



Published in final edited form as:

Cancer Res. 2020 June 15; 80(12): 2451–2460. doi:10.1158/0008-5472.CAN-19-2360.

A genome-wide association study identifies two novel susceptible regions for squamous cell carcinoma of the head and neck

Sanjay Shete^{1,2,**}, Hongliang Liu^{3,4,*}, Jian Wang^{2,*}, Robert Yu^{2,*}, Erich M. Sturgis⁵, Guojun Li⁵, Kristina R. Dahlstrom⁵, Zhensheng Liu^{3,4}, Christopher I. Amos⁶, Qingyi Wei^{3,7,**}

¹Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

²Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

³Duke Cancer Institute, Duke University Medical Center, Durham, NC, USA.

⁴Department of Medicine, Duke University School of Medicine, Durham, NC 27710, USA.

⁵Department of Head and Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

⁶The Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, 77030, TX, USA.

⁷Department of Population Health Sciences, Duke University Medical School, Durham, NC, USA.

Abstract

To identify genetic variants for risk of squamous cell carcinoma of the head and neck (SCCHN), we conducted a two-phase genome-wide association study consisting of 7,858,089 SNPs in 2,171 cases and 4,493 controls of non-Hispanic white, of which 434,839 typed and 7,423,250 imputed SNPs as the discovery. SNPs with $P < 1 \times 10^{-3}$ were further validated in the OncoArray study of oral and pharynx cancer (5,205 cases and 3,232 controls of European ancestry) from dbGaP. Meta-analysis of the discovery and replication studies identified one novel locus 6p22.1 ($P = 2.96 \times 10^{-9}$ for the leading rs259919) and two cancer susceptibility loci 6p21.32 (rs3135001, HLA-DQB1) and 6p21.33 (rs1265081, CCHCR1) associated with SCCHN risk. Further stratification by tumor site revealed four known cancer loci (5p15.33, 6p21.32, 6p21.33, and 2p23.1) associated with oral cavity cancer risk and oropharyngeal cancer risk, respectively. In addition, one novel locus 18q22.2 ($P = 2.54 \times 10^{-9}$ for the leading SNP rs142021700) was identified for hypo-pharynx and larynx cancer risk. For SNPs in those reported or novel loci, we also performed functional annotations by bioinformatics prediction and eQTL analysis. Collectively, our identification of

** Address for correspondence and reprints: Sanjay Shete, Department of Epidemiology, The University of Texas MD Anderson Cancer Center, 1400 Pressler Dr, FCT4.6002, Houston, TX 77030, USA, Phone: (713) 745-2483; Fax: (713) 563-4243; sshete@mdanderson.org and Qingyi Wei, Department of Population Health Sciences, Duke University School of Medicine, Durham, NC, Durham, NC 77030, USA, Phone: (919) 660-0562; FAX: (919) 660-0178; qingyi.wei@duke.edu.

*These authors contribute equally

Conflict of Interest: The authors declare no potential conflicts of interest.

four reported loci (2p23.1, 5p15.33, 6p21.32, and 6p21.33) and two novel loci (6p22.1 and 18q22.2) for SCCHN risk highlight the importance of HLA loci for oropharyngeal cancer risk, suggesting that immunologic mechanisms are implicated in the etiology of this subset of SCCHN.

Keywords

Head and neck cancer; Case-control; Genome-wide association study; Genetic susceptibility; Single-nucleotide polymorphism

Introduction

Squamous cell carcinoma of the head and neck (SCCHN) is the sixth most common malignancy world-wide and seventh leading cause of cancer-related deaths worldwide (1,2). In the United States, it is estimated that there will be approximately 65,410 new cases and 14,620 deaths to occur in 2019 (3). SCCHN includes cancers of the oral cavity (including the gums and tongue), pharynx and larynx with well-documented associations with exposure to tobacco and alcohol as well as infection with human papillomavirus (HPV) (4–6). However, the disease develops in only a small fraction of tobacco users, alcohol drinkers or individuals who contracted HPV (7); implying an important role of genetic susceptibility in the etiology of SCCHN (8,9). For example, genetic variants in alcohol-related genes (i.e., ADH1B and ADH7) have been reported to be associated with the risk of SCCHN and upper aerodigestive cancers (10,11). To date, there is only two published genome-wide association studies (GWASs) of SCCHN (12,13). One study was performed with 2,398 cases and 2,804 controls with Chinese ancestry and reported six loci (i.e., 5q14.3, 6p21.33, 6q16.1, 11q12.2, 12q24.21 and 16p13.2) associated with the laryngeal cancer risk (12). Another study investigated the genetic susceptibility of oral cavity and pharyngeal cancer with 6,034 cases and 6,585 controls of European ancestry, which reported three loci (i.e., 6p21.32, 10q26.13 and 11p15.4) associated with the overall cancer risk, four loci (i.e., 2p23.3, 5p15.33, 9p15.3 and 9q34.12) contribute to oral cancer, and the human leukocyte antigen (HLA) region 6p21.32 associated with oropharyngeal cancer (13). These limited risk loci represent only a small proportion of heritability, and no additional follow-up studies have been reported.

In the present GWAS, with the goal of identifying additional novel genetic risk loci for SCCHN, we conducted a two-phase GWA study in non-Hispanic whites. We first identified SNPs in the MDACC GWAS, followed by validating those SNPs with $P < 1 \times 10^{-3}$ using the published OncoArray GWAS data. As a result, we found three loci (6p22.1, 6p21.33 and 6p21.32) for overall SCCHN risk, three for oropharyngeal cancer risk (2p23.1, 6p21.33 and 6p21.32), two for oral cavity cancer risk (5p15.33 and 6p21.32), and one (18q22.2) for the hypo-pharyngeal and laryngeal cancer risk. Therefore, we identified two novel loci (6p22.1 and 18q22.2) for SCCHN risk in addition the replication of four known cancer susceptibility regions (2p23.1, 5p15.33, 6p21.33 and 6p21.32).

Material and Methods

Populations and genotyping

Discovery population: The SCCHN cases of the present GWAS were ascertained at Head and Neck Surgery Clinic through The University of Texas MD Anderson Cancer Center, Texas (MDACC) (14,15) between December 1996 and July 2011, whose genomic DNA was genotyped with Illumina HumanOmniExpress-12v1 BeadChip. All cases were individuals with newly diagnosed, histologically confirmed, previously untreated SCCHN of the oral cavity, pharynx or larynx (14–16). Cases were categorized by tumor site according to the International Classification of Diseases for Oncology (ICD-O, 2nd ed.) or ICD10 (5,17–19). We considered individuals with cancers of the oral cavity (codes C00.3–C00.9, C02.0–C02.3, C03.0, C03.1, C03.9, C04.0, C04.1, C04.8, C04.9, C05.0, C06.0–C06.2, C06.8, and C06.9), oropharynx (codes C01.9, C02.4, C05.1, C05.2, C09.0, C09.1, C09.8, C09.9, C10.0–C10.4, C10.8, and C10.9), hypopharynx (codes C12.9, C13.0–C13.2, C13.8, and C13.9), oral cavity or pharynx overlapping or not otherwise specified (codes C02.8, C02.9, C05.8, C05.9, C14.0, C14.2, and C14.8), larynx (codes C32.0–C32.3 and C32.8–C32.9). After quality control process, genotypes were available for 2,221 cases (Supplementary Figure. 1).

The controls were recruited from genetically unrelated visitors who accompanied cancer patients to MDACC outpatient clinics (14,15,20), or individuals recruited previously for the MDACC melanoma study (20), which was deposited in database of Genotypes and Phenotypes (dbGaP accession#: phs000187.v1.p1) or from Study of Addiction: Genetics and Environment (SAGE) study (SAGE; dbGaP accession #: phs000092.v1.p1) (21). Of these datasets, there were 1,188 cancer-free individuals recruited for the SCCHN study, whose genomic DNA was genotyped by using Illumina HumanOmniExpress-12v1 BeadChip; 1,026 cancer-free individuals previously recruited for the melanoma GWA study, in which the genomic DNA was genotyped by using Illumina Omni1-Quad_v1–0_B BeadChip, and 2,377 cancer-free individuals of European descent from the SAGE study(21), who have genotyping data generated by Illumina Human1Mv1 BeadChip. After quality control procedure, genotypes were available for 2,965 individuals. The genotyping data of the SCCHN GWAS has been deposited in the dbGaP (accession #: phs001173.v1.p1).

