Original Article

Integrative Genomic Analysis Identifies Ancestry-Related Expression Quantitative Trait Loci on DNA Polymerase β and Supports the Association of Genetic Ancestry With Survival Disparities in Head and Neck Squamous Cell Carcinoma

Meganathan P. Ramakodi, PhD[1,2,3,4,5]; Karthik Devarajan, PhD[6,7,8]; Elizabeth Blackman, MPH[1,5]; Denise Gibbs, BS[1,5]; Danièle Luce, PhD[5,9]; Jacqueline Deloumeaux, MD, PhD[5,10]; Suzy Duflo, MD, PhD[11]; Jeffrey C. Liu, MD[12,13]; Ranee Mehra, MD[14]; Rob J. Kulathinal, PhD[2,3,4,5]; and Camille C. Ragin, PhD, MPH[1,5,7,13]

BACKGROUND: African Americans with head and neck squamous cell carcinoma (HNSCC) have a lower survival rate than whites. This study investigated the functional importance of ancestry-informative single-nucleotide polymorphisms (SNPs) in HNSCC and also examined the effect of functionally important genetic elements on racial disparities in HNSCC survival. **METHODS:** Ancestry-informative SNPs, RNA sequencing, methylation, and copy number variation data for 316 oral cavity and laryngeal cancer patients were analyzed across 178 DNA repair genes. The results of expression quantitative trait locus (eQTL) analyses were also replicated with a Gene Expression Omnibus (GEO) data set. The effects of eQTLs on overall survival (OS) and disease-free survival (DFS) were evaluated. **RESULTS:** Five ancestry-related SNPs were identified as cis-eQTLs in the DNA polymerase β (*POLB*) gene (false discovery rate [FDR] < 0.01). The homozygous/heterozygous genotypes containing the African allele showed higher *POLB* expression than the homozygous white allele genotype ($P < .001$). A replication study using a GEO data set validated all 5 eQTLs and also showed a statistically significant difference in *POLB* expression based on genetic ancestry ($P = .002$). An association was observed between these eQTLs and OS ($P < .037$; FDR < 0.0363) as well as DFS ($P = .018$ to .0629; FDR < 0.079) for oral cavity and laryngeal cancer patients treated with platinum-based chemotherapy and/or radiotherapy. Genotypes containing the African allele were associated with poor OS/DFS in comparison with homozygous genotypes harboring the white allele. **CONCLUSIONS:** Analyses show that ancestry-related alleles could act as eQTLs in HNSCC and support the association of ancestry-related genetic factors with survival disparities in patients diagnosed with oral cavity and laryngeal cancer. ***Cancer* 2017;123:849-60.** © *2016 American Cancer Society.*

KEYWORDS: DNA polymerase β, expression quantitative trait locus (eQTL), genetic ancestry, head and neck squamous cell carcinoma, survival disparity.

INTRODUCTION

African Americans (AFR-AMRs) with head and neck squamous cell carcinoma (HNSCC) have consistently lower survival rates in comparison with white patients.[1,2] Previous studies have associated socioeconomic status with these survival differences.[3] However, recent literature suggests the contribution of genetic differences between populations to survival disparities in some cancer types.[4]

Corresponding author: Camille C. Ragin, PhD, MPH, Cancer Prevention and Control Program, Fox Chase Cancer Center–Temple Health, 333 Cottman Avenue, Philadelphia, PA 19111; Fax: (215) 214-1622; camille.ragin@fccc.edu

[1]Cancer Prevention and Control Program, Fox Chase Cancer Center–Temple Health, Philadelphia, Pennsylvania; [2]Department of Biology, Temple University, Philadelphia, Pennsylvania; [3]Center for Computational Genetics and Genomics, Temple University, Philadelphia, Pennsylvania; [4]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, Pennsylvania; [5]African-Caribbean Cancer Consortium, Philadelphia, Pennsylvania; [6]Biostatistics and Bioinformatics Facility, Fox Chase Cancer Center–Temple Health, Philadelphia, Pennsylvania; [7]Department of Epidemiology and Biostatistics, College of Public Health, Temple University, Philadelphia, Pennsylvania; [8]Center for High-Dimensional Statistics, Big Data Institute, Temple University, Philadelphia, Pennsylvania; [9]National Institute for Health and Medical Research (INSERM), Unit 1085; Institute for Research in Health, Environment, and Work (IRSET), Pointe-à-Pitre, Guadeloupe, French West Indies; [10]General Cancer Registry of Guadeloupe, University Hospital of Pointe-à-Pitre, Pointe-a-Pitre, Guadeloupe, French West Indies; [11]Department of Oto-Rhino-Laryngology and Head and Neck Surgery, University Hospital of Pointe à Pitre, Pointe-a-Pitre, Guadeloupe, French West Indies; [12]Head and Neck Surgery, Fox Chase Cancer Center–Temple Health, Philadelphia, Pennsylvania; [13]Department of Otolaryngology–Head and Neck Surgery, Lewis Katz School of Medicine, Temple University School of Medicine, Philadelphia, Pennsylvania; [14]Department of Hematology/Oncology, Fox Chase Cancer Center–Temple Health, Philadelphia, Pennsylvania

We thank the Biological Resource Center of Guadeloupe (Stanie Gaëte) as well as the Fox Chase Cancer Center Biorepository Facility (Philadelphia, PA) for managing and providing the patient samples for this study. We thank Erica Golemis (Fox Chase Cancer Center) for her valuable comments on this article. The results shown here are in whole or part based on data generated by The Cancer Genome Atlas Research Network (http://cancergenome.nih.gov/).
Additional supporting information may be found in the online version of this article

DOI: 10.1002/cncr.30457, **Received:** April 28, 2016; **Revised:** September 26, 2016; **Accepted:** October 28, 2016, **Published online** December 1, 2016 in Wiley Online Library (wileyonlinelibrary.com)

Survival from cancer depends on the success of common treatment methods such as chemotherapy, radiotherapy, and surgery. Treatment planning for potentially curable disease often requires a multidisciplinary approach.[5-7] Ionizing radiation generates free radicals, which damage cellular DNA; this results in apoptosis.[8] In chemotherapy, platinum-based drugs such as cisplatin and carboplatin are commonly used to treat HNSCC.[9,10] These platinum-based drugs bind to DNA and form DNA adducts,[11] and this leads to cell cycle arrest and cytotoxicity.[12] In response to DNA lesions caused by chemotherapy and/or radiotherapy, the cellular DNA damage response system repairs DNA aberrations and can reduce treatment sensitivity in cancer patients.[13,14] Thus, DNA repair genes play a key role in the treatment outcomes for many cancers, including HNSCC.

