

JAK3 Variant, Immune Signatures, DNA Methylation, and Social Determinants Linked to Survival Racial Disparities in Head and Neck Cancer Patients



Rafael Guerrero-Preston^{1,2}, Fahcina Lawson¹, Sebastian Rodriguez-Torres^{1,3}, Maartje G. Noordhuis^{1,4}, Francesca Pirini⁵, Laura Manuel⁶, Blanca L. Valle¹, Tal Hadar⁷, Bianca Rivera⁸, Oluwasina Folawiyi¹, Adriana Baez⁸, Luigi Marchionni⁹, Wayne M. Koch¹, William H. Westra¹⁰, Young J. Kim¹¹, James R. Eshleman¹², and David Sidransky¹

Abstract

To inform novel personalized medicine approaches for race and socioeconomic disparities in head and neck cancer, we examined germline and somatic mutations, immune signatures, and epigenetic alterations linked to neighborhood determinants of health in Black and non-Latino White (NLW) patients with head and neck cancer. Cox proportional hazards revealed that Black patients with squamous cell carcinoma of head and neck (HNSCC) with *PAX5* ($P = 0.06$) and *PAX1* ($P = 0.017$) promoter methylation had worse survival than NLW patients, after controlling for education, zipcode, and tumor-node-metastasis stage ($n = 118$). We also found that promoter methylation of *PAX1* and *PAX5* ($n = 78$), was correlated with neighborhood characteristics at the zip-code level ($P < 0.05$).

Analyses also showed differences in the frequency of *TP53* mutations ($n = 32$) and tumor-infiltrating lymphocyte (TIL) counts ($n = 24$), and the presence of a specific C → A germline mutation in *JAK3*, chr19:17954215 (protein P132T), in Black patients with HNSCC ($n = 73$; $P < 0.05$), when compared with NLW ($n = 37$) patients. TIL counts are associated ($P = 0.035$) with long-term (>5 years), when compared with short-term survival (<2 years). We show bio-social determinants of health associated with survival in Black patients with HNSCC, which together with racial differences shown in germline mutations, somatic mutations, and TIL counts, suggests that contextual factors may significantly inform precision oncology services for diverse populations.

Introduction

Molecular heterogeneity among patient populations plays an important role in determining cancer prognosis and can enable novel precision medicine applications (1–5). Precision medicine can provide powerful tools to overcome cancer disparities by taking into account how

contextual and environmental factors modulate genomic and epigenomic signatures of heterogeneity. In cancer, solid tumors derive from microscopic, clonal cellular proliferations that come about in progressive stages by the acquisition of somatic alterations (1–3). Accumulation of somatic mutations is associated with cellular divisions that

¹Department of Otolaryngology and Head and Neck Surgery, Johns Hopkins University, School of Medicine, Baltimore, Maryland. ²Department of Obstetrics and Gynecology, University of Puerto Rico, School of Medicine, San Juan, Puerto Rico. ³Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. ⁴Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. ⁵Biosciences Laboratory, Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST) IRCCS, Meldola, Italy. ⁶Department of Epidemiology and Biostatistics, The University of Texas Health Science Center at San Antonio, San Antonio, Texas. ⁷Breast Health Unit, Department of General Surgery, Shaare Zedek Medical Center, Jerusalem, Israel. ⁸Department of Otolaryngology, University of Puerto Rico, School of Medicine, San Juan, Puerto Rico. ⁹Department of Oncology, Johns Hopkins University, School of Medicine, Baltimore, Maryland. ¹⁰Department of Pathology, The Icahn School of Med-

icine at Mount Sinai Hospital, New York, New York. ¹¹Department of Otolaryngology, Vanderbilt University School of Medicine, Nashville, Tennessee. ¹²Department of Pathology, Johns Hopkins University, School of Medicine, Baltimore, Maryland.

Note: Supplementary data for this article are available at Cancer Prevention Research Online (<http://cancerprevres.aacrjournals.org/>).

Corrected online June 21, 2019.

Corresponding Author: Rafael Guerrero-Preston, University of Puerto Rico School of Medicine, San Juan, 00936, Puerto Rico. Phone: 7876307885; Fax: 410-514-1411; E-mail: rafael.guerrero@upr.edu

doi: 10.1158/1940-6207.CAPR-17-0356

©2019 American Association for Cancer Research.

progressively acquire stem cell characteristics. These characteristics are gained by replicative errors that arise in response to hereditary burden, environmental stressors, and random error (6, 7). Responses to environmental factors correlate closely with the acquisition of DNA methylation alterations and other DNA-based signatures, which may contribute to cancer heterogeneity (8–10). When heterogeneity is examined, the impact of racial differences in somatic and germline mutations, promoter methylation, and immuno-infiltration in cancer is largely unexplored (11). Most of the genomic sequencing studies have been performed on genomes of European descent (12–14). Only a handful of studies have reported racial and ethnic differences in the genomic landscape of patients with triple-negative breast cancer and head and neck cancer (15–18).

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer worldwide and clinical outcomes vary throughout racial/ethnic and socioeconomic groups. Out of all of the racial/ethnic groups, during the last 40 years the group who has had the best survival, clinical outcome, and earlier stage at diagnosis of HNSCC has been non-Latino Whites (NLW; ref. 19). Regardless of race, ethnicity, and other sociodemographic characteristics, novel molecular characterization tools and strategies based on the detection of genetic and epigenetic markers offer new hope for improved risk assessment, early cancer detection, and therapeutic intervention, as well as tumor surveillance in cancer for all patients (20). Contrary to the goal of benefiting all patients, the impact of these strategies in reducing HNSCC survival disparities has been limited by an incomplete understanding of population dynamics in HNSCC heterogeneity, particularly in cancer's early development stages, and subsequent clinical outcomes among racial/ethnic groups other than White, for whom social determinants continue to play critical roles in late-stage diagnosis, health care access, and worse clinical outcomes (21).

Genomic and epigenomic alterations that have been linked with HNSCC outcomes, such as *TP53* mutations (22–27), loss of heterozygosity (28), microsatellite instability (29), and differential DNA methylation (30) are associated with environmental exposures, such as cigarette smoke and alcohol consumption (31). Recent evidence points toward another exposure with differential survival, namely human papillomavirus-positive (HPV⁺) and HPV⁻ status in HNSCC. HPV status is possibly related to different pathway involvement in the initiation and progression of HPV-related disease (32). Moreover, HPV⁺ HNSCC tumors mostly occur in the oropharynx, show superior responses to chemotherapy and radiotherapy (33, 34), and are less prevalent in Blacks than in NLW patients (35).

To ascertain the landscape of differences between Black and NLW patients with HNSCC we surveyed genomic,

epigenomic, and immuno-oncology signatures in three separate cohorts: cohort 1, racial correlates of genomic and epigenomic differences; cohort 2, social correlates of genomic and epigenomic differences; and cohort 3, racial correlates of immuno-oncology differences.

To document frequency differences in HNSCC somatic mutations and epigenetic alterations between Black and NLW, we analyzed exome-sequencing data on approximately 18,000 protein-encoding genes from 32 HNSCC cases, together with genome-wide DNA methylation array and genome-wide mRNA array data on these 32 patients and frequency-matched 32 uvulopalatopharyngealplasty (UPPP) controls. The DNA methylation results were validated with genome-wide DNA methylation array and genome-wide mRNA array data from 279 patients with HNSCC from the Cancer Genome Atlas (TCGA) project and quantitative methylation specific PCR (qMSP) data in 76 patients with HNSCC from John Hopkins University (Baltimore, MD).

We then examined genome-wide DNA methylation and genome-wide mRNA alterations in a second discovery cohort of 24 patients with HNSCC and frequency-matched 27 UPPP controls, 70% of which were Black. To verify the robustness of our findings we used fluorogenic, qMSP in a second validation cohort of 118 patients with HNSCC, 55% of which were Black. The qMSP results were used to perform social determinants of health analyses at the zip-code level, using publicly available census data. We also performed targeted exome sequencing in 16 Black patients with HNSCC using a Hotspot Panel, which profiles 50 frequently mutated in-human cancer genes. We used Sanger sequencing to confirm the mutation frequency of specific gene alterations. We also used droplet digital PCR (ddPCR) to examine racial differences in the immune signatures of patients with HNSCC in discovery cohort 3 ($n = 24$; Supplementary Fig. S1).

Materials and Methods

Clinical samples

The head and neck cancer specimens collected for this study conform to the criteria we have previously used and to those established by the NIH TCGA Biospecimen Selection Process. All samples were obtained following patient consent. The study was approved by Johns Hopkins Hospitals Institutional Review Board (Baltimore, MD), and was performed in accordance with Health Insurance Portability and Accountability Act guidelines to safeguard protected health information and the U.S. Common Rule. Unique patient identifiers for each sample were created at study entry, allowing blind annotation and tracking of demographic, clinical, and molecular data. Tumor samples were obtained at the time of initial tumor resection. Clinical data elements were carefully collected for each patient using criteria modeled after the currently updated

TCGA clinical data requirements for other tumor types. Each tumor sample used in the study was obtained in accordance with the above protocol, and was subjected to pathologic review by a qualified pathologist in order to allow (i) confirmation of the diagnosis of squamous cell carcinoma; (ii) confirmation that the representative tumor section contains greater than 80% tumor cell nuclei; and (iii) confirmation that the sample is less than 40% necrotic. Forensic microsatellite repeat analysis and human leukocyte antigen genotyping was conducted on all tumor samples, as well as the DNA obtained from normal tissue from the same patient to confirm matching tumor and normal sample identities. In addition to the above quality control measures, each DNA sample was tested for efficiency of amplification using a stringent protocol developed in the laboratory. Any sample not meeting these criteria was not included in the study. Summaries of all quality control analyses of samples used in this project are available for databases containing the resulting sequencing data. The tumor DNA and RNA samples we used for molecular studies were selected to be of high purity and quality to ensure sensitive detection of genomic alterations; these are critical parameters for the success of such high-throughput DNA- and RNA-sequencing projects. Purity is defined as >80% tumor cells, and quality is defined by isolation of high-molecular weight DNA and RNA, assayed by gel electrophoresis, PicoGreen quantification, Bioanalyzer profiles, and by a quantitative functional assay for PCR amplification.

Our Johns Hopkins Head and Neck tumor bank database (HAND) stores over 70,000 specimens from over 14,000 patients recruited during the past 20 years. We are fortunate to have a large collection available of such high quality frozen sections of surgical samples of HNSCC, linked with clinical information. These samples are accompanied by paired normal tissue (lymphocyte DNA), which is critical for exome-sequencing studies. This sample collection has been maintained and curated by W.M. Koch, Department of Otolaryngology-Head and Neck Surgery, Johns Hopkins University School of Medicine (Baltimore, MD), with support from the Johns Hopkins HNSCC SPORE grant.