All participants in the discovery study signed a written informed consent form that permits us to collect blood samples and clinic-pathological information. The study protocols were approved by the Institutional Review Board of MDACC in accordance with tenets of the Declaration of Helsinki.

Replication Population: The replication dataset was part of a published study, which comprised 6,034 cases and 6,585 controls derived from 12 epidemiological studies, with the majority having been collected through a case-control design as part of the International Head and Neck Cancer Epidemiology Consortium (INHANCE) (13). We requested the related genotyping data and phenotype data from dbGaP (accession#: phs001202.v1.p1), in which data were available for 6034 cases and 4062 controls. Genomic DNA isolated from blood or buccal cells was genotyped at the Center for Inherited Disease Research (CIDR) with a novel genotyping tool, the Illumina OncoArray custom array designed for cancer

studies by the OncoArray Consortium part of the Genetic Associations and Mechanisms in Oncology (GAME-ON) Network. The majority of the samples that were genotyped had oral and pharynx cancer.

Quality control in both discovery and replication GWASs

For discovery study, we used genotypes to identify individuals with discordant sex information, duplicates and closely related individuals among all samples. We identified genetically related individuals by calculating genome-wide identity-by-state (IBS) distances on markers for each pair of individuals. For any pair with allele sharing of > 80%, we excluded the sample generating the lowest call rate from further analysis. Across data from both phases, we excluded 15 lacking consent, 101 duplicated individuals, 49 individuals because of discordant sex information, 8 because of cryptic relatedness, and 9 because of overall genotyping rate below 95% (Supplementary Figure. 1). For the combined data of 2,171 cases and 4,493 controls, 543,328 tagging SNPs were available. After applying quality control on genotype data, we retained 414,349 autosomal and 9,955 X chromosome SNPs showing minimal departure from Hardy-Weinberg equilibrium (HWE; $P > 10^{-6}$ in controls), genotyping typing call rate $\geq 95\%$ and minor allele frequencies (MAF) $\geq 1\%$ in cases and controls.

As described in a previous publication, a similar quality control process was applied in the validation GWAS (13). Briefly, this study first excluded samples with genotyping rate <95% and SNPs with call rate < 95%. After that, we also removed samples with unsolved genetic and reported sex discrepancies and individuals with outlying autosomal heterozygosity rate [± 4 standard deviations (SD)], as well as duplicate-pairs (IBD > 0.9) and relative pairs (showing IBD > 0.3). SNPs with MAF < 1% and deviation of HWE in controls ($P > 10^{-6}$) were also removed.

We also applied FastPop to estimate ancestry proportion (22). The final samples were those with the proportion of European ancestry ≥ 0.8 , which included 2,171 cases and 4,493 controls for the discovery study (MDACC, Supplementary Figure 2), and 5,205 cases and 3,232 controls for the replication study (OncoArray).

Imputation

To impute unknown genetic variation, we first performed strand flip using PLINK to convert all alleles to the forward genomic strand, and then used SHAPEIT for phasing and performed imputation with minimac4 on the Michigan imputation server (<https://imputationserver.sph.umich.edu>) with the HRC reference panel (Version r1.1 2016) consisting of 64,940 haplotypes of predominantly European ancestry. For imputation, we used a set of high-quality SNPs: an MAF > 0.01; a call rate > 95%, Hardy-Weinberg equilibrium test $P > 10^{-6}$; an allele frequency difference > 0.20 between the sample data and the reference panel. After imputation, SNPs with an MAF < 0.01 or imputation quality $r^2 < 0.3$ or a significant allele frequency difference ($P < 1 \times 10^{-3}$) among the controls of newly genotyped and those from the MDACC and SAGE GWAS, were excluded from the final association analysis. Thus, the final set included 7,858,089 SNPs on autosomes and X chromosome, of which 434,839 were typed and 7,423,250 were imputed SNP.

We also imputed HLA classical alleles and amino acids by using software SNP2HLA and the Type I Diabetes Genetics Consortium (TDGC) reference panel of 5,225 individuals of European descent. We divided the HN GWAS dataset into three subsets (each subset with around 1,100 samples), and the control samples from SAGE GWAS into two subsets (with 1,200 and 1,113 samples), and performed imputation separately for each subset. The final panel included 8,926 HLA alleles, of which 8,648 and 7,463 were imputed alleles with an info score 0.3 and 0.9, respectively. We then performed regional association analyses of binary markers followed by meta-analysis of imputed binary markers (SNPs, classical alleles and aminoacids) using PLINK.

Statistical methods and *in silico* functional annotations

To control for population confounding, for the two discovery datasets and replication dataset, we performed principal components analysis (PCA) in EIGENSTRAT using approximately 10,000 common markers in low linkage disequilibrium (LD) ($r^2 < 0.1$, MAF > 0.05). Significant PCs associated with disease status ($P < 0.05$) were adjusted as covariates in the further risk association analysis (including PC 1, 2, 5, 6, and 8 in the discovery GWAS and the top three PCs and the continent source in the replication dataset). As the distributions of age and sex are significantly different between cases and controls in the two discovery studies and the OncoArray replication study, we also adjusted them in the risk analysis. We performed an unconditional logistic regression to estimate odds ratios (ORs) and 95% confidence intervals (CIs) per effect allele by using PLINK (v2.0, <https://www.cog-genomics.org/plink/2.0/>) software with adjustment for the age, sex and top significant PCs. The association analysis between SNPs on X chromosome and cancer risk was performed by using the --xchr-model 1 option in PLINK as well as stratified analysis by sex. SNPs with $P < 10^{-3}$ were chosen to validate in the OncoArray GWAS. SNPs with combined $P < 5 \times 10^{-8}$ were considered to reach the genome-wide significance. We performed both random-effects and fixed-effects meta-analyses by using the inverse variance-weighted average method to combine the summary results of the discovery and replication studies. Heterogeneity was assessed as a Q-test $P < 0.10$ or $I^2 > 50.0\%$. For SNPs in the identified regions, we performed clump analysis to remove high LD SNPs with pairwise $r^2 > 0.1$ and then performed conditional analysis with PLINK 1.9 to identify SNPs with independent effects. For SNPs remained significant or marginally significant in the conditional analysis, we constructed a polygenic risk score (PRS) by summing risk alleles weighted by their corresponding effect sizes in the MDACC study by using the "--score" function in PLINK 1.90. The PRS was standardized by the mean and standard deviation and then estimated for its association with SCCHN risk. The odds ratio was reported as per standard deviation of the PRS.

To explore the possible functions of SNPs at the final identified regions, we applied the online tool HaploReg v4.1 (<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>), which integrated the Encyclopedia of DNA elements (ENCODE) data, to perform functional annotation. We also performed *in silico* expression quantitative trait loci (eQTL) analysis by using data from multiple sources: the lymphoblastoid cell lines of 358 European individuals from Genetic European Variation in Health and Disease Consortium (GEUVADIS) and the 1000 Genomes Project (23); eQTL data of multiple tissues from the Genotype-Tissue Expression (GTEx) project (24); SNP and mRNA expression data in primary tumor tissues

from 344 SCCHN patients of European ancestry in The Cancer Genome Atlas (TCGA) database (dbGaP accession#: phs000178.v1.p1) (25). Manhattan plots was generated in R using the package qqman; the regional association plots and LD plots were constructed based on the 1000 Genomes European (EUR) reference data (phase 3, release date October 2014) by using LocusZoom and Haploview v4.2, respectively. SNP pruning was applied, and SNPs with paired-wise $r^2 < 0.30$ were considered as independent. All other analyses were conducted with R (version 3.5.1) and SAS (version 9.4; SAS Institute, Cary, NC, USA), if not mentioned otherwise.

