HHS Public Access

Author manuscript

Int J Cancer. Author manuscript; available in PMC 2023 August 15.

Published in final edited form as:

Int J Cancer. 2022 August 15; 151(4): 553–564. doi:10.1002/ijc.34023.

Genetic variants in *CYP2B6* and *HSD17B12* associated with the risk of squamous cell carcinoma of the head and neck

Hongliang Liu[1,2],

Guojun Li[3],

Erich M. Sturgis[3],

Sanjay Shete[4,5],

Kristina R. Dahlstrom[5],

Mulong Du[6,7],

Christopher I. Amos[8],

David C. Christiani[6,9],

Philip Lazarus[10,*],

Qingyi Wei[1,2,*]

[1]Duke Cancer Institute, Duke University Medical Center, Durham, NC, USA.

[2]Department of Population Health Sciences, Duke University Medical School, Durham, NC, USA.

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

[4]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

[5]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[6]Department of Environmental Health, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA, 02115, USA.

[*]Address for correspondence and reprints: Philip Lazarus, Department of Pharmaceutical Sciences, College of Pharmacy and Pharmaceutical Sciences, Washington State University, Spokane, WA, 99210, USA, , Phone: 509-358-7947; phil.lazarus@wsu.edu 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.

Author Contributions

The work reported in the paper has been performed by the authors, unless clearly specified in the text.

1) Conception and design of the study: Philip Lazarus, Qingyi Wei

2) Data curation: Hongliang Liu, Guojun Li, Erich M. Sturgis, Sanjay Shete, Kristina R. Dahlstrom, Mulong Du, Christopher I. Amos, David C. Christiani, Qingyi Wei

3) Formal Analysis: Hongliang Liu

4) Writing – original draft: Hongliang Liu

5) Writing – review & editing: David C. Christiani, Philip Lazarus, Qingyi Wei

Conflict of Interest

All authors declare no conflict of interest.

Ethics Statement

All participants in the discovery dataset signed a written informed consent form that permitted the use of the collected blood samples and clinic-pathological information. The study protocols were approved by the Institutional Review Board of MDACC in accordance with the Declaration of Helsinki. The present study also used the data collected by the protocol approved by both the Internal Review Board of Duke University School of Medicine (#Pro00054575) and the dbGaP database administration (Project #7356).

[7]Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, 101 Longmian Avenue, Nanjing, 211166, Jiangsu, China.

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

[9]Department of Medicine, Massachusetts General Hospital, 55 Fruit Street, Boston, MA, 02114, USA.

[10]Department of Pharmaceutical Sciences, College of Pharmacy and Pharmaceutical Sciences, Washington State University, Spokane, WA, 99210, USA.

Abstract

Polycyclic aromatic hydrocarbons (PAHs) and tobacco-specific nitrosamines (TSNA) metabolism-related genes play important role in the development of cancers. We assessed the associations of genetic variants in genes involved in the metabolism of PAHs and TSNA, and the risk of squamous cell carcinoma of the head and neck (SCCHN) in European populations using two published genome-wide association study datasets. In the single-locus analysis, we identified two SNPs (rs145533669 and rs35246205) in *CYP2B6* to be associated with risk of SCCHN ($P = 1.57\times10^{-4}$ and 0.004), two SNPs (*EPHX1* rs117522494 and *CYP2B6* rs145533669) to be associated with risk of oropharyngeal cancer ($P = 0.001$ and 0.004), and one SNPs (rs4359199 in *HSD17B12*) to be associated with risk of oral cancer ($P = 0.006$). Significant interaction effect was found between rs4359199 and drinking status on risk of SCCHN and oropharyngeal cancer ($P < 0.05$). eQTL and sQTL analyzes revealed that two SNPs (*CYP2B6* rs35246205 and *HSD17B12* rs4359199) were correlated with alternative splicing or mRNA expression levels of the corresponding genes in liver cells ($P < 0.05$). *In-silico* functional annotation suggested that these two SNPs may regulate mRNA expression by affecting the binding of transcription factors. Results from phenome-wide association studies presented significant associations between these genes and risks of other cancers, smoking behavior, and alcohol dependence ($P < 0.05$). Our study provided insight into the underlying genetic mechanism of head and neck cancer and warranted future functional validation.

Keywords

PAH metabolism; tobacco-specific nitrosamines; SNP; squamous cell carcinoma of the head and neck; GWAS; molecular epidemiology

Introduction

Squamous cell carcinoma of the head and neck (SCCHN) starts from the mucosal epithelium in the oral cavity, pharynx and larynx. It is the sixth most common malignancy and seventh leading cause of cancer-related deaths worldwide [1, 2]. In the United States, there are approximately 65,630 new SCCHN cases and 14,500 SCCHN-related deaths in 2020 [3]. Epidemiology studies have revealed some important risk factors of SCCHN, such as exposure to tobacco and alcohol as well as infection with human papillomavirus (HPV) [4–6]. However, only a fraction of tobacco users, alcohol drinkers or individuals who contracted HPV develop SCCHN, suggesting an important role of genetic susceptibility in its etiology

[7, 8]. Two previous studies have reported that polymorphisms in alcohol-related genes including alcohol-dehydrogenase 1B (*ADH1B*) and *ADH7* were associated with risk of head and neck cancer. Recent genome-wide association studies (GWASs) have also revealed multiple loci to be associated with risk of SCCHN and its subtypes (i.e., 2p23.3, 5p15.33, 5q14.3, 6q16.1, 6p21.33, 6p21.32, 9p15.3, 9q34.12, 10q26.13, 11p15.4, 11q12.2, 12q24.21 and 16p13.2) [9–11].

Polycyclic aromatic hydrocarbons (PAHs) and tobacco-specific nitrosamines (TSNA) are two of the most carcinogenic components of tobacco smoke. Accumulating evidence has demonstrated that variability in metabolic enzymes important in the bioactivation or inactivation of PAHs and TSNA is linked to risk of many cancer types [12–14]. Recent GWASs with large sample sizes also reported associations between genetic variants in metabolism-related genes and risk for multiple cancers [i.e., lung cancer [15, 16], bladder cancer [17], colorectal cancer [18]], but these hypothesis-free GWASs have not reported any SNPs in some well-known metabolic genes that were identified for head and neck cancer in previous molecular epidemiology studies using a candidate gene approach [7, 19, 20]. In the present study, we used a hypothesis-driven approach to comprehensively assess associations of genetic variants in 43 PAH and TSNA metabolism-related genes and the risk of SCCHN as well as its subtypes (e.g., oral and oropharyngeal cancers) in European populations using two published head and neck GWAS datasets, and explored interaction effects between identified SNPs and smoking/drinking status, as well as possible functional annotation of the identified SNPs.