Higher level expression of DNA repair genes have been observed in many cancer types,[15,16] and increased expression levels of DNA repair genes are associated with reduced sensitivity to chemotherapy and radiation therapy.[17,18] The increased expression of DNA repair genes in some cancers arises from multiple factors, including signaling within the tumor or from the tumor microenvironment, which leads to epigenetic upregulation, and/or de novo somatic mutations. Some evidence suggests that host genomic factors such as germline variants in DNA repair–associated genes may be associated with individual differences in responses. Previous studies have shown that germline variants can act as expression quantitative trait loci (eQTLs) and affect the expression of genes in cancer.[19] Such germline variants may be preferentially found in certain populations. Indeed, a number of studies have investigated the effect of population-specific genetic variants on gene expression in normal, non-cancer samples from various human populations and identified differential gene expression levels regulated by ancestry-related alleles.[20]

AFR-AMR and white HNSCC patients possess distinct genetic ancestries. Our recent genomic analysis of laryngeal cancer in AFR-AMR and white patients revealed that distinctive genetic ancestry corresponds to molecular differences in the laryngeal cancers arising in these 2 populations.[21] To date, the functional role of population-specific genomics variants in HNSCC survival is unknown. In this study, the functional role of ancestry-related genomic factors in the expression of DNA damage response genes and the effect of ancestry-informative single-nucleotide polymorphisms (SNPs) on racial disparities in HNSCC survival were investigated with an integrative genomics approach. We tested the hypothesis that ancestry-related genomic elements are associated with HNSCC survival disparities between AFR-AMR and white populations because of altered gene expression in DNA damage response genes, which affects the sensitivity to chemotherapy and/or radiotherapy. This work, focusing on oral cavity and laryngeal cancer, is the first study to investigate the functional importance of ancestry-informative SNPs in HNSCC and to examine the effect of functionally important genetic elements on racial disparities in HNSCC survival.

MATERIALS AND METHODS

Data Source

The genotype data for 4802 genome-wide ancestry-informative SNPs for 316 oral cavity and laryngeal cancer patients (30 AFR-AMRs and 286 whites) were retrieved from The Cancer Genome Atlas (TCGA) for eQTL analyses. Detailed methodologies explaining how the 4802 ancestry-informative SNPs were retrieved from TCGA are provided in the online supporting information. The raw genome-wide methylation data for approximately 485,000 CpG sites, based on the Infinium HumanMethylation450 BeadChip Kit (Illumina, Inc), were retrieved for the tumors of the 316 HNSCC patients from TCGA. The *M* values (methylation signals) were calculated for each CpG site with the Bioconductor minfi package.[22] The somatic copy number variation (CNV) data and RNA sequencing–based gene expression data (normalized RSEM values) for the tumors of the 316 HNSCC patients in TCGA were also obtained from the Broad Institute.[23] A gene was included in the analyses only when 50% or more of the patients had expression values for that gene.

eQTL Analyses

The effect of each ancestry-informative SNP on the expression of nearby genes (±1 Mb from the SNP) was analyzed to find potential ancestry-related eQTL candidates. A multivariable regression model was used to identify eQTLs (equation 1). A negative binomial regression model, based on the empirical mean-variance relation for gene expression, was used:

$$\mathrm{Exp_a} = \mathrm{GT_a} + \mathrm{CNV_a} + \mathrm{M_a} + \mathrm{Pop_a} + \varepsilon_a \qquad (1)$$

where $\mathrm{Exp_a}$ is the expression of gene a, $\mathrm{GT_a}$ is the genotype of the single-nucleotide polymorphism under study, $\mathrm{CNV_a}$ is the somatic copy number variation of gene a, $\mathrm{M_a}$ indicates the methylation levels of CpG sites associated with gene a, $\mathrm{Pop_a}$ indicates the top 3 principal component values to adjust for population stratification, and ε_a is the residual error.

The regression analyses were performed with the MASS package in R (version 3.0.2). All tests were 2-sided, and *P* values were corrected for multiple tests according to the Benjamini-Hochberg method (false discovery rate [FDR]). Potential ancestry-informative SNPs affecting the expression of DNA repair genes with an FDR ≤ 0.01 were retrieved. A list of 178 DNA repair genes (updated on April 15, 2014) was obtained from the online resource of The University of Texas MD Anderson Cancer Center[24] for our study. This comprehensive set of DNA repair genes has been used widely in several recent publications.[25,26] We also tested the effects of eQTLs (found in the pooled data set [30 AFR-AMRs and 286 whites]) on the expression of DNA repair genes with the regression model (equation 1) using the data set for white patients (n = 286).

Other Statistical Analyses

Age distribution differences between AFR-AMR and white patients were studied with the Mann-Whitney-Wilcoxon test. Differences in the proportions of current smokers, ex-smokers, and never smokers, and each pathological stage between AFR-AMR and white patients were assessed with a test of proportions. Gene expression levels of AFR-AMR and white patients were compared with the Mann-Whitney-Wilcoxon test. The genotype data were retrieved for each eQTL on the basis of ancestry, and the allele frequencies of each eQTL were calculated. Also, the allele frequencies of each eQTL were retrieved for ASW (African ancestry in the Southwest United States) and CEU (Utah residents with Northern and Western European ancestry) populations from 1000 Genomes Project (1000G) data, and the proportions of allele frequencies were compared between TCGA and 1000G samples (ASW vs TCGA AFR-AMRs and CEU vs TCGA whites) with a test of proportions. All tests were 2-sided, and *P* values $\leq .05$ were considered to be statistically significant.

Replication of eQTL Analyses

The TCGA eQTL analyses were replicated with a Gene Expression Omnibus (GEO) data set (GSE39368) generated by Walter et al.[27] This data set contains genome-wide SNP and CNV data for 99 HNSCC patients and gene expression data for 138 HNSCC patients. The SNP, CNV, and gene expression data sets were combined into a single data set with missing values coded as Not Available (NA). Data were extracted if a patient's ancestry was either AFR-AMR or white and the anatomical site represented the oral cavity or larynx. After these filters, data for 96 HNSCC patients (73 whites and 23 AFR-AMRs) were retrieved for further analyses. In this data set, the expression of each gene was measured with multiple probes, and the mean of multiple probes of each gene was taken as the expression measure of that gene. The effect of each SNP identified as an eQTL in the TCGA data set on gene expression was tested with the linear regression model after adjustments for CNV.

Linkage Disequilibrium (LD) Analyses

Genomic data for all the eQTLs and SNPs within a distance of ± 50 kb from each eQTL were retrieved for the ASW population from the 1000G database, and LDs between each eQTL and its nearby SNPs were analyzed with VCFtools.[28] SNPs in strong LD (scaled linkage disequilibrium estimate $[D'] \geq 0.8$) with eQTLs were identified for further analyses. The LD heat map was generated with Haploview software.[29]

Encyclopedia of DNA Elements (ENCODE) Functional Analyses

The genome-wide deoxyribonuclease (DNase) I sensitivity assay data and transcription factor binding sites were retrieved from ENCODE.[30] Each eQTL and LD SNP was intersected with the ENCODE database with custom Perl and Shell scripts. The eQTL/LD SNP was thought to be functionally important if the eQTL/SNP was found in the regulatory region of a gene, and DNase I sensitivity and/or transcription factor binding site. The genomic position of functionally important eQTL/SNPs was visualized with the University of California Santa Cruz genome browser.[31]

Survival Analyses Based on eQTLs

The effect of eQTLs on overall survival (OS) and disease-free survival (DFS) for HNSCC patients with a history of platinum-based chemotherapy and/or radiation therapy was investigated. First, Kaplan-Meier plots were generated for each eQTL to visualize the effect of eQTL genotypes with Stata (version 14.01). Second, hazard ratios (HRs) for the risk of death (OS) according to the eQTL genotype were calculated with Cox proportional hazards regression models after adjustments for age and pathological stage with the survival package in R, and a goodness-of-fit test using Schoenfeld residuals was performed to test the appropriateness of the Cox proportional hazards model.[32] All tests were 2-sided, and a *P*-value threshold of .05 was used to determine statistical significance. The *P* values were corrected for multiple tests with the Benjamini-Hochberg method.

