Genome-wide association meta-analysis identifies pleiotropic risk loci for aerodigestive squamous cell cancers

Corina Lesseur, Aida Ferreiro-Iglesias, James D. McKay, Yohan Bossé, Mattias Johansson, Valérie Gaborieau, Maria Teresa Landi, David C. Christiani, Neil C. Caporaso, Stig E. Bojesen, Christopher I. Amos, Sanjay Shete, Geoffrey Liu, Gadi Rennert, Demetrius Albanes, Melinda C. Aldrich, Adonina Tardon, Chu Chen, Liloglou Triantafillou, John K. Field, Marion Dawn Teare, Lambertus A. Kiemeney, Brenda Diergaarde, Robert L. Ferris, Shanbeh Zienolddiny, Stephen Lamp, Andrew F. Olshan, Mark C. Weissler, Martin Lacko, Angela Risch, Heike Bickeböller, Andy R. Ness, Steve Thomas, Loïc Le Marchand, Matthew B. Schabath, Victor Wünsch-Filho, Eloïza H. Tajara, Angeline S. Andrew, Gary M. Clifford, Philipp Lazarus, Kjell Granvik, Mikael Johansson, Susanne Arnold, Olle Melander, Hans Brunström, Stefania Boccia, Gabriella Cadoni, Wim Timens, Ma’en Obeidat, Xiangjun Xiao, Richard S. Houlston, Rayjean J. Hung, Paul Brennan

1 Section of Genetics, Genetic Epidemiology Group, International Agency for Research on Cancer, World Health Organization, Lyon, France. 2 Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America. 3 Section of Genetics, Genetic Cancer Susceptibility Group, International Agency for Research on Cancer, World Health Organization, Lyon, France. 4 Department of Molecular Medicine, Institut universitaire de cardiologie et de pneumologie de Québec, Laval University, Quebec City, Canada. 5 Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America. 6 Department of Environmental Health, Harvard TH Chan School of Public Health, Massachusetts General Hospital, Boston, Massachusetts, United States of America. 7 Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen, Denmark. 8 Department of Medicine, Baylor college of Medicine, Houston, Texas, United States of America. 9 Department of Biostatistics, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, United States of America. 10 Lunenfeld-Tanenbaum Research Institute of Sinai Health System, University of Toronto, Toronto, Canada. 11 Cilăt National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel. 12 Department of Thoracic Surgery, Division of Epidemiology, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America. 13 Faculty of Medicine, University of Oviedo and CIBERESP, Oviedo, Spain. 14 Department of Epidemiology, University of Washington School of Public Health and Community Medicine, Seattle, Washington, United States of America. 15 Institute of Translational Medicine, University of Liverpool, Liverpool, United Kingdom. 16 School of Health and Related Research, University Of Sheffield, Sheffield, United Kingdom. 17 Radboud University Medical Center, Nijmegen, The Netherlands. 18 Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America. 19 Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America. 20 UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America. 21 National Institute of Occupational Health, Oslo, Norway. 22 British Columbia Cancer Agency, Vancouver, Canada. 23 Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America. 24 Department of Otolaryngology/Head and Neck Surgery, UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America. 25 Department of Otorhinolaryngology, Head and Neck Surgery, Maastricht University Medical Center, Maastricht, The Netherlands. 26 University of Salzburg, Department of Biosciences and Cancer Cluster Salzburg, Salzburg, Austria. 27 Division of Epigenomics, DKFZ – German Cancer Research Center, Heidelberg, Germany. 28 Translational Lung Research Center Heidelberg (TLRC-H), Member of the German Center for Lung Research (DZL), Heidelberg, Germany. 29 Department of Genetic Epidemiology, University Medical Center, Georg-August-University Göttingen, Göttingen, Germany. 30 National Institute for Health Research (NIHR) Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust, Bristol, United Kingdom. 31 Bristol Dental School, University of Bristol, Bristol, United Kingdom.
summary statistics for the lung squamous dataset are deposited in dbGaP (phs001273.v3.p2). The oral and pharyngeal GWAS summary statistics by cancer site and world region have been deposited in the IEU Open GWAS platform (https://gwas.mrcieu.ac.uk/) under the GWAs IDs: ieu-b-91, ieu-b-95 and 98.