Materials and Methods

Discovery population

The discovery dataset included 2,171 SCCHN cases ascertained at the Head and Neck Surgery Clinic at the University of Texas MD Anderson Cancer Center (MDACC; Houston, TX) between December 1996 and July 2011[11, 21, 22]. All cases were individuals with newly diagnosed, histologically confirmed, previously untreated squamous cell carcinoma of the oral cavity, oropharynx, hypopharynx, or larynx. Blood samples were collected for genomic DNA extraction and genotyping was performed with Illumina HumanOmniExpress-12v1 BeadChip[11]. This study included 4,493 cancer-free controls, who were recruited from genetically unrelated visitors who accompanied cancer patients to MDACC outpatient clinics, or from the MDACC melanoma study with GWAS data deposited in database of Genotypes and Phenotypes (with dbGaP accession#: phs000187.v1.p1) [21–23], or from the Study of Addiction, Genetics and Environment (SAGE) (SAGE; dbGaP accession #: phs000092.v1.p1) [24]. Of these controls, there were 1,149 cancer-free individuals recruited for the SCCHN study with genomic DNA genotyped by using Illumina HumanOmniExpress-12v1 BeadChip; 1,022 cancer-free individuals from the melanoma GWA study whose genomic DNA was genotyped by Illumina Omni1-Quad_v1–0_B BeadChip, and 2,322 cancer-free individuals of European descendent from the SAGE study [24] whose genotyping data were generated by Illumina Human1Mv1 BeadChip. All of these three GWAS datasets contained around 1 million genotyped SNPs. The SCCHN GWAS data is available in dbGaP (accession #: phs001173.v1.p1) [11].

Replication Population

The replication dataset was from the OncoArray study of oral and pharynx cancer, which is part of the International Head and Neck Cancer Epidemiology Consortium (INHANCE) and included 6,034 cases and 6,585 controls derived from 12 epidemiological studies in European populations [9]. Genomic DNA extracted from blood or buccal cells was genotyped at the Center for Inherited Disease Research (CIDR) with the Illumina OncoArray custom array. Related GWAS data were requested from dbGaP (accession#: phs001202.v1.p1) in which genotyping data were available for 6,034 cases and 4,062 controls. After quality control [11], 5,205 cases and 3,232 controls of European ancestry were included in the present study.

GWAS data extraction

Based on previous publications, we selected 43 genes encoding enzymes that play a key role in the metabolism of either PAHs or TSNAs [7, 14, 25]. These genes are listed in Supplementary Table 1. Imputation was performed on the Michigan imputation server (https://imputationserver.sph.umich.edu) with the Haplotype Reference Consortium (HRC) reference panel (Version r1.1 2016) consisting of 64,940 haplotypes for individuals of predominantly European ancestry. In the discovery stage, the genotyping/imputed data was extracted for 4,177 SNPs (555 genotyped and 3,622 imputed) with a minor allele frequency ≥ 0.01 and r-square ≥0.3 within 5 kb upstream and 5 kb downstream of the 43 metabolizing enzyme genes from the SCCHN GWAS of MDACC.

Statistical analysis

Single-locus analysis with an unconditional logistic regression model was performed to estimate odds ratios (ORs) and 95% confidence intervals (CIs) per effect allele using the PLINK (v2.00) software with adjustment for the age, sex and top five significant principal components (PCs). Stratified analysis by tumor site (i.e., oral, oropharyngeal, and hypo-pharyngeal and laryngeal cancer) was performed using the false discovery rate (FDR) followed by Bayesian false-discovery probability (BFDP) for multiple-testing correction (with the setting of prior probability = 0.1 and the upper bounder of detectable OR = 3). Because the SNPs under investigation were mostly from imputation (3,622 out of 4,177 SNPs), those SNPs with BFDP ≤ 0.8 were then examined in the OncoArray study. The summary results of both discovery and replication datasets were then examined by a meta-analyses with the inverse variance-weighted average method. A random-effects model was applied as a Q-test $P \leq 0.10$ or I^2 >50.0%; otherwise, a fixed-effects model was applied. eQTL and sQTL analyzes were applied to elucidate biological mechanisms of those identified genetic variants by using FIVEx (https://fivex.sph.umich.edu/). Phenome-wide association analysis (PheWAS) analysis was performed to investigate the association between identified genes and 600 traits from UK Biobank release 2 by using the online tool: https://atlas.ctglab.nl/PheWAS. The related methods can be found elsewhere [26]. Briefly, the MAGMA gene analysis was performed to test the gene-trait correlation. Firstly, K SNPs are assigned to genes with 1-kb window both sides. SNP-wise mean models were then used to combine SNP Z statistics into a gene test-statistics $\mathrm{T:T} = \sum_j^K Z_j^2$, where $Z_J = \Phi(p_j)$, Φ is the cumulative normal distribution function; p_j is the marginal P value of $\mathrm{SNP_j}$. Gene

based p-value: $P = Pr(T \geq T_{obs})$. In the evaluation of gene test-statistics, LD between SNPs in the gene is also be estimated to produce accurate results. We also used LocusZoom [27] and Haploview v4.2 [28] to construct the regional association plots and linkage disequilibrium (LD) plots, respectively.

Results

Characteristic distribution of the study populations

As shown in the study workflow depicted in Figure 1, this is a two-phase study design which included a discovery (MDACC) dataset and then a replication (OncoArray) dataset. Table 1 presents the distribution of characteristics in both the discovery and replication datasets. The mean age of cases and controls are 57.9 vs 50 years old in the discovery dataset and 59.7 vs 58.1 years old in the replication dataset, and there were more males in cases than in controls in both the discovery dataset(77.2% vs 55.1%) and replication dataset (74.2% vs 70.9%). There were 2,171 case and 2,169 controls with smoking and drinking data available in the discovery dataset. More smokers and drinkers were found in cases than in controls (68.4% vs. 47.0%, $P < 0.0001$; 73.5% vs. 54.6%, $P < 0.0001$, respectively). Of the cases in the discovery population, there were 631 (29.1%), 1,144 (52.7%), 316 (14.6%), and 78 (3.6%) patients with cancers of the oral cavity, oropharynx, laryngeal, and hypopharynx or overlapping cancer, respectively. In the replication dataset, there were 2,568 (49.5%), 2,328 (44.8%), 295 (5.7%) patients with oral cavity cancer, oropharyngeal cancer, and hypo-pharyngeal or overlapping cancer, respectively. Two and 14 patients in the discovery and replication datasets, respectively, were missing values for histological types.

Association analysis

Single-locus analysis was performed for each of the SNPs with an imputation quality $r^2 \geq 0.3$ and a minor allele frequency ≥ 0.01 in the MDACC discovery population. There were 212, 263, and 173 SNPs with $P < 0.050$ in the main effect analysis for SCCHN, and stratified analyses by oral and oropharyngeal, respectively (Supplementary Table 2-4). After multiple test correction, there are 110, 145, and 110 SNPs passed BFDP threshold ≤ 0.8 in both the main effect analysis and the stratified analyses by organ site (i.e., oral, and oropharyngeal), respectively (Supplementary Tables 2-4). No SNPs could pass the FDR correction (≤ 0.2). The Manhattan plots of these results are shown in Figure 2A-C.