Patient characteristics

Patient cohorts for the different phases of this project were drawn from 248 Black and 248 NLW patients with HNSCC treated in Johns Hopkins, for whom we had access to their medical record and had obtained consent to store tumor tissue. Their clinical and sociodemographic characteristics are listed in Table 1. We frequency matched the samples on sex (70.2% of Black and 80.6% of NLW are males); age (the median age for Black is 59.6 years and for NLW is 58.4 years); and site (25.8% of Black and 25% of NLW have oral cancer; 29% of Black and 32.3% of NLW have oropharyngeal cancer; 7.3% of Black and 8.1% of NLW have hypopharyngeal cancer; and 34.7% of Black and

32.3% of NLW patients have larynx cancer). A subset of oropharyngeal tumors was HPV⁺: 21% of Black and 49% of NLW. Furthermore, there were no differences in stage (10.5% of Black and 12.9% of NLW have stage I; 8.1% of Black and 11.3% of NLW have stage II; 13.7% of Black and 18.5% of NLW are stage III; and 57.3% of Black and 43.5% of NLW are stage IV) or ethanol use (41.1% of Black and 31.5% of NLW were heavy drinkers; 20.2% of Black and 25.8% of NLW were occasional drinkers; and 21.8% of Black and 14.5% of NLW did not drink). There were more active smokers among Black (71%) than NLW (50.8%). Consequently, there were more former smokers (NLW, 17.7% and Black, 9.7%) and nonsmokers, (NLW, 16.1% and Black, 4.8%), as listed in ($P = 0.02$). These are most likely a reflection of the racial differences and composition in our HNSCC tumor bank, by tumor subsite. We adjusted for smoking status as a possible confounder in our comparisons.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) of functional themes was performed to capture biological processes overrepresented in the various conditions under investigation using analysis of functional annotation. A χ^2 test was applied to test whether each functional gene set (FGS) was overrepresented in any of the gene list associated with any of the investigated contrasts/conditions (e.g.,

Table 1. Clinical and sociodemographic characteristics

	Total	Black	Non-Latino Whites
Total <i>n</i>	496	248	248
Sex, <i>n</i> (%)			
Female	122	74 (29.8)	48 (19.4)
Male	374	174 (70.2)	200 (80.6)
Age, <i>n</i> (%)			
Median (years)		59.6	58.4
<40 years	24	12 (4.8)	12 (4.8)
40–55 years	152	74 (29.8)	78 (31.5)
>55 years	318	160 (64.5)	158 (63.7)
Site, <i>n</i> (%)			
Oral cavity (%)	124 (25)	64 (25.8)	60 (25)
Oropharynx (%)	152 (30.6)	72 (29)	80 (32.3)
Nasopharynx (%)	6 (1.2)	2 (0.8)	4 (1.6)
Hypopharynx (%)	38 (7.7)	18 (7.3)	20 (8.1)
Larynx (%)	166 (33.5)	86 (34.7)	80 (32.3)
Unknown (%)	8 (1.6)	6 (2.4)	2 (0.8)
Stage, <i>n</i> (%)			
1	58 (11.7)	26 (10.5)	32 (12.9)
2	48 (9.7)	20 (8.1)	28 (11.3)
3	80 (16.1)	34 (13.7)	46 (18.5)
4	250 (50.4)	142 (57.3)	108 (43.5)
Unknown	60 (12.1)	26 (10.5)	34 (13.7)
Smoking, <i>n</i> (%)			
Active	302 (60.9)	176 (71)	126 (50.8)
Past	68 (13.7)	24 (9.7)	44 (17.7)
Nonsmokers	52 (10.5)	12 (4.8)	40 (16.1)
Unknown	74 (14.9)	36 (14.5)	38 (15.3)
EtOH, <i>n</i> (%)			
Heavy	180 (36.3)	102 (41.1)	78 (31.5)
Occasional	114 (23)	50 (20.2)	64 (25.8)
None	90 (18.1)	54 (21.8)	36 (14.5)
Unknown	102 (2.6)	42 (16.9)	70 (28.2)

gene associated with hypermethylated promoters in HNSCC). In this study, individual, nonredundant genes, as annotated in the NCBI Entrez gene database (R/Bioconductor package org.Hs.eg.db version 2.4.6) were used as the total gene space, and contingency tables were used to identify gene sets overrepresented in the investigated conditions.

Correction for multiple hypothesis testing was obtained separately for each FGS collection, by applying the Benjamini and Hochberg method (36–38) as implemented in the multtest R/Bioconductor package. Overall, this approach is analogous to GSEA (39), and has already been successfully applied in other studies (40, 41). The heatmaps color bar represents the negative \log_{10} FDR. For each gene set collection, the sets for which at least one condition showed $FDR < 0.01$ were reported. The top 150 conditions were reported when too many gene sets were retrieved.

Social determinants of health

We used qMSP to compare the differences in promoter DNA methylation of four genes, *NID2*, *EDNRB*, *PAX1*, and *PAX5*, identified as biomarkers of HNSCC risk in prior publications. The unbiased discovery of promoter methylation biomarkers of HNSCC using genome-wide DNA methylation arrays and next-generation sequencing, as well the design and optimization of qMSP primers and probes to validate these findings in separate cohorts, is described in previous publications from our laboratory (42, 43).

Neighborhood-level variables

Median household income, home ownership, home vacancy, and insurance coverage rates using corresponding (patient) 2010 zip-code tabulation areas (ZCTA) were provided by the U.S. Census Bureau's American Community Survey. ZCTAs are generalized areal representations of United States Postal Service ZIP Code service areas (44, 45). On the basis of this zip-code data, the census bureau aggregates ZCTAs from addresses contained within each block. This aggregation of data allows a point-based dataset (addresses) to be converted into an area feature dataset ZCTAs (46).

Statistical analysis

Three methods were implemented to assess the relationship between molecular alterations, clinical factors, and neighborhood zip-code-level interactions with HNSCC survival. First, we estimated HNSCC survival linked covariables using the Kaplan–Meier method. Second, we used Cox proportional hazards models to conduct univariate and bivariate linear and logistic regression models to determine whether there was a correlation between study variables and HNSCC survival was used to analyze the effects of covariables on survival. Third, a geographically weighted regression analysis was applied to evaluate

whether this correlation occurred after weighing for local neighbors.

Kaplan–Meier method

The log-rank test was used to test the significance of the association between survival and the following categorical variables: promoter methylation of *EDNRB*, *NID2*, *PAX5*, *PAX1*, and tumor stage; and the following determinants of health at the zip-code level: high school diploma, median family income, health insurance, home ownership rate, and home vacancy rate. Promoter methylation of *EDNRB*, *NID2*, *PAX5*, and *PAX1* categorical variables were recorded as follows: 1 = methylated; 2 = unmethylated. Data for high school diploma, median family income, health insurance, home ownership rate, and home vacancy rates categorical variables are as follows: 1 = above the median; 2 = below the median.

Cox proportional hazard regression

Cox proportional hazard regression was used to test the significance of the association between survival and the following continuous variables: promoter methylation of *EDNRB*, *NID2*, *PAX5*, and *PAX1* and the following determinants of health at the zip-code level: high school diploma, median family income, health insurance, home ownership rate, and home vacancy rate. We then used Cox proportional hazard models for univariate and multivariate analyses to estimate HRs with 95% confidence intervals.

Multiple regression models were developed to describe the relationship between HNSCC survival and differential promoter methylation, clinical factors, and neighborhood-level interactions within the entire study population and among racial strata. All statistical tests were two-sided, and $P < 0.05$ was considered as statistically significant. Analyses were performed using STATA version 13.

Geographically weighted regression analyses

To assess spatial relationships of racial/ethnic disparities between *EDNRB*, *NID2*, *PAX5*, and *PAX1* promoter methylation, primary site, median family income, health insurance rate, home ownership rate, and home vacancy rate, we conducted a geographically weighted regression (GWR). Because of the substantial geographic variation in the United States, GWR is performed to identify spatial heterogeneities in regression models of geo-referenced data. Local regression coefficients and associated statistics (i.e., proportion of variance explained, correlation coefficients) can then be mapped to visualize how the explanatory power of covariates changes spatially.

Spatial autocorrelation was tested for and found within the residuals from linear regressions using both a local and a global Moran I. The global Moran I is an analysis on whether or not variables seem to be spatially correlated (i.e., high values are next to high values) or not spatially correlated (i.e., values do not appear to affect one another).

The closer the value is to $-1/n-1$ ($-1/13$) the more random the data. The closer to 1 the more correlated, the closer to -1 the more negatively correlated. Closeness was determined by Euclidean distance from the center of each zipcode using threshold distance of 0.036529. To evaluate the strength of the correlation of racial/ethnic disparities between *EDNRB*, *NID2*, *PAX5*, and *PAX1* promoter methylation, primary site, median family income, health insurance rate, home ownership rate, and home vacancy rate accounting for the local geographic weight around a focal point, GWR models were constructed to weigh the neighborhoods using ArcGIS (Version 10.1).

However, logistic regression models with geographic weighting could not converge to estimate model coefficients using the maximum likelihood approach.

Thus, only geographically weighted linear regression was conducted in this study. The weighting scheme utilized in this study was the Gaussian kernel function (30). The bandwidth, or number of neighbors used for each local estimation, is perhaps the most important parameter for GWR, and the optimal bandwidth was determined by minimizing Akaike information criterion (AICc). The Gaussian kernel bandwidth varies across space with the decay function of weighting neighborhoods. The kernel type was set up as adaptive to account for the density of spatial features and the optimal bandwidth was determined by minimizing AICc (31). The parameter estimates were mapped in ArcInfo (Version 9.3).

Immuno-oncology determinants of survival

The QuanTILfy assay and ddPCR assay quantifies T lymphocytes using a multiplex ddPCR system, which amplifies rearranged TCR β loci from genomic DNA using 45 forward primers, each specific to one or multiple functional TCR V β segments, and 13 reverse primers, each specific to a TCR J β segment. The multiplex reaction also includes in each well one of a series of 35 minor groove binder 6-carboxyfluorescein (FAM) TaqMan probes complementary to 52 different V β gene segments, and a VIC probe complementary to ribonuclease (RNase) P protein subunit p30 (RPP30) per well, which serves as a reference gene to permit normalized quantification. To identify clonal expansion of TCR genes, ddPCR assays were designed for eight V β gene segment subgroups, each of which contains the forward primers and TaqMan probes specific to a nonoverlapping subset of V β gene segments. Each V β gene subgroup was combined with all 13 J β gene segment primers, as well as the RPP30 primers and probe, effectively creating a multiplex ddPCR assay for TCR β rearrangement detection. Each of the 52 possible TCR β gene segments was measured once in exactly one of the wells. Therefore, the sum of counts from all wells gives a precise digital quantification of the total number of rearranged TCR β s in the sample.

