

Immunogenetic Determinants of Susceptibility to Head and Neck Cancer in the Million Veteran Program Cohort

Yanhong Liu^{1,2,3}, Jennifer R. Kramer^{1,3}, Vlad C. Sandulache^{4,5,6}, Robert Yu⁷, Guojun Li⁸, Liang Chen^{1,3}, Zenab I. Yusuf^{1,3}, Yunling Shi⁹, Saiju Pyarajan⁹, Spyros Tsavachidis¹, Li Jiao¹, Michelle L. Mierzwa¹⁰, Elizabeth Chiao¹¹, Yvonne M. Mowery¹², Andrew Shuman^{13,14}, Sanjay Shete⁷, Andrew G. Sikora⁸, and Donna L. White^{1,2,3,6}



ABSTRACT

Increasing rates of human papillomavirus (HPV)-driven oropharyngeal cancer (OPC) have largely offset declines in tobacco-associated head and neck squamous cell carcinoma (HNSCC) at non-OPC sites. Host immunity is an important modulator of HPV infection, persistence, and clearance, and also of immune evasion in both virally- and nonvirally-driven cancers. However, the association between collective known cancer-related immune gene variants and HNSCC susceptibility has not been fully characterized. Here, we conducted a genetic association study in the multiethnic Veterans Affairs Million Veteran Program cohort, evaluating 16,050 variants in 1,576 immune genes in 4,012 HNSCC cases (OPC = 1,823; non-OPC = 2,189) and 16,048 matched controls. Significant polymorphisms were further examined in a non-Hispanic white (NHW) validation cohort (OPC = 1,206; non-OPC = 955; controls = 4,507). For overall HNSCC susceptibility in NHWs, we discovered and validated a novel 9q31.1 SMC2 association and replicated the known 6p21.32 HLA-DQ-DR association.

Six loci/genes for overall HNSCC susceptibility were selectively enriched in African-Americans (6p21.32 HLA-G, 9q21.33 GAS1, 11q12.2 CD6, 11q23.2 NCAM1/CD56, 17p13.1 CD68, 18q22.2 SOCS6); all 6 genes function in antigen-presenting regulation and T-cell activation. Two additional loci (10q26 DMBT1, 15q22.2 TPM1) were uncovered for non-OPC susceptibility, and three loci (11q24 CRTAM, 16q21 CDH5, 18q12.1 CDH2) were identified for HPV-positive OPC susceptibility. This study underscores the role of immune gene variants in modulating susceptibility for both HPV-driven and non-HPV-driven HNSCC. Additional large studies, particularly in racially diverse populations, are needed to further validate the associations and to help elucidate other potential immune factors and mechanisms that may underlie HNSCC risk.

Significance: Several inherited variations in immune system genes are significantly associated with susceptibility to head and neck cancer, which could help improve personalized cancer risk estimates.

Introduction

Head and neck squamous cell carcinoma (HNSCC), which includes cancers of the oral cavity, pharynx, and larynx, is the seventh most common type of cancer globally (1). HNSCCs are primarily diagnosed in men in their 60s and 70s and are commonly subdivided into human papillomavirus (HPV)-driven oropharyngeal cancers (OPC) or non-HPV-related cancers typically related to tobacco and/or alcohol use. These carcinogen-mediated cancers can also arise in the oropharynx but are more common at non-oropharyngeal sites (non-OPC; refs. 2, 3). The incidence of OPC has increased at a near exponential rate over the last few decades. Largely driven by epidemic increases in oral HPV infection in the general population, most OPCs (up to 80%) are HPV-positive, with OPC now overtaking cervical cancer as the leading cause of HPV-related cancer in the United States. In contrast, the rate of non-OPC HNSCC has steadily decreased, mirroring declining smoking rates in the U.S. population.

In addition to oncogenic risk factors (e.g., HPV infection, tobacco use, and alcohol consumption), a substantial heritable genetic component to HNSCC susceptibility is supported by large family studies demonstrating a 3- to 10-fold increased risk in first-degree relatives of patients with HNSCC (4, 5). The incidence of HPV-driven cancers (cervical, anogenital, and HNSCC) is dramatically higher in immunosuppressed individuals and people living with human immunodeficiency virus (HIV; refs. 6, 7). Prior candidate gene association studies have found polymorphisms in immune-mediator genes associated with susceptibility to persistent HPV infection and HPV-driven cancers. Several genome-wide association studies (GWAS; refs. 8–12) have identified over 14 HNSCC predisposing loci; all consistently identified the human leukocyte antigen (*HLA*) locus, highlighting the

¹Department of Medicine, Baylor College of Medicine, Houston, Texas. ²Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas. ³Veterans Affairs (VA) Health Services Research & Development Center of Innovations in Quality, Effectiveness, and Safety (IQESt), Michael E. DeBakey VA Medical Center, Houston, Texas. ⁴ENT Section, Operative Care Line, Michael E. DeBakey VA Medical Center, Houston, Texas. ⁵Bobby R. Alford Department of Otolaryngology-Head and Neck Surgery, Baylor College of Medicine, Houston, Texas. ⁶Center for Translational Research in Inflammatory Disease (CTRID), Michael E. DeBakey VA Medical Center, Houston, Texas. ⁷Department of Biostatistics, Division of Basic Sciences, The University of Texas MD Anderson Cancer Center, Houston, Texas. ⁸Department of Head and Neck Surgery, Division of Surgery, The University of Texas MD Anderson Cancer Center, Houston, Texas. ⁹Center for Data and Computational Sciences, VA Boston Healthcare System, Boston, Massachusetts. ¹⁰University Hospital, University of Michigan, Ann Arbor, Michigan. ¹¹Departments of Epidemiology and General Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas. ¹²Departments of Radiation Oncology and Head and Neck Surgery & Communication Sciences, Duke University Medical Center, Durham, North Carolina. ¹³Department of Otolaryngology, University of Michigan Medical School, Ann Arbor, Michigan. ¹⁴Veterans Affairs Ann Arbor Health System, Ann Arbor, Michigan.

Corresponding Authors: Donna L. White, Department of Medicine, Baylor College of Medicine, Houston, TX 77021. E-mail: dwhite1@bcm.edu; and Andrew G. Sikora, Department of Head and Neck Surgery, Division of Surgery, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030. E-mail: agsikora@mdanderson.org

Cancer Res 2023;83:386–97

doi: 10.1158/0008-5472.CAN-22-1641

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

©2022 The Authors; Published by the American Association for Cancer Research

importance of immunologic mechanisms in the etiology of HNSCC. However, most previous studies lacked detailed information on HPV status among OPC, and none were powered to evaluate the collective impact of immune gene variants on HNSCC risk.

We intend to close this knowledge gap using data from the multi-ethnic Million Veteran Program (MVP; ref. 13), one of the largest and most racial/ethnically diverse biobanks in the world, to examine the genetic associations of HPV-driven and HPV-negative HNSCC. The MVP was initiated by the Department of Veterans Affairs (VA) in 2011 to collect genomic, epidemiologic, and clinical data from at least one million veterans to study the role of genetics, lifestyle, and military experiences in the development of human diseases. We hypothesized that immune variants may confer an immune advantage to the virus or the host, and thus play a prominent role in HNSCC susceptibility. We believe that results from our study will have relevance to the development of HNSCC screening programs, genetic counseling, and targeted interventions (e.g., HPV vaccination and smoking cessation), which can ultimately reduce the burden of HNSCC.

Materials and Methods

Human subjects, sample collection, and genomic analyses of the MVP cohort

The MVP (RRID:SCR_021731) eligible study population consists of U.S. men and women veterans (age 18–90 years old) who were enrolled in the nationwide VA healthcare system, beginning in early 2011 (13). All participants in the parent MVP study cohort provided written informed consent in accordance with recognized ethical guidelines of U.S. Common Rule as the only inclusion criteria (see informed consent document at <http://www.research.va.gov/MVP/>). The parent MVP study and the current study protocol were both approved by the U.S. Veteran Affairs Central Institutional Review Board.

Study participants provided blood samples for genotyping and completed surveys to obtain basic demographic and lifestyle characteristics. DNA was extracted from blood and genotyped on a custom Axiom MVP 1.0 array (~723,000 markers). This MVP chip: (i) is enriched for exome polymorphisms and tag SNPs validated for diseases; (ii) augmented with biomarkers of specific interest to the VA population including African-American and Hispanic ancestry markers; and (iii) improves causal variant mapping, in particular covering a large proportion of the *HLA* region, which is one of the most variable regions in the human genome and technically challenging to assay directly on other commercial arrays.

Identification of HNSCC cases and controls in the MVP-HNSCC study

MVP participants who were enrolled between 2011 and December 2019, and whose genotype data were available as of June 2021, were eligible for this study. We identified all HNSCCs among this MVP cohort and further classified them as (i) primary OPC—cancer in the oropharynx, soft palate, or tonsils, and (ii) primary non-OPC—sites in the larynx, hypopharynx, and oral cavity, respectively using diagnostic code and keyword searches of the cancer site and histology variables in the VA's adjudicated Central Cancer Registry (VA CCR). Additional diagnostic ICD code searches were performed followed by physician Electronic Medical Records (EMR) review to confirm additional cases not captured by the VA CCR. Searches for HNSCC diagnoses were performed using VA's Corporate Data Warehouse (CDW). We excluded cases in the salivary gland and nasopharynx as a subset

attributable to Epstein–Barr virus (EBV). We also excluded patients if they received allogeneic (e.g., solid organ, bone marrow) transplants and if they had any cancer diagnosis other than nonmelanoma skin cancer prior to HNSCC diagnosis. We selected 4:1 frequency-matched cancer-free controls separately for OPC and non-OPC cases using approximate incidence density sampling including the following match criteria: age, sex, race/ethnicity, first visit date in VA healthcare and index date with controls having VA visit in the same year as case's index date (date of the first diagnosis of HNSCC). The MVP cohort used Harmonized Ancestry and Race/Ethnicity (HARE) to partition a multiethnic cohort into nonoverlapping strata: non-Hispanic whites, Hispanic-Americans and African-Americans. This enables most individuals to be included in the individual analyses, regardless of whether self-identified race/ethnicity is available.