SNPs that passed BFDP correction were then selected for replication with the OncoArray dataset, with 2, 2, and 1 SNPs replicated with $P < 0.05$ for an association with risk of SCCHN and oropharyngeal and oral cancer (Table 2). As there was no heterogeneity between the discovery and replication datasets (Q-test $P > 0.1$; I-squared < 50%), a meta-analysis of discovery and replication results was performed with the fixed-effects model (Table 2). In this model, 2 SNPs (rs145533669 and rs35246205) in *CYP2B6* within the 19p13.2 loci exhibited a significant association with SCCHN risk. The variant T allele of the leading SNP *CYP2B6* rs145533669 was associated with decreased SCCHN risk (OR = 0.60, 95% CI = 0.46–0.78, $P = 1.57 \times 10^{-4}$). Two SNPs (*EPHX1* rs117522494 within the 8p21.2-p21.1 loci and *CYP2B6* rs145533669 within the 19p13.2 loci) showed significant associations with oropharyngeal cancer risk, with the variant T alleles of SNPs rs117522494

and rs145533669 associated with increased (OR = 1.45, 95% CI: 1.16–1.82, $P = 0.001$) and decreased risk (OR = 0.62, 95% CI: 0.44–0.85, $P = 0.004$), respectively. In addition, one SNP (rs4359199) in *HSD17B12* exhibited significant associations with the risk of oral cancer, with the variant allele C associated with increases cancer risk (OR = 1.11, 95% CI: 1.04–1.18, $P = 0.0018$).

Considering only few SNPs identified in the above analyses, we calculated the observed power by using the online Genetic Association Study (GAS) Power Calculator (https://csg.sph.umich.edu/abecasis/gas_power_calculator/) to exclude the possibility of false negative results. As shown in Supplementary Table 5, the power in the oropharyngeal cancer group with 1,144 cases and 4,493 controls appears to be enough to detect the effects of SNPs with a minor allele frequency > 0.02 and a relative risk > 1.68. Otherwise, the power would be insufficient. For the analysis of the oral cancer group (with 631 cases and 4493 controls), the power was not adequate to exclude the type II error.

Regional association plots of the four SNPs were presented in Supplementary Figure 1A-D. Linkage disequilibrium (LD) analysis was performed for these four identified SNPs (Supplementary Figure 2). There is no LD observed for the two SNPs (rs145533669 and rs35246205) in *CYP2B6.* Two other SNPs (rs117522494 in *EPHX2* and rs4359199 in *HSD17B12*) were located at different chromosome regions (8p21.2 and 11p11.2).

Interaction analysis

As smoking and drinking are known risk factors for SCCHN, we also performed interaction analysis of the four identified SNPs with smoking and alcohol drinking. As shown in Supplementary Table 6, we found a marginal significance was found for the interaction effect between SNP rs117522494 in gene *EPHX2* and drinking status in oral cancer (interaction $P = 0.080$), and SNP rs4359199 in gene *HSD17B12* region exhibited significant interaction effects with drinking status in overall SCCHN (interaction $P = 0.028$) and oropharyngeal cancer (interaction $P = 0.033$). We then performed stratified analysis by smoking and drinking status for this SNP. As shown in Table 3, the variant C allele of SNP rs4359199 was associated with increased risk of SCCHN (OR = 1.10, 95%CI: 0.99–1.23, $P = 0.085$) and for oral cancer (OR = 1.22, 95%CI: 1.03–1.43, $P = 0.019$) in smokers. While a protective effect was observed for the T allele on oropharyngeal cancer risk in non-drinkers (OR = 0.82, 95%CI: 0.68–0.99, $P = 0.043$). No significant effects were observed in smokers or non-smokers for SCCHN or after stratification by oropharyngeal and oral cancers.

Gene-based test

To test the multiple-markers effect, gene-based analysis was performed with MAGMA, which is based on a multiple linear principal components regression model using an F-test to compute the gene p-value [29]. The results showed that the associations between ten metabolism-related genes (i.e., *AKR1A1, EPHX1, SULT1B1, UGT3A1, NAT2, AKR1C1, HSD17B12, SULT1A2, SULT1A1,* and *CYP2A6*) and SCCHN risk were statistically significant in the MDACC discovery dataset ($P < 0.05$, Supplementary Table 7). Two of the associations (for *AKR1A1* and *HSD17B12*) were replicated in the OncoArray dataset ($P = 0.053$ and 0.003, respectively).

In silico functional annotation (eQTL, sQTL and PheWAS)

For those replicated SNPs, the association results from eQTL analysis in multiple tissues were shown in Figure 3: A) rs35246205 and *CYP2B6*, C) rs117522494 and *EPHX2*, E) rs4359199 and *HSD17B12*, respectively. Multiple pairs were found with significant correlations in different tissues. As the PAHs and TSNA were mainly metabolized in liver and immunity capacity can influence the risk of SCCHN, we were more interested in the eQTL results in these related tissues. Significant correlations were found between rs4359199 and the mRNA expression of *HSD17B12* in blood cells (P = 4.9E-21), monocytes (P = 3.3E-9), and T cells (P = 1.74E-08) (Figure 3E). sQTL associations in multiple tissues were presented in Figure 3: B) rs35246205 and *CYP2B6*; D) of rs117522494 and *EPHX2*; F) rs4359199 and *HSD17B12*, respectively. Notable significant correlations were found between rs35246205 and *CYP2B6* in liver (P = 8.71E-04 and 0.011 in Figure 3B); rs4359199 and *HSD17B12* in LCL, T-cells, and liver cells (P = 1.0E-19, 1.1E-04, and 1.3E-04, respectively, in Figure 3F).

PheWAS results (with gene level $P < 0.05$) of the three identified genes and 600 traits in UK Biobank were presented in Figure 4A-C. As it shown, significant correlations were found between these genes and multiple traits. For example, there were significant association between *CYP2B6* and cigarettes used per day (P = 2.27E-11), squamous cell lung cancer risk (P = 0.004) (Figure 4A). *EPHX2* was also found to be correlated with traits including ever smoker (P = 2.73E-05), breast cancer risk (P = 0.012), Cadherin-15 contents in blood (P = 4.67E-06) (Figure 4B). *HSD17B12* was found to be associated with traits in the metabolic domain, including body Mass Index (P = 2.95E-28), type 2 Diabetes (P = 1.21E-10), number of cigarettes previously smoked daily (P = 4.35E-04), and lung cancer risk (P = 0.008) (Figure 4C).