**Funding:** The INTEGRAL-ILCOnOncArray was supported by the Centre for Inherited Disease Research (268201200086-0-26800068-1). Genotyping for the oral and oropharyngeal cancer OncoArray was funded through the U.S. National Institute of Dental and Craniofacial Research (NIDCR) grant 1X01HG007870-0. The Integrative Analysis of Lung Cancer Risk and (INTEGRAL) of the International Lung Cancer Consortium (ILCOnO) was supported by grants U19-CA148127 and CA148127S1 and more recently by the INTEGRAL grant U19CA203654. ILCOnO data harmonization is supported by the Canada Research Chair to R.J.H. and U19 CA203654. C.I.A is a Research Scholar of the Cancer Prevention Institute of Texas and supported by RR170048. The work of the Houston Laboratory is funded by Cancer Research UK. The CAPUA study was supported by FIS-FEDER/Spain grant numbers FIS-01/310, FIS-P103-0365, and FIS-07-B1060604, FICYT/Asturias grant numbers FICYT PB02-67 and FICYT IB09-133, and the University Institute of Oncology (UIOPA), of the University of Oviedo and the Ciber de Epidemiología y Salud Pública. CIBERESP, SPAIN. The work performed in the CARET study was supported by the National Institute of Health R01 CA111703 (PI Chen), National Institute of Dental and Craniofacial Research (NIDCR) grant 1X01HG007780 -0. The Integrative Innovation, St Paul’s Hospital, The University of British Columbia, Vancouver, Canada, Division of Genetics and Epidemiology, The Institute of Cancer Research, London, United Kingdom, 53 Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Canada

* rayjean.hung@lunenfeld.ca (RJH); gep@iarc.fr (PB)

**Abstract**

Squamous cell carcinomas (SqCC) of the aerodigestive tract have similar etiological risk factors. Although genetic risk variants for individual cancers have been identified, an agnostic, genome-wide search for shared genetic susceptibility has not been performed. To identify novel and pleotropic SqCC risk variants, we performed a meta-analysis of GWAS data on lung SqCC (LuSqCC), oro/pharyngeal SqCC (OSqCC), laryngeal SqCC (LaSqCC) and esophageal SqCC (ESqCC) cancers, totaling 13,887 cases and 61,961 controls of European ancestry. We identified one novel genome-wide significant ($P_{meta}<5x10^{-8}$) aerodigestive SqCC susceptibility loci in the 2q33.1 region (rs56321285, $TMEM273$). Additionally, three previously unknown loci reached suggestive significance ($P_{meta}<5x10^{-7}$): 1q32.1 (rs12133735, near $MDM4$), 5q31.2 (rs13181561, $TMEM173$) and 19p13.11 (rs61494113, $ABHD8$). Multiple previously identified loci for aerodigestive SqCC also showed evidence of pleiotropy in at least another SqCC site, these include: 4q23 ($ADH1B$), 6p21.32 ($STK19$), 6p21.32 ($HLA-DQB1$), 9p21.33 ($CDKN2B-AS1$) and 13q13.11 ($BRCA2$). Gene-based association and gene set enrichment identified a set of 48 SqCC-related genes rel to DNA damage and epigenetic regulation pathways. Our study highlights the importance of cross-cancer analyses to identify pleiotropic risk loci of histology-related cancers arising at distinct anatomical sites.
Introduction

The squamous cell carcinomas (SqCC) of the aerodigestive tract [1], lung squamous cell carcinoma (LuSqCC) and head and neck cancers (HNC, >90% SqCCs) including; oral/pharyngeal SqCC (OSqCC), larynx SqCC (LaSqCC), and esophageal SqCC (ESqCC); are strongly associated with common risk factors such as tobacco smoking, alcohol consumption and human papilloma virus (HPV) infection [2]. Similarly, recent molecular characterization studies across anatomically distinct SqCCs have shown that histology is more important than tissue of origin in defining tumor molecular profiles determined by shared features including somatic mutations, copy number alternations, deregulation of DNA methylation and/or gene expression [3-4].