HPV status and risk factor phenotyping in the MVP-HNSCC study

HPV/p16 status was determined using text-word searches of pathology/oncology relevant narrative notes followed by EMR review to confirm p16 IHC staining (HPV surrogate biomarker) or direct high-risk HPV testing (*in situ* hybridization) and their results (positive, negative, or indeterminate) for OPCs diagnosed 2010 onward, as testing was not routinely conducted before then. Smoking history (never, former, and current) and presence of alcohol abuse using the validated Alcohol Use Disorders Identification Test-Consumption (AUDIT-C) were obtained from nationwide mandated health factor screenings before the index date. Maximum body mass index (BMI) based on the highest value before the index date was categorized as normal (<25), overweight (25–30), and obese (>30).

Immune genes selection and genotype data acquisition for the MVP-HNSCC study

We assembled a list of 1,576 immune-related genes by undertaking an extensive literature search and by querying InnateDB (RRID:SCR_006714) and immunome databases that cover the key pathways that regulate immunity and inflammation such as antigen presentation, T-cell priming or activation, immune cell localization, recognition/killing of cancer cells, myeloid cell and natural killer (NK) cell activity, cell cycle and proliferation, tumor-intrinsic factors, immuno-metabolism, killer cell immunoglobulin-like receptor (KIR) cluster, and common signaling pathways (Wnt, TGF β , NF- κ B, TLR, and Jak/STAT; Supplementary Table S1). We also included the significant susceptibility genes identified in prior HNSCC-GWAS (8–12).

A total of 22,953 SNPs in these 1,576 immune-related genes were identified from the MVP Chip. Our analysis included common [minor allele frequency (MAF) \geq 5%] and low-frequency (1% \leq MAF < 5%) variants. To avoid analysis of highly correlated SNPs, we used Haploview (RRID:SCR_003076) software (14) to prune the dataset by imposing a linkage disequilibrium (LD) threshold of $r^2 < 0.8$. Subjects with call rates <95% and SNPs with call rates <95% were removed. SNPs with Hardy–Weinberg equilibrium tests $P < 10^{-3}$ among controls were removed. The subject's sex was verified using the sex check option in PLINK. Relationship checking was performed by estimating the proportion of alleles shared identically by descent for all pairs of subjects in PLINK (RRID:SCR_001757).

Association analysis in the MVP-HNSCC study

The association between each SNP and overall HNSCC risk and risk by tumor site (OPC and non-OPC) was assessed by the Cochran–Armitage trend test. To control for multiple testing, a FDR adjusted

P value threshold of 5×10^{-5} was used. Effect size odds ratio (OR) measuring risk change per copy of the minor allele was calculated in logistic regression assuming a log-additive genetic model. Regression models were adjusted for age, sex, race/ethnicity (HARE), smoking status, alcohol abuse, BMI, and eigenvectors. Because each HARE stratum is not genetically homogeneous, therefore, we have also accounted for genetic structure by adjusting principal components in our analysis. The selection of covariables was based on *a priori* knowledge regarding the relationships between these factors and HNSCC risk. LD blocks were defined based on the HapMap recombination rate using the recombination hotspots. Haplotypes were inferred using an Expectation–Maximization algorithm and analyzed based on the generalized linear model framework with adjustment of confounding variables. In the racial stratification, we focused on non-Hispanic whites (NHW) and African-Americans, because other ethnic groups were not well represented. We also performed stratified analyses based on smoking, alcohol abuse, and BMI.

Functional variant annotation, expression quantitative trait loci, and Kyoto Encyclopedia of Genes and Genomes pathway analysis

To explore possible functional implications of candidate variants identified in association analysis (overall, OPC, non-OPC, and stratified analyses), we used multiple public annotation databases including (i) Encyclopedia of DNA Elements (ENCODE; RRID:SCR_006793), to predict SNP functional regulatory features (transcription factor binding, open chromatin, and the presence of putative enhancer, promoter, enhancers); (ii) Genotype-Tissue Expression Project (GTEx v8; RRID:SCR_013042) expression quantitative trait loci (eQTL) summary statistics, based on RNA-seq analysis of lung and squamous esophagus ($n = 483$) that best represent squamous epithelium (like oropharyngeal and oral cavity); (iii) pathway enrichment analyses by Kyoto Encyclopedia of Genes and Genomes (KEGG; RRID:SCR_001120) and Gene Ontology (GO; RRID:SCR_002811); and (iv) functional interaction networks via Search Tool for the Retrieval of Interacting Genes (STRING; RRID:SCR_005223).

External validation and replication analysis

The replication cohort dataset was part of a published HNSCC-GWAS (9) in NHW with a diagnosis of HNSCC treated at The University of Texas MD Anderson Cancer Center (MDACC) between December 1996 and July 2011, whose genomic DNA was genotyped with Illumina HumanOmniExpress-12v1. The controls were recruited from genetically unrelated visitors who accompanied patients with cancer to MDACC outpatient clinics, or individuals from the MDACC melanoma study and the Study of Addiction: Genetics and Environment (SAGE). The candidate variants associated with HNSCC risk (overall and by site, OPC and non-OPC, respectively) identified from the discovery in the NHW were further examined in the MDACC validation study. We performed a random-effects meta-analysis of association statistics from the two studies to calculate a combined P value. Random-effects meta-analyses allow for heterogeneity caused by differences in study populations and imbalanced study sizes. Classical fixed-effects meta-analyses are not optimal for the analysis across study estimates where underlying allele frequencies are different between studies.

Data availability

All project-level data (anonymized individual genotypes, epidemiologic, and clinical data) from the MVP has been made available

to approved MVP researchers at the Genomic Information System for Integrative Science (GenISIS). The genotyping data of MDACC HNSCC cases were deposited in the NCBI dbGaP (RRID:SCR_002709) database of genotypes and phenotypes (dbGaP accession no.: phs001173.v1.p1). The controls including the MDACC melanoma study49 (dbGaP accession no.: phs000187.v1.p1) and Study of Addiction: Genetics and Environment (SAGE50; dbGaP accession no.: phs000092.v1.p1). Genotype-Tissue Expression dataset (GTEx v8) is publicly available and can be downloaded (<https://gtexportal.org/home/protectedDataAccess>).

Results

Characteristics of the MVP-HNSCC study and MDACC validation study

In the discovery study, we identified 20,060 study-eligible MVP participants, including 4,012 patients with HNSCC (45% OPC; 55% non-OPC) and 16,048 frequency-matched cancer-free controls (Table 1). As expected, there was a high predominance of males (98.7%) given the demographics of the Veteran population. African Americans represented 13.6% of the cohort. The non-OPC cases were more likely to be lifelong smokers and have a history of alcohol abuse. Among OPC cases from 2010 onward, 1,002 (55%) had HPV and/or p16 testing data available (Table 1). Of these, 793 (79%) were deemed HPV/p16 positive and 209 (21%) were deemed HPV/p16 negative. Among African-American OPC patients, 103 (51%) with HPV data and/or p16 testing data available, 84 (82%) were deemed HPV/p16 positive and 19 (18%) were deemed HPV/p16 negative.

The MDACC validation study is comprised of 2,185 cases (55% OPC; 45% non-OPC) and 4,507 controls. Compared with the MVP-HNSCC study, the MDACC study had a substantially lower prevalence of males (77% of men in cases, 55% of men in controls); the MDACC study only included NHW subjects (by study design).

Risk of all HNSCC types in MVP-HNSCC study

In the discovery set, after stringent QC exclusions, genotype data were available for 16,050 common and low-frequency SNPs in the 1,576 immune genes. We performed association analysis for these SNPs with overall HNSCC susceptibility in NHW and African Americans separately (Fig. 1). In NHW (3,266 cases vs. 13,064 controls), we identified five loci (53 SNPs) with P values $< 5 \times 10^{-5}$ (Fig. 1A). The strongest signal was seen in the *HLA* region on 6p21.32. Forty-one significant SNPs were detected in this *HLA* region, with the lowest P value occurring within *HLA* class-II variants, such as *HLA-DRB5* rs111834747, *HLA-DRB1* rs28724008, and *HLA-DQB1* rs6928482 with $P = 5.5 \times 10^{-7}$, 1.3×10^{-6} , and 2.2×10^{-6} , respectively (Table 2). The second strong association is 19p13.2 intercellular adhesion molecule 5 (*ICAM5*) rs11575074 ($P = 8.4 \times 10^{-6}$). The third locus is 16q24.3 Fanconi anemia complementation group A (*FANCA*) rs12931267 ($P = 1.4 \times 10^{-11}$), and two additional significant variants (rs17233497 and rs17232910) in *FANCA*. Haplotype analysis of these *FANCA* three variants showed a significant global association ($P_{\text{global}} = 1.4 \times 10^{-11}$, Table 3). The fourth locus is 5q35.1 Fibroblast Growth Factor 18 (*FGF18*) rs67585403 ($P = 4.6 \times 10^{-5}$). Additional suggestive candidates in *FGF18* are rs62383998 and rs78810186. The fifth locus is 9q31.1 structural maintenance of chromosomes 2 (*SMC2*) rs3818625 ($P = 4.1 \times 10^{-5}$). Additional suggestive (borderline significance after FDR correction) candidates in *SMC2* include rs10820599,

Table 1. Basic characteristics of the MVP-HNSCC study participants.

