

Published in final edited form as:

Nat Genet. 2016 December ; 48(12): 1544–1550. doi:10.1038/ng.3685.

Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer

A full list of authors and affiliations appears at the end of the article.

Abstract

We conducted a genome-wide association study of oral cavity and pharyngeal cancer in 6,034 cases and 6,585 controls from Europe, North America and South America; we detected 8 loci ($P < 5 \times 10^{-8}$), 7 of which are novel for these cancer sites. Oral and pharyngeal cancers combined were associated with loci at 6p21.32 (rs3828805, *HLA-DQB1*), 10q26.13 (rs201982221, *LHPP*) and 11p15.4 (rs1453414, *OR52N2/TRIM5*). Oral cancer was associated with two new regions 2p23.3 (rs6547741, *GPN1*) and 9q34.12 (rs928674, *LAMC3*), and with known cancer loci: 9p21.3 (rs8181047, *CDKN2B-AS1*) and 5p15.33 (rs10462706, *CLPTMIL*). Oropharyngeal cancer associations were limited to the human leukocyte antigen (HLA) region and classical HLA allele imputation revealed a protective association with the class II haplotype *DRB1*1301-DQA1*0103-DQB1*0603* (odds ratio (OR)=0.59, $P=2.7 \times 10^{-9}$). Stratified analyses on a subgroup of oropharyngeal cases with human papillomavirus (HPV) status indicated that this association was considerably stronger in HPV-positive (OR=0.23, $P=1.6 \times 10^{-6}$) compared to HPV-negative cancers (OR=0.75, $P=0.16$).

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding Author: Paul Brennan, Ph.D., Section of Genetics, International Agency for Research on Cancer (IARC), Lyon, France, gep@iarc.fr
⁵¹deceased

URLs. R, <http://www.r-project.org/>; PLINK, <https://www.cog-genomics.org/plink2>; University of Michigan Imputation Server, <https://imputationserver.sph.umich.edu/>; Haplotype Reference Consortium, <http://www.haplotype-reference-consortium.org/>; SNP2HLA, <https://www.broadinstitute.org/mpg/snp2hla/>; HaploReg v4.1, <http://www.broadinstitute.org/mammals/haploreg/haploreg.php>; GTEx Portal, <http://www.gtexportal.org/home/>; LocusZoom, <http://locuszoom.sph.umich.edu/locuszoom/>; metafor R package <http://www.metafor-project.org/doku.php>; LDlink, <http://analysisitools.nci.nih.gov/LDlink/>; INHANCE Consortium, <http://www.inhance.utah.edu/>; OncoArray Network, <http://epi.grants.cancer.gov/oncoarray/>; GAME-ON, <http://epi.grants.cancer.gov/gameon/>; GENCAPO, <http://www.gencapo.famerp.br>

Accession codes

dbGaP phs001202.v1.p1

Competing financial interests: The authors declare no competing financial interests.

Author contributions

P.Brennan and J.D.M. conceived and designed the project. C.L. undertook data harmonization, genotypes quality control, GWAS analysis, imputation and meta-analyses. X.X. performed genotype calling. V.G. and A.C. organized and supervised sample selection and DNA shipments at IARC. A.C. performed replication TaqMan genotyping. C.L. and V.G. analyzed data from replication genotyping. C.L. and P.Brennan drafted the first version of the manuscript. B.D., A.F.O., V.W.-F., A.R.N., G.L., M.L., J.E.-N., S.F., P.L., G.J.M., L.R., S.B., J.P., K.K., D.Z., M.J., A.M.M., M.P.C., M.R., W.A., C.C., A.Z., X.C., D.I.C., I.H., D.M., M.V., C.M.H., N.S.-D., E.F., J.L., J.R.G., M.C.W., E.H.T., F.D.N., M.B.Dc., S.T., R.J.H., W.H.M.P., R.H., G.C., A.S., A.A., O.S., H.B.B.Dm, P.Boffetta and D.A., contributed with reagent/samples/material and reviewed/approved the final manuscript. J.D.M. and C.I.A. designed and coordinated the Lung Cancer OncoArray. P.Brennan obtained funding for the project, provided overall supervision and management.

Cancers of the oral cavity (OC) and oropharynx (OPC) are predominantly caused by tobacco and alcohol use, although oral infection with HPV, particularly HPV16, is an increasingly important cause of OPC^{1,2}, especially in the US and northern Europe. The proportion of HPV-related OPCs varies widely and is estimated to be approximately 60% in the US, 30% in Europe and lower in South America^{2–5}. Genetic factors have also been implicated in OC and OPC susceptibility, especially polymorphisms within alcohol-related genes including alcohol-dehydrogenase 1B (*ADH1B*) and *ADH7*^{6,7}. In order to identify additional susceptibility loci, 13,107 individuals from 12 epidemiological studies (Supplementary Table 1) were genotyped using the Illumina OncoArray and after stringent quality-control steps (Supplementary Table 2, Online Methods) 6,034 cases and 6,585 cancer-free controls remained for analyses (Table 1). We next performed genome-wide imputation using the Haplotype Reference Consortium panel⁸ and obtained approximately 7 million high-quality imputed variants (Supplementary Fig. 1). Given the ethnic diversity of our study, we evaluated associations within continent (Europe, North and South America) using multivariate unconditional logistic regressions under a log-additive genetic model adjusted for age, sex and regional eigenvectors. Results by continent were combined using fixed-effect meta-analyses to derive associations for overall OC and pharynx cancer (oral, oropharynx, hypopharynx and overlapping cancers; $n = 6,034$), as well as site-specific OC ($n = 2,990$) and OPC ($n = 2,641$). Although, several ethnic groups are present in the study, supervised ancestry analyses indicated that >90% of participants were predominantly of European (>70%CEU) ancestry, although some population admixture was observed in South America (Supplementary Table 3).

GWA meta-analyses of overall and site-specific cancers identified 9 regions at genome-wide significance ($P < 5 \times 10^{-8}$) (Fig.1). Quantile-quantile (Q-Q) plots of observed and expected P -values showed moderate genomic inflation (λ) for the 3 meta-analyses (λ range = 1.04–1.06, Supplementary Fig. 2–3). Since λ increases with sample size, we scaled it to 1000 cases and controls resulting in ameliorated inflation (λ_{1000} range = 1.009–1.01)⁹. Overall OC and pharynx cancer were associated with rs79767424 (5p14.3), rs1229984 (4q23), rs201982221 (10q26.13), rs1453414 (11p15.4) and 123 SNPs at 6p21.32 (Supplementary Table 4). Twenty-six variants were associated ($P < 5 \times 10^{-8}$) with OC (Supplementary Table 5), 4 of which mapped to 2p23.3, 1 to 4q23, 3 to 9q34.12, 13 to 5p15.33 and 5 to 9p21.3. For OPC, novel significant variants were located at 6p21.32 (62 SNPs, Supplementary Table 6). Suggestive susceptibility variants ($P < 5 \times 10^{-7}$, Supplementary Tables 7–9) were associated with OC at 4 additional loci: 6p21.33, 6p21.32, 15q21.2, 15q26.2 and, with OPC at 2q36.1. Other genomic locations outside the HLA region showed promising associations ($P < 5 \times 10^{-6}$) with OPC (Supplementary Table 10). For susceptibility loci at $P < 5 \times 10^{-8}$, functional annotation of regulatory features with ENCODE and eQTL information, if available, are summarized in Supplementary Table 11 and 12. Given the geographical heterogeneity of our population, we performed sensitivity analyses after excluding individuals with <70% CEU ancestry and these showed similar results (Supplementary Table 13). To validate array genotypes and imputed dosages, we directly genotyped by a different platform (TaqMan) at least one variant within each locus ($P = 5 \times 10^{-7}$) in a subset of approximately 700 individuals. Concordance between genotyped/imputed genotypes and TaqMan results was >97% for all regions with the exception of rs2398180, an imputed

variant which had a concordance of 94% (Supplementary Table 14-15). For 2 rare variants, rs201982221 (10q26.13) and rs7976742 (5p14.3), TaqMan assays could not be designed and we used Sanger sequencing for validation (Supplementary Table 16). We were able to validate the rs201982221 deletion (Online Methods), but rs7976742 did not validate (Online Methods). The lead variant at each validated loci ($P < 5 \times 10^{-8}$) for overall and site-specific analyses are shown in Table 2. Results stratified by geographical region, smoking and alcohol status are displayed in Figure 2 (oral and pharynx cancer combined) and Figure 3 (oral cancer).