To validate the survival analyses results, existing germline DNA and clinical data for 20 additional oral

cavity or laryngeal cancer patients of African (AFR) ancestry who had platinum-based chemotherapy and/or radiation therapy were obtained.[33] Germline DNA for all 20 patients was genotyped for 1 of the eQTLs (rs2272733) with a TaqMan real-time polymerase chain reaction assay (Life Technologies/Thermo Fisher Scientific, Waltham, Mass). The genotype and clinical data for these 20 patients were combined with the data for 157 patients from TCGA to generate an enriched data set (data set 2) containing 177 HNSCC patients (36 AFR-AMR patients and 141 white patients) with a history of platinum-based chemotherapy and/or radiation therapy for survival analyses. All the human subject investigations were approved by institutional review boards of the Fox Chase Cancer Center.

Estimation of Admixture Proportions and Survival Analyses

Autosomal ancestry-informative marker data for TCGA patients along with YRI (Yoruba African), CEU, JPT (Japanese Tokyo), and CHB (Han Chinese) individuals from 1000G were retrieved. The genetic admixture proportions for each individual, including AFR-AMR and white patients from TCGA, were estimated with a model-based clustering approach implemented in STRUCTURE (version 2.3.4).[34] In STRUCTURE, the data were analyzed with different K values (genetic clusters) ranging from 3 to 10 under the admixture model. For each K value, 10 runs were performed with 10,000 burn-in and an additional 20,000 replicates. The best K value was estimated with the method of Evanno et al[35] as implemented in the STRUCTURE HARVESTER program.[36] The output of STRUCTURE based on the best K value was analyzed with CLUMPP.[37] The AFR-admixed fraction of each HNSCC patient (n = 157) with platinum-based chemotherapy and/or radiotherapy history was obtained from the CLUMPP output and used for survival analyses. The effects of the AFR-admixed proportion on OS and DFS were analyzed with Cox proportional hazards regression models from the survival package in R after adjustments for age and pathological stage. Goodness-of-fit tests using Schoenfeld residuals were performed to evaluate the appropriateness of the Cox proportional hazards model. All tests were 2-sided, and a P-value threshold of .05 was used to determine statistical significance.

RESULTS

TCGA Sample Characteristics

Summary statistics for the 316 HNSCC patients included in this analysis are provided in Table 1. The differences in age, smoking status, and pathological stage between AFR-AMR and white patients were not significant.

TABLE 1. Characteristics of Patients With Head and Neck Squamous Cell Carcinoma From The Cancer Genome Atlas

Characteristics	All (n = 316)	African Americans (n = 30)	Whites (n = 286)
Age, y	61.56 ± 11.5	58 ± 7.74	61.94 ± 11.8
Sex, No. (%)			
Male	226	24 (80.0)	202 (70.6)
Female	90	6 (20.0)	84 (29.4)
Tumor site, No. (%)			
Hypopharynx	5	1 (3.3)	4 (1.4)
Larynx	81	11 (36.7)	70 (24.5)
Oral cavity[a]	230	18 (60.0)	212 (74.1)
Smoking status, No. (%)			
Current smoker	115	16 (53.3)	99 (34.6)
Ex-smoker	128	9 (30.0)	119 (41.6)
Never smoker	64	2 (6.7)	62 (21.7)
Unknown	9	3 (10.0)	6 (2.1)
Pathological stage, No. (%)			
I	15	1 (3.3)	14 (4.9)
II	47	1 (3.3)	46 (16.1)
III	49	3 (10.0)	46 (16.1)
IV	177	22 (73.3)	155 (54.2)
Unknown	28	3 (10.0)	25 (8.7)

[a] The oral cavity includes the alveolar ridge, buccal mucosa, floor of the mouth, and oral tongue.

Effect of Ancestry-Informative SNPs on the Expression of DNA Repair Genes (eQTL Analysis)

The focus of this study was an analysis of the effect of ancestry-informative SNPs on an annotated set of 178 DNA repair genes. Our results showed that the expression of 1 DNA repair gene, DNA polymerase β (*POLB*), was significantly affected by nearby ancestry-informative SNPs with an FDR $\leq$ 0.01. Among 4802 ancestry-informative SNPs, 5 SNPs (rs2272733, rs3136790, rs6474387, rs2272732, and rs10096210) were found to be eQTLs affecting the expression of *POLB* (FDR $\leq$ 0.01). Each of the 5 SNPs was also observed to be an eQTL when the data set was limited to whites only. The P values for the 5 SNPs based on pooled (AFR-AMR and white) and white data sets are reported in Supporting Table 1 (see online supporting information). As an illustration, the effects of rs2272732 on *POLB* are shown in Figure 1 (TCGA panel). The effects of the other 4 SNPs on *POLB* were similar to the effects of rs2272732 (Supporting Fig. 1 [see online supporting information]). In this article, we use the terms *AFR allele* and *white allele* to