Characteristics N (%) ^a	Overall HNSCC N = 4,012	Controls N = 16,048	P value ^a	OPC n = 1,823	Controls n = 7,292	P value ^a	Non-OPC n = 2,189	Controls n = 8,756	P value ^a
HARE race/ethnicity ^b			1.00			1.00			1.00
Non-Hispanic Whites (NHW)	3,266 (81.4)	13,064 (81.4)		1,525 (83.6)	6,100 (83.6)		1,741 (79.5)	6,964 (79.5)	
Hispanic-Americans	161 (4.0)	644 (4.0)		79 (4.3)	316 (4.3)		82 (3.8)	328 (3.8)	
African-Americans	549 (13.6)	2196 (13.6)		201 (11.0)	804 (11.0)		348 (15.9)	1,392 (15.9)	
Age, years			0.97			0.66			0.86
Mean (SD)	66 (8.9)	66 (9.1)		64.9 (7.5)	65.0 (8.0)		66.9 (8.6)	66.9 (9.0)	
Sex			1.00			1.00			1.00
Male	3,958 (98.7)	15,832 (98.7)		1,796 (98.5)	7,184 (98.5)		2,162 (98.8)	8,648 (98.8)	
Female	54 (1.3)	216 (1.3)		27 (1.5)	108 (1.5)		27 (1.2)	108 (1.2)	
Smoking			<0.001			<0.001			<0.001
Never	547 (13.6)	3,914 (24.4)		359 (19.7)	1,796 (24.6)		188 (8.6)	2,118 (24.2)	
Former	1,722 (42.9)	7,701 (48.0)		849 (46.6)	3,414 (46.8)		873 (39.9)	4,287 (49.0)	
Current	1,702 (42.4)	3,985 (24.8)		592 (32.5)	1,841 (25.2)		1,110 (50.7)	2,144 (24.5)	
Alcohol abuse			<0.001			0.04			<0.001
No	2,683 (66.9)	11,660 (72.7)		1,328 (72.8)	5,405 (74.1)		1,355 (61.9)	6,255 (71.4)	
Yes	709 (17.7)	1,892 (11.8)		269 (14.8)	916 (12.6)		440 (20.1)	976 (11.2)	
Body mass index (BMI)			<0.001			<0.001			<0.001
Normal, BMI < 25	1,239 (30.9)	2,574 (16.0)		494 (27.1)	1,116 (15.3)		745 (34.0)	1,458 (16.7)	
Overweight, 25 ≤ BMI ≤ 30	1,419 (35.4)	5,676 (35.4)		636 (34.9)	2,506 (34.4)		783 (35.8)	3,170 (36.2)	
Obese, BMI > 30	1,323 (33.0)	6,985 (43.5)		674 (37.0)	3,284 (45.0)		649 (29.7)	3,701 (42.3)	
HPV status ^c									
Positive	—	—	—	793 (43.5)	—	—	—	—	—
Negative	—	—	—	209 (11.5)	—	—	—	—	—
Unknown	—	—	—	821 (45.0)	—	—	—	—	—

Note: Numbers do not add up to the column totals due to missing values. Control subjects were matched 4:1 to case subjects.

^aP value from the two-sided chi-square test (for categorical variables) and Student *t* test (for continuous variables).

^bMVP cohort used HARE to partition a multiethnic cohort into three nonoverlapping strata: non-Hispanic whites, Hispanic-Americans, and African-Americans. This enables most individuals to be included in the individual analyses, regardless of whether self-identified race/ethnicity is available.

^cHPV status was EMR confirmable for 55% of the total OPC cohort, all diagnosed from 2010 onward when HPV/P16 testing was first available in VA hospitals; all OPCs diagnosed pre-2010 or 2010 onward but not tested were classified as unknown.

rs10820605, rs966390, and rs7041529. A significant association was detected in haplotype tests of *SMC2* variants ($P_{\text{global}} = 7.2 \times 10^{-6}$, **Table 3**)

Because of the relatively small size of the African-Americans (549 cases vs. 2,196 controls), we relaxed the significance threshold to P value 10^{-4} and identified six loci for overall HNSCC risk (**Table 2; Fig. 1B**). The strongest association was 17p13.1 Macrophage Antigen (*CD68*) p.Q254K ($P = 1.1 \times 10^{-4}$). Other significant loci identified were 6p21.32 *HLA* class I gene, *HLA-G* p.H117H (OR, 0.70; 95% CI, 0.57–0.86; additional *HLA-G* candidates includes rs1736959 and rs1610707), 9q21.33 growth arrest-specific 1 (*GAS1*) rs111548894 (OR, 4.48; 95% CI, 2.03–9.89), 11q12.2 T-cell differentiation antigen (*CD6*) rs72928596 (OR, 2.81; 95% CI, 1.59–4.97), 11q23.2 neural cell adhesion molecule 1 (*NCAM1*, also called *CD56*) rs17510855 (OR, 4.99; 95% CI, 2.11–11.80), and 18q22.2 *SOCS6* rs11665533. Of these seven loci, three regions (6p21.32, 11q12.2, and 18q22.2) were previously reported in HNSCC-GWAS in European and Chinese populations (8–10).

Risk of OPC and HPV-positive OPC in MVP-HNSCC study

As shown in the *HLA* regional association plot (**Fig. 1C**), a total of 75 variants across *HLA* class II genes showed a strong association for OPC risk, with eight top-ranked hits displaying association $P < 5 \times 10^{-7}$. The most significant three were seen in *HLA-DQB1* rs3135006, *HLA-DRB1* rs28724008, and *HLA-DRB5* rs111834747 ($P = 2.9 \times 10^{-8}$, 3.4×10^{-8} , and 3.7×10^{-7} , respectively; **Table 2**). In addition, we identified a new locus 2p21 Protein Kinase C Epsilon (*PRKCE*)

rs2711286 ($P = 2.3 \times 10^{-5}$) and an HPV-driven cervical cancer GWAS locus, 4q12 proto-oncogene tyrosine-protein Kinase (*CKIT*, also called *CD117*) rs2646357 ($P = 2.3 \times 10^{-5}$; **Table 2**).

As the *HLA* region is characterized by a complex LD pattern and a considerable number of variants showing strong associations, we subsequently assessed the *HLA-DQA1*, *HLA-DRA*, and *HLA* class II (eight lead SNPs) haplotype associations with OPC. The global score test showed significant differences for haplotype effects in *HLA-DQA1*, *HLA-DRA*, and *HLA* class II ($P = 1.7 \times 10^{-5}$, 0.0077, and 6.5×10^{-4} , respectively; **Table 3**). For haplotypes derived from the eight lead SNPs cross *HLA* class II, three risk haplotypes that are carrying 2, 3, 4, and 5 at-risk alleles, also confer a steadily increased risk in OPC (ORs from 1.29, 1.32, 1.44–1.53).

When focusing on the HPV-positive OPC subgroup ($n = 793$), our analysis revealed an even stronger association in the *HLA* class-II genes. In particular, the OR for carriers of the risk alleles of *HLA-DRA* rs3135394 and *HLA-DQB2* was 3.8- and 2.2-fold greater than non-carriers (**Table 2**). We also identified three additional loci including 11q24 MHC-restricted cytotoxic and regulatory T-cell molecule (*CRTAM/CD355*) rs117421411, 16q21 vascular endothelial cadherin (*CDH5*, also called *CD144*) rs77789388, and 18q12.1 N-cadherin (*CDH2*, also called *CD325*) rs9965078.

Risk of non-OPC HNSCC in MVP-HNSCC study

On non-OPC subgroup stratification (2,189 cases), the association at 1p36.32 Tumor Protein 73 (*TP73*, member of *TP53* family) rs1122723 and 16q24.3 *FANCA* rs12931267 retained high significance

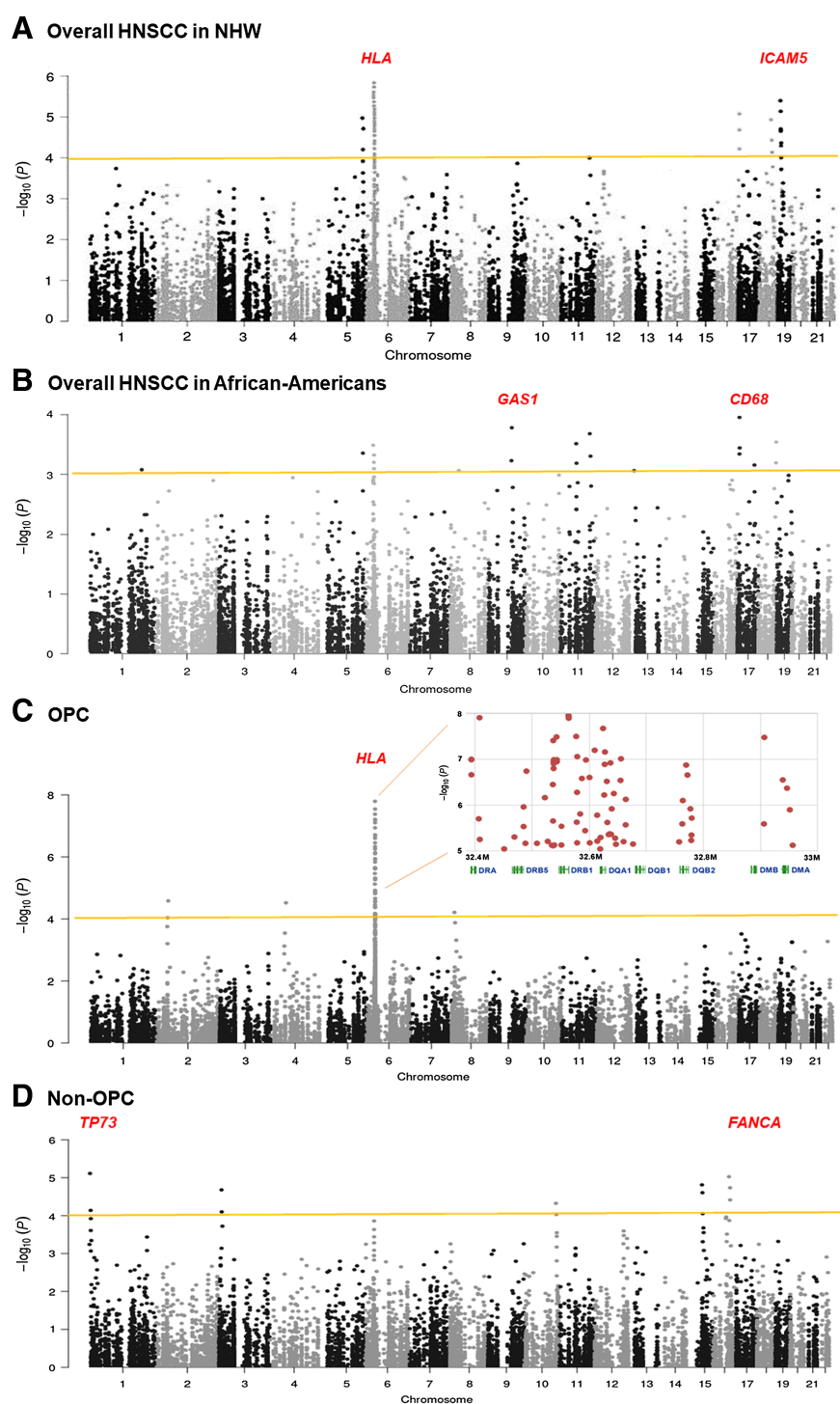


Figure 1.