Single-nucleotide variant in JAK3

Discovery with next-generation sequencing. We used the Ion AmpliSeqCancer Hotspot Panel v2 to profile 50 frequently mutated in-human cancer genes in a discovery cohort of HNSCC Black patients. Libraries for the discovery cohort were generated using the Ion AmpliSeq Library Kit 2.0 according to the manufacturer's instructions (Life Technologies). Included in this panel were primers for 207 amplicons covering 2,800 Catalogue of Somatic Mutations in Cancer (COSMIC, <http://cancer.sanger.ac.uk/cancergenome/projects/cosmic>) of 50 genes with known cancer associations (*ABL1*, *AKT1*, *ALK*, *APC*, *ATM*, *BRAF*, *CDH1*, *CDKN2A*, *CSF1R*, *CTNNB1*, *EGFR*, *ERBB2*, *ERBB4*, *EZH2*, *FBXW7*, *FGFR1*, *FGFR2*, *FGFR3*, *FLT3*, *GNA11*, *GNAS*, *GNAQ*, *HNF1A*, *HRAS*, *IDH1*, *JAK2*, *JAK3*, *IDH2*, *KDR*, *KIT*, *KRAS*, *MET*, *MLH1*, *MPL*, *NOTCH1*, *NPM1*, *NRAS*, *PDGFRA*, *PIK3CA*, *PTEN*, *PTPN11*, *RB1*, *RET*, *SMAD4*, *SMARCB1*, *SMO*, *SRC*, *STK11*, *TP53*, and *VHL*). DNA (10ng) from the tumor samples was used as the template to prepare the library. Amplified libraries were quantified using the Qubit 2.0 Fluorometer and the High Sensitivity Qubit Assay Kit (Life Technologies). Amplified libraries were assessed for quality (size and concentration) using the Agilent 2100 Bioanalyzer Instrument (Agilent Technologies) following the Bioanalyzer standard protocol. The AmpliSeq libraries were clonally amplified on to ion sphere particles (ISP) using emulsion PCR following standard ion torrent protocols. ISP preparation was performed using the automated Ion Torrent OneTouch2 system following the manufacturer's protocol (MAN0007220 Revision 4.0). The Qubit Fluorometer was used to assess ISP quality after ISP preparation but before ISP enrichment. Up to eight specimens were barcoded with Ion Xpress Barcode Adapters (Life Technologies), pooled, and run on a single Ion 318 chip. This includes multiple-patient samples and one control, which we rotate among water, normal, and a mix of positive control cell lines.

Validation with Sanger sequencing. The results were validated with Sanger sequencing in a separate cohort of 76 patients with HNSCC. The region surrounding the P132T amino acid site of *JAK3* were sequenced by Sanger sequencing. Amplification of *JAK3* region was performed by PCR performed in 12.5 μ L reactions containing 1 \times PCR buffer (67 mmol/L Tris- HCl, pH 8.8, 6.7 mmol/L MgCl₂, 16.6 mmol/L NH₄SO₄, 10 mmol/L 2-mercaptoethanol), 10 mmol/L dNTPs (Invitrogen), 10 μ mol/L forward (5'-GTAAAACGACGGCCAGTCTGTGAGGCCCTCCGCAG-A-3') and 10 μ mol/L reverse (5'-CAGGAAACAGCTAT-GACCGATTGCATGCCAGTCCCTCA-3') primers, 1.25 U Platinum Taq (Invitrogen), and 10 ng DNA. Reactions were performed in 364-well plates using ABI 7900 Thermocycler (Applied Biosystems) as follows: 1 cycle of 96°C for 30 seconds; 40 cycles of 95°C for 30 seconds, 66°C for 60 seconds, 72°C for 60 seconds; and 1 cycle of 72°C

for 5 minutes. Templates have been purified by QIAquick PCR Purification Kit (Qiagen). Sanger sequencing was performed by GENEWIZ using ABI 3730xl DNA analyzer for capillary electrophoresis and fluorescent dye terminator detection.

Quantitative real-time reverse transcription PCR. HNSCC RNA from a subsample of Black patients was assessed for *JAK3* and *GAPDH* expression levels using quantitative real-time reverse transcription (RT)-PCR (TaqMan). Reverse transcription was performed with random hexamer primers and Superscript II Reverse Transcriptase (Invitrogen Corp.) according to the manufacturer's instructions. Quantitative RT-PCR was then performed on the Applied Biosystems 7900 Sequence Detection Instrument (Applied Biosystems) using TaqMan Expression Assays (Life Technologies).

Results

HNSCC disparities in somatic mutations and DNA methylation

We observed disparities in the frequency of mutated and methylated events in the *PAX*, *NOTCH1*, and *TP53* pathways (Supplementary Table S1). We observed higher frequencies of *TP53* and *NOTCH1* mutations (*NOTCH1mut*) in Black patients with HNSCC and no differences in *PAX1* or *PAX5* methylation across all tumor sites, and interestingly, these patterns differed when we stratified by anatomic subsite. Outside the oropharynx, Black patients have a higher frequency of *PAX5* methylation (*PAX5met*) and *TP53* mutations (*p53mut*) than NLW, and no *NOTCH1mut*. Inside the oropharynx, Black patients have a lower frequency of *PAX5met* when compared with NLW. Inside the oropharynx, NLW also had a higher frequency of combined *p53mut* or *PAX5met*, while Blacks had a higher frequency of combined *NOTCH1mut* or *PAX1* methylation (*PAX1met*). Complex genetic and epigenetic interactions between *PAX*, *NOTCH*, and *p53* pathways may be differentially driving HNSCC initiation and progression events in Black and NLW patients with HNSCC.

Patients showed obvious differences in the genetic landscapes of HPV-associated and HPV⁻ HNSCC. In the HPV-associated tumors compared with those tumors not related to HPV, far fewer genes were mutated per tumor (4.8 ± 3 vs. 20.6 ± 16.7 ; $P < 0.05$, Welch two sample *t* test). These data are consistent with previous results on HNSCC, as well as on HPV-associated cervical cancers (10–12). More cancer-related mutations were identified in tumors from patients with a history of tobacco use compared with those from patients who did not use tobacco (21.6 ± 17.8 vs. 9.5 ± 6.5 ; $P < 0.05$, Welch two sample *t* test).

We also observed gene-specific gain and loss of methylation events correlated with etiologic factors. *PAX1* promoter is not methylated in HPV⁺ tumors, whereas *PAX1*

gains methylation in HPV⁻ tumors. *PAX1* promoter methylation is also observed in most patients with a history of tobacco exposure (71%), while only 33% of patients without tobacco exposure history exhibited methylation in the promoter region of *PAX1*. Most HPV⁻ tumors (83%) show promoter methylation of *PAX5* compared with only 25% of HPV⁺ tumors. On the contrary, tumors from patients with a history of tobacco exposure (57%) had similar frequency of promoter methylation in *PAX5* when compared with tumors from patients with no history of tobacco (67%). We also observed concurrent genomic and epigenomic associations with viral and tobacco exposures; patients with *TP53* mutations also had *PAX1* promoter methylation, a history of tobacco exposure, and were HPV⁻.

Somatic mutation profile in Black compared with NLW patients with HNSCC

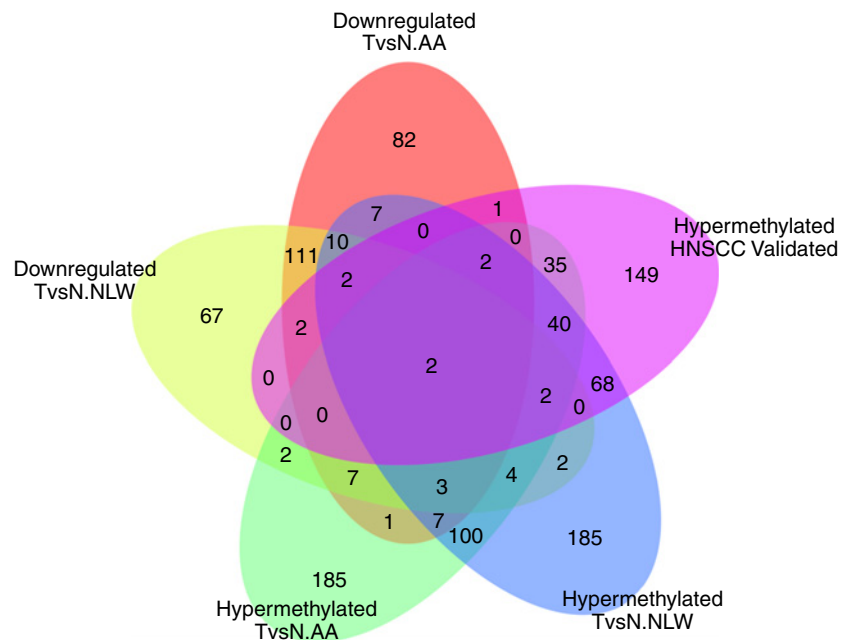
We examined the mutational profile in a separate cohort of Black patients with HNSCC to verify the findings observed in the discovery cohort. Black patients with HNSCC have different frequencies of somatic mutations in the 50 genes most commonly mutated in human cancers when compared with NLW patients with HNSCC, as well as across anatomic sites. Black patients with larynx cancer had a higher frequency of *TP53*, *PIK3CA*, *JAK3*, *KIT*, *APC*, and *MET* somatic mutations. *NOTCH1* had a higher frequency of somatic mutations in Black patients with oropharyngeal cancer and *CDKN2A* had a higher frequency of somatic mutations in Black patients with oral cancer (Supplementary Table S2).

Germline mutation in *JAK3* in Black patients with HNSCC

In addition to somatic mutations, we found a germline C ≥ A mutation in *JAK3*, chr19:17954215 (protein P132T) together with a paired polymorphism (at position chr19:17954149) in 25% of the samples in discovery cohort, of which all happened to be males with larynx cancer. In an additional cohort of 76 patients with HNSCC (39 Black and 37 NLW), we validated this finding with Sanger sequencing, where we found this same *JAK3* mutation in 8% of the Black patients while none of the NLW patients had it. Black males with larynx cancer (67%) and Black females with cancer of the oropharynx (33%) had this *JAK3* germline mutation. This is a known SNP (rs3212723) present in COSMIC (COSM34216) and the dbSNP databases. This SNP is present only in people from Africa or African descent in the 1,000 genome project: Masai (MKK NA 21737), American from the southwest (ASW NA 19834, NA19701) and Yoruba (YRI NA 18502, NA 18504, and NA 19238). Further sequencing of the cohorts showed that normal Black patients had the mutation at a higher proportion than Black patients with HNSCC.

