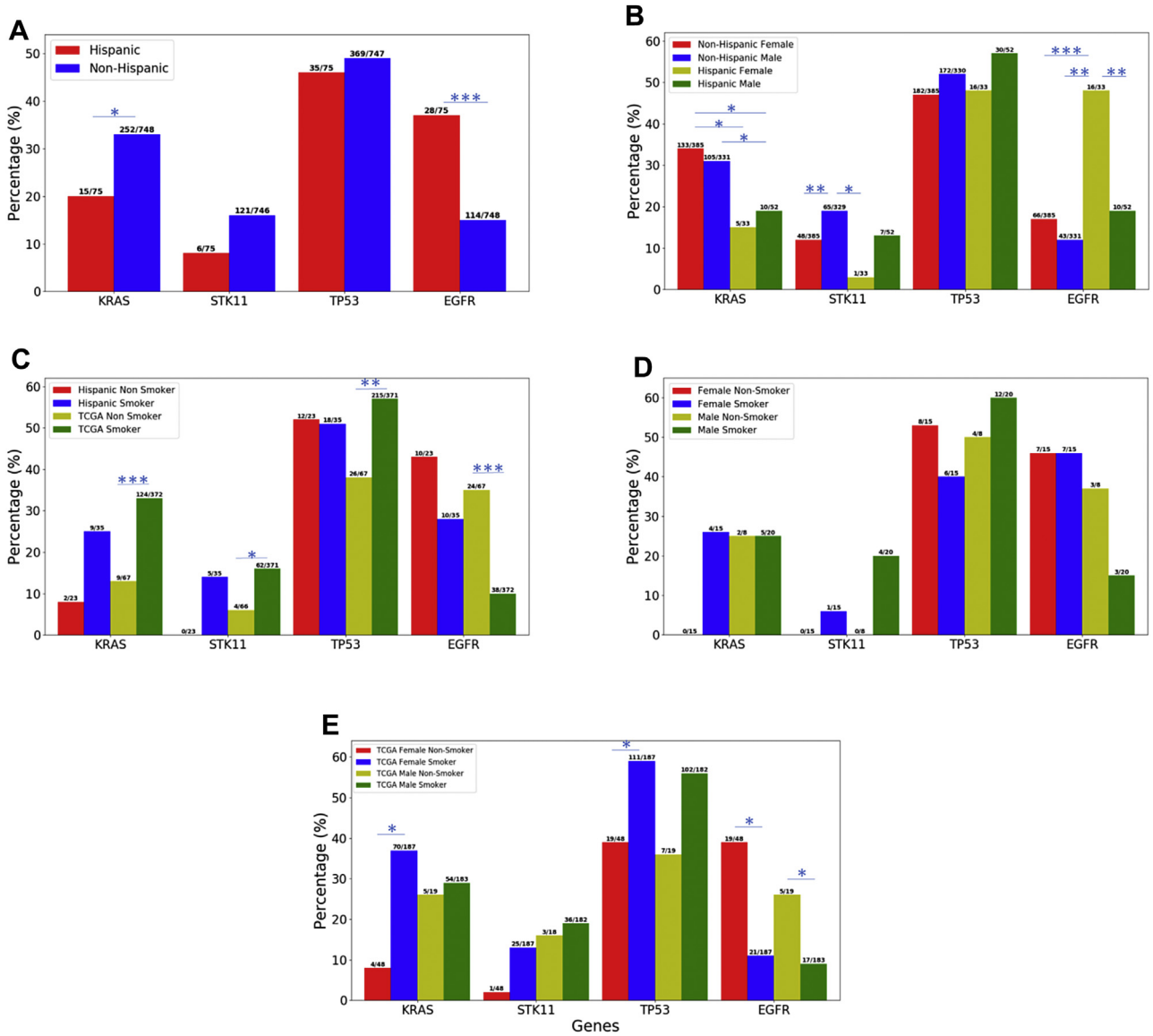

Figure 1. Mutation frequency comparisons using Fisher's exact test. (*A*) Mutation rate comparison between 75 lung adenocarcinomas in H/L patients and 748 in NHW patients. Histograms of the four most common mutations in adenocarcinomas in NHW patients as indicated as the percentage of patients with any detected variant. Number of mutants or number in cohort is indicated above the histograms. All significantly different rates are indicated by Fisher's exact test *p* values: **p* <0.05, ***p* <0.01, and ****p* <0.001. See Supplementary Table 2 for analysis of additional genes. (*B*) Mutation rate comparison between male and female patients and H/L and NHW patients. Same analysis as in (*A*), except that H/L and NHW patients were divided into male and female patients. See Supplementary Table 3 for *p* values. (*C*) Mutation rate comparison between H/L (and NHW) ever-smokers and never-smokers. Although statistically insignificant, the trend suggests that *EGFR* mutations are seen more commonly in nonsmokers, whereas *KRAS* and seronine/threonine kinase 11 gene (*STK11*) mutations are more frequently seen in smokers. See Supplementary Table 4 for *p* values. (*D*) Mutation rate comparison between H/L male and female patients and ever-smokers and never-smokers. Although statistically insignificant, the trend suggests that *EGFR* mutations are seen more commonly in nonsmokers, whereas *KRAS* and *STK11* mutations are more frequently mutated in smokers. See Supplementary Table 5 for *p* values. (*E*) Mutation rate comparison between NHW males and female patients and ever-smokers and never-smokers. See Supplementary Table 6 for *p* values. *TP53*, tumor protein p53 gene.

observed in NHW individuals for *STK11* and *KRAS* in that mutations in both genes increase with smoking in both groups (Fig. 1*C*). Unlike in NHWs, mutations in *TP53* occurred with nearly equal frequency in both smoking groups in the H/L cohort and the inverse correlation between smoking and *EGFR* mutations was dampened in H/L patients. These results suggest that smoking status is not the primary factor explaining the high rate of *EGFR* mutations in H/L patients. The observation that *KRAS* and *STK11* mutations are increased in H/L smokers addresses the objection that H/L ever-smokers are lighter smokers than NHW individuals are. There may be