Manhattan plots of $-\text{Log}_{10}(P)$ versus chromosomal position of the association results of immune gene variants in the MVP-HNSCC study. A total of 1,576 immune-related genes (16,050 common and low-frequency variants) were evaluated in this analysis. **A**, Risk of overall HNSCC in NHW: 3,266 cases versus 13,064 controls. **B**, Risk of overall HNSCC in African-Americans: 549 cases versus 2,196 controls. **C**, Risk of OPC and *HLA* regional plot: 1,823 OPC cases versus 7,292 controls. **D**, Risk of non-OPC: 2,189 cases versus 8,756 controls. The y-axis shows the $-\text{log}_{10} P$ values and the x-axis shows the chromosomal positions. The dashed horizontal lines represent the study-wide significant threshold of $P = 5 \times 10^{-4}$. The *HLA* association showed consistent effects on the overall HNSCC, OPC, and non-OPC analyses. The association results were based on the discovery study.

($P = 1.2 \times 10^{-6}$ and 9.1×10^{-6} , respectively; Table 2), whereas 6p21.32 *HLA* association was attenuated and with borderline significance (Fig. 1C). Moreover, we identified three new loci including 3p25.3 IL1 receptor associated-kinase 2 (*IRAK2*) rs6442161, 10q26.13 deleted in malignant brain tumors 1 (*DMBT1*) rs17103659, and 15q22.2 tropomyosin 1 (*TPM1*) rs72743223 ($P = 3.4 \times 10^{-5}$, 4.8×10^{-5} , and 4.2×10^{-5} , respectively).

Stratified analyses by smoking, alcohol abuse, and BMI in MVP-HNSCC study

As shown in Table 4, four loci were significantly enriched in smokers: 2q33.1 apoptosis-related cysteine peptidase (*CASP8*) p.K14R, *HLA-DOA* rs3135333, 11q13.1 NF- κ B transcription factor p65 (also called *RELA*) rs7115734, and *FANCA* rs12931267 ($P = 2.8 \times 10^{-9}$, 8.3×10^{-6} , 3.3×10^{-5} , and 5.6×10^{-9} ,

Table 2. Association results of leading SNPs in overall HNSCC, OPC, and non-OPC in MVP Discovery and MDAC Validation Cohort.

Chr. loci*	Gene	rs ID [#]	Genic location ^a Tissue-specific eQTL	MAF		Discovery		Validation		Meta-analysis	
				Case	Control	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
Overall HNSCC in NHWS											
5q35.1	FGF18	rs67585403	intergenic	0.23	0.24	0.90 (0.86-0.94)	4.6 × 10 ⁻⁵	0.99 (0.89-1.10)	0.7658	0.95 (0.84-1.05)	0.0098
6p21.32 ^{+h,c}	HLA-DRB5	rs11834747	promoter	0.37	0.29	1.35 (1.22-1.50)	5.5 × 10 ⁻⁷	1.19 (1.11-1.27)	2.5 × 10 ⁻⁷	1.28 (1.12-1.47)	2.6 × 10⁻⁹
	HLA-DRB1	rs28724008	intron	0.37	0.30	1.23 (1.13-1.33)	1.3 × 10 ⁻⁶	1.15 (1.08-1.23)	5.2 × 10 ⁻⁵	1.19 (1.06-1.32)	7.4 × 10⁻⁸
	HLA-DQB1	rs6928482	downstream, eQTL ^{-1,s}	0.32	0.29	1.24 (1.13-1.35)	2.2 × 10 ⁻⁶	1.24 (1.13-1.36)	4.5 × 10 ⁻⁶	1.24 (1.10-1.39)	2.2 × 10⁻⁸
9q31.1	SMC2	rs3818625	promoter, eQTL ^{-s}	0.43	0.44	0.92 (0.88-0.95)	4.1 × 10 ⁻⁵	0.91 (0.82-1.04)	0.1619	0.90 (0.82-1.02)	4.1 × 10⁻⁴
16q24.3	FANCA	rs12931267	intron, eQTL ^{-1,s}	0.08	0.07	1.17 (1.09-1.30)	2.3 × 10 ⁻⁵	0.99 (0.83-1.11)	0.7512	1.14 (0.98-1.28)	0.0099
19p13.2 ^{+c}	ICAM5	rs11575074	downstream	0.06	0.05	1.21 (1.10-1.31)	8.4 × 10 ⁻⁶	1.06 (0.95-1.18)	0.3757	1.16 (1.01-1.29)	0.0045
Overall HNSCC in African-Americans											
6p21.32 ^{+h,c}	HLA-G	rs1130356	p.H117H, eQTL ^{-1,s}	0.23	0.28	0.70 (0.57-0.86)	4.8 × 10 ⁻⁴	—	—	—	—
9q21.33	GAS1	rs111548894	promoter	0.02	0.01	4.48 (2.03-9.89)	2.1 × 10 ⁻⁴	—	—	—	—
11q12.2 ^{+h}	CD6	rs72928596	promoter	0.03	0.01	2.81 (1.59-4.97)	3.8 × 10 ⁻⁴	—	—	—	—
11q23.2	NCAM1/CD56	rs17510855	enhancer	0.01	0.00	4.99 (2.11-11.8)	2.6 × 10 ⁻⁴	—	—	—	—
17p13.1	CD68	rs9901673	p.Q254K, eQTL ^{-s}	0.14	0.19	0.61 (0.48-0.79)	1.1 × 10 ⁻⁴	—	—	—	—
18q22.2 ^{+h,c}	SOC56	rs11665533	promoter	0.54	0.49	1.36 (1.15-1.60)	3.6 × 10 ⁻⁴	—	—	—	—
OPC in all races/ethnicities											
2p21	PRKCE	rs2711286	intron	0.21	0.23	0.81 (0.72-0.91)	2.3 × 10 ⁻⁵	0.94 (0.83-1.07)	0.3489	0.83 (0.70-1.04)	0.0053
4q12 ^{+c}	CKIT/CD117	rs2646357	3' flanking	0.48	0.44	1.19 (1.08-1.30)	2.3 × 10 ⁻⁵	0.97 (0.89-1.07)	0.5762	1.11 (0.96-1.27)	0.0061
6p21.32 ^{+h,c}	HLA-DRB5	rs11834747	promoter	0.37	0.29	1.35 (1.20-1.52)	3.7 × 10 ⁻⁷	1.33 (1.22-1.46)	3.3 × 10 ⁻¹⁰	1.35 (1.20-1.55)	9.7 × 10⁻¹¹
	HLA-DRB1	rs28724008	intron	0.38	0.30	1.38 (1.23-1.55)	3.4 × 10 ⁻⁸	1.35 (1.22-1.48)	9.8 × 10 ⁻¹⁰	1.37 (1.22-1.58)	2.8 × 10⁻¹¹
	HLA-DQB1	rs3135006	intergenic, eQTL ^{-1,s}	0.09	0.12	0.70 (0.61-0.82)	2.9 × 10 ⁻⁸	0.71 (0.64-0.78)	4.6 × 10 ⁻¹¹	0.70 (0.60-0.83)	5.1 × 10⁻¹²
Non-OPC in all races/ethnicities											
1p36.32 ^{+h,c}	TP73	rs1122723	intron	0.40	0.44	0.84 (0.76-0.92)	9.1 × 10 ⁻⁶	0.93 (0.83-1.03)	0.2061	0.86 (0.75-1.03)	0.0031
3p25.3	IRAK2	rs6442161	intron, eQTL ^{-1,s}	0.40	0.37	1.18 (1.08-1.29)	3.4 × 10 ⁻⁵	1.09 (0.97-1.21)	0.1092	1.15 (1.00-1.29)	0.0022
10q26 ^{+h}	DMBT1	rs17103659	intergenic	0.12	0.16	0.81 (0.72-0.91)	4.8 × 10 ⁻⁵	0.81 (0.66-0.98)	0.0393	0.80 (0.64-0.98)	1.3 × 10⁻⁴
15q22.2	TPM1	rs72743223	intron	0.10	0.09	1.31 (1.13-1.52)	4.2 × 10 ⁻⁵	1.17 (1.00-1.36)	0.0534	1.26 (1.01-1.48)	2.0 × 10⁻⁴
16q24.3	FANCA	rs12931267	intron, eQTL ^{-1,s}	0.08	0.07	1.21 (1.11-1.32)	1.2 × 10 ⁻⁵	1.10 (0.85-1.40)	0.4726	1.15 (0.98-1.39)	0.0062

Note: *, Known GWAS loci for HNSCC as ^h, ⁺, ⁺, known loci for cervical cancer as ^c.

Abbreviations: ID, RefSNP identification number; UTR, untranslated region; TSS, transcriptional start site.

^aSNP functional regulatory features (transcriptional enhancer, promoter, transcriptional start site) were predicted by the Encyclopedia of DNA Elements (ENCODE). The eQTLs were retrieved from the GTEx catalog, based on RNA-seq analysis for normal lung (eQTL⁻¹) and squamous esophagus (eQTL⁻⁵) tissues, which best represent oropharyngeal and oral cavity tissue.