Figure 1.

Venn diagrams that show the intersection of genes in 49 African-American (AA) and NLW patients with HNSCC compared with 51 UPPP controls.



Gene expression signatures in Black patients with HNSCC

We then examined genome-wide expression and DNA methylation array data from 56 patients with HNSCC and 59 normal epithelium controls. Analysis in Black patients with HNSCC had GSEA show that promoter methylation correlates with downregulation in several genes (Fig. 1), many of which are in immune-related pathways: antigen processing; chemokine signaling; cytokine–cytokine receptor interactions; and natural killer cell cytotoxicity. The top altered pathways revealed by gene ontology clustering involved the loss of *IRF4*, *IRF8*, *PAX5*, *PAX1*, *CXCL12*, *EBF1*, and *PARP15*, as can be seen in Supplementary Table S3.

Social determinants of HNSCC survival disparities: neighborhood effects that may impact genetic and epigenetic changes

The overall majority of the patients with HNSCC in our cohort reside in the state of Maryland (MD; 76.1%). Most of the remaining patients reside in the following states: The District of Columbia 1.7%, Delaware 8.5%, Pennsylvania 4.3%, Virginia 4.2%, and West Virginia 4.2%. The proportion of patients residing in MD differed by race ($P = 0.002$): Black (61.8%) and NLW (38.2%). Most of the patients residing in the state of MD (51.7%) resided outside of Baltimore City, and most were NLW patients (73.5%). The majority of Black patients with MD residency lived in Baltimore City (62.3%).

There were notable racial differences in zip-code–level neighborhood factors across the six states. Median family income differed significantly between Black and NLW patients ($P < 0.0001$). NLW patients had a 38.2% higher

zip-code–level median family income (\$65,005) than Black patients (\$47,040). The median zip-code–level health insurance rate also differed significantly between Black (87.8%) and NLW patients (89.8%; $P = 0.001$). The median zip-code–level homeownership rate was 5% lower for Black (86.9%) compared with NLW patients (91.9%; $P = 0.02$). NLW patients (8.1%) had a median home vacancy rate, which was 61.7% higher among Black (13.1%) than NLW ($P = 0.032$).

Choropleth thematic maps were created for every Baltimore City zipcode, in which our patients with HNSCC resided, using the layer property in ArcGIS. Choropleth maps aggregate data in the form of counts across a defined geographic space (zipcodes in this instance), indicating differences by shaded or colored areas (47). The resulting choropleth thematic maps geographically depicted the spatial distribution of *PAX5* methylation (in quartiles) and home vacancy rates at the zip-code level, for all the patients with HNSCC in our cohort who reside in Baltimore. Darker colors represent a higher quartile *PAX5* promoter methylation in the *PAX5* methylation map (Fig. 2A), or a higher percentage of home vacancies per zipcode in the home vacancy map (Fig. 2B). Choropleth thematic maps were labeled with zip-code–level information for two zipcodes that represent opposite ends of the socioeconomic spectrum in Baltimore: Roland Park (21210) and Madison/East End (21205). The residents of Roland Park have a median annual income of \$90,492 (the highest in Baltimore City), a 3.4% unemployment rate, and a mean life expectancy of 83.1 years life expectancy rate. In contrast, the residents of Madison/East have a median annual income of \$30,389 (in the lowest quartile), a 14.4% unemployment rate, and a mean life expectancy of

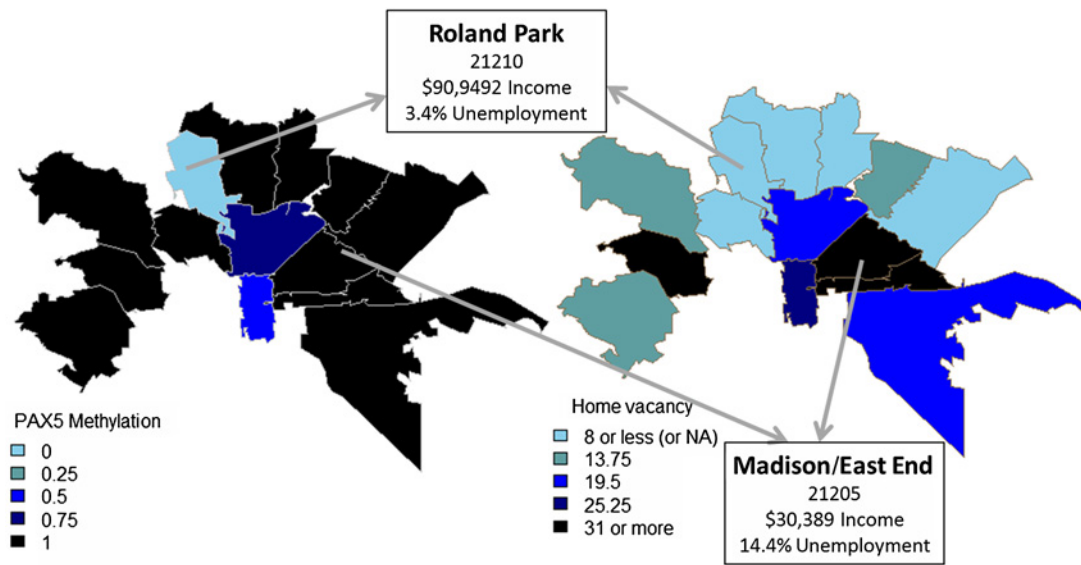


Figure 2.

Choropleth maps geographically depicted the spatial distribution of selected variables at the zip-code level for all the patients with HNSCC in our cohort who reside in Baltimore: *PAX5* promoter methylation map (in quartiles; **A**); and home vacancy rates map (zip-code level; **B**). Darker colors represent a higher quartile in the *PAX5* promoter methylation map, or a higher percentage of home vacancies per zipcode in the home vacancy map. Choropleth thematic maps were labeled with zip-code-level information for two zipcodes that represent opposite ends of the socioeconomic spectrum in Baltimore: Roland Park (21210) and Madison/East End (21205). The residents of Roland Park have a median annual income of \$90,492 (the highest in Baltimore City), while the residents of Madison/East End have a median annual income of \$30,389.

64.8 years life expectancy rate. Patients with HNSCC in our cohort who lived in the zipcode with the highest median annual income had the lowest levels of *PAX5* methylation and home vacancy rates, after adjusting for race. Conversely, patients with HNSCC in our cohort who lived in the zipcode with some of the lowest median annual income, had the highest levels of *PAX5* methylation and home vacancy rates, after adjusting for race. Home vacancy is defined as the proportion of homes that are not occupied in a specific area and is a known social determinant of health of the build environment. This environmental measure is correlated with undesirable health outcomes. This information suggests that there may be exposures to multiple environmental hazards and social stressors at the zip-code level, which may contribute to differences in HNSC risk, indolence and aggressiveness. Home vacancy rates at the zipcode level and *PAX5* promoter methylation may be part of the complex mechanism by which the exposome becomes biology, which is reflected in the different patterns of somatic mutation and promoter methylation changes between Black and NLW patients with HNSCC in our cohort.

We then examined whether survival disparities in HNSCC are linked to neighborhood determinants of health. Using data from the U.S. Census Bureau's American Community Survey, we examined the relationships of

social determinant of health variables at the neighborhood-level and promoter methylation of *PAX5*, *PAX1*, *EDNRB*, and *NID2* in the patients with HNSCC residing in Baltimore. We examined five domains at the zip-code level that have previously been identified as social determinants of health: educational attainment (percentage of High school diplomas), median household income, home ownership, vacancy status, and insurance coverage. Pooled logistic regression analysis showed an association between percentage of residents with high-school diploma at the zip-code level (HD) and *PAX5* ($P = 0.09$) and *PAX1* ($P = 0.05$) promoter methylation (Fig. 3A). Survival analyses revealed that patients with *PAX5* ($P = 0.001$) or *NID2* ($P = 0.05$) promoter methylation had a worse outcome than those without it. Cox proportional hazards regression multivariable analysis, revealed that Black patients with *PAX5* ($P = 0.06$) and *PAX1* ($P = 0.017$) methylation had worse survival than NLW, after controlling for HD, zip-code, and tumor-node-metastasis (TNM) stage. (Fig. 3B).

In subsequent multivariate Cox regression analyses of TCGA HNSCC data ($n = 279$), we confirmed that patients with *PAX5* methylation had worse survival than those without it (HR = 1.63), after adjusting for smoking, surgical margin, and p53 mutations. ($P = 0.03$). This is a very similar HR that for current or previous smoking (HR = 1.88; $P = 0.03$) and positive surgical margins

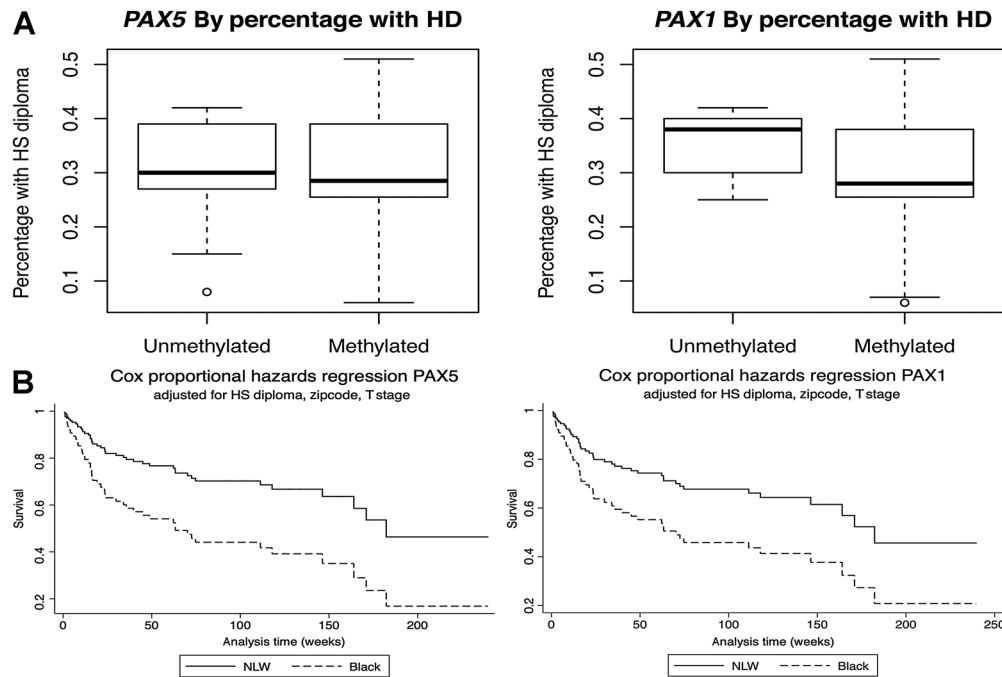


Figure 3.