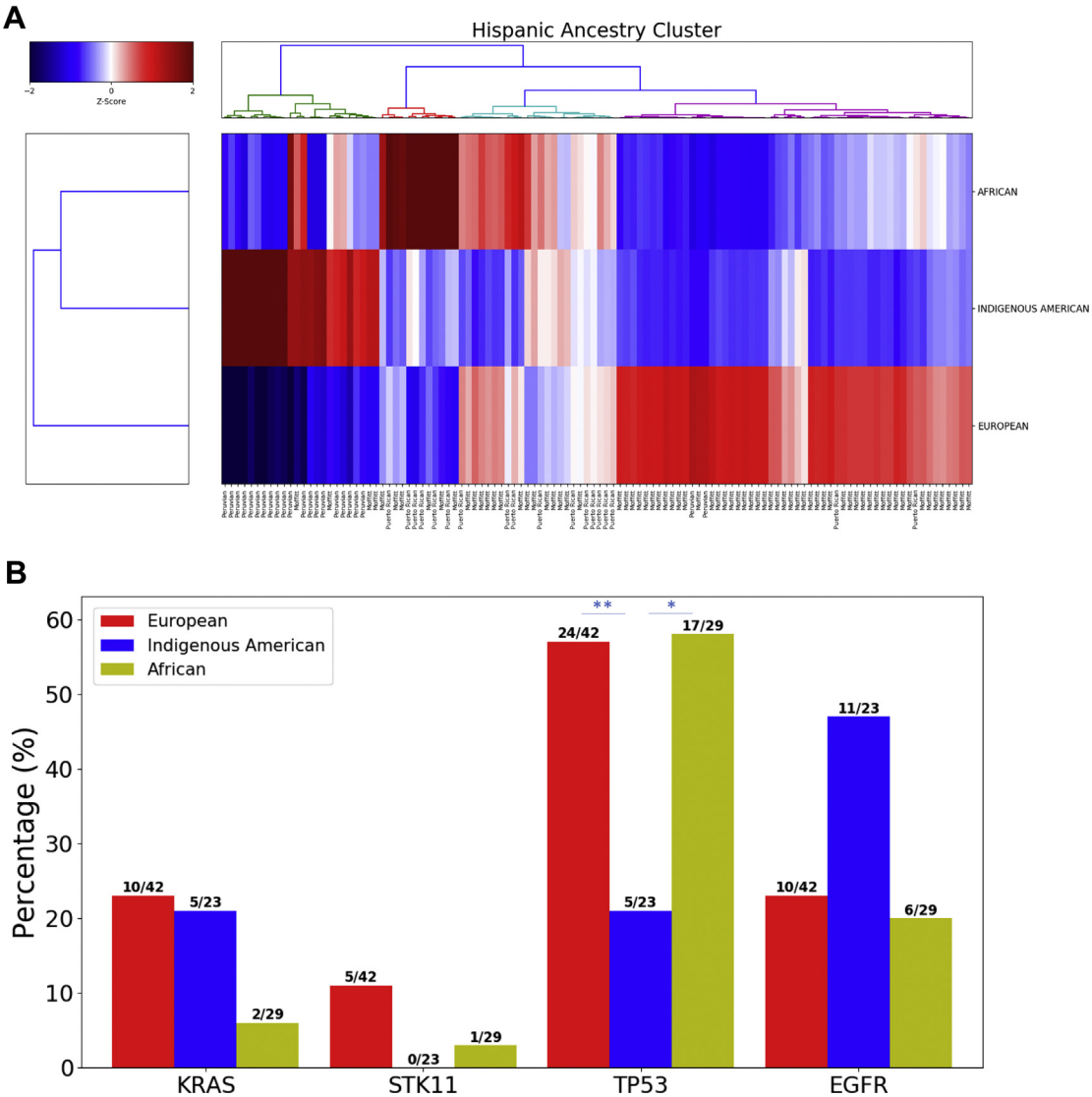

Figure 2. Mutation rate comparison by ancestry. (*A*) The Ward linkage method and Euclidean distance metric were used to cluster H/L patients by ancestry. The clusters were classified as follows: (1) indigenous American (*green cluster*), (2) African (*red* and *left blue clusters*), and (3) European (*right blue cluster*). (*B*) Mutation rate comparison between the three ancestral clusters. Mutation rates of four common drivers within the three ancestral clusters reveal a statistically significant difference between tumor protein p53 gene (*TP53*) mutations between indigenous American and European ($p = 0.0087$) and indigenous American and African ancestral clusters ($p = 0.011$). The same comparison for *EGFR* has p values of 0.058 and 0.072, respectively. See Supplementary Table 7 for p values. seronine/threonine kinase 11 gene.

differences in how smoking influences the mutations observed in men versus in women (Fig. 1*D* and *E*). However, drawing this conclusion firmly will require a larger data set. Regardless of the cause, the significant differences in specific mutations, and in mutation burden, suggest that special consideration be given to H/Ls patients considered for immune therapeutic approaches. For example, future studies should carefully examine the expression of targetable immune components in H/L patients.