^bIn the discovery (MVP-HNSCC study), the regression models and P values were adjusted for age, sex, genetically Harmonized Ancestry and Race/Ethnicity (HARE), smoking status, alcohol abuse, and BMI category. In the validation (MDACC-HNSCC study), only age, sex, and the principal components (eigenvectors) were adjusted. Variants remaining significant (P_{meta} ≤ 4.95 × 10⁻⁴) in the meta-analysis were bolded.

Table 3. Haplotypes and associations in the discovery MVP-HNSCC study.

Gene	Haplotype ^a	Case (%)	Control (%)	OR (95% CI) ^b	P value ^b	Global score test P value ^b
SMC2 haplotype and overall HNSCC risk in NHWs						7.2×10 ⁻⁶
rs3818625, rs10820599	GTCAG	0.52	0.50	1.00 (reference)		
rs10820605, rs966390	AGTGA	0.40	0.41	0.92 (0.88-0.97)	5.14×10 ⁻⁴	
rs7041529	Others	0.08	0.09	0.89 (0.76-1.01)	0.015	
FANCA haplotype and overall HNSCC risk in NHWs						1.4×10 ⁻¹¹
rs17233497	AGG	0.79	0.81	1.00 (reference)		
rs12931267	GGG	0.11	0.10	1.16 (1.09-1.22)	5.1×10 ⁻⁶	
rs17232910	ACC	0.08	0.07	1.20 (1.13-1.26)	3.5×10 ⁻⁹	
HLA-DQA1 haplotype and OPC risk in all races/ethnicities						1.7×10 ⁻⁵
rs17843619	GCC	0.62	0.63	1.00 (reference)		
rs28584179	GAC	0.25	0.24	1.22 (0.97-1.46)	0.0425	
rs6928482	TCT	0.09	0.07	1.36 (1.18-1.57)	4.4×10 ⁻⁴	
	TAT	0.07	0.05	1.43 (1.13-1.71)	0.0293	
HLA-DRA haplotype and OPC risk in all races/ethnicities						0.0077
	CCGGAC	0.50	0.51	1.00 (reference)		
rs2395166, rs3135352	TCGGAC	0.26	0.24	1.05 (0.99-1.1)	0.0957	
rs3129843, rs3135394	TAGGGC	0.12	0.12	1.15 (1.07-1.23)	4.7×10 ⁻⁵	
rs3135388, rs3763326	CCAAAC	0.09	0.09	1.21 (1.12-1.3)	4.8×10 ⁻⁷	
	others	0.03	0.04	0.92 (0.81-1.04)	0.178	
HLA class II haplotype and OPC risk in all races/ethnicities ^c						6.5×10 ⁻⁴
DRA rs3135394	GCGCTTAG	0.38	0.39	1.00 (reference)		
DRB5 rs111834747	AGGCTTTG	0.25	0.24	1.29 (0.99-1.54)	0.0452	
DRB1 rs3208409, rs28724008	GGGCTTTG	0.08	0.07	1.32 (1.02-1.65)	0.0443	
DQB1 rs3135006, rs6928482	GGGCTTTA	0.11	0.09	1.44 (1.17-1.78)	3.7×10 ⁻⁵	
DQB2 rs1383264	AGACCCAA	0.08	0.07	1.53 (1.21-1.85)	2.8×10 ⁻⁷	
DMB rs171329	Others ^a	0.05	0.08	0.87 (0.79-0.96)	0.0184	

^aHaplotypes with a frequency <0.03 were pooled into a combined group.

^bRegression models were adjusted for age, sex, genetically HARE, smoking status, alcohol abuse, and BMI.

^cThe HLA class II haplotypes were derived from eight variants displaying association $P < 5 \times 10^{-7}$ (Fig. 1C, regional plot).

respectively). Five loci enriched in patients with alcohol abuse: 1p31.1 leucine-rich repeat-containing 7 (*LRRC7*) rs954302, *CASP8* p.K14R, 5q33.1 HLA class II antigen γ chain (*CD74*, also called *HLA-DG*) rs79078220, *SMC2* rs3818625, *FANCA* rs12931267 ($P = 3.7 \times 10^{-4}$, 5.6×10^{-7} , 2.8×10^{-4} , 7.9×10^{-5} , and 1.7×10^{-6} , respectively). Five loci were evident in the overweight and obese patients, including *CASP8* p.K14R, *HLA-G* p.H117H, 8p23.3 myomesin-2 (*MYOM2*) rs17064865, 10p11.21 partitioning defective protein 3 (*PARD3*) rs9418121 and *FANCA* rs12931267 ($P = 8.8 \times 10^{-9}$, 5.2×10^{-6} , 1.1×10^{-5} , 8.3×10^{-5} , and 1.2×10^{-7} , respectively). Notably, *FANCA* rs12931267 and *CASP8* p.K14R were significantly enriched in smokers, alcohol-abuser, and overweight/obese patients.

Functional annotation, eQTL evidence, biological pathways, and networks

Among the 35 top-ranked noteworthy SNPs, we discovered (listed in Tables 2 and 4), 14 (40%) were predicted to reside within regulatory elements (promoters, enhancers, and transcription start site) by ENCODE and as strong eQTLs in lung and/or squamous esophagus (P value <0.001 in the GTEx catalog) that influence the expression of target genes.

As shown in Supplementary Table S2, GO enrichment and KEGG pathway analysis highlighted (i) overall HNSCC risk-associated genes were enriched in immunoglobulin complex, (ii) OPC-associated genes were involved in antigen presentation and signaling, (iii) non-OPC-associated susceptibility genes were enriched in the cellular response to stress, and (iv) risk-factor interacted genes were involved in the regulation of T-cell activation.

Figure 2 illustrates the gene interaction networks, either as experimentally determined or database curated (high confidence protein-protein interaction score ≥ 0.90). The networks consist of 18 genes in four distinct clusters: HLA, NF- κ B, TGF β receptor, and cKIT signaling. Detailed pair-wise interaction scores are listed in Supplementary Table S3.

MDACC validation and meta-analyses

To validate our findings from the discovery dataset, we performed replication analysis for associations of overall HNSCC in NHW patients with OPC or non-OPC, respectively. The MDACC study does not have African-American participants and HPV status is not available. There are 101 variants (38 for overall HNSCC, 37 for OPC, and 26 for non-OPC) available in the MDACC study for testing. Of these 101 variants, 90 had a consistent direction of effect, 35 showed at least nominal significance ($P < 0.05$), 18 remained significant after correction for multiple comparisons ($P \leq 4.95 \times 10^{-4}$), and nine HLA variants reached genome-wide significance level ($P < 5 \times 10^{-8}$).

Table 2 summarizes the top-ranked candidates with consistent associations from the random-effect meta-analysis. For HNSCC overall, two loci (six variants) were validated, namely 6p21.33 HLA ($P = 2.6 \times 10^{-9}$, 7.4×10^{-8} , and 2.2×10^{-8} for HLA-DRB5 rs111834747, HLA-DRB1 rs28724008, and HLA-DQB1 rs6928482, respectively), and 9q31.1 SMC2 rs3818625 ($P = 4.1 \times 10^{-4}$). For OPC, the HLA class II association was even more evident, with nine HLA variants reaching significance $P < 5 \times 10^{-4}$. Specifically, the three leading associations from the discovery were HLA-DRB1 rs28724008, HLA-DRB5 rs111834747, and HLA-DQB1 rs6928482 ($P = 2.8 \times 10^{-11}$, 9.7×10^{-11} , and 5.1×10^{-12} , respectively). For risk of non-OPC, two

Table 4. Overall HNSCC associated leading SNPs in HPV-positive OPC, smokers, alcohol abuse, and overweight/obese subgroups in the discovery MVP-HNSCC study.

Loci*	Gene	rs ID#	Genic location ^a tissue-specific eQTL	MAF		Logistic regression ^b	
				Case	Control	OR (95% CI)	P value
HPV-positive OPC							
6p21.32 ^{*h+c}	<i>HLA-DRA</i>	rs3135394	promoter, eQTL ^{-1s}	0.11	0.07	3.82 (2.41-6.07)	1.3×10 ⁻⁸
	<i>HLA-DRB1</i>	rs3208409	3' UTR	0.41	0.38	1.54 (1.31-1.82)	3.1×10 ⁻⁷
	<i>HLA-DQB2</i>	rs1383264	promoter, eQTL ^{-1s}	0.42	0.36	2.23 (1.60-3.11)	2.4×10 ⁻⁷
	<i>HLA-DMB</i>	rs171329	3 bp to TSS, eQTL ⁻⁵	0.55	0.48	1.48 (1.29-1.69)	1.2×10 ⁻⁸
11q24	<i>CRTAM</i>	rs11742141	enhancer	0.06	0.04	1.95 (1.41-2.69)	5.0×10 ⁻⁵
16q21	<i>CDH5</i>	rs77789388	promoter	0.20	0.15	1.40 (1.19-1.66)	6.3×10 ⁻⁵
18q12.1	<i>CDH2</i>	rs9965078	intergenic	0.12	0.08	1.54 (1.25-1.90)	6.2×10 ⁻⁵
Smokers							
2q33.1 ^{*c}	<i>CASP8</i>	rs3769823	p.K14R, eQTL ^{-1s}	0.29	0.28	1.14 (1.09-1.19)	2.8×10 ⁻⁹
6p21.32 ^{*h+c}	<i>HLA-DOA</i>	rs3135333	intergenic	0.28	0.27	1.11 (1.06-1.16)	8.3×10 ⁻⁶
	<i>NFkB-p65/RELA</i>	rs7115734	upstream, eQTL ^{-1s}	0.32	0.34	0.91 (0.88-0.95)	3.3×10 ⁻⁵
16q24.3 ^{*c}	<i>FANCA</i>	rs12931267	intron, eQTL ^{-1s}	0.08	0.07	1.24 (1.15-1.33)	5.6×10 ⁻⁹
Alcohol abuse							
1p31.1	<i>LRRC7</i>	rs954302	intergenic	0.17	0.20	0.81 (0.72-0.91)	3.7×10 ⁻⁴
2q33.1 ^{*c}	<i>CASP8</i>	rs3769823	p.K14R, eQTL ^{-1s}	0.33	0.31	1.12 (1.07-1.16)	5.6×10 ⁻⁷
5q33.1	<i>CD74/HLA-DG</i>	rs79078220	upstream	0.11	0.09	1.32 (1.14-1.53)	2.8×10 ⁻⁴
9q31.1	<i>SMC2</i>	rs3818625	promoter, eQTL ⁻⁵	0.42	0.44	0.91 (0.87-0.96)	7.9×10 ⁻⁵
16q24.3 ^{*c}	<i>FANCA</i>	rs12931267	intron, eQTL ^{-1s}	0.09	0.07	1.20 (1.11-1.29)	1.7×10 ⁻⁶
Overweight/obese							
2q33.1 ^{*c}	<i>CASP8</i>	rs3769823	p.K14R, eQTL ^{-1s}	0.30	0.27	1.14 (1.09-1.19)	8.8×10 ⁻⁹
6p21.32 ^{*h+c}	<i>HLA-G</i>	rs1130356	p.H117H, eQTL ^{-1s}	0.31	0.28	1.19 (1.09-1.28)	5.2×10 ⁻⁶
	<i>MYOM2</i>	rs17064865	promoter	0.10	0.09	1.16 (1.09-1.24)	1.1×10 ⁻⁵
10p11.21	<i>PARD3</i>	rs9418121	promoter, eQTL ⁻⁵	0.24	0.26	0.91 (0.87-0.95)	8.3×10 ⁻⁵
16q24.3 ^{*c}	<i>FANCA</i>	rs12931267	intron, eQTL ^{-1s}	0.08	0.07	1.22 (1.13-1.31)	1.2×10 ⁻⁷