Further functional annotation using HaploReg V4.1 also revealed that *CYP2B7P* rs117522494 is located in the potential promoter region with some evidence from DNAse, histone, and protein bound ChIP-Seq experiments (Supplementary Table 8). It should be noted that rs4359199 is located in the intron of gene *HSD17B12* but was correlated with 50 eQTL traits in GTEx.

Discussion

In this study, we performed a comprehensive candidate-pathway analysis of SNPs in 43 PAH- and TSNA-metabolizing genes and the risk of SCCHN by using a two-phase study design. We identified four SNPs in three gene regions (i.e., *CYP2B6, EPHX2, and HSD17B12*) associated with risk of SCCHN and its subtypes. We also revealed significant interaction effects between SNP rs4359199 in *HSD17B12* and drinking status. *In-silico* eQTL and sQTL analyses also provided functional evidence for the identified SNPs (rs35246205 and rs4359199) in *CYP2B6* and *HSD17B12, respectively*. Our results suggest that the variant alleles of the these identified SNPs might be associated with increased cancer risk of SCCHN and its subtypes by regulating corresponding gene expression or affecting mRNA alternative splicing.

Cytochrome P450 (CYP) enzymes are crucial for the metabolic activation of PAHs and TSNA, which are important carcinogens in cancer development. Dysfunctions of the relevant genes have been found to the increased risks of multiple cancers [25]. In this study, we found SNPs located in the *CYP2B6* region were associated with the risk of SCCHN and oropharyngeal cancer. Our eQTL and sQTL results showed that the identified genetic variant rs35246205 might be a tagging SNP of other functional variants influencing the alternative splicing of *CYP2B6* in liver. Polymorphisms in this gene have also been associated with the risk of breast cancer and acute myeloid leukemia [30, 31]. CYP2B6 and other family members, such as CYP2A6, are mainly responsible for converting of BaP and NNK to their intermediates that produce DNA adducts [32]. Although inconsistent results were also reported [33], several studies reported a potential link between this gene and nicotine metabolism [34–36], which were consistent with the PheWAS results and suggested that CYP2B6 might also contribute to squamous cell cancer risk by changing smoking behavior.

Two SNPs in two other gene (*EPHX2 and HSD17B12) were also* associated with the risk of oropharyngeal cancer and hypo-pharyngeal and laryngeal cancer, respectively, in the present study. EPHX2 is a member of the epoxide hydrolase family whose encoded enzyme can bind to specific epoxides (the intermediates from PAH) and convert them to the corresponding dihydrodiols [37]. One study has reported that the expression of *EPHX2* was higher in an oropharyngeal HPV-positive cohort than in the negative cohort from The Cancer Genome Atlas (TCGA) [38]. Although previous reports suggested that EPHX2 might be involved in the regulation of smoking induced inflammation and autophagy, the underlying function of EPHX2 involved in oropharyngeal is still unknown [39, 40]. Smoking and drinking are two important independent risk factor for the development of oral and pharyngeal cancers. In the present study, we found a significant interaction effect between *HSD17B12* SNP rs4359199 and drinking status in oropharyngeal cancer. The variant allele C presented a protective effect on oropharyngeal cancer risk in non-drinkers, but increased the risk of oral cancer in drinkers. However, few reported studies have investigated the interaction effect of this genes and the underlying mechanism is still unclear. HSD17B12 play an important role in the regulation of sex steroid metabolism by catalyzing the conversion between 17-keto and 17-hydroxysteroids. The dysregulation of this gene has been implicated in multiple estrogen and androgen-related diseases including breast cancer, endometriosis, prostate and non-small cell lung cancer [41, 42]. Polymorphisms in this gene have been linked to risk for multiple cancers including gastric cancer, endometrial cancer, breast cancer and prostate cancer [43–46]. PheWAS results showed that the identified SNP involves multiple metabolic-related phenotypes. This is consistent with the current findings that HSD17B12 may play its role in inflammation and cancer development by upregulating fatty acid synthesis in human cancers.[47] While the present study is the first to report on associations of genetic variants in these two gene and risk for SCCHN and its sub-types, further replications are required.

There are several limitations to this study. Firstly, the underlying biological mechanism of the identified associations are still unclear. Although we have provided some *in-silico* functional evidence, most of them were from whole blood or other tissues but not the target tissue; additional evidence in target tissues or cell lines will be required. Second, the smoking and alcohol data are not available in the OncoArray study. The identified interaction effects in the MDACC study need to be further replicated in another independent

study. Thirdly, HPV infection is one of the primary risk factors in oropharyngeal cancer. However, we did not have adequate HPV data available to detect the possible interactions between SNPs and HPV status. In addition, SNPs with minor effects or low frequency cannot be identified due to sample size limitations of the MDACC and OncoArray studies.

In conclusion, we found that two SCCHN risk-associated SNPs (rs35246205 in *CYP2B6* and rs4359199 in *HSD17B12*) were correlated with mRNA expression levels of the corresponding genes and that two SNPs, rs117522494 in *CYP2B6* and rs4359199 in *HSD17B12*, are located at the potential promoter regions and may regulate mRNA expression and alternative splicing. Further functional validation and population replication are warranted to substantiate the present findings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank all participants who have contributed their samples and clinical data for this study and we also thank ILCCO members, who provided access to samples and clinical data. This work was partially supported by R01 ES025460–01 from National Institute of Environmental Health Sciences (PI: P. Lazarus). 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). 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.

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.

Data Availability Statement

The GWAS dataset used in this study can be requested from dbGaP (https://www.ncbi.nlm.nih.gov/gap/) with Accession#: phs001173.v1.p1 for the SCCHN GWAS, phs000187.v1.p1 for melanoma GWAS, phs000092 for the SAGE GWAS, and phs001202.v1.p1 for the OncoArray study. Further information is available from the corresponding authors upon request.

Abbreviations:

BFDP	Bayesian false-discovery probability
CI	confidence intervals
eQTL	expression quantitative trait loci
FDR	false discovery rate
GWAS	genome-wide association study
MDACC	MD Anderson Cancer Center
OR	odds ratio
PAHs	polycyclic aromatic hydrocarbons
PheWAS	phenome-wide association study
SAGE	the Study of Addiction, Genetics and Environment
SCCHN	squamous cell carcinoma of the head and neck
sQTL	splicing quantitative trait loci
TSNA	tobacco-specific nitrosamines

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 RL, Miller KD, Jemal A. Cancer statistics, 2020. CA: a cancer journal for clinicians 2020;70: 7–30. [PubMed: 31912902]

4. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, Jiang B, Goodman MT, Sibug-Saber M, Cozen W, Liu L, Lynch CF, 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, Dal Maso L, Daudt AW, Fabianova E, Fernandez L, Wunsch-Filho V, Franceschi S, 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. 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]
8. 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]
9. Lesseur C, Diergaarde B, Olshan AF, Wunsch-Filho V, Ness AR, Liu G, Lacko M, Eluf-Neto J, Franceschi S, Lagiou P, Macfarlane GJ, Richiardi L, et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. Nature genetics 2016;48: 1544–50. [PubMed: 27749845]
10. Wei Q, Yu D, Liu M, Wang M, Zhao M, Liu M, Jia W, Ma H, Fang J, Xu W, Chen K, Xu Z, 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]
11. Shete S, Liu H, Wang J, Yu R, Sturgis EM, Li G, Dahlstrom KR, Liu Z, Amos CI, Wei Q. A Genome-Wide Association Study Identifies Two Novel Susceptible Regions for Squamous Cell Carcinoma of the Head and Neck. Cancer research 2020;80: 2451–60. [PubMed: 32276964]
12. Rodriguez-Antona C, Ingelman-Sundberg M. Cytochrome P450 pharmacogenetics and cancer. Oncogene 2006;25: 1679–91. [PubMed: 16550168]
13. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144: 646–74. [PubMed: 21376230]
14. Shultz CA, Quinn AM, Park JH, Harvey RG, Bolton JL, Maser E, Penning TM. Specificity of human aldo-keto reductases, NAD(P)H:quinone oxidoreductase, and carbonyl reductases to redox-cycle polycyclic aromatic hydrocarbon diones and 4-hydroxyequilenin-o-quinone. Chem Res Toxicol 2011;24: 2153–66. [PubMed: 21910479]
15. Patel YM, Park SL, Han Y, Wilkens LR, Bickeboller H, Rosenberger A, Caporaso N, Landi MT, Bruske I, Risch A, Wei Y, Christiani DC, et al. Novel Association of Genetic Markers Affecting CYP2A6 Activity and Lung Cancer Risk. Cancer research 2016;76: 5768–76. [PubMed: 27488534]
16. McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, Caporaso NE, Johansson M, Xiao X, Li Y, Byun J, Dunning A, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nature genetics 2017;49: 1126–32. [PubMed: 28604730]
17. de Maturana EL, Rava M, Anumudu C, Saez O, Alonso D, Malats N. Bladder Cancer Genetic Susceptibility. A Systematic Review. Bladder Cancer 2018;4: 215–26. [PubMed: 29732392]
18. Zeng C, Matsuda K, Jia WH, Chang J, Kweon SS, Xiang YB, Shin A, Jee SH, Kim DH, Zhang B, Cai Q, Guo X, et al. Identification of Susceptibility Loci and Genes for Colorectal Cancer Risk. Gastroenterology 2016;150: 1633–45. [PubMed: 26965516]
19. Cadoni G, Boccia S, Petrelli L, Di Giannantonio P, Arzani D, Giorgio A, De Feo E, Pandolfini M, Galli P, Paludetti G, Ricciardi G. A review of genetic epidemiology of head and neck cancer related to polymorphisms in metabolic genes, cell cycle control and alcohol metabolism. Acta Otorhinolaryngol Ital 2012;32: 1–11. [PubMed: 22500060]

20. Sabitha K, Reddy MV, Jamil K. Smoking related risk involved in individuals carrying genetic variants of CYP1A1 gene in head and neck cancer. Cancer Epidemiol 2010;34: 587–92. [PubMed: 20887941]
21. Neumann AS, Lyons HJ, Shen H, Liu Z, Shi Q, Sturgis EM, Shete S, Spitz MR, El-Naggar A, Hong WK, Wei Q. 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]
22. Li G, Sturgis EM, Wang LE, Chamberlain RM, Amos CI, Spitz MR, El-Naggar AK, Hong WK, Wei Q. 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]
23. Amos CI, Wang LE, Lee JE, Gershenwald JE, Chen WV, Fang S, Kosoy R, Zhang M, Qureshi AA, Vattathil S, Schacherer CW, Gardner JM, et al. Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. Human molecular genetics 2011;20: 5012–23. [PubMed: 21926416]
24. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, et al. The NCBI dbGaP database of genotypes and phenotypes. Nature genetics 2007;39: 1181–6. [PubMed: 17898773]
25. Modesto JL, Hull A, Angstadt AY, Berg A, Gallagher CJ, Lazarus P, Muscat JE. NNK reduction pathway gene polymorphisms and risk of lung cancer. Molecular carcinogenesis 2015;54 Suppl 1: E94–E102. [PubMed: 24976539]
26. Watanabe K, Stringer S, Frei O, Umicevic Mirkov M, de Leeuw C, Polderman TJC, van der Sluis S, Andreassen OA, Neale BM, Posthuma D. A global overview of pleiotropy and genetic architecture in complex traits. Nature genetics 2019;51: 1339–48. [PubMed: 31427789]
27. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26: 2336–7. [PubMed: 20634204]
28. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21: 263–5. [PubMed: 15297300]
29. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015;11: e1004219. [PubMed: 25885710]
30. Justenhoven C, Pentimalli D, Rabstein S, Harth V, Lotz A, Pesch B, Bruning T, Dork T, Schurmann P, Bogdanova N, Park-Simon TW, Couch FJ, et al. CYP2B6*6 is associated with increased breast cancer risk. International journal of cancer Journal international du cancer 2014;134: 426–30. [PubMed: 23824676]
31. Daraki A, Zachaki S, Koromila T, Diamantopoulou P, Pantelias GE, Sambani C, Aleporou V, Kollia P, Manola KN. The G(5)(1)(6)T CYP2B6 germline polymorphism affects the risk of acute myeloid leukemia and is associated with specific chromosomal abnormalities. PloS one 2014;9: e88879. [PubMed: 24586425]
32. Shimada T, Fujii-Kuriyama Y. Metabolic activation of polycyclic aromatic hydrocarbons to carcinogens by cytochromes P450 1A1 and 1B1. Cancer Sci 2004;95: 1–6. [PubMed: 14720319]
33. Lee AM, Jepson C, Shields PG, Benowitz N, Lerman C, Tyndale RF. CYP2B6 genotype does not alter nicotine metabolism, plasma levels, or abstinence with nicotine replacement therapy. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2007;16: 1312–4.
34. Al Koudsi N, Tyndale RF. Hepatic CYP2B6 is altered by genetic, physiologic, and environmental factors but plays little role in nicotine metabolism. Xenobiotica 2010;40: 381–92. [PubMed: 20307138]
35. Dicke KE, Skrlin SM, Murphy SE. Nicotine and 4-(methylnitrosamino)-1-(3-pyridyl)-butanone metabolism by cytochrome P450 2B6. Drug Metab Dispos 2005;33: 1760–4. [PubMed: 16174803]
36. Bloom AJ, Wang PF, Kharasch ED. Nicotine oxidation by genetic variants of CYP2B6 and in human brain microsomes. Pharmacol Res Perspect 2019;7: e00468. [PubMed: 30906561]
37. Gautheron J, Jeru I. The Multifaceted Role of Epoxide Hydrolases in Human Health and Disease. Int J Mol Sci 2020;22.