The rs1229984 (4q23, *ADH1B*) association has been previously reported as a susceptibility locus for OC and OPC6, and similar to previous findings this variant showed heterogeneity by region, smoking and alcohol drinking status^{10,11} (Fig. 2a). Three other 4q23 SNPs reached $P < 5 \times 10^{-8}$, although conditional analyses indicated these are not independent signals (Supplementary Table 17). The rs1573496 (*ADH7*) variant reported to be strongly associated with OC and OPC in the previous upper aerodigestive tract cancer GWAS⁷ was only moderately associated here (Supplementary Table 18). In the overall OC and pharynx cancer analysis, we identified rs201982221 at 10q26.13 (OR = 1.67, $P = 1.58 \times 10^{-9}$), that was also separately associated with OC (OR = 1.71, $P = 1.04 \times 10^{-7}$) and OPC (OR = 1.70, $P = 7.9 \times 10^{-7}$) (Fig. 2b). rs201982221 is located within the *LHPP* gene in a region with reported regulatory features (Supplementary Table 11). However, it is a rare intronic deletion in an area of low linkage disequilibrium (LD) (Supplementary Fig. 4), and thus warrants further validation in a different population. rs1453414, the lead signal at 11p15.4 (Supplementary Table 19), is an intronic variant that showed a borderline association in the overall (OR = 1.19, $P = 4.78 \times 10^{-8}$) and site-specific analyses [OC (OR = 1.19, $P = 1.65 \times 10^{-5}$) and OPC (OR = 1.22, $P = 4.26 \times 10^{-6}$)] (Fig. 2c, Supplementary Fig. 5). rs1453414 is upstream of *OR52N2*, an olfactory receptor, and within *TRIM5*, an E3-ubiquitin ligase, and is an eQTL for these genes in brain tissue¹² (Supplementary Table 12).

At 2p23.3, 4 SNPs showed evidence ($P < 5 \times 10^{-8}$) for an association with OC, and in conditional analyses did not appear to be independent (Supplementary Table 20). These signals map to a high LD area that includes *C2orf16*, *ZNF512*, *CCDC121* and *GPNI* (Supplementary Fig. 6). The lead SNP, rs6547741, was associated with OC but not with OPC, and maps to an intron of *GPNI*, a GTPase involved in RNA polymerase II transport and DNA repair¹³. Associations between rs6547741 and OC were homogenous across other stratified analyses by region, sex, smoking and drinking status (Fig. 3a).

Variation within 5p15.33 was also exclusively associated with OC (OPC, rs10462706, $P = 0.47$). The top signal, rs10462706, was associated with decreased OC risk (OR = 0.74, $P = 5.54 \times 10^{-10}$) and is in low LD ($r^2 = 0.15$, Supplementary Fig. 7) with the second strongest signal rs467095. These two variants are 7kb apart and map to intron 13 of *CLPTMIL* and in stratified analysis showed stronger effects in never smokers ($P_{\text{het}} = 0.07$ and $P_{\text{het}} = 0.0028$, respectively) and never drinkers ($P_{\text{het}} = 0.01$ and $P_{\text{het}} = 0.0025$, respectively) (Fig. 3b, Supplementary Fig. 8). Conditional analyses showed that these SNPs are not completely independent (Supplementary Table 21). *TERT* and *CLPTMIL* encode the telomerase reverse transcriptase (*TERT*) and the cleft-lip and palate-associated transmembrane 1-like protein (*CLPTMIL*), respectively. Notably, rs467095 is an esophageal *TERT* eQTL¹⁴

(Supplementary Table 12) and is in high LD with rs401681 (OR = 1.18, $P = 2.1 \times 10^{-7}$, $r^2 = 0.94$) a widely studied SNP associated with risk of several cancers including: lung^{15,16}, bladder, prostate, cervical, melanoma¹⁷, basal cell¹⁸, esophageal¹⁹, pancreatic²⁰ and nasopharyngeal cancer²¹. Multiple 5p15.33 variants have been reported to independently influence cancer risk in both an increasing and decreasing fashion. Interestingly, rs401681[A] was associated with an increased OC risk similar to previous melanoma associations, and in an opposite direction to previous lung cancer results¹⁷.

Several variants within the *CDKN2A–CDKN2B* locus (9p21.3) were found to be associated with OC. The lead SNP, rs8181047, is an intronic variant within the *CDKN2B1* antisense RNA 1 (*CDKN2B-AS1*) (Fig. 3c). rs8181047 is in LD ($r^2_{\text{range}} = 0.6-0.8$) with 4 other 9p21.3 variants strongly associated with OC (Supplementary Fig. 9) that in conditional analyses did not show independent associations (Supplementary Table 22). The *CDKN2A–CDKN2B* locus contains genes involved in cell-cycle regulation and senescence and has been associated with multiple malignancies including melanoma²², glioma²³, basal cell¹⁸, breast²⁴, lung²⁵, nasopharyngeal²⁶ and esophageal cancer²⁷. Notably, *CDKN2A* is frequently mutated in HPV-negative head and neck cancers²⁸.

The OC associated variants at 9q34.12 mapped to an intron of *LAMC3*, a laminin involved in cortical development²⁹. rs928674, the peak signal, showed consistent effects across strata and a weaker association with OPC ($P = 0.003$) (Fig. 3d). rs928674 is in high LD with 3 other robustly associated 9q34.12 SNPs ($r^2_{\text{range}} = 0.82-0.96$, Supplemental Fig. 10, Supplementary Table 23) and is an esophageal mucosa cis-eQTL for a downstream gene *AIF1L* (Allograft Inflammatory Factor 1-Like)¹⁴.