Finally, the role that ancestry might play in this cohort was examined by clustering patients into ancestral groups on the basis of AIM analyses (Fig. 2*A*). Whereas the analysis results are statistically significant only for the *TP53* mutations, the data (Fig. 2*B*) suggest that relative to European ancestry, indigenous American ancestry correlates with low rates of *TP53* and *STK11* mutations and high rates of *EGFR* mutations and African ancestry correlates to low rates of *KRAS* mutations. At this point, our data set is too small to make claims with respect other mutations that are even less common than these four.

There are a few notable study limitations worth addressing. One of the most significant was the lack of matched normal samples for each of the sequenced tumors. Likewise, we utilized various sources of tissue

and various sequencing technologies. These issues should be considered when interpreting these data. Other limitations include incomplete clinical data for several samples in our cohort (which affected our ability to determine statistical significance) and lack of age, tumor stage, and outcome data. Future studies will be necessary to provide statistical power for additional comparisons. Nonetheless, this study is the first addressing the lack of comprehensive genomic data in H/L patients with lung cancer. The data suggest that *EGFR* mutations in H/L patients with lung adenocarcinoma are associated with their indigenous American ancestry. This observation may point to connection to a genetic component from Asian-Pacific migration because it is known that the *EGFR* mutation rate is high among Asian patients.[5] Future studies will need to expand this study to fully elucidate the role that ancestry is playing in the mutation driving lung cancer.