Results

Characteristics of the study populations

The workflow of the present GWAS is depicted in Supplementary Figure 1. The distributions of age and sex were statistically different between cases and controls (Table 1, $P < 0.001$), in the discovery dataset, case group included much older males (mean age 57.9 (SD \pm 11.2) for cases and 50.0 (12.3) for controls with 22.8% of males in cancers and 44.9% in controls). Of the cases, there were 631 (29.1%) patients with oral cavity cancer, 1,144 (52.7%) with oropharyngeal cancer, 394 (28.2%) with hypo-pharyngeal, laryngeal or overlapping cancer sites, two samples with missing values for histological types.

Similar to the discovery population, case group in the replication study had a higher proportion of males (74.2%) and older subjects (mean age of 59.7) than control group (70.9% and mean age of 58.1, respectively) (Table 1). Of the cases, there were 2,568 (49.5%) patients with oral cavity cancer, 2,328 (44.8%) with oropharyngeal cancer, and 295 (5.7%) with hypo-pharyngeal, laryngeal or overlapping cancer sites.

Association analysis

We performed association analysis for SNPs with imputation quality $r^2 \geq 0.3$ and minor allele frequency ≥ 0.01 , and quality distributions have been shown in Supplementary Figure 3a–f for those SNPs with MAF ≥ 0.01 from the SCCHN GWAS, and the using controls from melanoma GWAS and SAGE studies, respectively. The overall results of the discovery results are presented in Figure 1a. There were 10,714 SNPs (i.e., 10,218 SNPs on autosomes and 496 SNPs on X chromosome) with $P \leq 1.00 \times 10^{-3}$ and 25 SNPs with $P \leq 5.00 \times 10^{-8}$ in the MDACC discovery study. Quantile-quantile (Q-Q) plots of observed and expected P -values showed a moderate genomic inflation (λ) for discovery results ($\lambda = 1.035$) (Supplementary Figure 4a). We then replicated the associations of these SNPs in the OncoArray study and found 94 and 87 SNPs located at three loci (6p22.1, 6p21.33 and 6p21.32) associated with SCCHN risk with $P \leq 5 \times 10^{-8}$ in the fixed-effects or random-effects model of the meta-analysis, respectively (Supplementary Table 1). We have also provided the results of the leading SNPs in each region in Table 2 (in the meta-analysis of the discovery and replication studies, $P = 2.96 \times 10^{-9}$, 3.75×10^{-10} , and 1.44×10^{-16} for SNP rs259919 at 6p22.1, rs1265081 at 6p21.33, and rs3135001 at 6p21.32 in random-effects model, respectively).

Further stratified analysis by tumor site revealed that there were 23, 75 and one SNPs with $P = 5.00 \times 10^{-8}$ in association with risk of oral cavity cancer, oropharyngeal cancer, and hypo-pharyngeal/laryngeal cancers in the discovery study, respectively (Figure 1b, 1c and 1d). We then selected SNPs with $P = 1.00 \times 10^{-3}$ (i.e., 8,658, 12,454 and 9,062 SNPs in the three sub-populations of the discovery study, respectively) for replication with the OncoArray dataset (Supplementary Table 2–4). We found in the random-effects model of meta-analysis that two loci associated with oral cavity cancer risk reached genome-wide significance (the leading SNPs rs10462706 in *CLPTMIL* at 5p15.33 region and rs1049055 in *HLA-DQB1* at 6p21.32 with $P=7.87 \times 10^{-11}$ (OR = 0.73 and 95% CI: 0.66–0.80) and $P=2.96 \times 10^{-9}$ (OR = 0.78 and 95% CI: 0.72–0.85), respectively]. Three loci (2p23.1, 6p21.33 and 6p21.32) were found to be associated with oropharyngeal cancer risk, with leading SNP rs4318431 ($P = 3.13 \times 10^{-9}$, OR = 1.39 and 95% CI: 1.25–1.55) nearby gene *GALNT14*; SNP rs13211972 in *MUC21* ($P=1.04 \times 10^{-10}$, OR = 1.55 and 95%CI: 1.36–1.77); and SNP rs34518860 ($P=2.61 \times 10^{-17}$, OR = 0.61 and 95%CI: 0.54–0.68) in gene *HLA-DQA1*, respectively. We also identified one novel locus (18q22.2) associated with the risk of hypo-pharyngeal and laryngeal cancers ($P=2.54 \times 10^{-9}$, OR = 3.95 and 95%CI: 3.51–6.21 for leading SNP rs142021700 in gene *RTTN*) (Supplementary Tables 2 and 4). Quantile-quantile (Q-Q) plots of the stratified results are shown in Supplementary Figure 4b–4d, and regional association plots for each identified locus are presented in Figure 2a–2i.

To identify independent SNPs, we also performed clump analysis and revealed four low LD clumps consisted of the 87 SNPs associated with the overall SCCHN risk. In the following conditional analysis, we found that each of four SNPs (rs3129726 and rs62404579 at 6p21.32, rs1265081 at 6p21.33, and rs259919 at 6p22.1) remained significant in the presence of three other SNPs ($P<0.05$) (Supplementary Table 5). Similar, we found one SNP (i.e., SNP rs7713218 at 5q15.33) remained significant after conditioning on two lead SNPs at 5p15.33 and 6p21.32 for oral cavity cancer (Supplementary Table 6); five SNPs in 6p21.33 remained significantly associated with oropharyngeal cancer risk after conditioning on SNPs at the two reported loci 6p21.32 and 2p23.1; while no SNP was significant with hypo-pharyngeal and laryngeal cancer risk after conditioning on the leading SNP at 18q22.2 (Supplementary Table 7–8). These results suggested that independent signals at three loci (6p22.1, 6p21.33, and 6p21.32) contribute to the risks of overall cancer and oropharyngeal cancer.

The two previous GWAs and candidate gene based studies have reported 31 loci (including 33 leading SNPs) associated with the risk of oral, oropharyngeal, pharyngeal and laryngeal cancers (12,13). We extracted the results from GWAS catalog (<https://www.ebi.ac.uk/gwas/home>) and investigated their association in the MDACC GWAS. As a result, we found five loci could be replicated in the current study (i.e., rs10462706 at locus 5p15.33 and rs1800628 at 6p21.33 with oral cancer risk, rs2216824 at 2p23.1 and rs1453414 at 11p15.4 with oropharyngeal cancer risk, and rs1229984 at 4q23 with hypo-pharyngeal and laryngeal cancer risk) (Supplementary Table 9).

Since multiple SNPs in the HLA region have been associated with overall cancer risk and oral/oropharyngeal cancer risk, we then performed a HLA imputation to reveal the exact HLA alleles associated with cancer risk. For those with an imputation info > 0.3 , we found

53 HLA alleles associated with overall cancer risk with $P < 0.05$, of which three alleles reached genome-wide significance (*HLA-B*37*, *HLA-B*3701*, and *HLA-DQB1*06*) (Supplementary Table 10); 28 HLA alleles associated with oral cancer risk, of which two alleles reached genome-wide significance (*HLA-B*37* and *HLA-B*3701*) (Supplementary Table 11); 59 HLA alleles associated with oropharyngeal cancer risk (Supplementary Table 12), of which four alleles (i.e., *HLA-B*37*, *HLA-B*3701*, *HLA-DQB1*06*, and *HLA-DRB1*13*) reached genome-wide significant level; and 23 HLA variants associated with hypo-pharyngeal and laryngeal cancer risk (Supplementary Table 13). As shown in Supplementary Table 12 and 14, we also replicated the association results of three reported HLA specific alleles (*DRB1*1301*, *DQA1*0103* and *DQB1*0603*) and their haplotype with decreased oropharyngeal cancer risk ($P = 6.5 \times 10^{-6}$, 4.16×10^{-7} , 6.2×10^{-7} , and 1.95×10^{-7} , respectively) (13).

We also constructed PRS by summing the effects of the 12 SNPs remained significant or marginally significant in the conditional analysis of oropharyngeal cancer (i.e., rs73730372, rs28366328, rs9469220, rs13211972, rs17207190, rs114202986, rs144112342, rs2194452, rs41258944, rs114949918, rs3131013 and rs147748716), and analyzed its association with oropharyngeal cancer risk. As a results (Supplementary Table 15), the PRS showed significant association ($P < 2.00E-16$) with oropharyngeal cancer risk with an OR per standard deviation of the PRS of 1.49 (95% CI: 1.39–1.60) and 1.38 (95% CI: 1.30–1.46) in the MDACC study and OncoArray study, respectively.