The most prominent finding in the overall and OPC meta-analyses was a large association signal at 6p21.32 within the HLA class II region. The lead variant in both analyses, rs3828805, maps 1.7kb 5' of *HLA-DQB1* (Fig. 2d and Supplementary Fig. 11) and similar to other 6p21.32 variants (Supplementary Table 4 and 6), showed heterogeneity by geographical region ($P_{\text{het}} = 0.007$) with no effect in South America ($P = 0.62$). Association analyses of 6p21.32 variants ($P < 5 \times 10^{-8}$) conditioned on rs3828805 did not reveal multiple independent signals (Supplementary Table 24), suggesting a common haplotype. To further investigate HLA associations, we imputed classical alleles in 11,436 individuals (>70% Caucasian ancestry) (Online Methods). Three classical HLA alleles *DRB1*1301*, *DQA1*0103* and *DQB1*0603* reached $P < 5 \times 10^{-8}$ in the overall analysis and were also strongly associated with OPC (Supplementary Table 25). These alleles are in high LD ($r^2 > 0.9$) and are part of the HLA class II haplotype, *DRB1*1301-DQA1*0103-DQB1*0603*, which is common in Europeans and previously reported to be associated with decreased cervical cancer risk³⁰. *DRB1*1301-DQA1*0103-DQB1*0603* was strongly associated with reduced OPC risk (OR = 0.59, $P = 2.7 \times 10^{-9}$) and more weakly with OC risk (OR = 0.75, $P = 1.7 \times 10^{-4}$). Further conditional analysis on this haplotype and 6p21.32 variants did not reveal evidence of additional independent effects (Supplementary Table 26). Given the importance of HPV infection in the etiology of cervical and oropharyngeal cancer³¹, we conducted post-hoc analyses to examine the effect of *DRB1*1301-DQA1*0103-DQB1*0603* in a subset of 576 cases with available HPV-status and 3,662 controls. *DRB1*1301-DQA1*0103-DQB1*0603* was associated with a strong reduced risk of HPV-positive OPC

(OR = 0.23, $P = 1 \times 10^{-6}$, $n = 336$) with no significant association in 240 OPC HPV-negative cases (OR = 0.75, $P = 0.16$) (Table 3). These results indicate that the class II HLA region is implicated in at least two HPV-driven cancers, namely HPV-positive OPC and cervical cancer. The lack of an association between 6p21.3 SNPs and OPC risk in South America could relate to previous findings that less than 10% of OPC are HPV-positive in this region^{4,5}. Moreover, a weaker association with OC could be due to a smaller proportion of these cases being HPV-positive, as well as possibly some misclassified OPC cases, especially for base of the tongue tumors. Further evaluation of the extent and specificity of this HLA effect in HPV-associated cancers is important given the strength of the observed association. This may help elucidate why some individuals are at higher risk of HPV-positive OPC after HPV infection and may also have implications for cancer immunotherapies targeting the HLA class II antigen presentation pathway³².

In summary, we identified seven oral and pharyngeal cancer susceptibility loci including a strong HLA signal narrowed to a class II haplotype. Future replication of these findings in an independent population is warranted as well as fine-mapping and functional studies necessary to establish the biological framework underneath these associations.

Online Methods

Study population and genotyping

The study population comprised 6,034 cases and 6,585 controls derived from 12 epidemiological studies, the majority of case-control design and part of the International Head and Neck Cancer Epidemiology Consortium (INHANCE). Additionally, cases and controls from a European cohort study (EPIC) and cases from a United Kingdom case-series (HN5000) were also included. Characteristics and references for each study are summarized in Supplementary Table 1. Informed consent was obtained for all participants and studies were approved by respective Institutions Review Boards. Cancer cases comprised the following ICD codes: oral cavity (C02.0-C02.9, C03.0-C03.9, C04.0-C04.9, C05.0-C06.9) oropharynx (C01.9, C02.4, C09.0-C10.9), hypopharynx (C13.0-C13.9), overlapping (C14 and combination of other sites) and 25 oral or pharyngeal cases with unknown ICD code (other).

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 designed for cancer studies by the OncoArray Consortium³³ part of the Genetic Associations and Mechanisms in Oncology (GAME-ON) Network. The majority of the samples were genotyped as part of the oral and pharynx cancer OncoArray, with the exception of 2,476 shared controls (1,453 from the EPIC study and 1,023 from the Toronto study) that were genotyped at CIDR but as part of the Lung OncoArray. Genotype calls were made by the Dartmouth team in GenomeStudio software (Illumina, Inc.) using a standardized cluster file for OncoArray studies. Cluster plots for top SNPs for individuals genotyped as part the oral and pharynx cancer OncoArray are shown in Supplementary Figure 12.

GWAS quality control

We used PLINK 1.934 to conduct systematic quality-control (QC) steps on genotypes calls. An initial filtering step on the complete dataset excluded samples with genotyping rate <80% and SNPs with call rate <80%, then we excluded samples and SNPs <95% call rates. During the individuals QC, we removed samples with unsolved genetic and reported sex discrepancies and individuals with outlying autosomic heterozygosity rate (± 4 standard deviations (SD)). Identity-by-descent (IBD) analysis performed on the LD pruned dataset ($r^2 < 0.1$) identified; 103 expected experimental duplicate-pairs (IBD > 0.9), from these we excluded the sample with lower genotyping rate, and 74 unexpected duplicate-pairs that were excluded. Additionally, we identified 44 unexpected relatives pairs (IBD > 0.3) and excluded one sample from each pair, prioritizing cases over controls and for pairs with same status we excluded according to genotyping rate. After the initial SNP filtering calling rate and exclusion of duplicated and zeroed probes, 513,311 probes remained for analysis. Next and given the heterogeneity of the study population, we divided the dataset by geographical region (Europe, North America and South America) and excluded SNPs within each region with deviation of Hardy-Weinberg Equilibrium (HWE) in controls ($P < 1 \times 10^{-7}$). To account for potential population stratification within geographical area, we performed principal components analysis (PCA) in EIGENSTRAT35 using approximately 10,000 common markers in low LD ($r^2 < 0.004$, MAF > 0.05), subsequently we excluded population outliers ($n = 139$) and derived regional eigenvectors to adjust regional GWAS analyses (Supplementary Fig. 13). PCA analyses on each individual study are displayed in Supplementary Fig. 14. To inquire about ancestry within regions, we used STRUCTURE 2.3.436 under a supervised approach with HapMap samples to determine individual CEU, YRI and CHB/JPT ancestries. All coordinates refer to genome build HG19/GRCh37.

Imputation

Imputation of unknown genetic variation was performed using the Michigan Imputation Server³⁷ (<https://imputationserver.sph.umich.edu/>), a free service for largescale population studies. We used SHAPEIT³⁸ for phasing, Minimac³⁹ for imputation and the first release of the Haplotype Reference Consortium panel8 (HRC) a large collection of human haplotypes ($n = 64,976$) that combines data from multiple initiatives including the 1000 genomes project. For imputation we used a set of high-quality SNPs; MAF > 0.01, call rate >98%, we imputed in randomized batches of approximately 3000 individuals and analyzed imputation quality statistics together, SNPs with MAF < 0.01 or $R^2 < 0.3$ in any of the batches were excluded before associations analysis. Thus, the final set of 7,099,472 imputed variants used in association analysis had very high quality: 86% of the variants with MAF 0.05 had $R^2 > 0.9$ and 52% of the less common variants (MAF < 0.05) had $R^2 > 0.9$ (Supplementary Figure 1).

GWAS and meta-analyses

We undertook a by geographic region GWAS and meta-analysis approach to evaluate the relation between SNPs and overall OC and pharynx cancer, and site-specific OC and OPC risk; in total we tested we evaluated 7,574,753 SNPs (genotyped and imputed variants). By region GWAS were performed in PLINK and R glm function for genotype dosages, using

multivariable unconditional logistic regression assuming a log-additive genetic or dosage model with age, sex and principal components as covariates. Next, association statistics were included in a fixed-effects meta-analysis performed in PLINK³⁴. The *P*-value for heterogeneity was calculated using Cochran's *Q*. Conditional analyses within associated regions were performed in R (glm) and meta-analysis of regional results were done in PLINK.