Note: *, Known GWAS loci for HNSCC as ^{*h}; **, known loci for cervical cancer as ^{*c}.

Abbreviation: ID, RefSNP identification number.

^aSNP functional regulatory features (enhancer, promoter) were predicted by the ENCODE. The eQTLs were retrieved from the GTEx catalog, based on RNA-seq analysis for normal lung (eQTL⁻¹) and squamous esophagus (eQTL⁻⁵) tissues, which best represent oropharyngeal and oral cavity tissue.

^bRegression models were adjusted for age, sex, genetically HARE, smoking status, alcohol abuse, and BMI.

novel loci were replicated: 10q26.13 *DMBT1* rs17103659 and 15q22.2 *TPM1* rs72743223 ($P = 1.3 \times 10^{-4}$ and 2.0×10^{-4} , respectively).

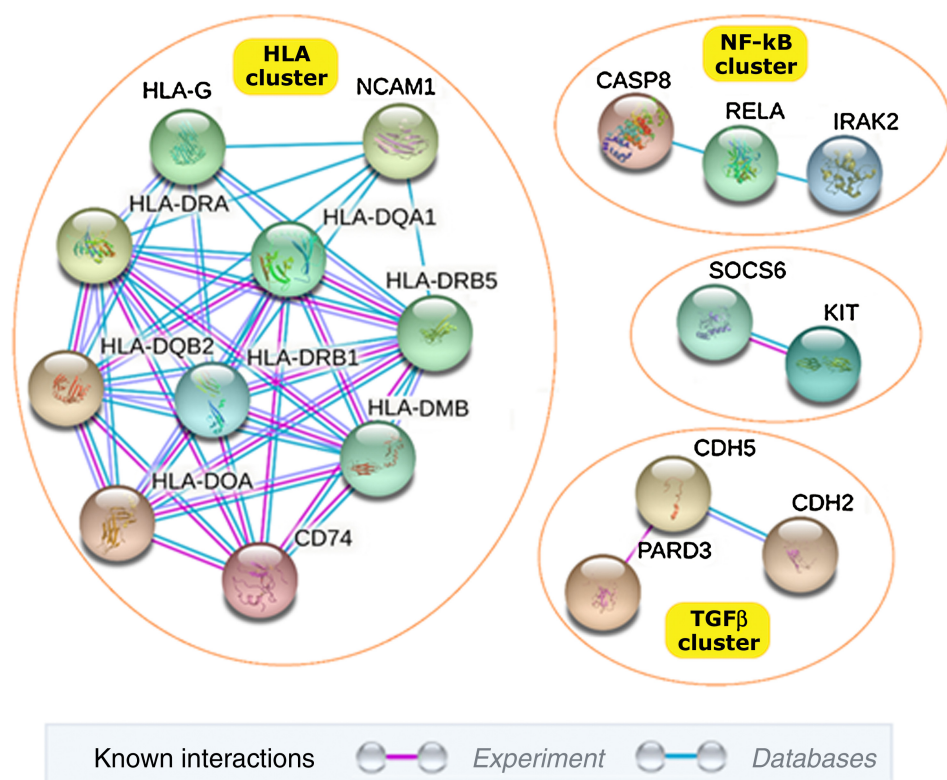
Discussion

We present the first HNSCC association study utilizing the large, multiethnic MVP cohort and an independent validation study. We identified and validated two loci for overall HNSCC susceptibility, including the novel 9q31.1 *SMC2* and the known 6p21.32 *HLA*. We uncovered novel loci selectively enriched in African-Americans for overall HNSCC risk and reveal ethnic/racial susceptibility differences. Our large sample sizes allowed us to conduct well-powered analyses for distinct HNSCC sites and revealed striking and distinct association signatures across HPV-positive OPC and non-OPC. Although clinical studies of tumor biology and preclinical models of HPV-associated disease have long suggested distinct immune features of this virally mediated malignancy, the data summarized here provide critical complementary support for this hypothesis and have significant implications for both cancer prevention and cancer therapeutics.

We have provided insight into the interaction between gene variants and risk factors in the risk of overall HNSCC. Gene variants were identified in both immune-related genes (the primary focus of the study) including HLAs, *CD68*, *NCAM1/CD56*, *CD6*, and genes associated with aggressive tumor behavior, which may be unrelated to antitumor or systemic immunity including

FANCA, *FGF18*, and *SMC2*. Taken together, our data support an integrated cancer development model, in which both altered immune status and tumor cell-intrinsic mechanisms (e.g., cellular dedifferentiation, proliferation, and invasion) lead to increased risk of HNSCC overall and risk of HPV-associated disease more specifically. Overall, 40% of candidate variants are tissue-specific eQTLs within the lung and squamous esophagus tissues, providing additional biological evidence (regulating gene expression) for putative susceptibility loci.

A major finding in this study was the consistent association in both discovery and validation analyses of overall HNSCC susceptibility with a novel locus, 9q31.1 *SMC2*, and known locus 6p21.32 *HLA*. *SMC2* plays a dual role not only in tumor growth but also in invasion and tumor metastasis (15, 16). The 9q31.1 region, where *SMC2* resides, is a known risk GWAS locus for pancreatic cancer (17–19). There have been no prior reports of this locus on HNSCC susceptibility. In agreement with published HNSCC-GWAS (8–12), our result further reinforces the central functional role of *HLA* molecules in the presentation of foreign antigens, immune recognition, and subsequent serological responsiveness to HNSCC (20, 21). The strong effect of the *HLA* region is also commonly seen in GWAS of HPV-driven cervical cancer and autoimmune disorders (22–25). We speculate that a possible underlying mechanism may be related to risk-associated *HLA* variations that affect the antigen-presenting environment. Specifically, carriers of specific *HLA* alleles may have loss or downregulation of

**Figure 2.**

Protein-protein interaction network of candidate genes. The protein-protein interaction analysis demonstrates the high degree of significant interaction (high-confidence score ≥ 0.90) among the top candidate immune genes associated with overall HNSCC in NHW and African-Americans, OPC, non-OPC, and risk factors (i.e., HPV-positive OPC, smoking, overweight). These interactions were either experimentally determined (fuchsia line) or database curated (blue line). The interaction networks consist of 18 genes in four distinct clusters: (i) HLA cluster: eight *HLA* genes and two antigen-related genes, *CD74/HLA-DG* and *NCAM1/CD56*; (ii) NF- κ B cluster: *CASP8* - *NFKB-p65* - *IRAK2*; (iii) TGF β cluster: *PARD3* - *CDH5* - *CDH2*; and (iv) c-KIT signaling: *SOCS6* - *cKIT*. The figure was generated using the STRING. The detailed pairwise interaction score is listed in Supplementary Table S3.

MHC molecules expression, impaired or reduced viral antigen-presenting capacity or binding affinity, thus affecting the clearance of viral infections and activation of antiviral immune responses, and subsequently the susceptibility to HPV-related cancer.

African-Americans have been understudied in genetic association studies of HNSCC. Several candidate genes-based association studies have been performed in African populations but were underpowered (with sample sizes ranging from 137 to 392 for HNSCC cases; refs. 26–31); previous HNSCC-GWAS were largely based on homogeneous ethnic backgrounds, either European ancestry or East Asian populations (8–12). Our study is among the first to examine the overall HNSCC susceptibility and immunogenic associations in African-Americans. Our subgroup analysis suggests that immune gene variants profiles in HNSCC susceptibility vary across racial/ethnic groups. One novel finding is that the association of a nonclassical HLA-class I molecule, *HLA-G* was specific to African-Americans, whereas the association of class-II *HLA-DQ-DR* genes was exclusively observed in NHW. *HLA-G* has been termed “nonclassical” due to its low frequency of polymorphisms and immunoinhibitory properties, which are different from the properties of classical HLA-class I molecules. It has become increasingly evident that *HLA-G* is involved in modulating immune responses and in promoting immune escape in various types of cancers and infectious diseases (32–34). Accumulating evidence has supported the concept that *HLA-G* polymorphisms are genetic susceptibility and/or protection-relevant factors for cervical HPV infections and viral persistence (35–38). A SNP (rs1633038) in the 3'UTR of *HLA-G* gene was significantly related to higher HPV clearance rates among African-Americans with HIV/HPV co-infection, but this association was not observed in Hispanics or European Americans (35). Other interesting novel loci include 9q21.33 *GAS1* (a tumor suppressor), 11q23.2 *NCAM1/CD56* (immunoglobulin

superfamily), and 17p13.1 *CD68* (Macrophage Antigen). These markers have not been noted in HNSCC-GWAS of European and Asian populations, but more research on a larger African-American population is needed.