38. Reddy RB, Khora SS, Suresh A. Molecular prognosticators in clinically and pathologically distinct cohorts of head and neck squamous cell carcinoma-A meta-analysis approach. PloS one 2019;14: e0218989. [PubMed: 31310629]

39. Pedro NF, Biselli JM, Maniglia JV, Santi-Neto D, Pavarino EC, Goloni-Bertollo EM, Biselli-Chicote PM. Candidate Biomarkers for Oral Squamous Cell Carcinoma: Differential Expression of Oxidative Stress-Related Genes. Asian Pac J Cancer Prev 2018;19: 1343–9. [PubMed: 29802697]

40. Li Y, Yu G, Yuan S, Tan C, Lian P, Fu L, Hou Q, Xu B, Wang H. Cigarette Smoke-Induced Pulmonary Inflammation and Autophagy Are Attenuated in Ephx2-Deficient Mice. Inflammation 2017;40: 497–510. [PubMed: 28028752]

41. He W, Gauri M, Li T, Wang R, Lin SX. Current knowledge of the multifunctional 17beta-hydroxysteroid dehydrogenase type 1 (HSD17B1). Gene 2016;588: 54–61. [PubMed: 27102893]

42. Hilborn E, Stal O, Jansson A. Estrogen and androgen-converting enzymes 17beta-hydroxysteroid dehydrogenase and their involvement in cancer: with a special focus on 17beta-hydroxysteroid dehydrogenase type 1, 2, and breast cancer. Oncotarget 2017;8: 30552–62. [PubMed: 28430630]

43. Shi L, Yang X, Dong X, Zhang B. Polymorphism of HSD17B1 Ser312Gly with Cancer Risk: Evidence from 66,147 Subjects. Twin Res Hum Genet 2016;19: 136–45. [PubMed: 26923943]

44. Cho LY, Yang JJ, Ko KP, Ma SH, Shin A, Choi BY, Han DS, Song KS, Kim YS, Chang SH, Shin HR, Kang D, et al. Genetic susceptibility factors on genes involved in the steroid hormone biosynthesis pathway and progesterone receptor for gastric cancer risk. PloS one 2012;7: e47603. [PubMed: 23110082]

45. Setiawan VW, Hankinson SE, Colditz GA, Hunter DJ, De Vivo I. HSD17B1 gene polymorphisms and risk of endometrial and breast cancer. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2004;13: 213–9.

46. Kraft P, Pharoah P, Chanock SJ, Albanes D, Kolonel LN, Hayes RB, Altshuler D, Andriole G, Berg C, Boeing H, Burtt NP, Bueno-de-Mesquita B, et al. Genetic variation in the HSD17B1 gene and risk of prostate cancer. PLoS genetics 2005;1: e68. [PubMed: 16311626]

47. Nagasaki S, Suzuki T, Miki Y, Akahira J, Kitada K, Ishida T, Handa H, Ohuchi N, Sasano H. 17Beta-hydroxysteroid dehydrogenase type 12 in human breast carcinoma: a prognostic factor via potential regulation of fatty acid synthesis. Cancer research 2009;69: 1392–9. [PubMed: 19190350]

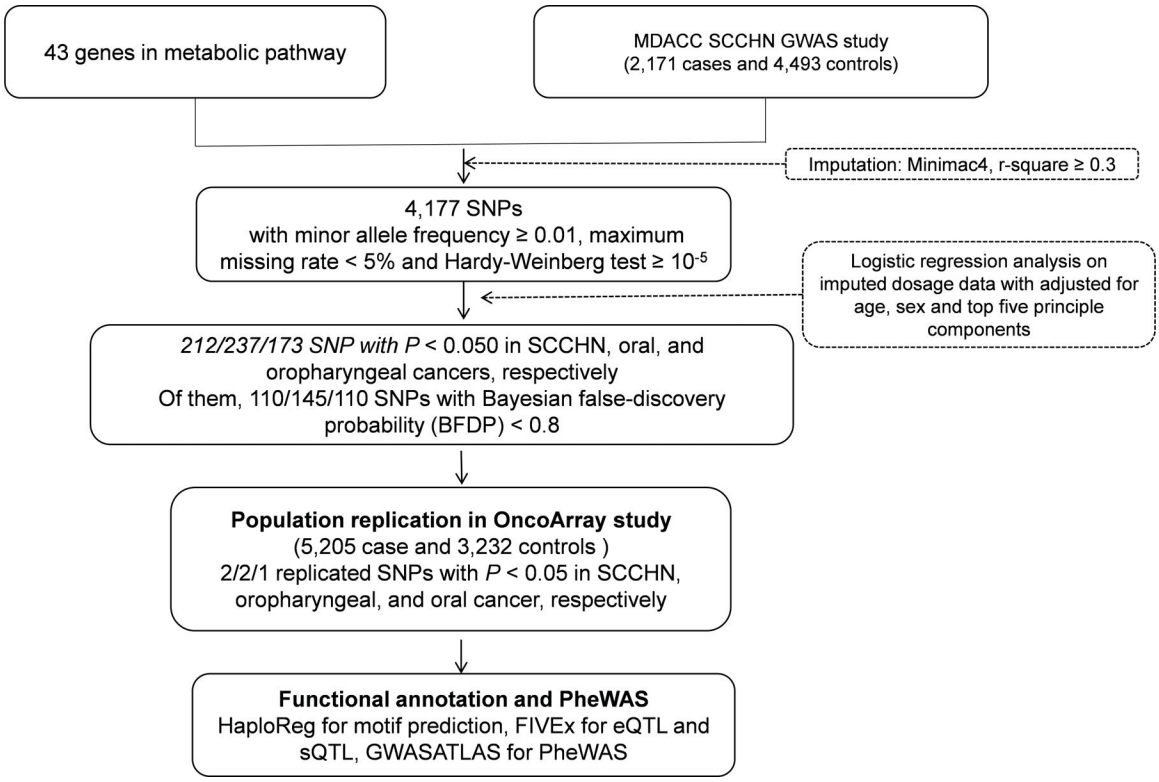

Figure 1.
Study flowchart.