HLA imputation and haplotype analyses

Using SNP2HLA40 and the Type I Diabetes Genetics Consortium reference panel of 5,225 individuals of European descent, we imputed HLA classical alleles and aminoacids in 11,436 study participants with >70% CEU ancestry. Individuals were randomized into 12 groups approximately of 1000 each and quality imputation scores from beagle examined across groups. Markers with $R^2 \geq 0.7$ in all groups and average $R^2 > 0.9$ (8,286 markers of 8,961 included in the panel) were carried on for analyses and 98% of these had $R^2 > 0.95$. We then performed regional association analyses of binary markers followed by meta-analysis of imputed binary markers (SNPs, classical alleles and aminoacids) as previously described using PLINK and R. DRB1*1301-DQA1*0103-DQB1*0603 haplotype information was extracted from beagle files of phased best-guess genotypes and used to estimate ORs and 95% CIs from multivariable unconditional logistic regression models with each phenotype. Haplotype and HPV stratified analyses in a subset of cases with available data, was performed with unconditional logistic regression with haplotype as predictor and oropharyngeal cancer HPV positive or negative as outcomes. The measure of HPV positivity was derived by HPV16E6 serology in the ARCAGE and EPIC studies, tumor HPV16 DNA and p16 overexpression in the CHANCE study and for the Pittsburgh cases a combination of tumor HPV in situ hybridization and p16 immunohistochemistry.

Technical validation genotyping

In order to confirm array genotypes and imputed dosages, we genotyped at least one SNP within each associated ($P < 5 \times 10^{-7}$) genomic region using TaqMan assays (Thermo Fisher Scientific) in a subset of approximately 900 individuals from ARCAGE, IARC Latin America, EPIC, and IARC oral cancer studies (Supplementary Table 14). Taqman genotyping included standard positive and negative controls; assay numbers and cycling conditions are described in Supplementary Table 15. For 10 of the 12 SNPs genotyped by TaqMan there was no evident departure from Hardy-Weinberg equilibrium ($P > 0.05$). However, two variants rs467095 and rs3731239 were slightly out of HWE only in samples from the EPIC study ($P < 0.01$), which might relate to the small sample size genotyped in the technical validation. The rare variants at 5p14.3 (rs79767424) and 10q26.13 (rs201982221, Supplementary Figure 15) could not be genotyped with TaqMan assays, thus for these, validation was performed with Sanger sequencing on the on a small subset of samples available within IARC studies; rs201982221 (15 wild-type, 15 heterozygous and 2 homozygous) and for rs79767424 (15 wild-type and 15 heterozygous). Primer sequences and conditions for PCR are described in Supplementary Table 16; amplification products were sequenced on the Applied Biosystems 3730 DNA Analyzer at Biofidal, France.

Bioinformatic annotation and analyses

To explore possible functional implications of variants within associated regions, we used an internet tool HaploReg v4.141 that includes annotations from the Encyclopedia of DNA elements (ENCODE). Functional annotation for variants reaching $P < 5 \times 10^{-8}$ is summarized Supplemental Table 11. We also retrieved eQTL information from several sources including the Genotype-Tissue Expression Project (GTEx)¹⁴ and HaploReg⁴¹. Manhattan and quantile-quantile plots were generated in R using the package qqman⁴², and forest plots were generated using the metafor R library⁴³. Regional association plots were generated using LocusZoom⁴⁴ with LD and recombination rates data from the 1000 Genomes Project November 2014 release. To search for additional, pairwise LD information we used the online tool LD link⁴⁵.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Corina Lesseur¹, Brenda Diergaarde^{2,3}, Andrew F Olshan^{4,5}, Victor Wünsch-Filho⁶, Andrew R Ness^{7,8}, Geoffrey Liu⁹, Martin Lacko¹⁰, José Eluf-Neto¹¹, Silvia Franceschi¹, Pagona Lagiou¹², Gary J Macfarlane¹³, Lorenzo Richiardi¹⁴, Stefania Boccia¹⁵, Jerry Polesel¹⁶, Kristina Kjaerheim¹⁷, David Zaridze¹⁸, Mattias Johansson¹, Ana M Menezes¹⁹, Maria Paula Curado²⁰, Max Robinson²¹, Wolfgang Ahrens²², Cristina Canova²³, Ariana Znaor^{1,24}, Xavier Castellsagué^{25,51}, David I Conway^{26,27}, Ivana Holcátová²⁸, Dana Mates²⁹, Marta Vilensky³⁰, Claire M Healy³¹, Neonila Szeszenia-D browska³², Eleonóra Fabiánová³³, Jolanta Lissowska³⁴, Jennifer R Grandis^{35,36}, Mark C Weissler³⁷, Eloiza H Tajara³⁸, Fabio D Nunes³⁹, Marcos B de Carvalho⁴⁰, Steve Thomas⁸, Rayjean J Hung⁴¹, Wilbert H M Peters⁴², Rolando Herrero¹, Gabriella Cadoni⁴⁴, H Bas Bueno-de-Mesquita^{44,45,46}, Annika Steffen⁴⁷, Antonio Agudo⁴⁸, Oxana Shangina¹⁸, Xiangjun Xiao⁴⁹, Valérie Gaborieau¹, Amélie Chabrier¹, Devasena Anantharaman¹, Paolo Boffetta⁵⁰, Christopher I Amos⁴⁹, James D McKay¹, and Paul Brennan¹

Affiliations

¹International Agency for Research on Cancer (IARC/WHO), Lyon, France
²Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA ³University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA ⁴Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, NC, USA ⁵UNC Lineberger Comprehensive Cancer Center, Chapel Hill, NC, USA ⁶Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo, SP, Brazil ⁷National Institute for Health Research (NIHR) Biomedical Research Unit in Nutrition, Diet and Lifestyle at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol, Bristol, UK ⁸University of Bristol and School of Oral and Dental Sciences, University of Bristol, Bristol, UK ⁹Princess Margaret Cancer Centre, Toronto, Ontario, Canada ¹⁰Department of Otorhinolaryngology, Head and Neck Surgery, Maastricht University Medical Center,