The *in silico* functional annotation

Functional annotations for the identified representative genetic variants reaching $P < 5 \times 10^{-8}$ are summarized in Supplementary Table 16. There were 179 SNPs at 6p21.32 and 6p21.33 with potential effects on the promoter or enhancer activities with a significant eQTL evidence. We also retrieved the eQTL results of multiple tissues from the Genotype-Tissue Expression Project (GTEx) for the lead SNPs or LD SNPs significantly correlated with corresponding mRNA expression levels. For example, the variant allele A of rs259919 at 6p22.1 and the wild allele T of rs1049055 at 6p21.32 were significantly correlated with increased mRNA expression levels of *ZFP57* and *HLA-DQB1*, respectively (Supplementary Figure 5a and 5c), while the variant allele of rs13211972 at 6p21.33 was found to be associated with the decreased mRNA expression levels of *MICA* in multiple tissues (Supplementary Figure 5b). The effect of rs10462706 in 5p15.33 on mRNA expression levels of *CLPTMIL* was different by tissues (Supplementary Figure 5d). However, we did not find any evidence for the effects of rs78082910 in the 2p23.1 region on the expression of nearby genes (<https://www.gtexportal.org/home/snp/rs78082910>). We also performed eQTL analysis in the primary tumor tissues of 344 SCCHN patients from TCGA which were with both genotyping/imputation data and mRNA expression data available. As one of the lead SNP rs73730372 was not included in the TCGA data, we used one of its high LD SNP rs115625939 in the eQTL analysis. Of the five loci, we found that the variant alleles of SNP rs115625939 (which had high LD with one representative SNP rs73730372 with $r^2 = 0.89$) in 6p21.32, and rs27069 (which has high LD with representative SNP rs2447853 with $r^2 = 0.69$) in 5p15.33 were correlated with the upregulated mRNA expression of *HLA-DQB1* and *CLPTMIL* in the tumor tissues of head and neck cancer, respectively (Supplementary Figure

6a–b: $P = 0.012$ and 0.008 , respectively). No significance was found in the eQTL analyzes for *MICA* and *GALNT14* in TCGA, and the mRNA expression data was unavailable for *ZFP57*. We also found that the identified SNP rs73730372 in 6p21.32 was significantly correlated with the mRNA expression of *HLA-DQB1* (Supplementary Figure 6c: $P = 3.67 \times 10^{-10}$) in the lymphoblastoid cell lines of 358 European individuals from 1000 Genomes project.

We have also performed differential expression analysis for the five genes in the identified regions (Table 2) by using the mRNA expression data in 520 SCCHN tumor tissues and 44 adjacent normal tissues from TCGA (<http://ualcan.path.uab.edu/cgi-bin/ualcan-res.pl>). As a result, we found that the mRNA expression of the five genes were significantly higher in the primary tumor tissues than in the adjacent normal tissues with $P < 0.05$ (Supplementary Figure 7a–e for *ZFP57*, *MICA*, *HLA-DQB1*, *CLPTMIL*, and *GALNT14*, respectively). Eight other genes at the five loci (*HLA-DRB1*, *HLA-DQA1*, *HLA-DQA2*, *PSORS1C3*, *HCG27*, *MICA*, *HCP5*, *HLA-DRB5*, *DPCR1* and *MUC21*) also showed significant difference on mRNA expression between the tumor tissues and adjacent normal tissues (<http://ualcan.path.uab.edu/cgi-bin/TCGAExResultNew2.pl?genenam=HLA-DRB1,HLA-DQA1,HLA-DQA2,PSORS1C3,HCG27,MICA,HCP5,HLA-DRB5,DPCR1,MUC21&ctype=HNSC>).

Discussion

In the present GWAS study, we aimed to identify additional genetic loci associated with risk of SCCHN and its sub-types by using a discovery study, followed by another independent replication study. In the meta-analysis of all the two GWAS datasets, we identified SNPs at six genomic regions (i.e., 2p23.1, 5p15.33, 6p21.32, 6p21.33, 6p22.1 and 18q22.2) to be associated with risk of SCCHN or its sub-types at a GWAS significance level. Four of the regions are the known SCCHN risk loci (i.e., 2p23.1, 5p15.33, 6p21.32 and 6p21.33), while two other loci (i.e., 6p22.1 and 18q22.2) are novel findings for SCCHN risk. Functional annotation revealed that multiple SNPs in these regions were potentially functional, because they may affect their mRNA expression. The most prominent finding in the overall and stratified meta-analyses was a strong association signal at 6p21.32 within the HLA class II region. SNPs in this region showed significant associations with risk of SCCHN and all three sub-types, especially oropharyngeal cancer characterized by HPV-infection in the etiology.

The HLA system has long been recognized in humans as a very important genomic region relating to infection, inflammation, autoimmunity and transplantation medicine (26). The HLA system is categorized into class I, II and III regions and consists of more than 200 genes. Genes in the HLA system have multiple biological functions with an emphasis on immunological functions (27). Specifically, SNP rs9273448/rs1049225 maps to the 3' UTR of the gene MHC, class II, DQ beta-1, also called *HLA-DQB1*, which belongs to HLA class II beta chain paralogs. The protein encoded by this gene is one of two proteins that are required to form the DQ heterodimer, a cell surface receptor essential to the function of the immune system. The identified SNPs for all SCCHN, oral and mixed hypo-pharyngeal and laryngeal cancers are mainly located around *HLA-DQB1/DQA1*, while the identified SNPs

for oropharyngeal cancer are distributed in a wide range, covering multiple HLA genes, including *HLA-B*, *HCP5*, *HLA-DRA1*, *HLA-DRB1*, *HLA-DRB1*, *HLA-DQA1* and *HLA-DQB1*. Previous studies have reported significant associations between *HLA-DQB1* polymorphisms and the risks of HPV-related oropharyngeal (13), cervical cancers (13,28), cutaneous melanoma (29), gastric adenocarcinoma (30,31), breast cancer (32), and nasopharyngeal carcinoma. In the present study, we replicated previous report that three HLA alleles (*DRB1*1301*, *DQA1*0103* and *DQB1*0603*) as well as their haplotype were associated with the risks of SCCHN and oropharyngeal cancer (13). In addition, we have shown that the variant allele of rs73730372 was associated with both higher mRNA expression levels of *HLA-DQB1* and lower risk of oropharyngeal cancer, which is consistent with previous reports that HLA-DQB1 may be involved in the HPV-specific immune response (33). We also observed similar results by using the TCGA expression data, which indicated those HLA genes had higher expression in the SCCHN tumor tissues than in adjacent normal tissues. Further functional studies are warranted to illuminate the underlying biological mechanisms.

The closest gene at the replicated loci 2p23.1 is *GALNT14* that encodes a Golgi protein, which catalyzes the transfer of N-acetyl-D-galactosamine (GalNAc) to large proteins like mucins (34). Aberrant glycosylation is a hallmark of most human cancers and affects many cellular properties, including cell proliferation, apoptosis, differentiation, transformation, migration, invasion, and immune responses (35). *GALNT14* has been reported to be involved in the initial step of mucin-type O-glycosylation and thus plays a critical role in the invasion and migration of breast cancers by regulating the activity of MMP-2 and expression of some EMT genes (36). Another study also suggested that *GALNT14* might contribute to ovarian carcinogenesis through aberrant glycosylation of MUC13, whose expression was dysregulated in many human cancers (37,38). In this study, we also observed that *GALNT14* had higher mRNA expression in the SCCHN tumor than in the adjacent normal tissues. Interestingly, we also found that SNPs in *MUC21/MUC22* (6p21.33) were associated with risk of the overall cancer and oropharyngeal cancer. Splicing variants and mutations in mucin genes have been observed in various cancers and shown to participate in cancer progression and metastasis (39). The nearby gene *MUC21* is localized on chromosome 6 (6p21.33) closing to the *HLA* class I region, which is a membrane-associated mucin belonging to the mucin family (40). Clinically, mucins are used as carcinoma markers and therapeutic targets for cancer treatment (40,41). The protein encoded by *MUC21* has been shown to be expressed by adenocarcinomas of the lung (40). However, by using the TCGA data, we observed that *MUC21* had higher expression in the normal tissues than in the SCCHN tumor tissues, which implied this gene might play a different role in the development of SCCHN. Previous association studies of genetic variants have linked the *MUC21* gene to non-cancer diseases (e.g., Stevens-Johnson syndrome / toxic epidermal necrolysis, and pulmonary function) (42,43). However, few studies have investigated the functions of the *MUC22* gene. In addition, we revealed that the variant allele of SNP rs13211972 in the 6p21.33 region was significantly correlated with decreased mRNA expression levels of *MICA* and an increased risk of SCCHN. *MICA* encodes a membrane-bound protein, acting as a ligand of natural killer (NK) group 2D (NKG2D) to trigger NK cell-mediated cytotoxicity. *MICA* has an antitumor property as its expression is induced in