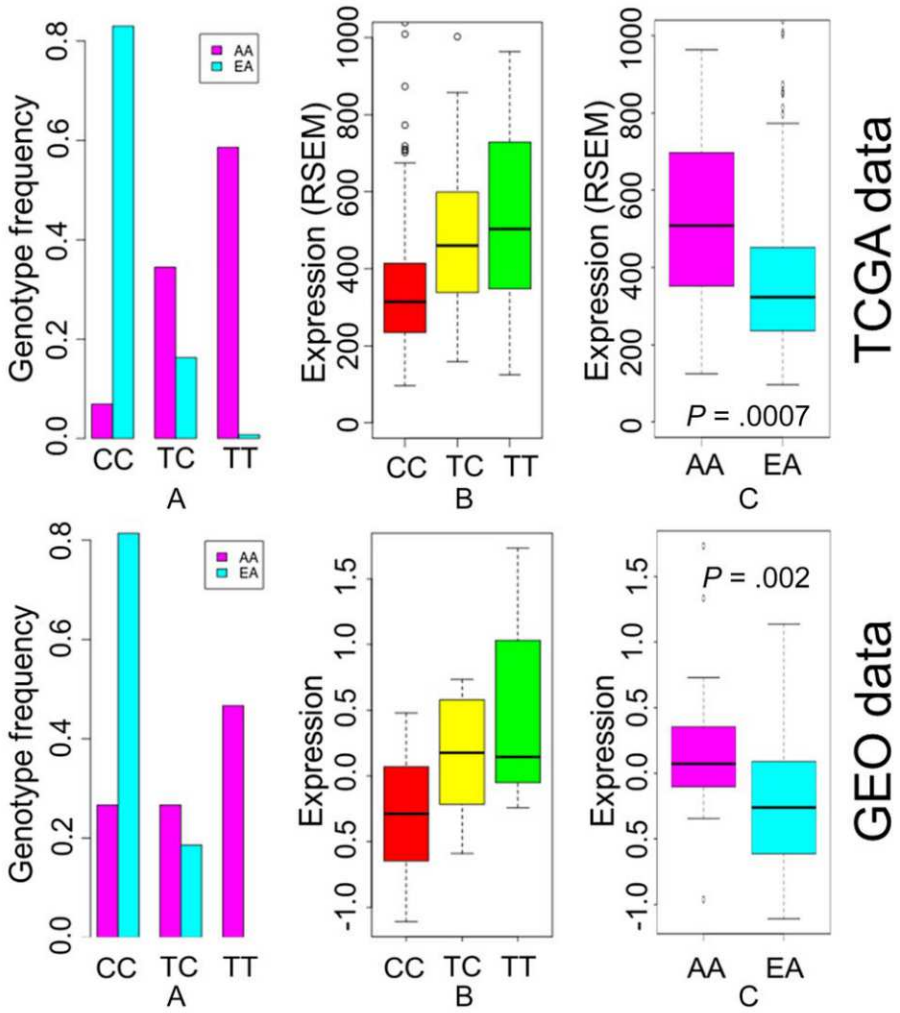


Figure 1. Results of TCGA and GEO data analyses: (A) genotype frequencies of rs2272732 for AA and EA patients, (B) effect of rs2272732 on DNA polymerase β gene expression, and (C) DNA polymerase β gene expression for AA and EA patients. AA indicates African American; EA, white; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas.

denote the major alleles specific to AFR-AMR and white populations, respectively. Homozygous genotypes containing the AFR allele were associated with higher levels of *POLB*, whereas homozygous white-allele genotypes had decreased *POLB* expression. Heterozygous genotypes containing an AFR allele and a white allele were associated with moderately higher levels of *POLB* expression (TCGA panel in Fig. 1B). A comparison of *POLB* messenger RNA expression data showed higher level expression of *POLB* in AFR-AMR patients (quartile 1 [Q1], 355.2; median, 508.3; quartile 3 [Q3], 693.7) versus white patients (Q1, 236.1; median, 323.2; Q3, 451.2). There was a statistically significant difference in *POLB* expression between AFR-AMR and white patients ($P < .001$; TCGA panel in Fig. 1C).

Replication of eQTL Analyses

The results of eQTL analyses based on TCGA data were replicated with a GEO data set and were consistent with the results observed from the TCGA data analyses. The regression analyses identified all 5 SNPs as eQTLs affecting the expression of the *POLB* gene ($P < .007$). The effects of rs2272732 on *POLB* expression in the GEO data set are shown in Figure 1 (GEO panel). As with the TCGA data, homozygous AFR-allele genotypes of all 5 SNPs were associated with higher expression levels of *POLB* in comparison with the homozygous white-allele genotype from the GEO data set. Also, the heterozygous genotypes had moderately higher levels of *POLB* expression than the homozygous genotypes of the white allele. An evaluation of GEO data confirmed that AFR-AMR

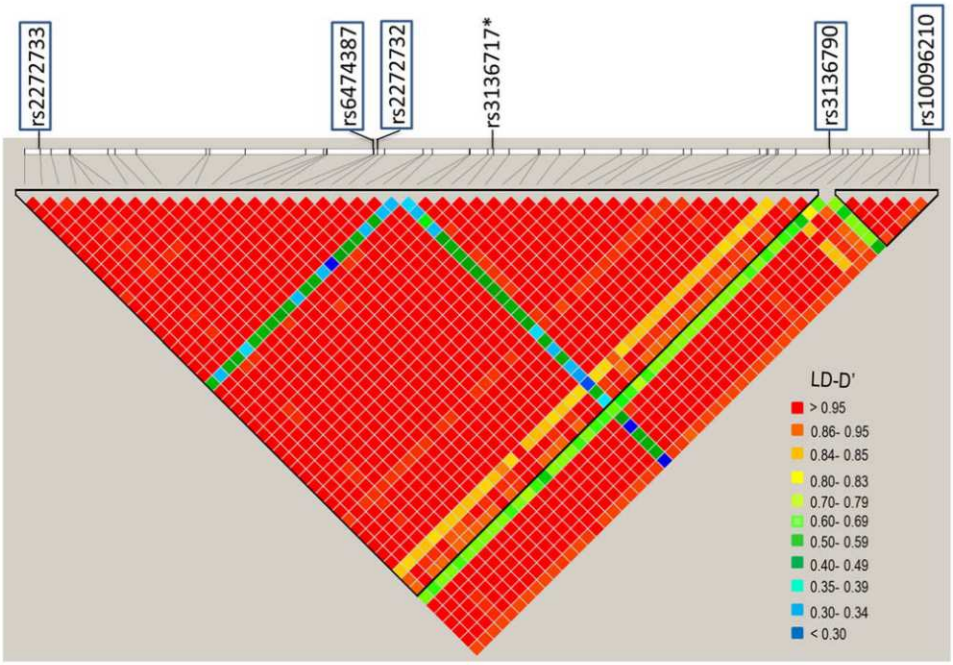


Figure 2. LD analyses show that all 5 eQTLs are in strong LD. In addition, another ancestry-informative single-nucleotide polymorphism, rs3136717, is in strong LD with all 5 eQTLs. *D'* indicates scaled linkage disequilibrium estimate; eQTL, expression quantitative trait locus; LD, linkage disequilibrium.

patients had a higher level of *POLB* expression (Q1, –0.10; median, 0.07; Q3, 0.35) than white patients (Q1, –0.61; median, –0.26; Q3, 0.09), and there was a significant difference in *POLB* expression levels between the 2 populations ($P = .002$).

LD and ENCODE Analyses

The results of the LD analyses are shown in Figure 2. Each of the 5 eQTLs was in strong LD ($D' \geq 0.9$) with the others. In addition, another ancestry-informative SNP, rs3136717, had a strong association with all 5 eQTLs ($D' > 0.8$). The functional importance of the 5 eQTLs and rs3136717 was investigated with the ENCODE data. None of the 5 eQTLs were found to be in the *POLB* gene region with strong DNase I sensitivity/TF binding signals. However, the associated ancestry-informative SNP, rs3136717, was located in the regulatory region of *POLB* and in the DNase I sensitivity region of *POLB* in all 125 cell lines assayed in the ENCODE project. In addition, rs3136717 is located in the binding site of several transcription factors, specifically polymerase (RNA) II subunit A (POLR2A). The genomic position and associated ENCODE annotations for rs3136717 are shown in Figure 3.