A, Pooled logistic regression analysis showed a marginally significant association between the percentage of residents with high-school (HS) diploma at the zip-code level (HD) with *PAX5* ($P = 0.09$) and *PAX1* ($P = 0.05$) promoter methylation. **B**, Multivariable Cox proportional hazards regression analysis shows that AA patients with *PAX5* ($P = 0.06$) and *PAX1* ($P = 0.017$) promoter methylation had worse survival than NLW, after controlling for HD, zipcode, and TNM stage (T stage).

($HR = 1.76$; $P = 0.02$). We also found that combined somatic *TP53* mutations and *PAX5* promoter methylation are linked to worse outcomes when compared with patients with either alteration alone, after adjusting for smoking, and surgical margin ($HR = 2.16$; $P < 0.001$; Fig. 4). This is a slightly higher ratio than for *p53* mutation alone, after adjusting for smoking, surgical margin, and *PAX5* ($HR = 2.06$; $P = 0.004$; Supplementary Table S4).

Immuno-oncology determinants of survival in HNSCC

When we used the QuanTILfy digital PCR assay (48) to examine TILs in HNSCC tumor tissue, the number of TILs was found to be increased in long-term survivors (>5 years) compared with short-term survivors (<5 years; $P < 0.05$; Fig. 5A). We also compared TIL counts by site and observed more in larynx cancers (Fig. 5B). Cancers had larger TIL counts and clonality in Black patients, compared with NLWs (Fig. 5C).

Bio-social markers for personalized medicine workflows

A precision medicine index that integrates contextual, bio-psychosocial and molecular data can be a useful tool for clinicians, patients, public health practitioners, behavioral therapists, and policy makers. DNA methylation, the best understood epigenomic mark, is a molecular marker that can summarize multiple external and internal factors

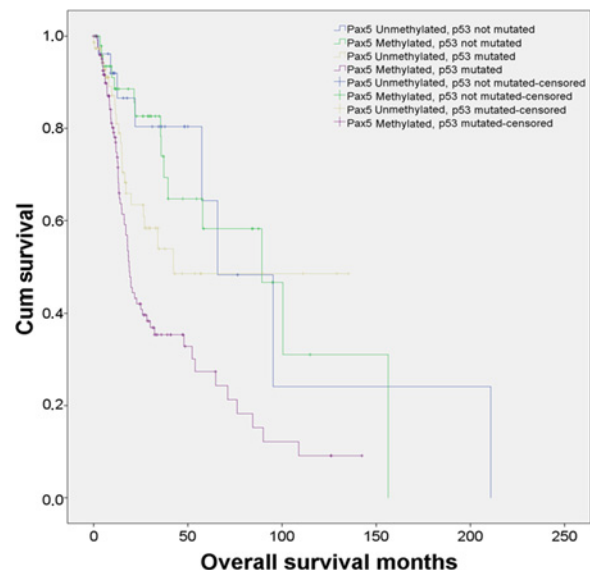


Figure 4.

Results from a Kaplan–Meier analysis of 279 patients with HNSCC from TCGA. Patients with promoter methylation of *PAX5* have worse outcomes than patients without *PAX5* methylation ($P = 0.026$). Patients with combined somatic *TP53* mutations and *PAX5* promoter methylation have poorer outcomes when compared with patients with *TP53* mutations, who do not have *PAX5* methylation ($P < 0.012$).

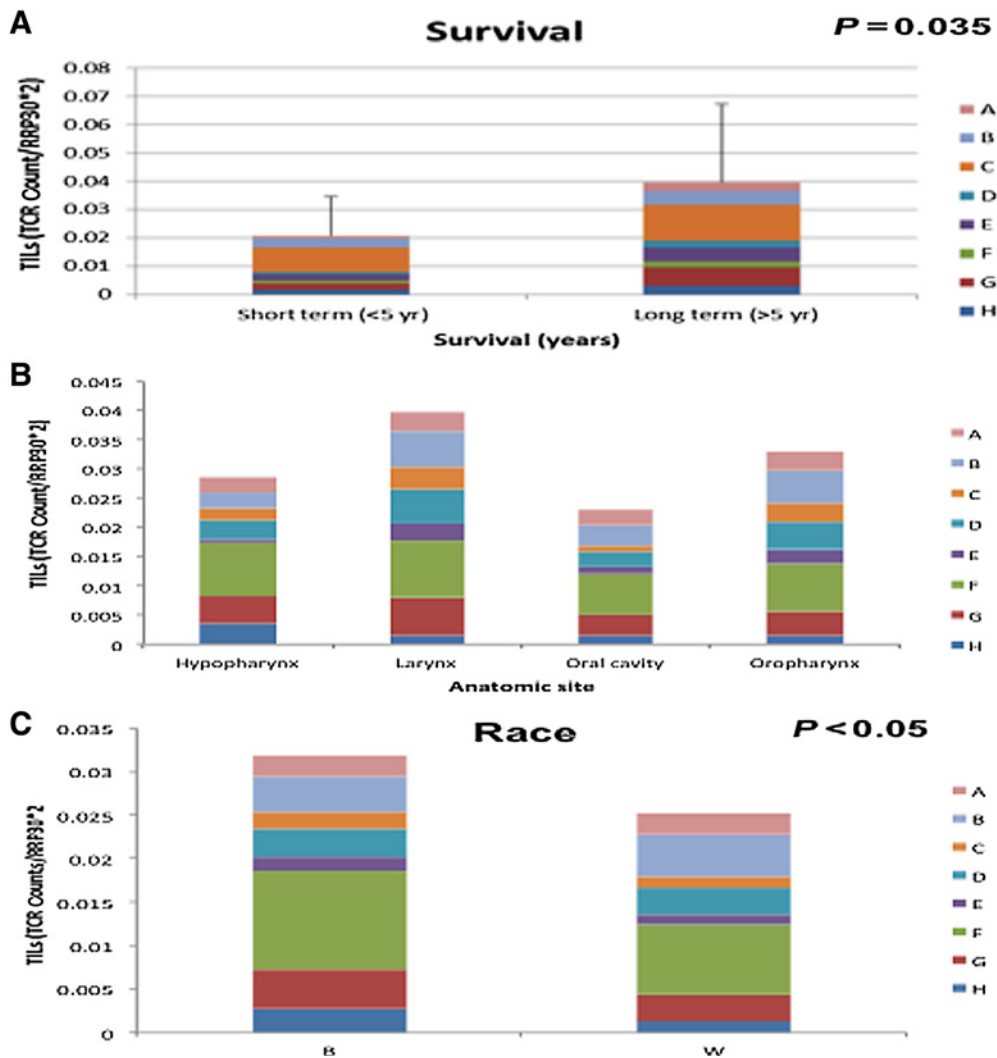


Figure 5. TILs as immuno-oncology determinants in HNSCC, quantified with the QuanTILfy assay, a ddPCR assay. Larger numbers of TILs are associated with survival (A), larynx cancer (B), and Black race (C).

that modulate health and disease. It is a tangible measurement of how the environment, social forces, emotions, and psychologic processes modulate gene expression, and eventually impinge upon normal and disrupted, intracellular and extracellular processes, via molecular pathways that are slowly being mapped and validated. In Fig. 6A we depict how contextual, demographic, lifestyle, and molecular markers can be useful markers for a variety of HNSCC endpoints: (i) molecular markers of clean surgical margins; (ii) bio-social markers of outcome disparities; and (iii) precision medicine markers of therapeutic response. These markers can also help us to create a better biological understanding of the complex relationships and interactions between different biological systems and functions. For instance, Fig. 6B shows that *PAX5* methylation levels inversely correlate ($r = -0.83$) with TIL counts in patients

with HNSCC. These data are consistent with both, the inverse association between *PAX5* methylation and survival and the direct association between TILs and survival we have observed. We have also observed that the patients with HNSCC in our cohort who lived in the zipcode with the lowest median annual income had the highest levels of *PAX5* methylation and home vacancy rates, after adjusting for race. Together these data suggests that *PAX5* methylation may be a bio-social marker, linked to both adverse social determinants of health and poor survival outcomes in head and neck cancer. Figure 6C shows that *JAK3* expression in Black patients with HNSCC differs by anatomic location. Figure 6D shows that Notch1 mutations and *PAX1* methylation levels also differ by anatomic location and race in patients with HNSCC. Together these data suggests that Black patients with HNSCC have different

Downloaded from <http://aacrjournals.org/cancerpreventionresearch/article-pdf/12/4/255/2242389/255.pdf> by guest on 27 February 2024