stressed cells, such as transformed tumor cells for the detection by NK cells (44). Several studies have reported that SNPs in *MICA* have been associated with risk of cervical squamous cell carcinoma (45,46) and HCV induced hepatocellular carcinoma (47). The association between the *MICA* STR polymorphism and risk of oral squamous cell carcinoma had been investigated in several candidate studies but with conflicting results, which might be due to small sample size (48–50). Considering the relatively large sample of the present GWAS, our results provided a strong evidence that individuals with SNPs associated with lower mRNA expression levels of *MICA* might have an increased risk of SCCHN and oropharyngeal cancer. Genetic variants in this region have also been reported to be associated with risk of lung cancer and follicular lymphoma, and the susceptibility gene *BAT3* was found to be involved in DNA damage-induced apoptosis and to modulate the acetylation of p53 during autophagy (51–53). In addition, in the HLA allele analysis, we observed that the *HLA-B*37* allele was associated with the risks of SCCHN, oral cancer and oropharyngeal cancer. Further functional validation for those susceptibility genes is warranted.

We also found that the variant allele of SNP rs259919 in the 6p22.1 region was significantly correlated with decreased mRNA expression levels of *ZFP57* in multiple tissues and an increased risk of SCCHN. *ZFP57* is an important transcriptional regulator involved in DNA methylation and genomic imprinting during development (54). In addition, previous studies have reported that *ZFP57* plays an important role in DNA methylation and epigenetic regulation and has important potential implications for diseases.

In the present GWAS, we also showed a significant association between the variant allele of SNP rs2447853 at 5p15.33 and an increased risk of oral cavity cancer, which confirms the previous finding (13). This locus was also reported to be associated with lung cancer risk, and the genetic variants in *TERT_CLPTMIL* have been reported to be associated with DNA-adduct levels in lung (51,55). By using the GTEx data, we found that the variant allele of rs2447853 was significantly correlated with increased mRNA expression levels of *CLPTMIL* in multiple tissues (e.g., small intestine, colon, and esophagus), which provides some biological evidence for the identified association. However, further functional studies are warranted.

It should be noted that the rs1229984 (4q23, *ADH1B*) had been previously reported as a susceptibility locus for oral cancer and oropharyngeal cancer in several studies. In the present study, we found this SNP was only associated with risk of hypo-pharyngeal and laryngeal cancers, but not oral cavity and oropharyngeal cancer. Such discrepancies might be due to population heterogeneity.

In summary, in the present GWAS of SCCHN in non-Hispanic whites, we identified two novel common loci that might influence SCCHN risk and replicated some loci previously reported, which highlights the importance of genetic variation of genes (e.g., *HLA-DQB1*, *HLA-DQA1*, and *MUC21*) in the *HLA* system in the development of head and neck cancer. These findings suggest that the immunologic mechanism is implicated in the etiology of SCCHN, particularly in oropharyngeal cancer. Future replication of these findings in other

independent populations is warranted with additional functional studies necessary to establish the biological framework underlying the observed associations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

Sanjay Shete was supported in part by the NIH grants 1R01CA131324 and R01DE022891; the Cancer Prevention Research Institute of Texas grants RP170259; the Barnhart Family Distinguished Professorship in Targeted Therapy; and Betty B. Marcus Chair in Cancer Prevention. Sanjay Shete and Jian Wang were supported in part by the Cancer Center Support Grant P30CA016672. Qingyi Wei was supported by NIH grants 2R01 ES011740 and 1R01CA 131274 and the Duke Cancer Institute as part of the P30 Cancer Center Support Grant (Grant ID: NIH/NCI CA014236).

Melanoma GWAS study

Part of the controls were from the melanoma GWAS study of MDACC, which was deposited in dbGaP (Accession#: phs000187.v1.p1). Research support to collect data and develop an application to support this project was provided by 3P50CA093459, 5P50CA097007, R01CA100264, and 5R01CA133996.

SAGE study

Part of the control were requested from the Study of Addiction: Genetics and Environment (SAGE) in dbGaP. Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative [GEI] (U01 HG004422). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract "High throughput genotyping for studying the genetic contributions to human disease" (HHSN268200782096C). The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1 through dbGaP accession number phs000092.v1.p.

OncoArray: Oral and Pharynx Cancer

The replication data was from the study of OncoArray: Oral and Pharynx Cancer (dbGaP Study Accession#: phs001202.v1.p1) in dbGaP. Genotyping performed at the Center for Inherited Disease Research (CIDR) was supported through contract number HHSN268201200008I: funds were provided by the U.S. National Institute of Dental and Craniofacial Research (NIDCR) grant X01HG007780; funds were also provided by the U.S. National Cancer Institute (NCI) for genotyping for shared controls with the Lung OncoArray initiative (grant X01HG007492). University of Pittsburgh head and neck cancer study: grants P50 CA097190 and P30 CA047904. Carolina Head and Neck Cancer Study (CHANCE): R01-CA90731. GENCAPO: FAPESP, grant numbers 04/12054-9 and 10/51168-0. HN5000 study: NIHR RP-PG-0707-10034. Toronto study: the Canadian Cancer Society Research Institute (020214) and NCI U19 CA148127. ARCAGE study: European Commission's 5th Framework Program (QLK1-2001-00182), FIRMS, Region Piemonte, and Padova University (CPDA057222). Rome Study: AIRC IG 2011 10491 and IG2013 14220, and Fondazione Veronesi. IARC Latin American study: European Commission INCO-DC IC18-CT97-0222, Fondo para la Investigacion Cientifica y Tecnologica (Argentina) and Fundação de Amparo à Pesquisa do Estado de São Paulo (01/01768-2). IARC Central Europe study: INCO-COPERNICUS Program (IC15- CT98-0332), NCI CA92039 and WCRF99A28. IARC Oral Cancer Multicenter study: Europe against Cancer (S06 96 202489 05F02), Spain FIS 97/0024, FIS 97/0662, BAE 01/5013, UICC Yamagiwa-Yoshida, National Cancer Institute of Canada, AIRC and PAHO/WHO. EPIC study: European Commission (DG SANCO) and IARC.

TCGA

eQTL analysis had been performed by using the genotyping data and mRNA expression data in the primary tumor tissues of 344 SCCHN patients of European ancestry from the Cancer Genome Atlas (TCGA) database (dbGaP accession#: phs000178.v1.p1). The results published here are in whole or part based upon data generated by The Cancer Genome Atlas managed by the NCI and NHGRI. Information about TCGA can be found at <http://cancergenome.nih.gov>.