HPV infection drives most new OPC (>70%) diagnoses in the United States, and HPV-positive OPC exhibits distinct genetic and immune profiles from HPV-negative HNSCC. Our findings suggest that the hypothesized and observed distinct immune profile of this virally mediated malignancy may reflect not only the impact of the virus itself, but intrinsic host immune characteristics, which may facilitate infection, subsequent malignant transformation, or both. Our findings were consistent with previous OPC-GWAS (10, 12), which underscore the critical role of *HLA* class II variants and haplotypes in HPV-driven OPC susceptibility. In addition, we uncovered three novel loci (*CRTAM/CD355*, *CDH2*, and *CDH5*). We have previously identified *CRTAM* as contributing to the risk of HPV-positive OPC and cervical cancer (39). Intriguingly, *CRTAM* promotes protective immunity against viral infection and appears to protect from autoimmunity (40, 41). Dysregulation and genomics alterations of cadherin genes *CDH2* and *CDH5* were reported in HPV-driven OPC and cervical cancer (42, 43).

Another interesting finding is the two novel non-OPC susceptibility loci: 10q26.13 *DMBT1* and 15q22.2 *TPM1*. Of particular significance, *DMBT1* expression was recently shown 3-fold higher in HNSCC who were ever smokers versus never smokers. *DMBT1* is known as a pattern recognition receptor in innate immunity involved in infection, inflammation, and cancer (44, 45). Another novel locus, *TPM1* is a crucial tumor-suppressing gene in many solid tumors including HNSCC (46). It has been demonstrated that reduced *TPM1* expression correlates with poor prognosis in non-OPC patients (46). Further studies are warranted for in-depth

functional characterization of these novel loci for non-OPC HNSCC susceptibility and tumorigenesis.

The stratified analysis revealed the association of *FANCA* and *CASP8* is most prominent in patients who are current/former smokers, with heavy alcohol use and overweight/obese. Reports suggest germline variants in both genes have been associated with the risk of multiple cancer sites (i.e., pleiotropy). *FANCA* is a key member of the Fanconi anemia (FA)/BRCA pathway including the core FA genes and *BRCA2* (*FANCD1*). *FANCA* mutations account for the majority (~66%) of FA disorder cases, which is associated with a 500- to 700-fold increased risk of developing HNSCC, especially in the oral cavity (i.e., non-OPC; ref. 47). Data from TCGA-HNSCC cohort confirmed the correlation of genomic gains on 16q24.3 with *FANCA* overexpression and reduced progression-free survival in HPV-negative HNSCC (48).

Despite the strengths and biological plausibility of the associations observed in our study, there are limitations inherent in our study. First, the demographics of the MVP cohort diverge from the general U.S. population. We repeated the analyses for the top associated variants in males (98.7%) and found similar association results. Second, many of the novel variants that we did not validate in the MDACC-HNSCC study are probably due to different ethnic backgrounds (multiethnic vs. NHW only), sex distribution, and other population-specific differences (i.e., higher rates of tobacco use and alcohol abuse in the veteran population). In addition, HPV status is not available for the MDACC-HNSCC cohort, because testing was not part of the standard of care during that time. As such, we can only infer the relative proportion of HPV-positive/negative patients from general population statistics that are unreliable, particularly, as the relative proportion of HPV-positive is now known to have been rapidly changing. Because much of the immune signal observed is driven by HPV-positive disease, a differential frequency of HPV-positive disease in the validation cohort may explain why some of the loci do not achieve statistical significance. Third, in the absence of comprehensive validation and a modest sample size of the African-American population and HPV-positive OPCs, these subgroup analyses have limited statistical power. Future research will be needed to generalize our results by performing validation studies in racially diverse populations with matching gender distributions and HPV information.

In summary, our results underscore the role of immunogenic variants in modulating susceptibility to both HPV-driven and non-HPV-related HNSCC. The *HLA* association showed consistent effects on the overall HNSCC, OPC, and non-OPC risk, but these risks differ by ethnicity (*HLA-G* in African Americans vs. *HLA-DQ-DR* genes in NHW). These insights may serve as a starting point for the development of genetically-informed approaches to HNSCC screening and risk assessment in the veteran population.

Authors' Disclosures

J. Kramer reports grants from U.S. Department of Veterans Affairs during the conduct of the study. S. Pyarajan reports employment with US Federal government and VA. E.Y. Chiao reports grants from VA during the conduct of the study and from NIH outside the submitted work. Y.M. Mowery reports grants from NIDCR during the conduct of the study; grants from Damon Runyon Cancer Foundation, NIDCR, NCI, SU2C, and Radiation Oncology Institute; other support from John R. Flanagan Charitable Foundation, UpToDate, and Oakstone

CME; grants and nonfinancial support from Merck; and nonfinancial support from Bayer outside the submitted work. A.G. Sikora reports grants from Veteran's Administration during the conduct of the study and personal fees from Roche outside the submitted work. D.L. White reports grants from U.S. Department of Veteran Affairs during the conduct of the study. No disclosures were reported by the other authors.

Authors' Contributions

Y. Liu: Conceptualization, data curation, formal analysis, funding acquisition, validation, methodology, writing—original draft, writing—review and editing. **J.R. Kramer:** Conceptualization, data curation, formal analysis, funding acquisition, validation, methodology, writing—original draft, writing—review and editing. **V.C. Sandulache:** Conceptualization, data curation, formal analysis, funding acquisition, validation, methodology, writing—original draft, writing—review and editing. **R. Yu:** Resources, formal analysis, validation, writing—review and editing. **G. Li:** Resources, formal analysis, validation, writing—review and editing. **L. Chen:** Data curation, formal analysis, visualization, writing—review and editing. **Z.I. Yusuf:** Data curation, formal analysis, visualization, writing—review and editing. **Y. Shi:** Resources, data curation, software, methodology, writing—review and editing. **S. Pyarajan:** Resources, data curation, software, methodology, writing—review and editing. **S. Tsavachidis:** Formal analysis, writing—review and editing. **L. Jiao:** Conceptualization, data curation, formal analysis, funding acquisition, validation, methodology, writing—original draft, writing—review and editing. **M.L. Mierzwa:** Conceptualization, funding acquisition, writing—review and editing. **E. Chiao:** Funding acquisition, methodology, writing—review and editing. **Y.M. Mowery:** Validation, methodology, writing—review and editing. **A. Shuman:** Conceptualization, funding acquisition, writing—review and editing. **S. Shete:** Resources, formal analysis, validation, writing—review and editing. **A.G. Sikora:** Conceptualization, data curation, formal analysis, funding acquisition, validation, methodology, writing—original draft, writing—review and editing. **D.L. White:** Conceptualization, data curation, formal analysis, funding acquisition, validation, methodology, writing—original draft, writing—review and editing.

Acknowledgments

This project is supported in part by the Biomedical Laboratory Research & Development, MVP Gamma Pilot Award (1 I01 BX004420; PIs: A.G. Sikora and D.L. White), and Department of Veterans Affairs, Office of Research and Development, MVP Core (#MVP000). Dr. D.L. White also received support from VA Clinical Science Research and Development (CSR&D) Merit Review (I01CX001430) and Dr. V.C. Sandulache from VA CSR&D Career Development Award (1IK2CX001953) and an American Cancer Society Research Scholar Grant (RSG-21-182-01). This study was also supported by grants from the NIH (R03 CA262911 to Y. Liu; 5R01CA131324 and 5P30CA016672 to S. Shete; U01DE028233 to A.G. Sikora; R01CA236859 to G. Li; and 5K08-DE029887-02 to Y. Mowery). The authors would like to thank all the participants in the MVP and MDACC for donating their samples, information, and time to this project. In addition, the authors would like to thank the hard work of all the MVP staff working in the various operational domains like the biorepository, recruitment sites, VA central office, and clinicians who together work tirelessly to make MVP operational. The content of this manuscript does not represent the views of the Department of Veterans Affairs or the U.S. Government.

The publication costs of this article were defrayed in part by the payment of publication fees. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

Note

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

Received May 27, 2022; revised September 2, 2022; accepted November 10, 2022; published first November 15, 2022.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and

mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71:209–49.