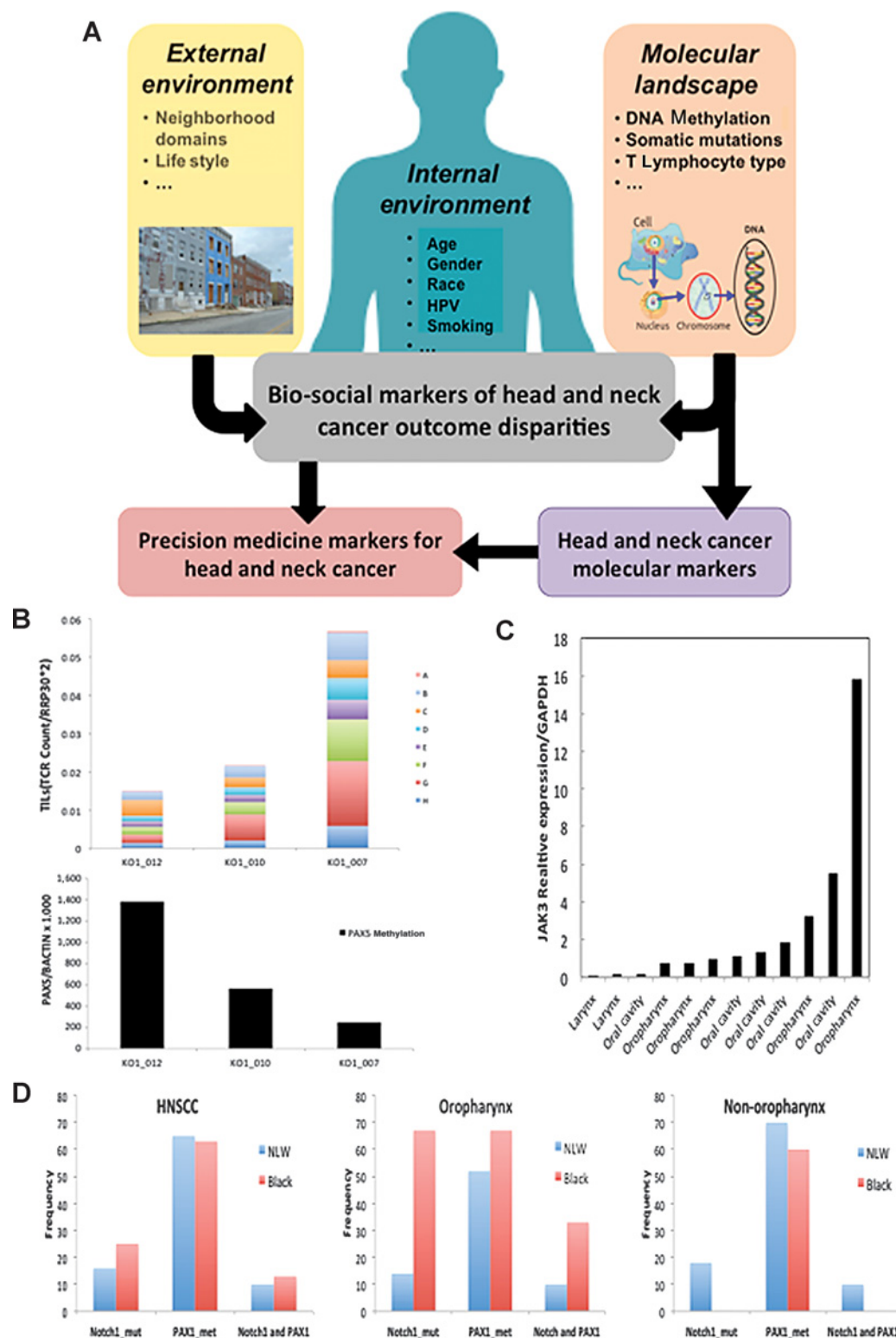


Figure 6.

A, Molecular markers, combined with external and internal environment variables, can be used as biosocial markers of head and neck cancer outcome disparities, head and neck cancer markers, and precision medicine markers for patients with head and neck cancer. **B**, Shows that *PAX5* methylation levels inversely correlate ($r = -0.83$) with TIL counts in patients with HNSCC. **C**, Shows that *JAK3* expression in Black patients with HNSCC differs by anatomic location. **D**, Shows that *Notch1* mutations and *PAX1* methylation levels also differ by anatomic location and race in patients with HNSCC.

methylation and mutation profiles than NLW HNSCC, which may underlie some of the unexplained survival disparities in HNSCC. Notch1 mutations are higher in Black HNSCC, but they are only observed in the oropharynx, where Black patients with HNSCC also show a higher frequency of *PAX1* methylation.

Discussion

This is the first study to examine molecular differences in NLW and Black patients with HNSCC. We performed genome-wide bioinformatics analyses and validated the frequency of selected somatic mutations, germline mutations, promoter methylation, gene expression, and TILs using conventional and droplet PCR primers and probes. We also examined the association of promoter methylation with well understood social determinants of health, identified as neighborhood characteristics at the zip-code level. To our knowledge this is the most comprehensive characterization of molecular disparities in any tumor type.

Genomic and epigenomic inactivation of tumor suppressor genes may partly explain survival disparities in HNSCC (47, 49–51). We show genetic, epigenetic, and TILs frequency differences between Black and NLW patients with HNSCC. Black patients with larynx cancer had a higher frequency of *TP53*, *PIK3CA*, *JAK3*, *KIT*, *APC*, and *MET* somatic mutations. *NOTCH1* had a higher frequency of somatic mutations in Black patients with oropharyngeal cancer and *CDKN2A* had a higher frequency of somatic mutations in Black patients with oral cancer. We also found racial differences in the frequency of a germline C ≥ A mutation in *JAK3*, chr19:17954215 (protein P132T) in larynx cancer. This is the first germline mutation associated to racial differences in HNSCC.

JAKs are a family of tyrosine kinases that are involved in cytokine receptor-mediated intracellular signal transduction. Upon receptor activation JAKs phosphorylate the transcription factors known as STATs and initiate the JAK-STAT signaling pathway. Four JAK family members have been identified (*JAK1*, *JAK2*, *JAK3*, and *Tyk2*), which share a similar protein domain structure: a kinase domain, a regulatory pseudo-kinase domain, a SH2 domain, and a FERM domain. The FERM domain of JAK family members mediates the association of JAK with other enzymes and cytokine receptors.

JAK3 associates with the IL2 receptor gamma chain. It is predominantly expressed in immune cells and transduces a signal in response to its activation via tyrosine phosphorylation by IL receptors. Mutations in this gene are associated with autosomal SCID. *JAK3* is a nonreceptor tyrosine kinase involved in various processes such as cell growth, development, or differentiation. *JAK3* mediates essential signaling events in both innate and adaptive immunity and plays a crucial role in hematopoiesis during T-cell devel-

opment. In the cytoplasm, *JAK3* plays a pivotal role in signal transduction via its association with type I receptors sharing the common subunit gamma such as IL2R, IL4R, IL7R, IL9R, IL15R, and IL21R. Following ligand binding to cell surface receptors, *JAK3* phosphorylates specific tyrosine residues on the cytoplasmic tails of the receptor, creating docking sites for STATs proteins. Subsequently, *JAK3* phosphorylates the STATs proteins once they are recruited to the receptor. Phosphorylated STATs then form homodimer or heterodimers and translocate to the nucleus to activate gene transcription. For example, upon IL2R activation by IL2, *JAK3* molecules bind to IL2R beta (IL2RB) and gamma chain (IL2RG) subunits inducing the tyrosine phosphorylation of both receptor subunits on their cytoplasmic domain. Then, *STAT5A* AND *STAT5B* are recruited, phosphorylated, and activated by *JAK3*. Once activated, dimerized *STAT5* translocates to the nucleus and promotes the transcription of specific target genes in a cytokine-specific fashion.

Survival analyses revealed that patients with *PAX5* ($P = 0.001$) or *NID2* ($P = 0.05$) promoter methylation had a worse outcome than those without it. We also found an association of *PAX5* and *PAX1* promoter methylation with social determinants of health at the zip-code level. Cox proportional hazards regression multivariable analysis, revealed that Black patients with HNSCC with *PAX5* ($P = 0.06$) and *PAX1* ($P = 0.017$) methylation had worse survival than NLW, after controlling for HD, zipcode, and TNM stage.

PAX5 maintains cellular identity by repressing gene expression throughout B-cell differentiation (52–54). Dysregulated expression of *PAX5* is involved in differentiation block (55), somatic hypermutation and immunoglobulin heavy chain class switch recombination (56), leukemogenesis (57), and positive regulation of c-Met transcription (58). *PAX5* association to different tumor types suggests it has diagnostic or prognostic utility in lung cancer (59), Hodgkin lymphoma (60), acute lymphocytic leukemia (61), breast cancer (61), oral cancer (62, 63), gastric cancer (64), head and neck cancer (65, 66), and esophageal cancer (66). A recent study demonstrated that the lack of expression of *PAX5* in lymphoid neoplasms is associated with promoter hypermethylation, leading to *PAX5* silencing in cases characterized by poor clinical outcome (67). Aberrant *PAX5* promoter methylation is also associated with HNSCC (68). *PAX5* promoter methylation is also significantly associated with poor survival in gastric cancer. *PAX5* can induce cell apoptosis through direct upregulation of *TP53*, impacting downstream targets in gastric cancer (69). When expressed, *PAX5* binds to the *TP53* promoter, inducing *TP53* expression, and consequently suppressing cell proliferation. But when *PAX5* is methylated, this tumor suppressor function is abrogated. Spatial analyses identified survival differences and *PAX5* methylation in Black and NLW patients with HNSCC, after

adjusting for socioeconomic characteristics at the zip-code level.

HNSCC disparities are partly mediated by differences in access to care, stage at diagnosis, insurance status, attitudes of health providers, as well as HPV infection status (70, 71) and measures of economic advantage (72). Variations in tumor-associated immune responses between Black and NLWs have also been implicated in cancer health disparities (73–74). Molecular alterations linked to HNSCC survival, such as *TP53* mutations (75), loss of heterozygosity (28), microsatellite instability (29), and differential DNA methylation (30) have been associated to HNSCC main risks factors: tobacco smoking, alcohol consumption, and HPV status (31). Most HPV⁺ tumors occur in the oropharynx, show superior responses to chemotherapy and radiotherapy (32–34), and are less prevalent in Black patients with HNSCC (36, 72). They are also characterized by TILs of the stroma and tumor nests, suggesting HPV⁺ HNSCC tumors do not evade immune surveillance (76). The results we obtained with the QuanTILfy assay support the notion that disparities in HNSCC may be related to differences in TIL's quantity and clonality.

Our results suggest a complex interrelationship between molecular, clinical, and social factors with racial differences in HNSCC survival. Our results underscore the importance of increasing the number of molecular and genomic studies that examine health disparities in HNSCC. HNSCC, like all solid tumors, is thought to be initiated and to progress through a series of clonal and subclonal genetic alterations. The frequency and timing of these alterations driving head and neck tumorigenesis in mostly NLW patients are the focus of several research groups' efforts, including ours. But relatively few efforts have been devoted to study the genetic and epigenetic alterations critical to the development of HNSCC in Black or Latinos.

The canonical clonal hypothesis of cancer posits that a tumor is initiated by a driver mutation, defined as a genetic alteration that increases the ratio of cell birth to cell death. Once these initiated cells avoid death to expand, successive clonal expansions occur with each new driver gene mutation. While there is no full consensus, mathematical models predict that at least 3–5 driver mutations in the same cell are needed to initiate an oncogenic process. It is now understood there is heterogeneity throughout this process, with clonal bottlenecks appearing as some clones predominate over other clones during tumor initiation and progression. Therefore, at any given point in tumor development, there will be at least some heterogeneity within the tumor as a result of anatomic constraints coupled with competing clonal growth, and some due to other factors that are not well understood. These observations reinforce the idea that most somatic mutations found in common adult tumors have arisen as a result of serial subclonal developments, each characterized by a few potential dri-

vers. These drivers can function as the genetic fingerprints of different cancerous tumors.

Together these results suggests that somatic and germline mutations, epigenetic alterations, and TILs may provide a unique molecular fingerprint associated with the higher levels of exposure to biopsychosocial stressors that many racial and ethnically diverse groups are exposed to, mainly by living in disadvantaged neighborhoods. Molecular and social determinants of health at the neighborhood level can potentially inform precision oncology services for racial and ethnically diverse populations.