Allele Frequencies of eQTLs Between TCGA and 1000G Data

The allele frequencies of the 5 eQTLs for AFR-AMR and white patients in TCGA were estimated and compared with the allele frequencies of their respective populations (ASW for AFR-AMR patients [Figure 4A] and CEU for white patients [Figure 4B]) from the 1000G data. The allele frequencies from the TCGA and 1000G data sets were not significantly different.

Survival Analyses Based on eQTLs

The effect of each eQTL on the OS of HNSCC patients who were treated with platinum-based chemotherapy and/or radiotherapy (n = 157; 16 AFR-AMRs and 141 whites) was examined. All 5 eQTLs were found to be significantly associated with OS (log-rank test $P < .037$; FDR < 0.0363). The Kaplan-Meier plot for a representative eQTL, rs2272733, is shown in Figure 5A. The Kaplan-Meier plots for the other 4 eQTLs are shown in Supporting Figure 2 (see online supporting information). The DFS analyses found that rs2272732 and rs2272733 were significantly associated with DFS ($P < .05$), whereas rs3136790 and rs10096210 were associated with DFS with moderate significance ($P = .0544$ to $.0629$). The Kaplan-Meier plots of DFS for each of the 5 eQTLs are

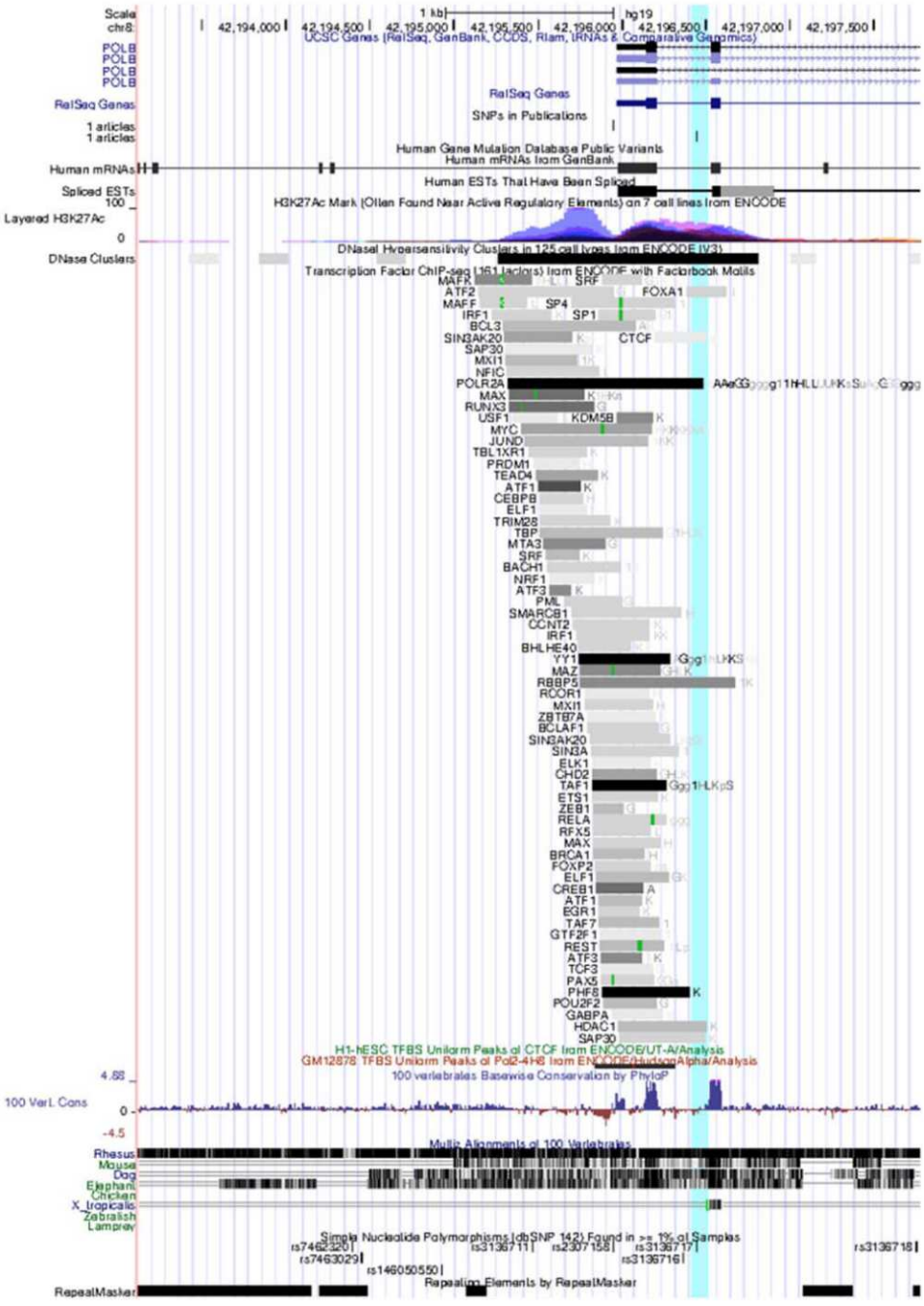


Figure 3. Genomic position of rs3136717 based on ENCODE data. The single-nucleotide polymorphism rs3136717 (shaded in cyan) is located in a known regulatory region, a deoxyribonuclease I sensitivity region, and intersects several transcription factor binding sites of the *POLB* gene. ENCODE indicates Encyclopedia of DNA Elements; POLB, DNA polymerase β.

shown in Supporting Figure 3 (see online supporting information). The OS HRs for each genotype of the 5 eQTLs were calculated with the Cox proportional hazards model after adjustments for age and pathological stage (Table 2). The goodness-of-fit test confirmed the appropriateness of the Cox proportional hazards model for OS analyses. For 4 of the 5 eQTLs, patients with the homozygous genotype of the white allele had a significantly lower risk of death ($P < .0003$; FDR < 0.0008) than patients with the homozygous AFR allele. Also, patients with the