Maastricht, Netherlands ¹¹Departamento de Medicina Preventiva, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil ¹²Department of Hygiene, Epidemiology and Medical Statistics, School of Medicine, National Kapodistrian University of Athens, Athens, Greece ¹³School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK ¹⁴Department of Medical Sciences, University of Turin, Turin, Italy ¹⁵Section of Hygiene, Institute of Public Health, Università Cattolica del Sacro Cuore, Fondazione Policlinico 'Agostino Gemelli', Rome, Italy ¹⁶Unit of Cancer Epidemiology, CRO Aviano National Cancer Institute, Aviano, Italy ¹⁷Cancer Registry of Norway, Oslo, Norway ¹⁸Department of Cancer Epidemiology and Prevention, Institute of Carcinogenesis, N.N. Blokhin Russian Cancer Research Centre of the Russian Ministry of Health, Moscow, Russian Federation ¹⁹Programa de Pós-Graduação em Epidemiologia, Universidade Federal de Pelotas (UFPel), Pelotas, Brazil ²⁰Epidemiology, International Center for Research (CIPE), A.C.Camargo Cancer Center, Sao Paulo, Brazil ²¹Centre for Oral Health Research, Newcastle University, Newcastle, UK ²²Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany ²³Department of Molecular Medicine, University of Padova, Padova, Italy ²⁴Croatian National Institute of Public Health, Zagreb, Croatia ²⁵Institut Català d'Oncologia (ICO)-DIBELL, CIBER-ESP, L'Hospitalet de Llobregat, Catalonia, Spain ²⁶School of Medicine, Dentistry and Nursing, University of Glasgow, University of Glasgow, Glasgow, UK ²⁷NHS NSS National Services Scotland, Edinburgh, UK ²⁸Institute of Hygiene & Epidemiology 1st Faculty of Medicine, Charles University, Prague, Czech Republic ²⁹National Institute of Public Health, Bucharest, Romania ³⁰Instituto de Oncologia "Angel H Roffo", Universidad de Buenos Aires, Buenos Aires, Argentina ³¹Trinity College School of Dental Science, Dublin, Ireland ³²Department of Environmental Epidemiology, Nofer Institute of Occupational Medicine, Lodz, Poland ³³Regional Authority of Public Health, Banská Bystrica, Slovakia ³⁴The Maria Skłodowska-Curie Memorial Cancer Centre and Institute of Oncology (MCMCC), Warsaw, Poland ³⁵Department of Otolaryngology-Head and Neck Surgery, University of California San Francisco, San Francisco, California, USA ³⁶Clinical Translational Science Institute, University of California at San Francisco, San Francisco, CA, USA ³⁷Department of Otolaryngology/Head and Neck Surgery, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA ³⁸Department of Molecular Biology, School of Medicine of São José do Rio Preto, São José do Rio Preto, SP, Brazil ³⁹Department of Stomatology, School of Dentistry, University of São Paulo, São Paulo, SP, Brazil ⁴⁰Department of Head and Neck Surgery, Heliópolis Hospital, São Paulo, SP, Brazil ⁴¹Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada ⁴²Department of Gastroenterology, Radboud University Nijmegen, Medical Center, Nijmegen, the Netherlands ⁴³Institute of Otorhinolaryngology, Università Cattolica del Sacro Cuore, Fondazione Policlinico 'Agostino Gemelli', Rome, Italy ⁴⁴Department for Determinants of Chronic Diseases (DCD), National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands ⁴⁵Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

⁴⁶Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia ⁴⁷German Institute of Human Nutrition in Potsdam-Rehbruecke (DIFE), Nuthetal, Germany ⁴⁸Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Institut Català d'Oncologia (ICO)-IDIBELL, l'Hospitalet de Llobregat, Barcelona, Spain ⁴⁹Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire, USA ⁵⁰Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA

Acknowledgments

Genotyping performed at the Center for Inherited Disease Research (CIDR) was funded through the U.S. National Institute of Dental and Craniofacial Research (NIDCR) grant 1X01HG007780-0. Genotyping for shared controls with the Lung OncoArray initiative was funded through the grant X01HG007492-0. Corina Lesseur undertook this work during the tenure of a Postdoctoral Fellowship awarded by the International Agency for Research on Cancer. The funders did not participate in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. We acknowledge all of the participants involved in this research and the funders and support. We thank Dr. Leticia Fernandez (Instituto Nacional de Oncología y Radiobiología, La Habana, Cuba) for her contribution to the IARC ORC multicenter study. We are also grateful to Sergio Koifman (Escola Nacional de Saúde Pública, Rio de Janeiro, Brazil) for his contribution to the IARC Latin America multicenter study (Sergio Koifman passed away in May 2014) and to Xavier Castellsagué from the ARCAGE Barcelona Center who recently passed away (June 2016).

The University of Pittsburgh head and neck cancer case-control study is supported by National Institutes of Health grants P50 CA097190 and P30 CA047904. The Carolina Head and Neck Cancer Study (CHANCE) was supported by the National Cancer Institute (R01-CA90731). The Head and Neck Genome Project (GENCAPO) was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Grant numbers 04/12054-9 and 10/51168-0). The authors thank all the members of the GENCAPO team. The HN5000 study was funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0707-10034), the views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The Toronto study was funded by the Canadian Cancer Society Research Institute (020214) and the National Cancer Institute (U19 CA148127) and the Cancer Care Ontario Research Chair. The alcohol-related cancers and genetic susceptibility study in Europe (ARCAGE) was funded by the European Commission's 5th Framework Program (QLK1-2001-00182), the Italian Association for Cancer Research, Compagnia di San Paolo/FIRMS, Region Piemonte, and Padova University (CPDA057222). The Rome Study was supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC) IG 2011 10491 and IG2013 14220 to SB, and Fondazione Veronesi to SB. The IARC Latin American study was funded by the European Commission INCO-DC programme (IC18-CT97-0222), with additional funding from Fondo para la Investigacion Científica y Tecnológica (Argentina) and the Fundação de Amparo à Pesquisa do Estado de São Paulo (01/01768-2). The IARC Central Europe study was supported by European Commission's INCO-COPERNICUS Program (IC15-CT98-0332), NIH/National Cancer Institute grant CA92039, and the World Cancer Research Foundation grant WCRF 99A28. The IARC Oral Cancer Multicenter study was funded by: grant S06 96 202489 05F02 from Europe against Cancer; Grants FIS 97/0024, FIS 97/0662, and BAE 01/5013 from Fondo de Investigaciones Sanitarias, Spain; UICC Yamagiwa-Yoshida Memorial International Cancer Study; National Cancer Institute of Canada; Italian Association for Research on Cancer; and the Pan American Health Organization. The coordination of EPIC study is financially supported by the European Commission (DG SANCO) and the International Agency for Research on Cancer.

References

1. Gillison ML, Chaturvedi AK, Anderson WF, Fakhry C. Epidemiology of Human Papillomavirus-Positive Head and Neck Squamous Cell Carcinoma. *J Clin Oncol.* 2015; 33:3235–42. [PubMed: 26351338]
2. Chaturvedi AK, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *Journal of Clinical Oncology.* 2011; 29:4294–4301. [PubMed: 21969503]

3. Kreimer AR, Clifford GM, Boyle P, Franceschi S. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. *Cancer Epidemiol Biomarkers Prev.* 2005; 14:467–75. [PubMed: 15734974]
4. Ribeiro KB, et al. Low human papillomavirus prevalence in head and neck cancer: results from two large case-control studies in high-incidence regions. *Int J Epidemiol.* 2011; 40:489–502. [PubMed: 21224273]
5. Lopez RV, et al. Human papillomavirus (HPV) 16 and the prognosis of head and neck cancer in a geographical region with a low prevalence of HPV infection. *Cancer Causes Control.* 2014; 25:461–71. [PubMed: 24474236]
6. Hashibe M, et al. Multiple ADH genes are associated with upper aerodigestive cancers. *Nature genetics.* 2008; 40:707–709. [PubMed: 18500343]
7. McKay JD, et al. A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. *PLoS genetics.* 2011; 7:e1001333. [PubMed: 21437268]
8. McCarthy S, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016; 48:1279–1283. [PubMed: 27548312]
9. de Bakker PI, et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet.* 2008; 17:R122–8. [PubMed: 18852200]
10. Hashibe M, et al. Evidence for an important role of alcohol- and aldehyde-metabolizing genes in cancers of the upper aerodigestive tract. *Cancer Epidemiol Biomarkers Prev.* 2006; 15:696–703. [PubMed: 16614111]
11. Chang JS, Straif K, Guha N. The role of alcohol dehydrogenase genes in head and neck cancers: a systematic review and meta-analysis of ADH1B and ADH1C. *Mutagenesis.* 2012; 27:275–86. [PubMed: 22042713]
12. Ramasamy A, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci.* 2014; 17:1418–28. [PubMed: 25174004]
13. Carre C, Shiekhattar R. Human GTPases associate with RNA polymerase II to mediate its nuclear import. *Mol Cell Biol.* 2011; 31:3953–62. [PubMed: 21768307]
14. GTEx-Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015; 348:648–60. [PubMed: 25954001]
15. McKay JD, et al. Lung cancer susceptibility locus at 5p15.33. *Nat Genet.* 2008; 40:1404–6. [PubMed: 18978790]
16. Wang Y, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet.* 2008; 40:1407–9. [PubMed: 18978787]
17. Rafnar T, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet.* 2009; 41:221–7. [PubMed: 19151717]
18. Stacey SN, et al. New common variants affecting susceptibility to basal cell carcinoma. *Nat Genet.* 2009; 41:909–14. [PubMed: 19578363]
19. Yin J, et al. TERT-CLPTM1L Rs401681 C>T polymorphism was associated with a decreased risk of esophageal cancer in a Chinese population. *PLoS One.* 2014; 9:e100667. [PubMed: 25007268]
20. Petersen GM, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat Genet.* 2010; 42:224–8. [PubMed: 20101243]
21. Bei JX, et al. A GWAS Meta-analysis and Replication Study Identifies a Novel Locus within CLPTM1L/TERT Associated with Nasopharyngeal Carcinoma in Individuals of Chinese Ancestry. *Cancer Epidemiol Biomarkers Prev.* 2016; 25:188–92. [PubMed: 26545403]
22. Law MH, et al. Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. *Nat Genet.* 2015; 47:987–95. [PubMed: 26237428]
23. Shete S, et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet.* 2009; 41:899–904. [PubMed: 19578367]
24. Turnbull C, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet.* 2010; 42:504–7. [PubMed: 20453838]
25. Timofeeva MN, et al. Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Hum Mol Genet.* 2012; 21:4980–95. [PubMed: 22899653]