References

1. Parkin DM, Pisani P, Ferlay J. Global cancer statistics. *CA: a cancer journal for clinicians* 1999;49:33–64, 1. [PubMed: 10200776]
2. Wang M, Chu H, Zhang Z, Wei Q. Molecular epidemiology of DNA repair gene polymorphisms and head and neck cancer. *Journal of biomedical research* 2013;27:179–92. [PubMed: 23720673]
3. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA: a cancer journal for clinicians* 2013;63:11–30. [PubMed: 23335087]
4. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011;29:4294–301. [PubMed: 21969503]
5. Hashibe M, Brennan P, Benhamou S, Castellsague X, Chen C, Curado MP, et al. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Journal of the National Cancer Institute* 2007;99:777–89. [PubMed: 17505073]
6. Sturgis EM, Cinciripini PM. Trends in head and neck cancer incidence in relation to smoking prevalence: an emerging epidemic of human papillomavirus-associated cancers? *Cancer* 2007;110:1429–35. [PubMed: 17724670]
7. 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. *International journal of cancer Journal international du cancer* 2009;124:394–401. [PubMed: 18814262]
8. Ho T, Wei Q, Sturgis EM. Epidemiology of carcinogen metabolism genes and risk of squamous cell carcinoma of the head and neck. *Head & neck* 2007;29:682–99. [PubMed: 17274053]
9. Neumann AS, Sturgis EM, Wei Q. Nucleotide excision repair as a marker for susceptibility to tobacco-related cancers: a review of molecular epidemiological studies. *Molecular carcinogenesis* 2005;42:65–92. [PubMed: 15682379]
10. Wei S, Liu Z, Zhao H, Niu J, Wang LE, El-Naggar AK, et al. A single nucleotide polymorphism in the alcohol dehydrogenase 7 gene (alanine to glycine substitution at amino acid 92) is associated with the risk of squamous cell carcinoma of the head and neck. *Cancer* 2010;116:2984–92. [PubMed: 20336794]
11. Hashibe M, McKay JD, Curado MP, Oliveira JC, Koifman S, Koifman R, et al. Multiple ADH genes are associated with upper aerodigestive cancers. *Nature genetics* 2008;40:707–9. [PubMed: 18500343]
12. 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. *Nature genetics* 2014;46:1110–4. [PubMed: 25194280]
13. Lesueur 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. *Nature genetics* 2016;48:1544–50. [PubMed: 27749845]
14. Neumann AS, Lyons HJ, Shen H, Liu Z, Shi Q, Sturgis EM, et al. Methylenetetrahydrofolate reductase polymorphisms and risk of squamous cell carcinoma of the head and neck: a case-control analysis. *International journal of cancer Journal international du cancer* 2005;115:131–6. [PubMed: 15688408]
15. Li G, Sturgis EM, Wang LE, Chamberlain RM, Amos CI, Spitz MR, et al. Association of a p73 exon 2 G4C14-to-A4T14 polymorphism with risk of squamous cell carcinoma of the head and neck. *Carcinogenesis* 2004;25:1911–6. [PubMed: 15180941]

16. Chen X, Sturgis EM, Lei D, Dahlstrom K, Wei Q, Li G. Human papillomavirus seropositivity synergizes with MDM2 variants to increase the risk of oral squamous cell carcinoma. *Cancer research* 2010;70:7199–208. [PubMed: 20736372]
17. Organization WH. *International Statistical Classification of Diseases, Injuries, and Causes of Death*. Geneva, Switzerland: World Health Organization; 1992.
18. Percy C, Van Holten VD, Muir C. *International Classification of Diseases for Oncology*. Geneva, Switzerland: World Health Organization; 1990.
19. Wyss A, Hashibe M, Chuang SC, Lee YC, Zhang ZF, Yu GP, et al. Cigarette, cigar, and pipe smoking and the risk of head and neck cancers: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *American journal of epidemiology* 2013;178:679–90. [PubMed: 23817919]
20. Amos CI, Wang LE, Lee JE, Gershenwald JE, Chen WV, Fang S, et al. Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Human molecular genetics* 2011;20:5012–23. [PubMed: 21926416]
21. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. *Nature genetics* 2007;39:1181–6. [PubMed: 17898773]
22. Li Y, Byun J, Cai G, Xiao X, Han Y, Cornelis O, et al. FastPop: a rapid principal component derived method to infer intercontinental ancestry using genetic data. *BMC Bioinformatics* 2016;17:122. [PubMed: 26961892]
23. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013;501:506–11. [PubMed: 24037378]
24. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;348:648–60. [PubMed: 25954001]
25. Cancer Genome Atlas N Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* 2015;517:576–82. [PubMed: 25631445]
26. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. *Journal of human genetics* 2009;54:15–39. [PubMed: 19158813]
27. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, et al. Gene map of the extended human MHC. *Nature reviews Genetics* 2004;5:889–99.
28. Zhang X, Lv Z, Yu H, Wang F, Zhu J. The HLA-DQB1 gene polymorphisms associated with cervical cancer risk: A meta-analysis. *Biomed Pharmacother* 2015;73:58–64. [PubMed: 26211583]
29. Lee JE, Reveille JD, Ross MI, Platsoucas CD. HLA-DQB1*0301 association with increased cutaneous melanoma risk. *International journal of cancer Journal international du cancer* 1994;59:510–3. [PubMed: 7960221]
30. Wu MS, Hsieh RP, Huang SP, Chang YT, Lin MT, Chang MC, et al. Association of HLA-DQB1*0301 and HLA-DQB1*0602 with different subtypes of gastric cancer in Taiwan. *Japanese journal of cancer research : Gann* 2002;93:404–10. [PubMed: 11985790]
31. Lee JE, Lowy AM, Thompson WA, Lu M, Loflin PT, Skibber JM, et al. Association of gastric adenocarcinoma with the HLA class II gene DQB10301. *Gastroenterology* 1996;111:426–32. [PubMed: 8690208]
32. Chaudhuri S, Cariappa A, Tang M, Bell D, Haber DA, Isselbacher KJ, et al. Genetic susceptibility to breast cancer: HLA DQB*03032 and HLA DRB1*11 may represent protective alleles. *Proceedings of the National Academy of Sciences of the United States of America* 2000;97:11451–4. [PubMed: 11027344]
33. 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. [PubMed: 17438108]
34. Wang H, Tachibana K, Zhang Y, Iwasaki H, Kameyama A, Cheng L, et al. Cloning and characterization of a novel UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase, pp-GalNAc-T14. *Biochemical and biophysical research communications* 2003;300:738–44. [PubMed: 12507512]

35. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* 2015;15:540–55. [PubMed: 26289314]
36. Huanna T, Tao Z, Xiangfei W, Longfei A, Yuanyuan X, Jianhua W, et al. GALNT14 mediates tumor invasion and migration in breast cancer cell MCF-7. *Molecular carcinogenesis* 2015;54:1159–71. [PubMed: 24962947]
37. Wang R, Yu C, Zhao D, Wu M, Yang Z. The mucin-type glycosylating enzyme polypeptide N-acetylgalactosaminyltransferase 14 promotes the migration of ovarian cancer by modifying mucin 13. *Oncol Rep* 2013;30:667–76. [PubMed: 23708057]
38. Maher DM, Gupta BK, Nagata S, Jaggi M, Chauhan SC. Mucin 13: structure, function, and potential roles in cancer pathogenesis. *Mol Cancer Res* 2011;9:531–37. [PubMed: 21450906]
39. Kumar S, Cruz E, Joshi S, Patel A, Jahan R, Batra SK, et al. Genetic variants of mucins: unexplored conundrum. *Carcinogenesis* 2017;38:671–79. [PubMed: 27838635]
40. Itoh Y, Kamata-Sakurai M, Denda-Nagai K, Nagai S, Tsuiji M, Ishii-Schrade K, et al. Identification and expression of human epiglycanin/MUC21: a novel transmembrane mucin. *Glycobiology* 2008;18:74–83. [PubMed: 17977904]
41. Hollingsworth MA, Swanson BJ. Mucins in cancer: protection and control of the cell surface. *Nature reviews Cancer* 2004;4:45–60. [PubMed: 14681689]
42. Genin E, Schumacher M, Roujeau JC, Naldi L, Liss Y, Kazma R, et al. Genome-wide association study of Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis in Europe. *Orphanet journal of rare diseases* 2011;6:52. [PubMed: 21801394]
43. Hancock DB, Artigas MS, Gharib SA, Henry A, Manichaikul A, Ramasamy A, et al. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS genetics* 2012;8:e1003098.
44. Chen D, Gyllensten U. MICA polymorphism: biology and importance in cancer. *Carcinogenesis* 2014;35:2633–42. [PubMed: 25330802]
45. 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. [PubMed: 23482656]
46. Chen D, Hammer J, Lindquist D, Idahl A, Gyllensten U. A variant upstream of HLA-DRB1 and multiple variants in MICA influence susceptibility to cervical cancer in a Swedish population. *Cancer Med* 2014;3:190–8. [PubMed: 24403192]
47. Jiang X, Zou Y, Huo Z, Yu P. Association of major histocompatibility complex class I chain-related gene A microsatellite polymorphism and hepatocellular carcinoma in South China Han population. *Tissue Antigens* 2011;78:143–7. [PubMed: 21644931]
48. Chung-Ji L, Yann-Jinn L, Hsin-Fu L, Ching-Wen D, Che-Shoa C, Yi-Shing L, et al. The increase in the frequency of MICA gene A6 allele in oral squamous cell carcinoma. *J Oral Pathol Med* 2002;31:323–8. [PubMed: 12190814]
49. Reinders J, Rozemuller EH, van der Ven KJ, Caillat-Zucman S, Slootweg PJ, de Weger RA, et al. MHC class I chain-related gene a diversity in head and neck squamous cell carcinoma. *Hum Immunol* 2006;67:196–203. [PubMed: 16698442]
50. Tamaki S, Sanefuzi N, Ohgi K, Imai Y, Kawakami M, Yamamoto K, et al. An association between the MICA-A5.1 allele and an increased susceptibility to oral squamous cell carcinoma in Japanese patients. *J Oral Pathol Med* 2007;36:351–6. [PubMed: 17559497]
51. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nature genetics* 2008;40:1407–9. [PubMed: 18978787]
52. Sasaki T, Gan EC, Wakeham A, Kornbluth S, Mak TW, Okada H. HLA-B-associated transcript 3 (Bat3)/Scythe is essential for p300-mediated acetylation of p53. *Genes Dev* 2007;21:848–61. [PubMed: 17403783]
53. Skibola CF, Bracci PM, Halperin E, Conde L, Craig DW, Agana L, et al. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. *Nature genetics* 2009;41:873–5. [PubMed: 19620980]