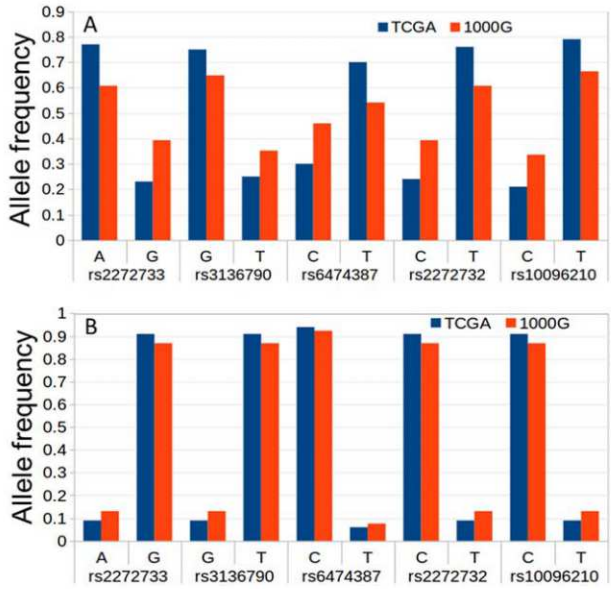


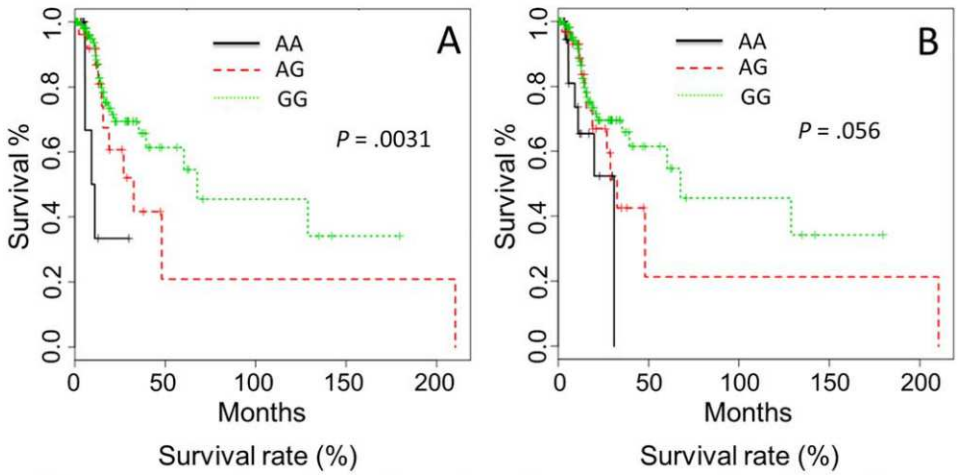
Figure 4. Allele frequency differences between TCGA and 1000G populations: (A) allele frequency differences between TCGA African American patients and the 1000G ASW population data set and (B) allele frequency differences between TCGA white patients and the 1000G CEU population data set. 1000G indicates 1000 Genomes Project; ASW, African ancestry in the Southwest United States; CEU, Utah residents with Northern and Western European ancestry; TCGA, The Cancer Genome Atlas.

heterozygous genotypes for rs2272733 and rs3136790 were found to have a significantly lower risk of death ($P < .002$; FDR < 0.0016) than patients who had homozygous genotypes consisting of the AFR-AMR major allele. The HR for the genotypes of rs6474387 was not found to be significant.

We also evaluated the effect of the 5 eQTLs on the OS and DFS of patients not treated with cisplatin/carboplatin/radiotherapy. None of the eQTL genotypes were found to be significantly associated with OS ($P > .6$) or DFS ($P > .8$).

Validation of eQTL-Based Survival Analyses

The survival analysis results from data set 2 for SNP rs2272733 were consistent with the results for rs2272733 from the TCGA data set. The Kaplan-Meier plot and survival rates are shown in Figure 5B. The homozygous genotype of the AFR allele (AA) was associated with lower OS in comparison with the other 2 genotypes (AG and GG; log-rank test $P = .056$). The Cox proportional hazards model, after adjustments for age, clinical stage, and cohort, revealed that patients with heterozygous genotype AG had a significantly lower risk of death than patients with homozygous AFR allele genotype AA (HR for AA vs AG, 0.26; 95% confidence interval [CI], 0.08-0.84; $P = .024$). Also, patients with the homozygous genotype



Survival rate (%)

Year	AA	AG	GG
1 y	33.3	86.7	88.3
2 y	--	60.7	69.4
5 y	--	20.8	61.4

Survival rate (%)

Year	AA	AG	GG
1 y	65.4	88.6	87.8
2 y	52.3	67.0	69.6
5 y	--	21.3	61.5

Figure 5. Kaplan-Meier plots of overall survival for head and neck squamous cell carcinoma patients treated with platinum-based chemotherapy and/or radiation therapy according to the rs2272733 genotypes: (A) TCGA data (n = 157) and (B) enriched data set (data set 2 [n = 177]). TCGA, The Cancer Genome Atlas.

TABLE 2. Hazard Ratios for 5 eQTLs Based on Overall Survival After Adjustments for the Age and Clinical Pathological Stage

eQTL	Major Allele		Genotype	Hazard Ratio	95% Confidence Interval		FDR
	African American	White			Lower	Upper	
rs2272733			AA	1.00 (reference)	—	—	—
	A	G	AG	0.079	0.016	0.380	0.0033
			GG	0.075	0.018	0.310	0.0008
rs3136790			GG	1.00 (reference)	—	—	—
	G	T	GT	0.048	0.008	0.281	0.0016
			TT	0.064	0.015	0.274	0.0008
rs6474387			CC	1.00 (reference)	—	—	—
	T	C	TC	1.869	0.695	5.026	0.27
			TT	4.693	0.560	39.354	0.22
rs2272732			CC	1.00 (reference)	—	—	—
	T	C	TC	0.670	0.200	2.260	0.53
			TT	13.500	3.260	55.876	0.0008
rs10096210			CC	1.00 (reference)	—	—	—
	T	C	TC	0.679	0.202	2.284	0.53
			TT	13.478	3.258	55.755	0.0008

Abbreviations: eQTL, expression quantitative trait locus; FDR, false discovery rate.

containing the white allele (GG) had a significantly lower risk of death than patients with the homozygous AFR-AMR genotype (AA; HR for AA vs GG, 0.15; 95% CI, 0.04-0.48; $P = .0012$).

Survival Analyses Based on Genetic Proportions

The main objective of this article was to understand the effect of AFR ancestry on survival disparities in HNSCC. Thus, we analyzed the effect of AFR admixture on survival in HNSCC patients with a history of platinum-based chemotherapy and/or radiotherapy. The HRs for OS and DFS were 8.99 (95% CI, 1.53-52.95; $P = .015$) and 7.12 (95% CI, 1.46-34.77; $P = .015$), respectively. The goodness-of-fit test confirmed the appropriateness of the Cox proportional hazards model for OS and DFS analyses.

DISCUSSION

Ancestry-Informative SNPs Act as eQTLs

Our stringent criteria (FDR$\leq$0.01) found 5 ancestry-informative SNPs to be eQTLs affecting *POLB* expression. This is the first study to demonstrate the impact of ancestry-informative SNPs on the expression of a gene involved in head and neck cancer tissues with a subsequent effect on survival disparities. The expression level of *POLB* significantly differed between AFR-AMR and white patients ($P < .001$), with AFR-AMR patients possessing higher levels of *POLB* expression in their tumors than white patients because of ancestry-related eQTLs. In general, major factors that could alter gene expression are SNPs (eQTLs), CNVs, somatic point mutations, methylation, and population differences. In our analyses, we included CNV, methylation, and the top 3 principal components in our regression model. Therefore, the effects of CNV, methylation, and population structure on *POLB* expression were adjusted, and this suggests an independent association of SNPs with gene expression. In addition, our candidate SNPs were also found to be eQTLs when they were tested exclusively with a white patient data set. We also checked somatic point mutations in our TCGA cohort based on exome data and did not find any somatic point mutations in the *POLB* gene.