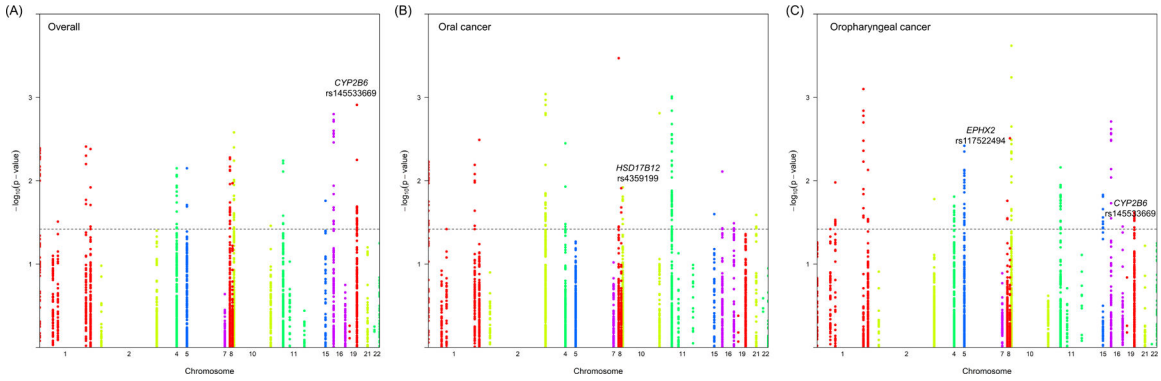

Figure 2.

Manhattan plots of the association results of (A) SCCHN; (B) Oral cancer; and (C) Oropharyngeal cancer.

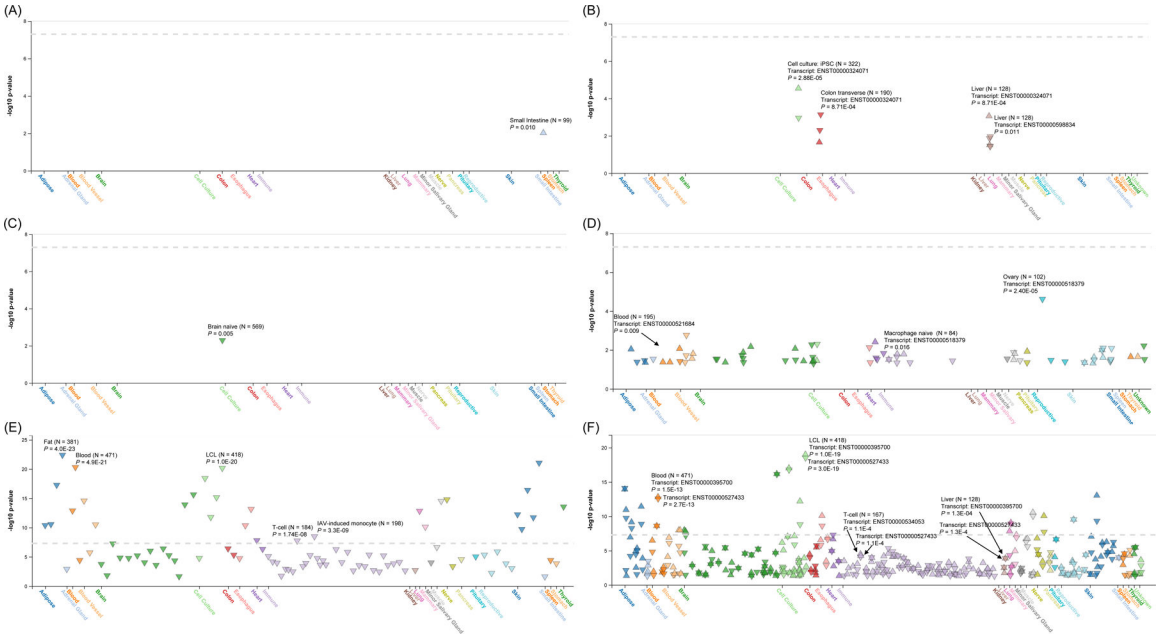

Figure 3.

Multiple-tissue eQTL and sQTL analyses.

eQTL results were presented in A) rs35246205 and *CYP2B6*, C) of rs117522494 and *EPHX2*, E) rs4359199 and *HSD17B12,* respectively. SNP rs4359199 showed significant correlation with mRNA expression in lymphocytes (i.e., T cell, monocytes)

sQTL results were presented in B) rs35246205 and *CYP2B6*, D) of rs117522494 and *EPHX2*, F) rs4359199 and *HSD17B12,* respectively. Notable sQTL associations were found between rs35246205 and two transcripts of *CYP2B6* in liver cells (*P* = 8.71E-04 and 0.011, respectively), rs4359199 and *HSD17B12* in lymphoblastoid cell lines (LCL) and liver cells (*P* = 1E-19 and 1.3E-04, respectively).