54. Plant K, Fairfax BP, Makino S, Vandiedonck C, Radhakrishnan J, Knight JC. Fine mapping genetic determinants of the highly variably expressed MHC gene ZFP57. *Eur J Hum Genet* 2014;22:568–71. [PubMed: 24193346]
55. Zienolddiny S, Skaug V, Landvik NE, Ryberg D, Phillips DH, Houlston R, et al. The TERT-CLPTM1L lung cancer susceptibility variant associates with higher DNA adduct formation in the lung. *Carcinogenesis* 2009;30:1368–71. [PubMed: 19465454]

Significance:

Two novel risk loci for SCCHN in non-Hispanic white individuals highlight the importance of immunologic mechanism in the disease etiology.

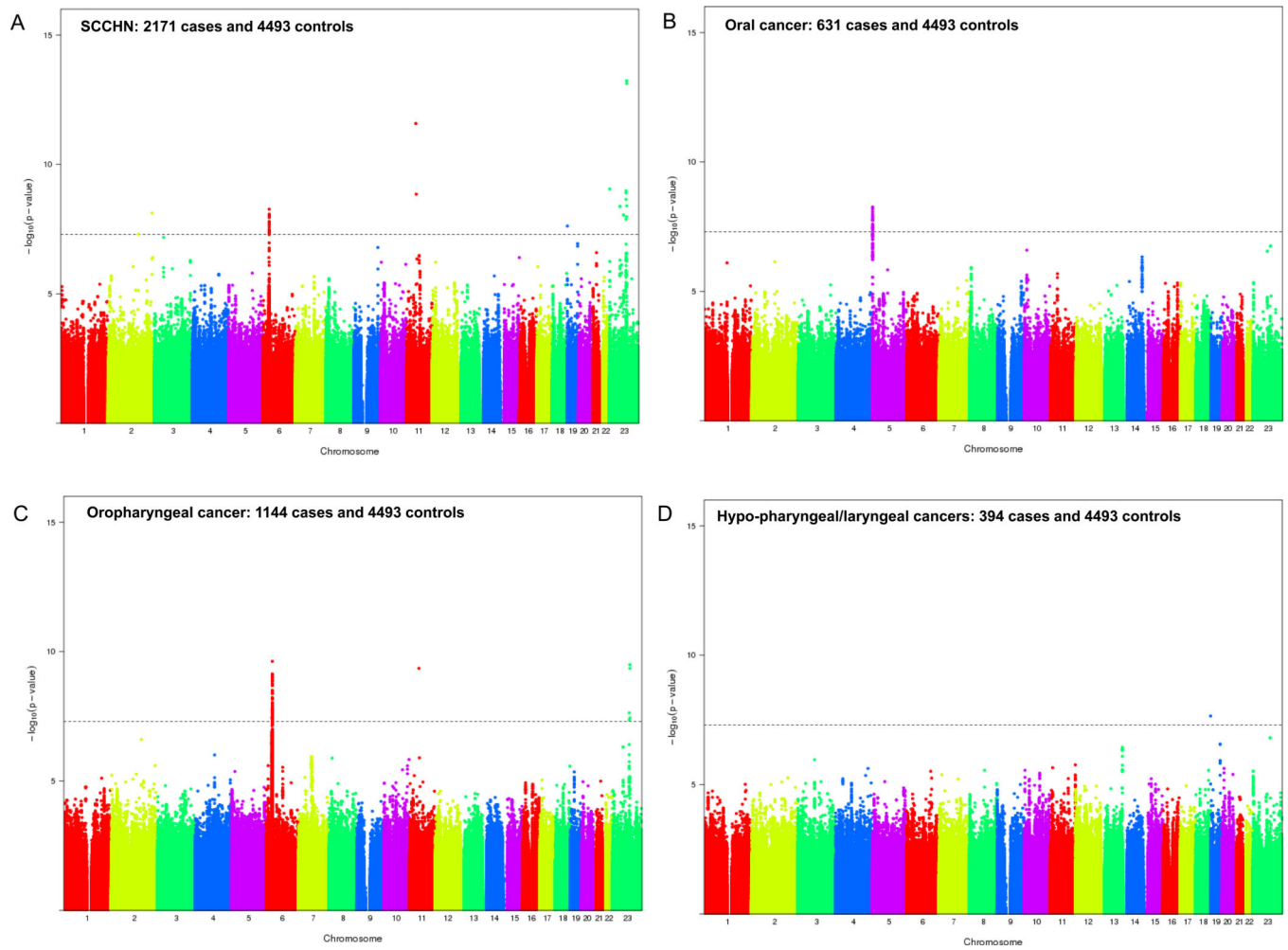


Figure 1. Manhattan plots of the association results in the discovery study: a) overall SCCHN risk; b) Oral cancer; c) Oropharyngeal cancer and d) Hypo-pharyngeal/laryngeal cancers. The dot line represents $P = 5 \times 10^{-8}$. The y-axis represents the $-\log_{10} P$ -values.