26. Bei JX, et al. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. *Nat Genet.* 2010; 42:599–603. [PubMed: 20512145]
27. Wu C, et al. Joint analysis of three genome-wide association studies of esophageal squamous cell carcinoma in Chinese populations. *Nat Genet.* 2014; 46:1001–6. [PubMed: 25129146]
28. The Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature.* 2015; 517:576–82. [PubMed: 25631445]
29. Barak T, et al. Recessive LAMC3 mutations cause malformations of occipital cortical development. *Nat Genet.* 2011; 43:590–4. [PubMed: 21572413]
30. Chen D, et al. Genome-wide association study of susceptibility loci for cervical cancer. *J Natl Cancer Inst.* 2013; 105:624–33. [PubMed: 23482656]
31. Bouvard V, et al. A review of human carcinogens—Part B: biological agents. *The lancet oncology.* 2009; 10:321–322. [PubMed: 19350698]
32. Thibodeau J, Bourgeois-Daigneault MC, Lapointe R. Targeting the MHC Class II antigen presentation pathway in cancer immunotherapy. *Oncoimmunology.* 2012; 1:908–916. [PubMed: 23162758]
33. Consortium launches genotyping effort. *Cancer Discov.* 2013; 3:1321–2. [PubMed: 24327678]
34. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–75. [PubMed: 17701901]
35. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38:904–9. [PubMed: 16862161]
36. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics.* 2003; 164:1567–87. [PubMed: 12930761]
37. Das S, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016; doi: 10.1038/ng.3656
38. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods.* 2012; 9:179–81.
39. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinformatics.* 2015; 31:782–4. [PubMed: 25338720]
40. Jia X, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One.* 2013; 8:e64683. [PubMed: 23762245]
41. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012; 40:D930–4. [PubMed: 22064851]
42. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv.* 2014
43. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software.* 2010; 36:48.
44. Pruim RJ, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010; 26:2336–7. [PubMed: 20634204]
45. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics.* 2015; 31:3555–7. [PubMed: 26139635]

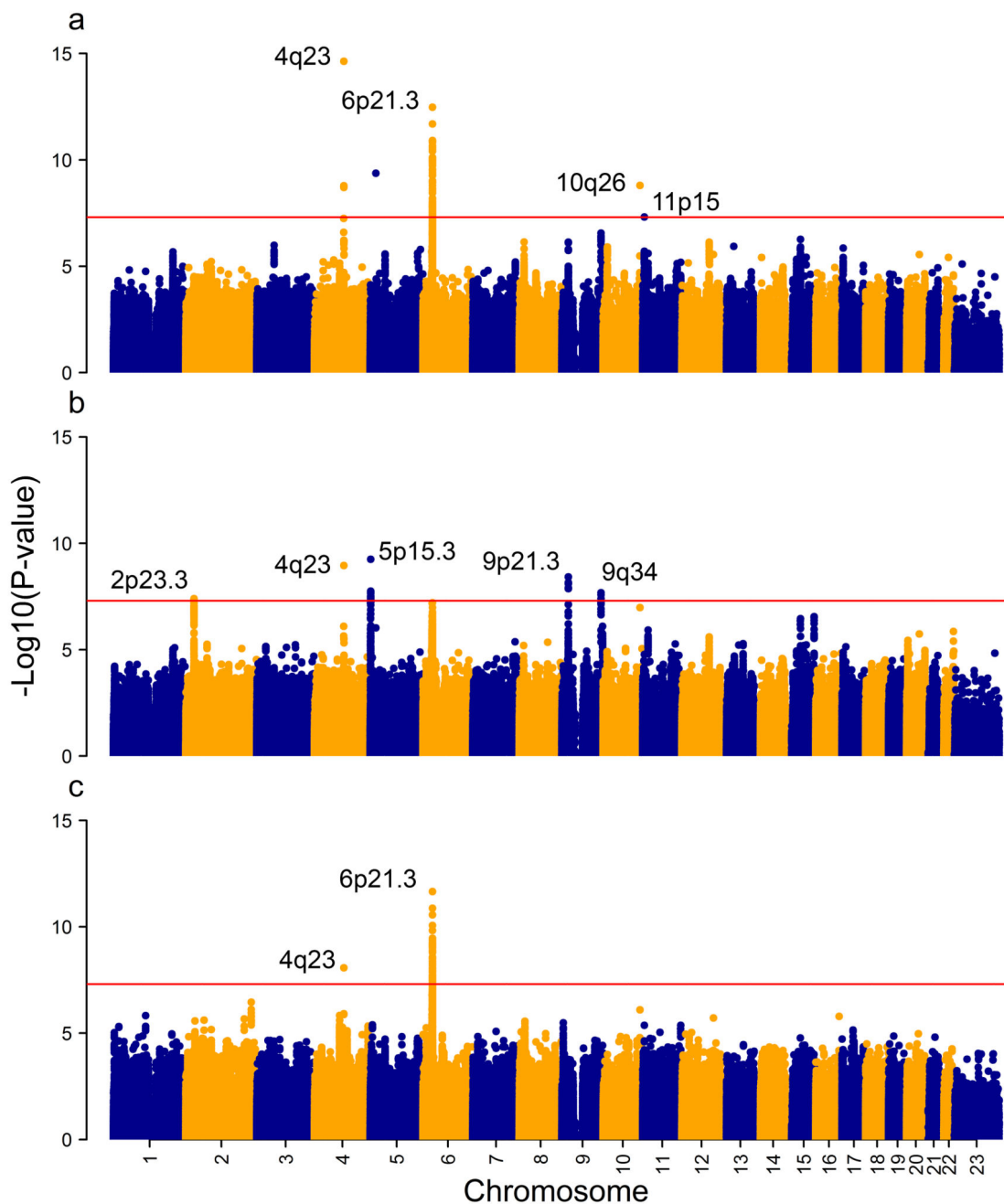


Figure 1. Genome-wide associations meta-analyses results.