The main limitation of this study is sample size. Approximately 20% of patient samples in the head and neck tumor bank at Hopkins are from African American patients and some of these samples have been depleted throughout the years. Therefore, we were unable to measure all markers in all patients. We were also limited by the lack of funding for sequencing projects that aim to characterize racial and ethnic disparities in cancer in general and HNSCC specifically. This has certainly not been a priority as evidenced by the small frequency of TCGA samples from Black, Latino/Hispanic, and Asian patients. Measurement of molecular alterations in these populations may increase our understanding of the role they play in cancer disparities, and their interplay with other molecular, psychosocial, and clinical factors related to cancer indolence and aggressiveness across diverse populations.

This is a discovery project, reporting molecular differences that had never been identified in HNSCC racial disparities research. We are only showing associations, not cause and effect relationships. We did not set out to prove the molecular reasons for health disparities in HNSCC. This would require another study design with a much larger sample size to confirm the observed associations in this initial study, followed by work on patient-derived tumor grafts (77, 78).

Ultimately, molecular analysis in precision medicine may be helpful to better elucidate whether observed genomic and epigenomic alterations represent a distinct entity with clinical, immunophenotypic, and molecular characteristics or an incidental phenomenon during malignant transformation. Larger sample sizes will also provide the opportunity to systematically close the existing gap between exposure data and diagnostic/prognostic data. Data science tools can be set in place to integrate demographic, social, economic, and behavioral data, at the individual and neighborhood levels, with molecular features, clinical information, and health services organization data.

This challenge can be addressed with big data analytic strategies, which may include machine learning and computer algorithms to integrate patient demographic, psychosocial, clinical, pathology, and molecular profiles with treatment recommendations, health insurance coverage, clinical, and health information data at the medical

center level, and publicly available geocoded socioeconomic, environmental, social determinants of health data at neighborhood levels. We show schematic representations of the data integration modules and data/systems modules (Supplementary Fig. S2) that can be added to an existing or newly created biospecimen bank, using the HAND as an example. This integration can become the foundation for precision medicine platforms that track the health information and status of individual from cradle to grave, using the similar mathematical modeling used by air traffic controllers to follow the trajectory or flight path of an airplane flying from one airport to another. We show elements that highlight the similarities between both analogies, as well as a framework in which contextual, social, biological, and psychologic process can determine and in turn influence epigenomic changes in Supplementary Fig. S3. In sum, we now have the understanding and capabilities of combining multiple big data streams into next-generation precision medicine tools with machine learning and quantum entanglement capabilities, which will allow us to obtain precise depictions of molecular, clinical, psychosocial, and contextual portraits to track health trajectories across the life span. The markers described in this article may be useful to improve our understanding of HNSCC survival, guide HNSCC treatment options, and inform public health strategies designed to reduce HNSCC cancer disparities.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Wu ES, Park JY, Zeitouni JA, Gomez CR, Reis IM, Zhao W, et al. Effect of actionable somatic mutations on racial/ethnic disparities in head and neck cancer prognosis. *Head Neck* 2016;38:1234–41.
2. Bandera EV, Lee VS, Rodriguez-Rodriguez L, Powell CB, Kushi LH. Racial/ethnic disparities in ovarian cancer treatment and survival. *Clin Cancer Res* 2016;22:5909–14.
3. Jones S, Anagnostou V, Lytle K, Parpart-Li S, Nesselbush M, Riley DR, et al. Personalized genomic analyses for cancer mutation discovery and interpretation. *Sci Transl Med* 2015;7:283ra53.
4. Dayem Ullah AZ, Oscanoa J, Wang J, Nagano A, Lemoine NR, Chelala C. SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Res* 2018;46:W109–13.
5. Finotti A, Allegretti M, Gasparello J, Giacomini P, Spandidos DA, Spoto G, et al. Liquid biopsy and PCR-free ultrasensitive detection systems in oncology (review). *Int J Oncol* 2018;53:1395–434.
6. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature* 2009;458:719–24.
7. Lin J, Gan CM, Zhang X, Jones S, Sjöblom T, Wood LD, et al. A multidimensional analysis of genes mutated in breast and colorectal cancers. *Genome Res* 2007;17:1304–18.
8. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, et al. Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci U S A* 2010;107:18545–50.
9. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 2015;347:78–81.
10. Tomasetti C, Li L, Vogelstein B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* 2017;355:1330–4.
11. Tan SH, Petrovics G, Srivastava S. Prostate cancer genomics: recent advances and the prevailing underrepresentation from racial and ethnic minorities. *Int J Mol Sci* 2018;19:1255.
12. Sawan C, Vaissière T, Murr R, Herceg Z. Epigenetic drivers and genetic passengers on the road to cancer. *Mutat Res* 2008;642:1–13.
13. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, et al. Mutational signatures associated with tobacco smoking in human cancer. *Science* 2016;354:618–22.
14. Spratt DE, Chan T, Waldron L, Speers C, Feng FY, Ogunwobi OO, et al. Racial/ethnic disparities in genomic sequencing. *JAMA Oncol* 2016;2:1070–4.
15. Ademuyiwa FO, Tao Y, Luo J, Weilbaecher K, Ma CX. Differences in the mutational landscape of triple-negative breast cancer in African Americans and Caucasians. *Breast Cancer Res Treat* 2017;161:491–9.
16. Keenan T, Moy B, Mroz EA, Ross K, Niemierko A, Rocco JW, et al. Comparison of the genomic landscape between primary breast cancer in African American versus white women and the

Authors' Contributions

Conception and design: R. Guerrero-Preston

Development of methodology: R. Guerrero-Preston, S. Rodriguez-Torres, F. Pirini, O. Folawiyo, Y.J. Kim

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): R. Guerrero-Preston, B.L. Valle, T. Hadar, B. Rivera, A. Baez, W.M. Koch, W.H. Westra, J.R. Eshleman

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R. Guerrero-Preston, F. Lawson, S. Rodriguez-Torres, M.G. Noordhuis, L. Manuel, B. Valle, L. Marchionni, W.H. Westra, Y.J. Kim, J.R. Eshleman

Writing, review, and/or revision of the manuscript: R. Guerrero-Preston, F. Lawson, S. Rodriguez-Torres, A. Baez, L. Marchionni, W.M. Koch, Y.J. Kim, J.R. Eshleman, D. Sidransky

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): R. Guerrero-Preston, F. Pirini, B.L. Valle, B. Rivera, O. Folawiyo, D. Sidransky

Study supervision: R. Guerrero-Preston

Acknowledgments

This research was supported by NCI grants U01CA84986 and K01CA164092 and CA121113; National Institute of Dental and Craniofacial Research grants P50DE019032 Head and Neck Cancer SPORE, and RC2 DE20957.

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 November 3, 2017; revised October 30, 2018; accepted February 11, 2019; published first February 18, 2019.

- association of racial differences with tumor recurrence. *J Clin Oncol* 2015;33:3621–7.
17. Parada H Jr, Sun X, Fleming JM, Williams-DeVane CLR, Kirk EL, Olsson LT, et al. Race-associated biological differences among luminal A and basal-like breast cancers in the Carolina Breast Cancer Study. *Breast Cancer Res* 2017;19:131.
 18. D'Souza G, Westra WH, Wang SJ, van Zante A, Wentz A, Kluz N, et al. Differences in the prevalence of human papillomavirus (HPV) in head and neck squamous cell cancers by sex, race, anatomic tumor site, and HPV detection method. *JAMA Oncol* 2017;3:169–77.
 19. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature* 2016;538:161–4.
 20. Boden G, Duan X, Homko C, Molina EJ, Song W, Perez O, et al. Increase in endoplasmic reticulum stress-related proteins and genes in adipose tissue of obese, insulin-resistant individuals. *Diabetes* 2008;57:2438–44.
 21. Green RC, Goddard KAB, Jarvik GP, Amendola LM, Appelbaum PS, Berg JS, et al. Clinical sequencing exploratory research consortium: accelerating evidence-based practice of genomic medicine. *Am J Hum Genet* 2016;98:1051–66.
 22. Goodwin WJ, Thomas GR, Parker DF, Joseph D, Levis S, Franzmann E, et al. Unequal burden of head and neck cancer in the United States. *Head Neck* 2008;30:358–71.
 23. Jiron J, Sethi S, Ali-Fehmi R, Franceschi S, Struijk L, van Doorn LJ, et al. Racial disparities in human papillomavirus (HPV) associated head and neck cancer. *Am J Otolaryngol* 2014;35:147–53.
 24. Liu GF, Ranck MC, Solanki AA, Cao H, Kolokythas A, Wenig BL, et al. Racial parities in outcomes after radiotherapy for head and neck cancer. *Cancer* 2014;120:244–52.
 25. Molina MA, Cheung MC, Perez EA, Byrne MM, Franceschi D, Moffat FL, et al. African American and poor patients have a dramatically worse prognosis for head and neck cancer: an examination of 20,915 patients. *Cancer* 2008;113:2797–806.
 26. Schrank TP, Han Y, Weiss H, Resto VA. Case-matching analysis of head and neck squamous cell carcinoma in racial and ethnic minorities in the United States—possible role for human papillomavirus in survival disparities. *Head Neck* 2011;33:45–53.
 27. Worsham MJ, Stephen JK, Lu M, Chen KM, Havard S, Shah V, et al. Disparate molecular, histopathology, and clinical factors in head and neck squamous cell carcinoma racial groups. *Otolaryngol Head Neck Surg* 2012;147:281–8.
 28. Poeta ML, Manola J, Goldwasser MA, Forastiere A, Benoit N, Califano JA, et al. TP53 mutations and survival in squamous-cell carcinoma of the head and neck. *N Engl J Med* 2007;357:2552–61.
 29. Tan D, Wiseman S, Zhou Y, Li Q, Ward P, Slocum HK, et al. Definition of a region of loss of heterozygosity at chromosome 11q23.3–25 in head and neck squamous cell carcinoma using laser capture microdissection technique. *Diagn Mol Pathol* 2004;13:33–40.
 30. Zuo C, Zhang H, Spencer HJ, Vural E, Suen JY, Schichman SA, et al. Increased microsatellite instability and epigenetic inactivation of the hMLH1 gene in head and neck squamous cell carcinoma. *Otolaryngol Head Neck Surg* 2009;141:484–90.
 31. Tan HK, Saulnier P, Auperin A, Lacroix L, Casiraghi O, Janot F, et al. Quantitative methylation analyses of resection margins predict local recurrences and disease-specific deaths in patients with head and neck squamous cell carcinomas. *Br J Cancer* 2008;99:357–63.
 32. Namazie A, Alavi S, Olopade OI, Pauletti G, Aghamohammadi N, Aghamohammadi M, et al. Cyclin D1 amplification and p16 (MTS1/CDK4I) deletion correlate with poor prognosis in head and neck tumors. *Laryngoscope* 2002;112:472–81.
 33. Marsit CJ, Posner MR, McClean MD, Kelsey KT. Hypermethylation of E-cadherin is an independent predictor of improved survival in head and neck squamous cell carcinoma. *Cancer* 2008;113:1566–71.
 34. Stanbery L, D'Silva NJ, Lee JS, Bradford CR, Carey TE, Prince ME, et al. High SEPT9_v1 expression is associated with poor clinical outcomes in head and neck squamous cell carcinoma. *Transl Oncol* 2010;3:239–45.
 35. Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, et al. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst* 2008;100:261–9.
 36. Reichert TE, Day R, Wagner EM, Whiteside TL. Absent or low expression of the zeta chain in T cells at the tumor site correlates with poor survival in patients with oral carcinoma. *Cancer Res* 1998;58:5344–7.
 37. Watanabe Y, Katou F, Ohtani H, Nakayama T, Yoshie O, Hashimoto K. Tumor-infiltrating lymphocytes, particularly the balance between CD8(+) T cells and CCR4(+) regulatory T cells, affect the survival of patients with oral squamous cell carcinoma. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2010;109:744–52.
 38. Badoual C, Hans S, Merillon N, Van Ryswick C, Ravel P, Benhamouda N, et al. PD-1-expressing tumor-infiltrating T cells are a favorable prognostic biomarker in HPV-associated head and neck cancer. *Cancer Res* 2013;73:128–38.
 39. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001;125:279–84.
 40. Kim SY, Volsky DJ. PAGE: parametric analysis of gene set enrichment. *BMC Bioinformatics* 2005;6:144.
 41. Mootha VK, Lepage P, Miller K, Bunkenborg J, Reich M, Hjerrild M, et al. Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc Natl Acad Sci U S A* 2003;100:605–10.
 42. Ross AE, Marchionni L, Vuica-Ross M, Cheadle C, Fan J, Berman DM, et al. Gene expression pathways of high grade localized prostate cancer. *Prostate* 2011;71:1568–77.
 43. Tyekuceva S, Marchionni L, Karchin R, Parmigiani G. Integrating diverse genomic data using gene sets. *Genome Biol* 2011;12:R105.
 44. Michailidi C, Soudry E, Brait M, Maldonado L, Jaffe A, Ili-Gangas C, et al. Genome-wide and gene-specific epigenomic platforms for hepatocellular carcinoma biomarker development trials. *Gastroenterol Res Pract* 2014;2014:597164.
 45. Guerrero-Preston R, Soudry E, Acero J, Orera M, Moreno-López L, Macía-Colón G, et al. NID2 and HOXA9 promoter hypermethylation as biomarkers for prevention and early detection in oral cavity squamous cell carcinoma tissues and saliva. *Cancer Prev Res* 2011;4:1061–72.
 46. U.S. Census Bureau. American FactFinder, Glossary. Census 2010. Available at <https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>.
 47. Settle K, Posner MR, Schumaker LM, Tan M, Suntharalingam M, Golubeva O, et al. Racial survival disparity in head and neck cancer results from low prevalence of human papillomavirus infection in black oropharyngeal cancer patients. *Cancer Prev Res* 2009;2:776–81.
 48. Aronson RE, Wallis AB, O'Campo PJ, Schafer P. Neighborhood mapping and evaluation: a methodology for participatory