Along with behavioral risk factors, it is increasingly recognized that inherited factors also play a role in aerodigestive SqCC risk. Previous genome-wide association studies (GWAS) have identified multiple genetic risk variants for individual aerodigestive SqCC types; notably variants in smoking-related genes at 15q25.1 for LuSqCC [5-7] and 4q23 variants in alcohol-related genes for upper aerodigestive tract (UADT) cancers [8]. Importantly, candidate-gene and GWAS studies have previously described rare genetic variants linked to aerodigestive SqCC risk; including variants near BRCA2 (13q13.1), first identified as a risk factor for ESqCC in Middle Eastern populations [9] and later described to increase risk of LuSqCC [10] and UADT SqCCs in Europeans [11]. Similarly, at 22q12.1 another rare missense variant within CHEK2 (rs17879961, p.Ile157Thr) has been linked to reduced risk of lung and UADT SqCCs [10-13]. Such studies provide evidence of genetic pleiotropy across aerodigestive SqCCs, as these variants exert cross-cancer effects possibly related to similar underlying mechanisms (i.e. DNA repair). Furthermore, a recent large-scale genome-wide genetic correlation analysis across six solid tumors (breast, colorectal, head/neck, lung, ovary and prostate cancer), highlighted that the strongest genetic correlation was between lung and head and neck cancers [14].

Collectively aerodigestive SqCC are an important public health issue; not only because as a group are amongst the most common type of solid tumors [2], but also due to the increasing global incidence of HPV-related head and neck SqCCs [15]. Identifying genetic risk loci that can have pleiotropic effects across aerodigestive SqCC sites is important for gaining insight into shared or divergent molecular basis of different tumors. To further examine pleiotropic risk genomic regions across LuSqCC [16], OSqCC [17], LaSqCC [8] and ESqCC [8] and to

Author summary

Squamous cell carcinomas are specific type of cancer that can arise in multiple organs of the aerodigestive tract including the lung, oral cavity, oropharynx, larynx and esophagus. Previous studies have shown that aerodigestive squamous cell carcinomas share common environmental risk factors (tobacco smoking and alcohol intake). Here, we investigate genetic factors involved in the risk of aerodigestive squamous cell carcinomas as a group in a large genetic association study involving 13,887 cancer cases and 61,961 controls. We identified one genome-wide significant region within 2q33.1 and 3 other suggestive regions at 1q32.1, Sq31.2 and 19p13.11. Gene-based analyses also identify a list of SqCC-related genes that are involved in DNA damage response and epigenetic regulation. Our results suggest some overlap in the genetic factors influencing the risk of aerodigestive squamous cell carcinomas in European populations and highlights the importance of cross-cancer studies.
Results