The allele frequencies of the 5 eQTLs for AFR-AMR patients and the 1000G ASW population and for white patients and the 1000G CEU population were not significantly different (Fig. 2; $P > .6$). Thus, we expect to observe the same effect for these eQTLs that we found in TCGA patients in any AFR-AMR and white populations. However, the effects of eQTLs in AFR-AMR and white controls (non-cancer samples) remain to be determined. Figure 4 reveals that candidate eQTL allele frequencies were different between AFR-AMR and white individuals, regardless of which data set (TCGA or 1000G) was used. In addition, for all 5 eQTLs, each specific allele was enriched in AFR-AMR patients versus white patients.

We also validated the results of the TCGA data set by replicating the eQTL analyses with a GEO data set. If the results obtained from the TCGA data set occurred by chance or were due to some unknown factors associated with TCGA samples, we would expect to see different

results with the GEO data set. Indeed, the results for the 5 ancestry-informative SNPs based on the GEO data set were consistent with the results of TCGA data analyses. Each of the 5 eQTLs observed in the TCGA data were identified as eQTLs and affected *POLB* expression in the GEO data set (Fig. 1). Thus, we provide evidence that ancestry-informative SNPs could act as eQTLs and alter the expression of genes in HNSCC. Although these results were observed on the basis of RNA expression, further study is required to confirm the effect of eQTLs on protein expression.

Increased levels of *POLB* expression have been shown to be associated with tumorigenesis.[38,39] In addition, an increased level of *POLB* expression has been observed in many cancer types,[40] and this suggests that a higher level of *POLB* expression could be associated with the risk of HNSCC. Compared with other groups, AFR-AMRs have a higher incidence of HNSCC, particularly in the larynx. These observations support the association between HNSCC incidence disparities and ancestry-informative SNPs. However, further studies are needed to confirm this association.

Identification of Potentially Functionally Important eQTLs

All 5 eQTLs are located within or near ($\pm$ 1Mb) the *POLB* gene region, and all 5 eQTLs are in strong LD. In addition, we have identified another ancestry-informative SNP, rs3136717, for which data are not available in TCGA and GEO data sets, in strong LD with the 5 eQTLs. It is expected that rs3136717 will be associated with levels of *POLB* expression similar to those observed with the other 5 eQTLs. Unfortunately, we could not test the effect of rs3136717 on *POLB* because of the unavailability of genotype data for rs3136717 in TCGA. Our analyses have shown that rs3136717 is found at the 5′ end of the *POLB* gene and in the DNase I sensitivity region in all 125 surveyed cell lines. In addition, rs3136717 is located in the region where the transcription factor POLR2A binds to the *POLB* gene (Fig. 4). Thus, rs3136717 is likely in an active regulatory region of the *POLB* gene. POLR2A is the major subunit of RNA polymerase II, which is required for RNA transcription. Because rs3136717 is in an active regulatory region and on the POLR2A binding site, we speculate that alternative alleles of rs3136717 could affect the binding affinity of POLR2A or other transcription factors to alter *POLB* gene expression. It is interesting to note that Figueroa el al[41] have already identified rs3136717 as being associated with the risk of bladder cancer in a case-control study. Thus, rs3136717 could be functionally important and could be associated with the risk of many cancers, including HNSCC. The role of rs3136717 in *POLB* expression in HNSCC needs to be further evaluated experimentally.

Effect of Ancestry-Related Genomic Variants on the Treatment Outcome

Higher levels of *POLB* expression decrease the sensitivity of platinum-based chemotherapy and/or radiotherapy and thus may be associated with poor survival.[42,43] Therefore, the 5 eQTLs shown to modulate *POLB* expression are expected to affect the outcomes of patients treated with platinum-based chemotherapy and/or radiation therapy. Among patients treated with platinum-based drugs and/or radiotherapy, the genotypes of eQTLs containing the AFR allele (homozygous/heterozygous) were associated with poor OS and DFS in comparison with the homozygous genotypes containing the white allele. Even after adjustments for age and pathological stage, statistically significant associations persisted. However, the results of OS analyses were slightly different among eQTLs (Table 2), even though these eQTLs were in strong LD. In addition, the effect of these eQTLs on DFS differed among eQTLs. This is not surprising because it is known that independent cis-eQTLs in LD can have different functional effects.[44] We also studied the effect of eQTLs among patients who were not treated with cisplatin/carboplatin/radiotherapy, and the results did not show any significant effect of eQTLs on the OS ($P > .6$) or DFS of these patients ($P > .8$). These findings provide evidence showing that the associations of these 5 eQTLs with OS/DFS were limited to patients with a cisplatin/carboplatin/radiotherapy treatment history and support the important role that *POLB* expression plays in treatment response.

We validated our survival analyses by analyzing 1 of the 5 eQTLs, rs2272733, in 141 white HNSCC patients and 36 HNSCC patients of AFR ancestry with a history of platinum-based chemotherapy and/or radiation therapy (data set 2). The results of TCGA and data set 2 for rs2272733 were consistent. The genotype containing at least 1 AFR allele (A) was associated with poorer OS in comparison with the homozygous white-allele (G) genotype.

To further test the idea that an AFR genetic ancestry is indeed related to survival disparities, we assayed the effect of AFR-admixed proportions on OS and DFS. The HRs for OS and DFS were >1.0, and this indicates that a higher AFR admixture is associated with poorer OS and DFS in HNSCC patients with a platinum-based chemotherapy and/or radiotherapy history. These results are

statistically significant; however, it is worth noting the wide CIs, which are partly due to the limited amount of data available for these endpoints. Moreover, a similar analysis using data from the entire cohort of subjects in this study showed similar tendencies and indicated poorer OS and DFS in patients with increasing AFR genetic admixture (data not shown). Thus, our study reveals a clear association between AFR ancestry–related genetic factors and poor treatment outcomes in AFR-AMR patients with HNSCC who were treated with platinum-based chemotherapy and/or radiotherapy. The validation of our findings in a larger independent cohort of subjects would help to further strengthen and establish their significance.

A limitation of this study is that socioeconomic status and environmental factors were not included in the analyses because such data were not available. However, we cannot ignore the effects of environmental factors and socioeconomic status on survival disparities. Thus, this study needs to be extended further to analyze the interactions between genetics and environmental factors/socioeconomic status with respect to survival disparities.

FUNDING SUPPORT

This work was supported by the American Cancer Society (RSG-14-033-01-CPPB to Camille C. Ragin) and the National Cancer Institute (CA006927). Meganathan P. Ramakodi is partially supported by the William J. Avery Postdoctoral Research Fellowship (Fox Chase Cancer Center). Danièle Luce acknowledges the support of the French National Cancer Institute.

CONFLICT OF INTEREST DISCLOSURES

Ranee Mehra reports consulting for Bristol-Myers Squibb, Novartis, Genentech, and Innate Pharma outside the submitted work.