The red line represents $P = 5 \times 10^{-8}$. The y-axis represents the $-\log_{10} P$ -values. (a) Overall OC and pharyngeal cancer 6,034 cases and 6585 controls (b) Oral cancer analysis with 2,990 cases and 6,585 controls (c) Oropharyngeal cancer analysis with 2,641 cases and 6,585 controls. Loci with GWA significant SNPs and technically validated are tagged with genomic location.

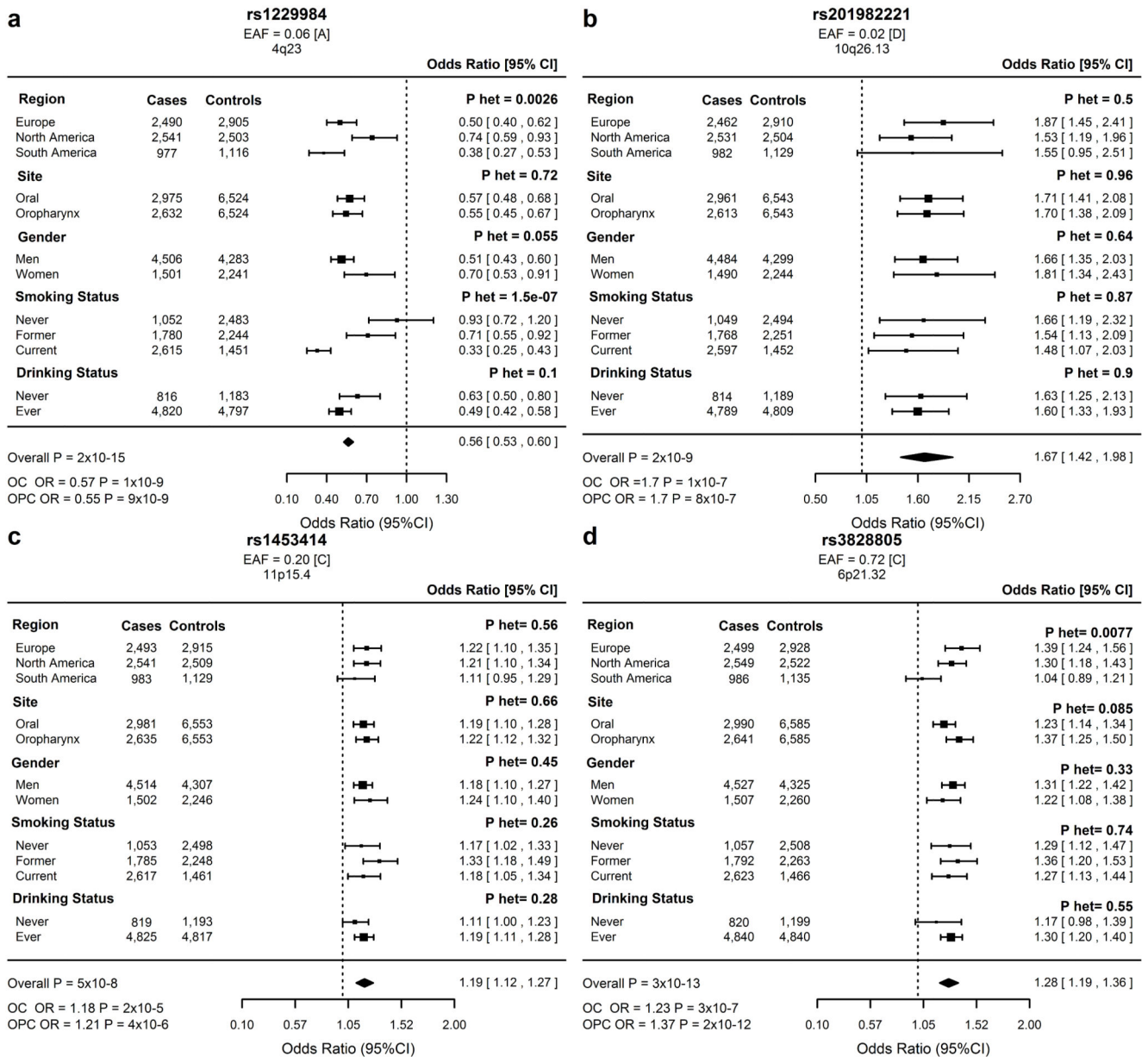


Figure 2. Forest plots of odds ratios for the lead SNP at each genome-wide significant loci in the overall oral and pharyngeal cancer meta-analysis.
(a)4q23, rs1229984 (b)10q26.13, rs20198222 (c)11p15.4 rs1453414. (d)6p21.32 rs3828805. EAF = effect allele frequency in 6585 controls. Effect allele in square brackets. OC = oral cancer; OPC = oropharynx cancer.