2. Villa A, Hanna GJ. Human papillomavirus and oropharyngeal cancer. *Curr Probl Cancer* 2018;42:466–75.
3. Koo HY, Han K, Shin DW, Yoo JE, Cho MH, Jeon KH, et al. Alcohol drinking pattern and risk of head and neck cancer: a nationwide cohort study. *Int J Environ Res Public Health* 2021;18:11204.
4. Negri E, Boffetta P, Berthiller J, Castellsague X, Curado MP, Dal Maso L, et al. Family history of cancer: pooled analysis in the international head and neck cancer epidemiology consortium. *Int J Cancer* 2009;124:394–401.
5. Garavello W, Foschi R, Talamini R, La Vecchia C, Rossi M, Dal Maso L, et al. Family history and the risk of oral and pharyngeal cancer. *Int J Cancer* 2008;122:1827–31.
6. Grulich AE, van Leeuwen MT, Falster MO, Vajdic CM. Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. *Lancet (London, England)* 2007;370:59–67.
7. Chaturvedi AK, Madeleine MM, Biggar RJ, Engels EA. Risk of human papillomavirus-associated cancers among persons with AIDS. *J Natl Cancer Inst* 2009;101:1120–30.
8. Wei Q, Yu D, Liu M, Wang M, Zhao M, Liu M, et al. Genome-wide association study identifies three susceptibility loci for laryngeal squamous cell carcinoma in the Chinese population. *Nat Genet* 2014;46:1110–4.
9. Shete S, Liu H, Wang J, Yu R, Sturgis EM, Li G, et al. A genome-wide association study identifies two novel susceptible regions for squamous cell carcinoma of the head and neck. *Cancer Res* 2020;80:2451–60.
10. Lesseur C, Diergaarde B, Olshan AF, Wunsch-Filho V, Ness AR, Liu G, et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nat Genet* 2016;48:1544–50.
11. Lesseur C, Ferreiro-Iglesias A, McKay JD, Bosse Y, Johansson M, Gaborieau V, et al. Genome-wide association meta-analysis identifies pleiotropic risk loci for aerodigestive squamous cell cancers. *PLoS Genet* 2021;17:e1009254.
12. Ferreiro-Iglesias A, McKay JD, Brenner N, Virani S, Lesseur C, Gaborieau V, et al. Germline determinants of humoral immune response to HPV-16 protect against oropharyngeal cancer. *Nat Commun* 2021;12:5945.
13. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, et al. Million veteran program: a mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* 2016;70:214–23.
14. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–5.
15. Davalos V, Suarez-Lopez L, Castano J, Messent A, Abasolo I, Fernandez Y, et al. Human SMC2 protein, a core subunit of human condensin complex, is a novel transcriptional target of the WNT signaling pathway and a new therapeutic target. *J Biol Chem* 2012;287:43472–81.
16. Han YH, Wan Y, Xiong H, Sun GL. Structural maintenance of chromosomes 2 is identified as an oncogene in bladder cancer *in vitro* and *in vivo*. *Neoplasma* 2020;67:364–70.
17. Zhong J, Jermusyk A, Wu L, Hoskins JW, Collins I, Mocchi E, et al. A transcriptome-wide association study identifies novel candidate susceptibility genes for pancreatic cancer. *J Natl Cancer Inst* 2020;112:1003–12.
18. Klein AP, Wolpin BM, Risch HA, Stolzenberg-Solomon RZ, Mocchi E, Zhang M, et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat Commun* 2018;9:556.
19. Feng Y, Liu H, Duan B, Liu Z, Abbruzzese J, Walsh KM, et al. Potential functional variants in SMC2 and TP53 in the AURORA pathway genes and risk of pancreatic cancer. *Carcinogenesis* 2019;40:521–8.
20. Breitburd F, Ramoz N, Salmon J, Orth G. HLA control in the progression of human papillomavirus infections. *Semin Cancer Biol* 1996;7:359–71.
21. Peng S, Trimble C, Wu L, Pardoll D, Roden R, Hung CF, et al. HLA-DQB1*02-restricted HPV-16 E7 peptide-specific CD4+ T-cell immune responses correlate with regression of HPV-16-associated high-grade squamous intraepithelial lesions. *Clin Cancer Res* 2007;13:2479–87.
22. Leo PJ, Madeleine MM, Wang S, Schwartz SM, Newell F, Pettersson-Kymmer U, et al. Defining the genetic susceptibility to cervical neoplasia—a genome-wide association study. *PLoS Genet* 2017;13:e1006866.
23. Matsumoto K, Maeda H, Oki A, Takatsuka N, Yasugi T, Furuta R, et al. Human leukocyte antigen class II DRB1*1302 allele protects against cervical cancer: at which step of multistage carcinogenesis? *Cancer Sci* 2015;106:1448–54.
24. Ivansson EL, Magnusson JJ, Magnusson PK, Erlich HA, Gyllensten UB. MHC loci affecting cervical cancer risk: distinguishing the effects of HLA-DQB1 and non-HLA genes TNF, LTA, TAP1 and TAP2. *Genes Immun* 2008;9:613–23.
25. Chen D, Juko-Pecirep I, Hammer J, Ivansson E, Enroth S, Gustavsson I, et al. Genome-wide association study of susceptibility loci for cervical cancer. *J Natl Cancer Inst* 2013;105:624–33.
26. Moumad K, Khaali W, Benider A, Ben Ayoub W, Hamdi-Cherif M, Boualga K, et al. Joint effect of smoking and NQO1 C609T polymorphism on undifferentiated nasopharyngeal carcinoma risk in a North African population. *Mol Genet Genomic Med* 2018;6:933–40.
27. Makni L, Ben Hamda C, Al-ansari A, Souiai O, Gazouani E, Mezlini A, et al. Association of common IL-10 promoter gene variants with the susceptibility to head and neck cancer in Tunisia. *Turk J Med Sci* 2019;49:123–8.
28. Khelifi R, Chakroun A, Hamza-Chaffai A, Rebai A. Association of CYP1A1 and CYP2D6 gene polymorphisms with head and neck cancer in Tunisian patients. *Mol Biol Rep* 2014;41:2591–600.
29. Ben Nasr H, Chahed K, Bouaouina N, Chouchane L. PTGS2 (COX-2) -765 G >C functional promoter polymorphism and its association with risk and lymph node metastasis in nasopharyngeal carcinoma. *Mol Biol Rep* 2009;36:193–200.
30. Ben Chaaben A, Busson M, Douik H, Boukouaci W, Mamoghli T, Chaouch L, et al. Association of IL-12p40 +1188 A/C polymorphism with nasopharyngeal cancer risk and tumor extension. *Tissue Antigens* 2011;78:148–51.
31. Aouf S, Laribi A, Gabbouj S, Hassen E, Bouaouina N, Zakhama A, et al. Contribution of Nitric oxide synthase 3 genetic variants to nasopharyngeal carcinoma risk and progression in a Tunisian population. *Eur Arch Otorhinolaryngol* 2019;276:1231–9.
32. Wlasiuk P, Putowski M, Giannopoulos K. PD1/PD1L pathway, HLA-G and T regulatory cells as new markers of immunosuppression in cancers. *Postepy Hig Med Dosw (Online)* 2016;70:1044–58.
33. Lin A, Yan WH. Intercellular transfer of HLA-G: its potential in cancer immunology. *Clin Transl Immunology* 2019;8:e1077.
34. Amiot L, Vu N, Samson M. Immunomodulatory properties of HLA-G in infectious diseases. *J Immunol Res* 2014;2014:298569.
35. Sudenga SL, Wiener HW, King CC, Rompalo AM, Cu-Uvin S, Klein RS, et al. Dense genotyping of immune-related loci identifies variants associated with clearance of HPV among HIV-positive women in the HIV epidemiology research study (HERS). *PLoS One* 2014;9:e99109.
36. Smith MA, Tellier PP, Roger M, Coutlee F, Franco EL, Richardson H. Determinants of human papillomavirus coinfections among Montreal university students: the influence of behavioral and biologic factors. *Cancer Epidemiol Biomarkers Prev* 2014;23:812–22.
37. Alves BM, Prellwitz IM, Siqueira JD, Meyrelles AR, Bergmann A, Seuanez HN, et al. The effect of human leukocyte antigen G alleles on human papillomavirus infection and persistence in a cohort of HIV-positive pregnant women from Brazil. *Infect Genet Evol* 2015;34:339–43.
38. Xu HH, Yan WH, Lin A. The role of HLA-G in human papillomavirus infections and cervical carcinogenesis. *Front Immunol* 2020;11:1349.
39. Levovitz C, Chen D, Ivansson E, Gyllensten U, Finnigan JP, Alshawish S, et al. TGFbeta receptor 1: an immune susceptibility gene in HPV-associated cancer. *Cancer Res* 2014;74:6833–44.
40. Landin D, Ahrlund-Richter A, Mirzaie L, Mints M, Nasman A, Kolev A, et al. Immune related proteins and tumor infiltrating CD8+ lymphocytes in hypopharyngeal cancer in relation to human papillomavirus (HPV) and clinical outcome. *Head Neck* 2020;42:3206–17.
41. Fuchs YF, Sharma V, Eugster A, Kraus G, Morgenstern R, Dahl A, et al. Gene expression-based identification of antigen-responsive CD8(+) T cells on a single-cell level. *Front Immunol* 2019;10:2568.
42. Mezi S, Chiappetta C, Carletti R, Nardini A, Cortesi E, Orsi E, et al. Clinical significance of epithelial-to-mesenchymal transition in laryngeal carcinoma: its role in the different subsites. *Head Neck* 2017;39:1806–18.
43. Wang KH, Lin CJ, Liu CJ, Liu DW, Huang RL, Ding DC, et al. Global methylation silencing of clustered proto-cadherin genes in cervical cancer: serving as diagnostic markers comparable to HPV. *Cancer Med* 2015;4:43–55.
44. Malamud D, Abrams WR, Barber CA, Weissman D, Rehtanz M, Golub E. Antiviral activities in human saliva. *Adv Dent Res* 2011;23:34–37.
45. Deng H, Gao YB, Wang HF, Jin XL, Xiao JC. Expression of deleted in malignant brain tumours 1 (DMBT1) relates to the proliferation and malignant transformation of hepatic progenitor cells in hepatitis B virus-related liver diseases. *Histopathology* 2012;60:249–60.

46. Pan H, Gu L, Liu B, Li Y, Wang Y, Bai X, et al. Tropomyosin-1 acts as a potential tumor suppressor in human oral squamous cell carcinoma. *PLoS One* 2017;12: e0168900.
47. Furquim CP, Pivovar A, Amenabar JM, Bonfim C, Torres-Pereira CC. Oral cancer in fanconi anemia: review of 121 cases. *Crit Rev Oncol Hematol* 2018;125:35–40.
48. Hess J, Unger K, Orth M, Schotz U, Schuttrumpf L, Zangen V, et al. Genomic amplification of Fanconi anemia complementation group A (FancA) in head and neck squamous cell carcinoma (HNSCC): cellular mechanisms of radioresistance and clinical relevance. *Cancer Lett* 2017;386: 87–99.