AUTHOR CONTRIBUTIONS

Meganathan P. Ramakodi: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing–original draft, writing–review and editing, and visualization. **Karthik Devarajan**: Methodology, formal analysis, writing–review and editing, and visualization. **Elizabeth Blackman**: Data curation, writing–review and editing, and project administration. **Denise Gibbs**: Validation, investigation, data curation, and writing–review and editing. **Danièle Luce**: Resources, data curation, and writing–review and editing. **Jacqueline Deloumeaux**: Resources, data curation, and writing–review and editing. **Suzy Duflo**: Resources, data curation, and writing–review and editing. **Jeffrey C. Liu**: Writing–review and editing and visualization. **Ranee Mehra**: Writing–review and editing and visualization. **Rob J. Kulathinal**: Conceptualization, resources, writing–review and editing, and visualization. **Camille C. Ragin**: Conceptualization, formal analysis, resources, writing–review and editing, visualization, supervision, and funding acquisition

REFERENCES

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin.* 2014;64:9-29.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin.* 2016;66:7-30.
3. Ghafoor A, Jemal A, Cokkinides V, et al. Cancer statistics for African Americans. *CA Cancer J Clin.* 2002;52:326-341.
4. Tan DS, Mok TS, Rebbeck TR. Cancer genomics: diversity and disparity across ethnicity and geography. *J Clin Oncol.* 2016;34:91-101.
5. Vermorken JB, Specenier P. Optimal treatment for recurrent/metastatic head and neck cancer. *Ann Oncol.* 2010;21(suppl 7):vii252-vii261.
6. Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. *Lancet.* 2008;371:1695-1709.
7. Moeller BJ, Yordy JS, Williams MD, et al. DNA repair biomarker profiling of head and neck cancer: Ku80 expression predicts locoregional failure and death following radiotherapy. *Clin Cancer Res.* 2011;17:2035-2043.
8. Borek C. Antioxidants and radiation therapy. *J Nutr.* 2004;134:3207S-3209S.
9. Fung C, Grandis JR. Emerging drugs to treat squamous cell carcinomas of the head and neck. *Expert Opin Emerg Drugs.* 2010;15:355-373.
10. Kelland L. The resurgence of platinum-based cancer chemotherapy. *Nat Rev Cancer.* 2007;7:573-584.
11. Eastman A. The formation, isolation and characterization of DNA adducts produced by anticancer platinum complexes. *Pharmacol Ther.* 1987;34:155-166.
12. Jordan P, Carmo-Fonseca M. Molecular mechanisms involved in cisplatin cytotoxicity. *Cell Mol Life Sci.* 2000;57:1229-1235.
13. Russell JS, Brady K, Burgan WE, et al. Gleevec-mediated inhibition of Rad51 expression and enhancement of tumor cell radiosensitivity. *Cancer Res.* 2003;63:7377-7383.
14. Martin LP, Hamilton TC, Schilder RJ. Platinum resistance: the role of DNA repair pathways. *Clin Cancer Res.* 2008;14:1291-1295.
15. Mathews LA, Cabarcas SM, Hurt EM, Zhang X, Jaffee EM, Farrar WL. Increased expression of DNA repair genes in invasive human pancreatic cancer cells. *Pancreas.* 2011;40:730-739.
16. Kauffmann A, Rosselli F, Lazar V, et al. High expression of DNA repair pathways is associated with metastasis in melanoma patients. *Oncogene.* 2008;27:565-573.
17. Pitroda SP, Pashtan IM, Logan HL, et al. DNA repair pathway gene expression score correlates with repair proficiency and tumor sensitivity to chemotherapy. *Sci Transl Med.* 2014;6:229ra242.
18. Dabholkar M, Vionnet J, Bostick-Bruton F, Yu JJ, Reed E. Messenger RNA levels of XPAC and ERCC1 in ovarian cancer tissue correlate with response to platinum-based chemotherapy. *J Clin Invest.* 1994;94:703-708.
19. Li Q, Seo JH, Stranger B, et al. Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell.* 2013;152:633-641.
20. Zhang W, Duan S, Kistner EO, et al. Evaluation of genetic variation contributing to differences in gene expression between populations. *Am J Hum Genet.* 2008;82:631-640.
21. Ramakodi MP, Kulathinal RJ, Chung Y, Serebriiskii I, Liu JC, Ragin CC. Ancestral-derived effects on the mutational landscape of laryngeal cancer. *Genomics.* 2016;107:76-82.
22. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 2014;30:1363-1369.
23. Broad Institute. http://gdac.broadinstitute.org/. Accessed April 5, 2015.
24. The University of Texas MD Anderson Cancer Center. Human DNA repair genes. http://sciencepark.mdanderson.org/labs/wood/dna_repair_genes.html. Accessed February 5, 2015.
25. Naccarati A, Pardini B, Stefano L, et al. Polymorphisms in miRNA-binding sites of nucleotide excision repair genes and colorectal cancer risk. *Carcinogenesis.* 2012;33:1346-1351.
26. Jacoby MA, De Jesus Pizarro RE, Shao J, et al. The DNA double-strand break response is abnormal in myeloblasts from patients with therapy-related acute myeloid leukemia. *Leukemia.* 2014;28:1242-1251.

27. Walter V, Yin X, Wilkerson MD, et al. Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes. *PLoS One.* 2013;8:e56823.
28. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156-2158.
29. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21:263-265.
30. Rosenbloom KR, Sloan CA, Malladi VS, et al. ENCODE data in the UCSC genome browser: year 5 update. *Nucleic Acids Res.* 2013; 41:D56-D63.
31. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res.* 2002;12:996-1006.
32. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika.* 1994;81:515-526.
33. Luce D. An epidemiological study on head and neck cancers in the French West Indies: rationale and study protocol. Paper presented at: 5th International African-Caribbean Cancer Consortium Conference; October 9-11, 2014; Schœlcher, Martinique.
34. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155:945-959.
35. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 2005;14:2611-2620.
36. Earl DA, Vonholdt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 2012;4:359-361.
37. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics.* 2007;23: 1801-1806.
38. Sweasy JB, Lang T, Starcevic D, et al. Expression of DNA polymerase β cancer-associated variants in mouse cells results in cellular transformation. *Proc Natl Acad Sci U S A.* 2005;102:14350-14355.
39. Bergoglio V, Pillaire MJ, Lacroix-Triki M, et al. Deregulated DNA polymerase beta induces chromosome instability and tumorigenesis. *Cancer Res.* 2002;62:3511-3514.
40. Srivastava DK, Husain I, Arteaga CL, Wilson SH. DNA polymerase beta expression differences in selected human tumors and cell lines. *Carcinogenesis.* 1999;20:1049-1054.
41. Figueroa JD, Malats N, Real FX, et al. Genetic variation in the base excision repair pathway and bladder cancer risk. *Hum Genet.* 2007; 121:233-242.
42. Vens C, Dahmen-Mooren E, Verwijs-Janssen M, et al. The role of DNA polymerase beta in determining sensitivity to ionizing radiation in human tumor cells. *Nucleic Acids Res.* 2002;30:2995-3004.
43. Iwatsuki M, Mimori K, Yokobori T, et al. A platinum agent resistance gene, POLB, is a prognostic indicator in colorectal cancer. *J Surg Oncol.* 2009;100:261-266.
44. Bryois J, Buil A, Evans DM, et al. Cis and trans effects of human genomic variants on gene expression. *PLoS Genet.* 2014;10: e1004461.