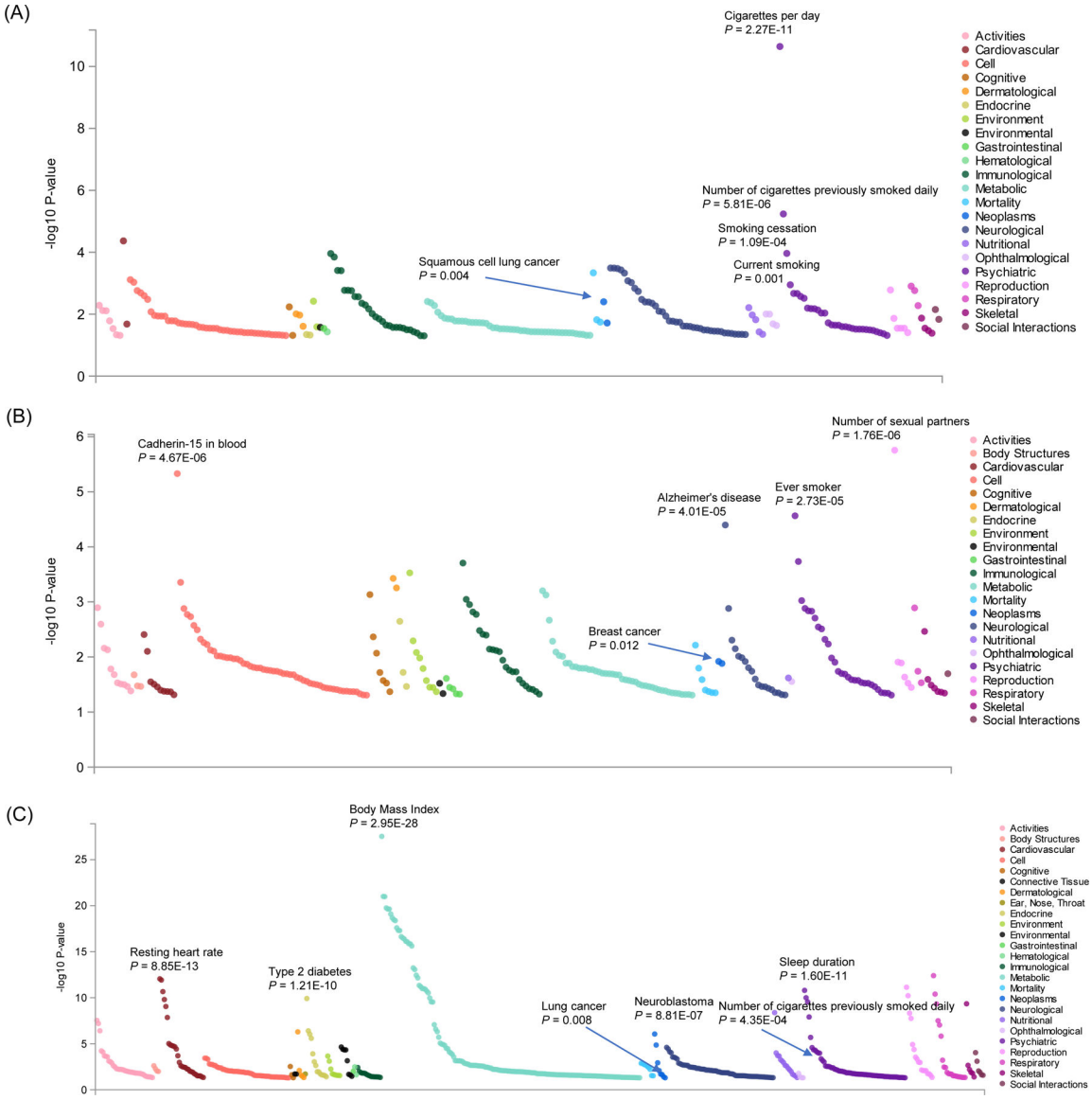

Figure 4.

PheWAS analyses showed significant associations of the three identified genes with multiple traits: (A) *CYP2B6:* related traits including cigarettes used per day (P = 2.27E-11), squamous cell lung cancer risk (P = 0.004); (B) *EPHX2:* related traits including ever smoker (P = 2.73E-05), breast cancer risk (P = 0.012), Cadherin-15 content in blood (P = 4.67E-06); (C) *HSD17B12:* related traits including Body Mass Index (P = 2.95E-28), Type 2 Diabetes (P = 1.21E-10), number of cigarettes previously smoked daily (P = 4.35E-04), and lung cancer risk (P = 0.008).

Table 1.

Distributions of population characteristics in the two-phase study.

	Discovery study (MDACC)					Validation study (OncoArray)				
	# cases[1]		# controls			# cases[2]		# controls		
Variables	**(N=2,171)**	**%**	**(N=4,493)**	**%**	***P***	**(N=5,205)**	**%**	**(N=3,232)**	**%**	***P***
Age					< 0.0001					< 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					<0.0001					0.001
Female	494	22.8	2,018	44.9		1,344	25.8	940	29.1	
Male	1,677	77.2	2,475	55.1		3,861	74.2	2,292	70.9	
Smoking status					<0.0001					
Non-smoking	686	31.6	1,150	53.0						
Smoking	1,485	68.4	1,019	47.0						
Drinking status					<0.0001					
Non-drinking	575	26.5	985	45.4						
Drinking	1,596	73.5	1,184	54.6						
Tumor sites										
Oral cavity	631	29.1				2,568	49.5			
Oropharynx	1,144	52.7				2,328	44.8			
Larynx	316	14.6				NA	NA			
Hypo-pharynx & other sites	78	3.6				295[3]	5.7			

1. Two cases were missing tumor site information in the discovery dataset from the MDACC (The University of Texas MD Anderson Cancer Center) genome-wide association study (GWAS).

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

3. Hypopharynx and overlapping cancers.

Table 2.

Association of SNPs in metabolic pathway-related genes and head and neck cancer risk.

					MDACC study				OncoArray study		Combined analysis			
SNP	**Location**	**Pos (hg19)**	**Gene**	**REF/ EFF**	**OR (95%CI)[1]**	***Pval*[1]**	**FDR**	**BFDP**	**OR (95%CI)[2]**	***Pval*[2]**	**OR (95%CI)[3]**	***Pval*[3]**	**Q-test**	**I^2**
All cases														
rs145533669	19q13.2	41520684	*CYP2B6*	A/T	0.46 (0.29–0.74)	0.001	0.884	0.307	0.68 (0.49–0.94)	0.018	0.60 (0.46–0.78)	1.57E-04	0.181	44.2
rs35246205	19q13.2	41527798	*CYP2B6*	C/T	1.26 (1.03–1.54)	0.024	0.898	0.757	1.15 (1.00–1.32)	0.049	1.18 (1.06–1.33)	0.004	0.464	0
Oropharyngeal cancer														
rs117522494	8p21.2	27350400	*EPHX2*	C/T	1.68 (1.19–2.37)	0.003	0.756	0.377	1.31 (0.98–1.75)	0.072	1.45 (1.16–1.82)	0.001	0.279	14.8
rs145533669	19q13.2	41520684	*CYP2B6*	A/T	0.55 (0.31–0.96)	0.036	0.935	0.792	0.65 (0.44–0.98)	0.039	0.62 (0.44–0.85)	0.004	0.62	0
Oral cancer														
rs4359199	11p11.2	43721500	*HSD17B12*	T/C	1.18 (1.04–1.33)	0.009	0.740	0.657	1.09 (1.00–1.18)	0.038	1.11 (1.03–1.2)	0.006	0.255	22.9

Abbreviations: CHR =chromosome; Pos = position, REF/EFF = Reference allele/ effect allele; *P*val = *P* value; FDR = false discovery rate; BFDP = Bayesian false discovery rate.

[1]Adjusted for age, sex and top five significant principal components.

[2]Adjusted for age, sex and top three principal components and continents.

[3]Results from fixed-effects model.

Table 3.

Stratified analysis of SNP rs4359199 (T>C) with SCCHN risk by smoking /drinking statuses.

	Overall[3]		Oropharyngeal[3]		Oral cavity[3]	
Variables	**OR (95%CI)**	**P**	**OR (95%CI)**	**P**	**OR (95%CI)**	**P**
Smoking status[1]						
Non-smoker	1.05 (0.92–1.21)	0.471	0.96 (0.82–1.13)	0.656	1.20 (0.97–1.48)	0.100
Smoker	1.10 (0.99–1.23)	0.091	0.97 (0.85–1.12)	0.706	1.14 (0.97–1.35)	0.110
Drinking status[2]						
Non-drinker	0.90 (0.78–1.04)	0.154	0.82 (0.68–0.99)	**0.043**	1.06 (0.85–1.31)	0.627
Drinker	1.10 (0.99–1.23)	0.085	1.04 (0.92–1.19)	0.517	1.22 (1.03–1.43)	**0.019**

[1]Adjusted for age, sex, drinking status, and top five significant principal components.

[2]Adjusted for age, sex, smoking status and top five significant principal components.

[3]There were 4,484 controls and 2,171 SCCHN cases with smoking and drinking data available in the MDACC dataset, which includes 1,144 oropharyngeal cancer and 631 non-oropharyngeal cancer.