- community health initiatives. *Matern Child Health J* 2007;11:373–83.
49. Arbes SJ Jr, Olshan AF, Caplan DJ, Schoenbach VJ, Slade GD, Symons MJ. Factors contributing to the poorer survival of black Americans diagnosed with oral cancer (United States). *Cancer Causes Control* 1999;10:513–23.
 50. Murdock JM, Gluckman JL. African-American and white head and neck carcinoma patients in a university medical center setting. Are treatments provided and are outcomes similar or disparate? *Cancer* 2001;91:279–83.
 51. Gourin CG, Podolsky RH. Racial disparities in patients with head and neck squamous cell carcinoma. *Laryngoscope* 2006;116:1093–106.
 52. Carotta S, Holmes ML, Pridans C, Nutt SL. Pax5 maintains cellular identity by repressing gene expression throughout B cell differentiation. *Cell Cycle* 2006;5:2452–6.
 53. Holmes ML, Pridans C, Nutt SL. The regulation of the B-cell gene expression programme by Pax5. *Immunol Cell Biol* 2008;86:47–53.
 54. Holmes ML, Carotta S, Corcoran LM, Nutt SL. Repression of Flt3 by Pax5 is crucial for B-cell lineage commitment. *Genes Dev* 2006;20:933–8.
 55. Douglas WG, Tracy E, Tan D, Yu J, Hicks WL Jr, Rigual NR, et al. Development of head and neck squamous cell carcinoma is associated with altered cytokine responsiveness. *Mol Cancer Res* 2004;2:585–93.
 56. Dege C, Hagman J. Activation of Aicda gene transcription by Pax5 in plasmacytoma cells. *Immunol Res* 2013;55:155–61.
 57. Medvedovic J, Ebert A, Tagoh H, Buslinger M. Pax5: a master regulator of B cell development and leukemogenesis. *Adv Immunol* 2011;111:179–206.
 58. Kanteti R, Nallasura V, Loganathan S, Tretiakova M, Kroll T, Krishnaswamy S, et al. PAX5 is expressed in small-cell lung cancer and positively regulates c-Met transcription. *Lab Invest* 2009;89:301–14.
 59. Ren Y, Hou J, Xu A, Pan Y. Diagnostic utility of PAX2 and PAX5 in distinguishing non-small cell lung cancer from small cell lung cancer. *Int J Clin Exp Pathol* 2015;8:14709–16.
 60. Ranuhardy D, Suzanna E, Sari RM, Hadisantoso DW, Andalucia R, Abdillah A. CD30, CD15, CD50, and PAX5 expressions as diagnostic markers for Hodgkin lymphoma (HL) and systemic anaplastic large cell lymphoma (sALCL). *Acta Med Indones* 2018;50:104–9.
 61. Moafi A, Zojaji A, Salehi R, Najafi Dorcheh S, Rahgozar S. The correlation between Pax5 deletion and patients survival in Iranian children with precursor B-cell acute lymphocytic leukemia. *Cell Mol Biol* 2017;63:19–22.
 62. Ribeiro IP, Caramelo F, Marques F, Domingues A, Mesquita M, Barroso L, et al. WT1, MSH6, GATA5 and PAX5 as epigenetic oral squamous cell carcinoma biomarkers - a short report. *Cell Oncol* 2016;39:573–82.
 63. Ribeiro IP, Caramelo F, Esteves L, Oliveira C, Marques F, Barroso L, et al. Genomic and epigenetic signatures associated with survival rate in oral squamous cell carcinoma patients. *J Cancer* 2018;9:1885–95.
 64. Deng J, Liang H, Zhang R, Dong Q, Hou Y, Yu J, et al. Applicability of the methylated CpG sites of paired box 5 (PAX5) promoter for prediction the prognosis of gastric cancer. *Oncotarget* 2014;5:7420–30.
 65. Hayashi M, Guerrero-Preston R, Sidransky D, Koch WM. Paired box 5 methylation detection by droplet digital PCR for ultra-sensitive deep surgical margins analysis of head and neck squamous cell carcinoma. *Cancer Prev Res* 2015;8:1017–26.
 66. Hayashi M, Guerrero-Preston R, Okamura J, Michailidi C, Kahn Z, Li X, et al. Innovative rapid gene methylation analysis of surgical margin tissues in head and neck cancer. *Ann Surg Oncol* 2014;21:3124–31.
 67. Lazzi S, Bellan C, Onnis A, De Falco G, Sayed S, Kostopoulos I, et al. Rare lymphoid neoplasms coexpressing B- and T-cell antigens. The role of PAX-5 gene methylation in their pathogenesis. *Hum Pathol* 2009;40:1252–61.
 68. Guerrero-Preston R, Michailidi C, Marchionni L, Pickering CR, Frederick MJ, Myers JN, et al. Key tumor suppressor genes inactivated by "greater promoter" methylation and somatic mutations in head and neck cancer. *Epigenetics* 2014;9:1031–46.
 69. Li X, Cheung KF, Ma X, Tian L, Zhao J, Go MY, et al. Epigenetic inactivation of paired box gene 5, a novel tumor suppressor gene, through direct upregulation of p53 is associated with prognosis in gastric cancer patients. *Oncogene* 2012;31:3419–30.
 70. Robins HS, Ericson NG, Guenthoer J, O'Briant KC, Tewari M, Drescher CW, et al. Digital genomic quantification of tumor-infiltrating lymphocytes. *Sci Transl Med* 2013;5:214ra169.
 71. Peterson CE, Khosla S, Chen LF, Joslin CE, Davis FG, Fitzgibbon ML, et al. Racial differences in head and neck squamous cell carcinomas among non-Hispanic black and white males identified through the National Cancer Database (1998–2012). *J Cancer Res Clin Oncol* 2016;142:1715–26.
 72. Peterson CE, Khosla S, Jefferson GD, Davis FG, Fitzgibbon ML, Freels S, et al. Measures of economic advantage associated with HPV-positive head and neck cancers among non-Hispanic black and white males identified through the National Cancer Database. *Cancer Epidemiol* 2017;48:1–7.
 73. Whitworth A. New research suggests access, genetic differences play role in high minority cancer death rate. *J Natl Cancer Inst* 2006;98:669.
 74. Lippman SM, Lee JJ. Reducing the "risk" of chemoprevention: defining and targeting high risk–2005 AACR cancer research and prevention foundation award lecture. *Cancer Res* 2006;66:2893–903.
 75. Yang HC, Liang YJ, Wu YL, Chung CM, Chiang KM, Ho HY, et al. Genome-wide association study of young-onset hypertension in the Han Chinese population of Taiwan. *PLoS One* 2009;4:e5459.
 76. Westra WH. The changing face of head and neck cancer in the 21st century: the impact of HPV on the epidemiology and pathology of oral cancer. *Head Neck Pathol* 2009;3:78–81.
 77. Izumchenko E, Paz K, Ciznadija D, Sloma I, Katz A, Vasquez-Dunddel D, et al. Patient-derived xenografts effectively capture responses to oncology therapy in a heterogeneous cohort of patients with solid tumors. *Ann Oncol* 2017;28:2595–605.
 78. Ben-David U, Ha G, Tseng YY, Greenwald NF, Oh C, Shih J, et al. Patient-derived xenografts undergo mouse-specific tumor evolution. *Nat Genet* 2017;49:1567–75.