Acknowledgments

This work was funded by multiple sources, including the National Cancer Institute (grants U54 CA163068 to Drs. Muñoz-Antonia and Cress and CA163071 to Drs. Santiago-Cardona and Dutil), the Miles for Moffitt Foundation (grant 09-33661-15-06 to Drs. Muñoz-Antonia and Cress), the James and Esther King Biomedical Research Program Grant (5JK06) from the Florida Department of Health (to Dr. Cress), and Total Cancer Care, which is supported in part by the DeBartolo Family. Our study also received valuable assistance from the Tissue Core, Cancer Informatics, and Genomics Core Facilities at the H. Lee Moffitt Cancer Center and Research Institute, a National Cancer Institute–designated Comprehensive Cancer Center, supported under National Institutes of Health grant P30-CA76292. We are grateful to Idhaliz Flores, director of the Puerto Rico Bio-Bank, the many patients who graciously provided data and tissue, and those health care workers who took the extra time to make these samples available for research. We also thank Tania Mesa, Marek Wloch, and many others who contributed expert technical support.

Supplementary Data

Note: To access the supplementary material accompanying this article, visit the online version of the *Journal of Thoracic Oncology* at www.jto.org and at http://dx.doi.org/10.1016/j.jtho.2017.08.019.

References

1. Colby SL, Ortman JM. *Projections of the size and composition of the U.S. population: 2014 to 2060. Current Population Reports*. Washington, DC: U.S. Census Bureau; 2014:25-1143.
2. Spratt DE, Chan T, Waldron L, et al. Racial/ethnic disparities in genomic sequencing. *JAMA Oncol*. 2016;2:1070-1074.
3. Schabath MB, Cress D, Muñoz-Antonia T. Racial and ethnic differences in the epidemiology and genomics of lung cancer. *Cancer Control*. 2016;23:338-346.
4. Cress WD, Chiappori A, Santiago P, Muñoz-Antonia T. Lung cancer mutations and use of targeted agents in Hispanics. *Rev Recent Clin Trials*. 2014;9:225-232.
5. Steuer CE, Behera M, Berry L, et al. Role of race in oncogenic driver prevalence and outcomes in lung adenocarcinoma: results from the Lung Cancer Mutation Consortium. *Cancer*. 2016;122:766-772.
6. Soh J, Toyooka S, Matsuo K, et al. Ethnicity affects EGFR and KRAS gene alterations of lung adenocarcinoma. *Oncol Lett*. 2015;10:1775-1782.
7. Pao W, Girard N. New driver mutations in non-small-cell lung cancer. *Lancet Oncol*. 2011;12:175-180.
8. Campbell JD, Lathan C, Sholl L, et al. Comparison of prevalence and types of mutations in lung cancers among black and white populations. *JAMA Oncol*. 2017;3:801-809.
9. Reinersman JM, Johnson ML, Riely GJ, et al. Frequency of EGFR and KRAS mutations in lung adenocarcinomas in African Americans. *J Thorac Oncol*. 2011;6:28-31.
10. Leduc N, Atallah V, Agossou M, et al. Lung adenocarcinoma survival in EGFR-mutated African-Caribbean patients: a multicenter study in the French West Indies [e-pub ahead of print]. *Target Oncol*. http://dx.doi.org/10.1007/s11523-017-0512-7, accessed September 21, 2017.
11. Arrieta O, Cardona AF, Martin C, et al. Updated frequency of EGFR and KRAS mutations in nonsmall-cell lung cancer in Latin America: the Latin-American Consortium for the Investigation of Lung Cancer (CLICaP). *J Thorac Oncol*. 2015;10:838-843.
12. Lopez-Chavez A, Thomas A, Evbuomwan MO, et al. EGFR mutations in Latinos from the United States and Latin America. *J Glob Oncol*. 2016;2:259-267.
13. Raez LE, Mourafetis J, Abrahams N, et al. Ethnic differences in non-small cell lung cancer (NSCLC) genetics. *J Thorac Oncol*. 2014;9:S159.
14. McQuitty E, Zhang W, Hendrickson H, et al. Lung adenocarcinoma biomarker incidence in Hispanic versus non-Hispanic white patients. *Arch Pathol Lab Med*. 2014;138:390-394.
15. Chen L, Engel BE, Welsh EA, et al. A sensitive NanoString-based assay to score STK11 (LKB1) pathway disruption in lung adenocarcinoma. *J Thorac Oncol*. 2016;11:838-849.