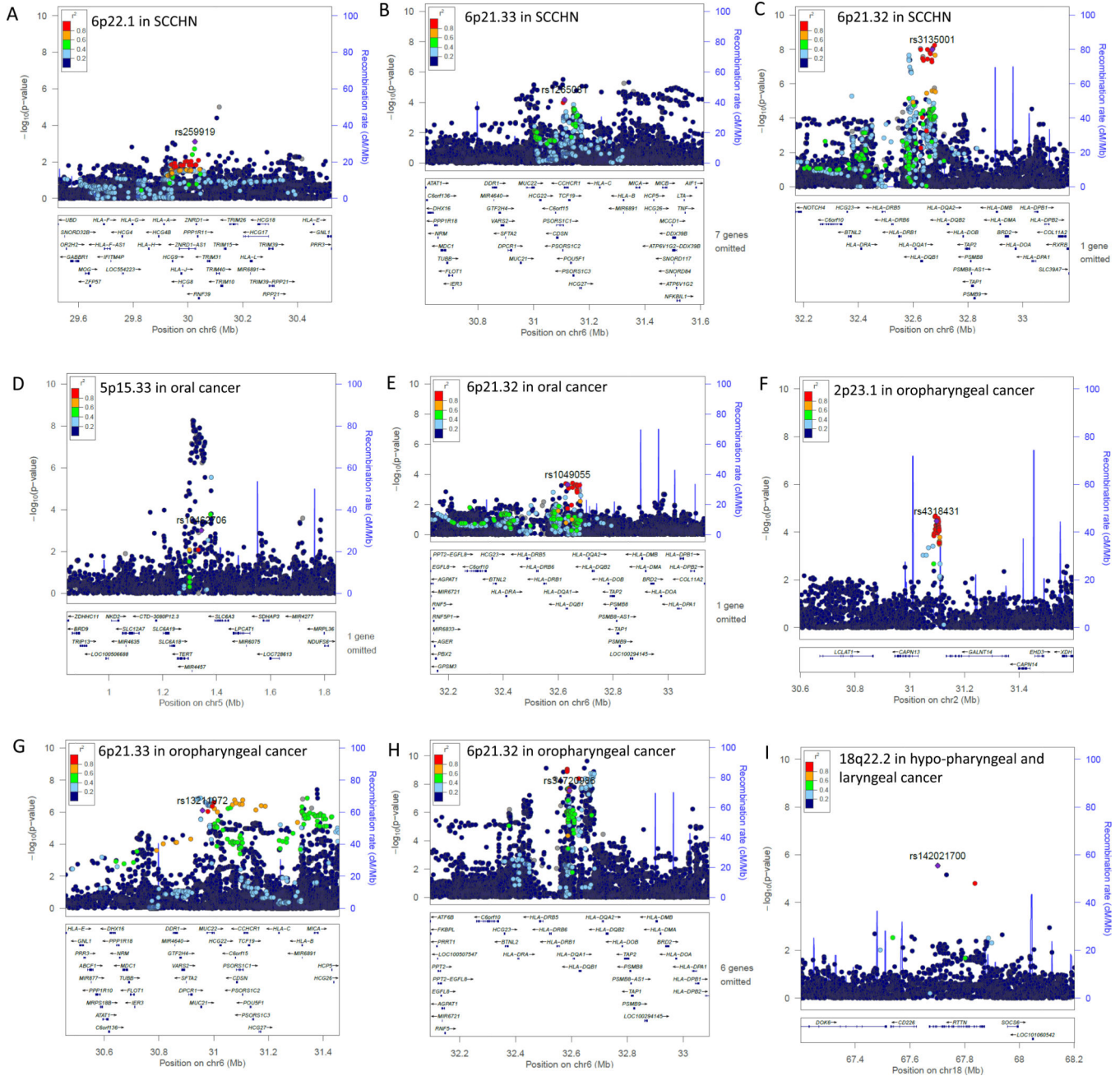


Figure 2. The genetic regions associated with SCCHN and three sub-types. a) 6p22.1 in overall SCCHN; b) 6p21.33 in overall SCCHN; c) 6p21.32 in overall SCCHN; d) 5p15.33 region in oral cancers; e) 6p21.32 in oral cancer; f) 2p23.1 in oropharyngeal cancer; g) 6p21.33 region in oral cancer; h) 6p21.32 region in oropharyngeal cancer; i) 18q22.2 in hypo-pharyngeal and laryngeal cancer. The association results were based the discovery study.

Table 1.

Distributions of population characteristics in the two-phase study

Variables	Discovery study (MDACC)			Replication study (OncoArray)			
	Cases ¹ # (N =2,171)	Controls # (N =4,493)	%	Cases ² # (N =5,205)	Controls # (N =3,232)	%	P
Age							< 0.0001
Median (Range)	57 (18–94)	49 (18–89)		59 (18–94)	58 (17–89)		
Mean (SD)	57.9 (11.2)	50.0 (12.3)		59.7 (10.9)	58.1 (11.5)		
Sex							0.001
Female	494	2,018	77.2	1,344	940	25.8	29.1
Male	1,677	2,475	22.8	3,861	2,292	74.2	70.9
Tumor sites							
Oral cancer	631	29.1		2,568	49.5		
Oropharynx	1,144	52.7		2,328	44.8		
Hypo-pharynx & larynx & other sites	394	28.2		295	5.7		

¹Two cases was missing the tumor site information in the discovery study from the MDACC (The University of Texas MD Anderson Cancer Center) GWAS

²In the replication study from the OncoArray GWAS, there are 14 cases with missing site information.

Table 2.

Association results of leading SNPs with P -value 1×10^{-3} in the discovery dataset and $P < 5 \times 10^{-8}$ in the random-effects model of the final meta-analysis

Region	SNP	Chr:Pos (hg19)	Gene	Eff/Ref	MDACC			OncoArray			Meta-analysis		
					Cases/ controls ¹	OR (95% CI) ²	P^2	Cases/ controls ¹	OR (95% CI) ³	P^3	OR (95% CI) ⁴	P^4	
Overall SCCHN													
6p22.1	rs259919	6:30025503	ZNRD1-AS1	A/G	0.34/0.31	1.15(1.06–1.25)	7.65E-04	0.34/0.28	1.19 (1.11–1.28)	8.51E-07	1.17 (1.11–1.24)	2.96E-09	
6p21.33	rs1265081	6:311111675	CCHCR1	C/A	0.47/0.5	0.85 (0.79–0.92)	6.71E-05	0.45/0.48	0.85 (0.80,91)	1.35E-06	0.85 (0.81–0.9)	3.75E-10	
6p21.32	rs3135001	6: 32670136	HLA-DQB1	T/C	0.21/0.25	0.76 (0.69–0.83)	9.89E-09	0.19/0.21	0.78 (0.72–0.85)	2.35E-09	0.77 (0.73–0.82)	1.44E-16	
Oral cavity cancer													
5p15.33	rs10462706	5:1343794	CLPTMIL	T/C	0.12/0.15	0.72(0.60–0.88)	9.65E-04	0.14/0.16	0.73 (0.65–0.81)	2.10E-08	0.73 (0.66–0.80)	7.87E-11	
6p21.32	rs1049055	6:32634387	HLA-DQB1	C/T	0.23/0.27	0.76(0.66–0.89)	4.86E-04	0.19/0.22	0.79 (0.72–0.87)	1.45E-06	0.78 (0.72–0.85)	2.96E-09	
Oropharyngeal cancer													
2p23.1	rs4318431	2:31098065	GALNT14	T/C	0.10/0.08	1.43 (1.21–1.69)	3.40E-05	0.10/0.08	1.37 (1.18–1.58)	2.14E-05	1.39 (1.25–1.55)	3.13E-09	
6p21.33	rs13211972	6:30959001	MUC21	A/G	0.08/0.05	1.64 (1.35–1.99)	7.79E-07	0.06/0.04	1.48 (1.23–1.77)	2.35E-05	1.55 (1.36–1.77)	1.04E-10	
6p21.32	rs34518860	6:32594103	HLA-DQA1	A/G	0.06/0.11	0.57 (0.48–0.69)	6.6E-09	0.07/0.11	0.63 (0.54–0.73)	5.40E-10	0.61 (0.54–0.68)	2.61E-17	
Hypo-pharyngeal and laryngeal cancer													
18q22.2	rs142021700	18:67701583	RITN	C/T	0.03/0.01	4.03 (2.25–7.21)	2.83E-06	0.03/0.01	3.84 (1.88–7.85)	2.28E-04	3.95 (2.51–6.21)	2.54E-09	

Abbreviations: Chr:Pos = Chromosome: position; Eff/Ref = Effect allele/reference allele; OR = odds ratio; CI = confidence interval; MDACC = The discovery study from the MD Anderson Cancer Center; OncoArray = the replication study from the OncoArray study

¹ Minor allele frequency in cases and controls

² Adjusted for top five significant principal components, age and sex in the MDACC study with 2,171 SCCHN cases, 631 patients with oral cancer, 1144 patients with oropharyngeal cancer, 394 patients with hypo-pharyngeal and laryngeal cancer, vs 4493 controls

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

³ Adjusted for top three significant principal components, age, sex and continent in the OncoArray study with 5205 SCCHN cases, 2568 patients with oral cavity cancer, 2368 patients with oropharyngeal cancer, 295 patients with hypo-pharyngeal and laryngeal cancer vs. 3232 controls

⁴ Meta-analysis with random-effects model.