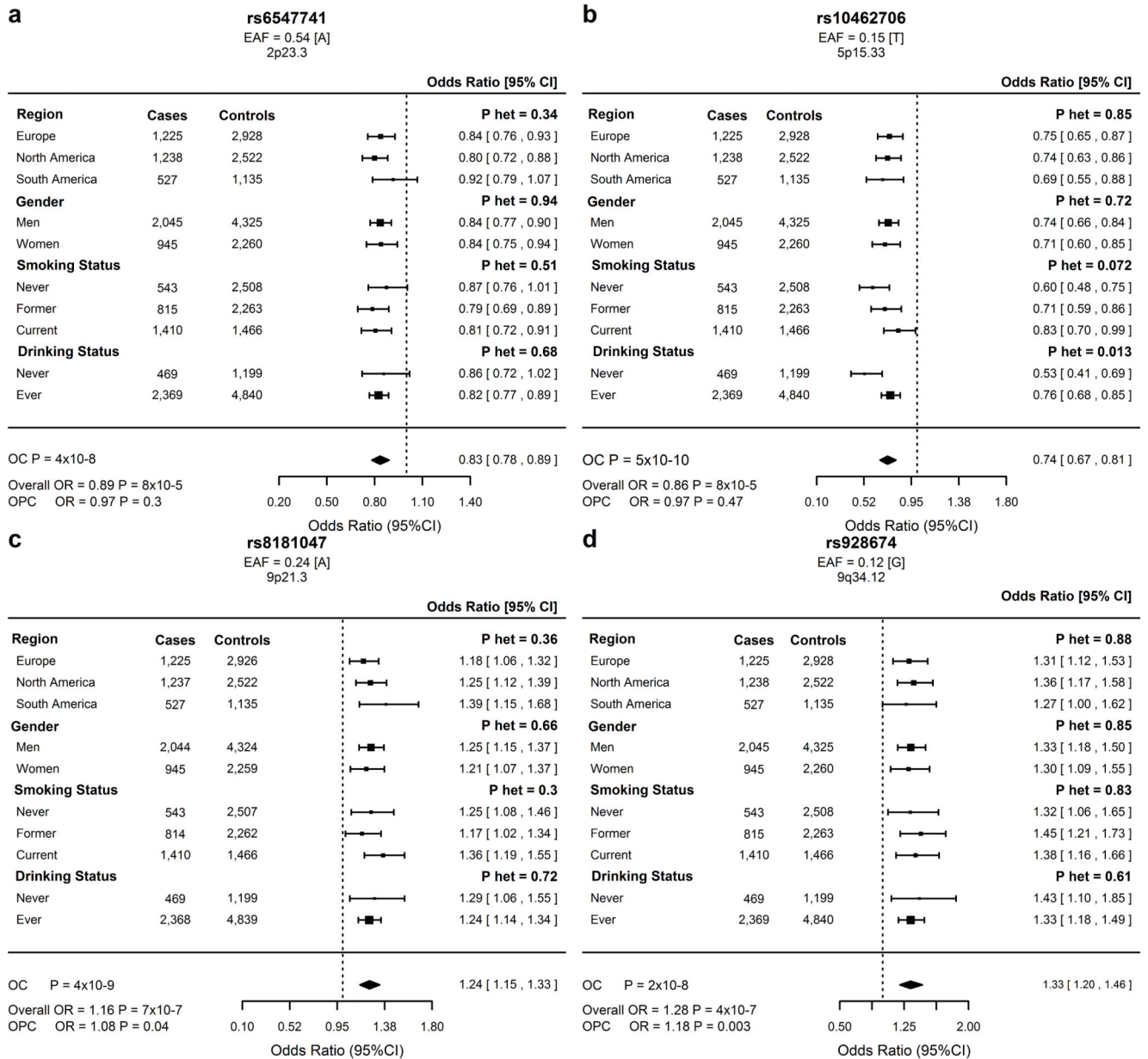


Figure 3. Forest plots of odds ratios for the lead SNP at each genome-wide significant loci in the oral cancer (OC) meta-analysis.

(a)2p23.3 rs6547741 (b)5p15.33, rs10462706 (c)9p21.3, rs8181047 (d)9q34.12, rs928674.

EAF = effect allele frequency in 6585 controls. Effect allele in square brackets. Overall = oral and pharyngeal cancer; OPC = oropharynx cancer.

Table 1

Epidemiological and clinical characteristics of cases and controls.

	Cases		Controls	
	N	%	N	%
Total	6,034		6,585	
Tumor site				
Oral cavity	2,990	49.6		
Oropharynx	2,641	43.8		
Hypopharynx	305	5.1		
Overlapping	73	1.2		
Other	25	0.4		
Geographic Region				
Europe	2,499	41.4	2,928	44.5
North America	2,549	42.2	2,522	38.3
South America	986	16.3	1,135	17.2
Sex				
Male	4,527	75.02	4,325	65.68
Female	1,507	24.98	2,260	34.32
Age, years				
≤50	1,315	21.8	1,355	20.6
50-60	2,006	33.2	1,954	29.7
60-70	1,748	29.0	1,983	30.1
≥70	964	16.0	1,293	19.6
Unknown	1	0.02		
Smoking Status				
Never	1,057	17.5	2,508	38.1
Former	1,792	29.7	2,263	34.4
Current	2,623	43.5	1,466	22.3
Unknown	562	9.3	348	5.3
Drinking Status				
Never	820	13.6	1,199	18.2
Ever	4,840	80.2	4,840	73.5
Unknown	374	6.2	546	8.3

Table 2

Lead genome-wide significant SNP for each validated locus from the regional meta-analyses of oral and pharyngeal cancers combined and analysis of each cancer separately^a and pharynx cancer combined, as well as OC^a and OPC^a separately.

Region	SNP	chr:pos ^b	Gene	EA/OA ^c	Info (Rsq) ^d	AF ^e case/control	OR	P	P _{het}
Oral and pharyngeal cancer									
4q23	rs1229984	4:100239319	<i>ADH1B</i>	A/G	Geno	0.03/0.06	0.56	2.29x10 ⁻¹⁵	0.002
6p21.32	rs3828805	6:32636120	<i>HLA-DQB1</i>	C/T	0.88	0.75/0.72	1.28	3.35x10 ⁻¹³	0.007
10q26.13	rs201982221	10:126157446	<i>LHPP</i>	D/I	Geno	0.03/0.02	1.67	1.58x10 ⁻⁹	0.50
11p15.4	rs1453414	11:5829084	<i>OR52N2/TRIM5</i>	C/A	Geno	0.23/0.20	1.19	4.78x10 ⁻⁸	0.55
Oral cancer									
2p23.3	rs6547741	2:27855924	<i>GPN1</i>	A/G	0.98	0.50/0.54	0.83	3.97x10 ⁻⁸	0.34
4q23	rs1229984	4:100239319	<i>ADH1B</i>	A/G	Geno	0.03/0.06	0.57	1.09x10 ⁻⁹	0.02
5p15.33	rs10462706	5:1343794	<i>CLPTMIL</i>	T/C	0.97	0.12/0.15	0.74	5.54x10 ⁻¹⁰	0.84
9p21.3	rs8181047	9:22064465	<i>CDKN2B-AS1</i>	A/G	Geno	0.29/0.24	1.24	3.80x10 ⁻⁹	0.37
9q34.12	rs928674	9:133952024	<i>LAMC3</i>	G/A	0.89	0.14/0.12	1.33	2.09x10 ⁻⁸	0.88
Oropharyngeal cancer									
4q23	rs1229984	4:100239319	<i>ADH1B</i>	A/G	Geno	0.02/0.06	0.55	8.53x10 ⁻⁹	0.05
6p21.32	rs3828805	6:32636120	<i>HLA-DQB1</i>	C/T	0.88	0.75/0.72	1.37	2.21x10 ⁻¹²	0.07

^aOC=oral cancer, OPC=oropharyngeal cancer;

^bSNP position according to NCBI genome build 37 (Hg19);

^cEA=Effect allele; OA=other allele;

^dGeno=genotyped, SNP, INFO, R² is the average across imputation batches;

^eAF=allele frequency of the effect allele

Table 3

Associations of *DRB1*1301-DQA1*0103-DQB1*0603* haplotype in individuals with >70% Caucasian ancestry

	haplotype ^a	case/control	case/control	HF ^b case	HF ^b control	OR	P	Meta-analysis ^c	
								OR	P
Oral and pharynx cancer								0.68	3.32x10 ⁻¹⁰
Europe	207/422		2,497/2,928	0.04	0.07	0.60	4.04x10 ⁻⁸		
North America	207/276		2,342/2,329	0.04	0.06	0.74	1.68x10 ⁻³		
South America	74/101		613/727	0.06	0.07	0.86	0.35		
Oral cancer								0.75	1.72x10 ⁻⁴
Europe	106/422		1,231/2,928	0.04	0.07	0.60	1.52x10 ⁻⁵		
North America	128/276		1,135/2,329	0.06	0.06	0.92	4.67x10 ⁻¹		
South America	41/101		351/727	0.06	0.07	0.80	0.26		
Oropharynx cancer								0.59	2.73x10 ⁻⁹
Europe	84/422		1,098/2,928	0.04	0.07	0.57	2.69x10 ⁻⁵		
North America	72/276		1,119/2,329	0.03	0.06	0.52	3.49x10 ⁻⁶		
South America	31/101		216/727	0.07	0.07	1.05	0.81		
Oropharynx cancer by HPV status^d									
HPV-positive	11/505		336/3,686	0.01	0.07	0.23	1.6x10 ⁻⁶		
HPV-negative	25/505		240/3,686	0.05	0.07	0.75	0.16		

^aNumber of copies of the haplotype in cases and controls

^bHaplotype frequency calculated as total number of copies of haplotype in the population (haplotype copies/2n).

^cFixed-effects meta-analysis of regional associations adjusted for age, sex and eigenvectors.

^dHPV-status available in a subset of cases from ARCAGE, EPIC, CHANCE and Pittsburgh studies