Overview

We performed GWAS meta-analysis on aerodigestive SqCC risk including 13,887 cancer cases and 61,961 non-overlapping controls of European ancestry. The SqCC cases comprised 7,426 LuSqCC, 5,452 OsSqCC, part of the OncoArray Consortium [16–18], and additional 693 LaSqCC and 316 ESqCC previously included in a upper aerodigestive cancer GWAS [8] (Table 1). Summary associations statistics were used to perform fixed-effects (F-E) and a subset-based meta-analyses using the ASSET software [19]. This approach allows exploration of all possible subsets of studies to identify the strongest association signal, while accounting for subset search multiple testing, and adjusting standard errors to account for overlapping controls between analyses; partial overlap (N = 2,500) between LuSqCC and OsSqCC and complete overlap between the EsqCC and LaSqCC. After quality control steps, 8,468, 885 genetic variants with summary statistics in at least three of the four interrogated SqCC types were used for analyses. The quantile–quantile plot (F-E meta-analysis, S1 Fig) shows little evidence of genomic inflation after correcting for sample size (λ = 1.006). Loci that reached $P_{\text{meta}} < 5 \times 10^{-7}$ were considered noteworthy; meta-analysis results for all SNPs below $P < 5 \times 10^{-8}$ are shown in S1 Table. From noteworthy loci, those not previously reported in the lung [16] or oral/pharyngeal [17] analyses (single-cancer $P > 5 \times 10^{-7}$) were considered as novel SqCC regions. We identified one novel aerodigestive SqCC loci at genome-wide significance (F-E meta-analysis, $P_{\text{meta}} < 5 \times 10^{-8}$) within 2q33.1. We detected suggestive associations with SqCC risk at 1q32.1, 5q31.2 and 19p13.11, not detected in previous analyses (Fig 1 and Table 2). Other loci that reached $P_{\text{meta}} < 5 \times 10^{-7}$ were considered pleiotropic if these had at least two cancer sites at $P < 5 \times 10^{-4}$ and the same effect direction in all tumor sites. Using these criteria, the loci categorized as pleiotropic (4q23, 6p21.32, 6p21.33, 6p22, 9p21.3 and 13q13.1) included 108 SNPs (S2 Table), the lead SNP (lowest $P_{\text{meta}}$) for each of these regions is shown in Table 3. In contrast, other known cancer regions that reached the GWAS threshold (12p13.33, 15q25, 19q13.2) or $P_{\text{meta}} < 5 \times 10^{-7}$ (4p14, 9q34.1, 10q24.31, 11q21, 15q13.5) in the SqCC meta-analysis were not pleiotropic (Fig 1). We did not observe additional associations reaching the GWAS threshold or suggestive significance in the ASSET subset-based meta-analysis, indicating that at least for the strongest associations, the effects have consistent direction across the examined aerodigestive SqCC types. For noteworthy SNPs, we performed expression quantitative trait (eQTL) analyses with normal lung tissues from the multicenter Lung Microarray Study (S3 Table). We also query these variants in multiple public genomic annotation databases. The Genotype-Tissue Expression (GTEx) for lung and esophageal eQTLs (S4 Table). ROADMAP and the Encyclopedia of DNA Elements (ENCODE) for epi/genomic annotations (S5 Table). The NHGRI EBI GWAS Catalog (S6 Table) for disease/phenotype associations and the COSMIC catalogue for cancer somatic mutation information (S7 Table). Lastly, we performed a genome-wide gene-based association analysis (GWGAS) of the SqCC meta-analyses results using MAGMA (Multi-marker Analysis of GenoMic Annotation) [20] (S8 Table). To map individual SNPs to genes we used the Functional Mapping and Annotation (FUMA, S9 Table) [21]. Overlapping genes from these were used to assemble a list of aerodigestive SqCC genes (S10 Table) used in enrichment analyses (S11 Table).
Novel loci with pleiotropic aerodigestive SqCC associations

At 2q33.1, the intronic variant rs56321285[A] within the transmembrane protein 237 (TMEM237) gene was associated with reduced risk of aerodigestive SqCC (OR = 0.90, $P_{\text{meta}} = 6.99 \times 10^{-9}$). This association showed little heterogeneity across cancer sites LuSqCC: OR = 0.92, $P = 2.51 \times 10^{-4}$; OSqCC: OR = 0.89, $P = 2.34 \times 10^{-4}$; LaSqCC: OR = 0.79, $P = 3.83 \times 10^{-3}$; ESqCC: OR = 0.80, $P = 3.56 \times 10^{-2}$ (Fig 2A). rs56321285 is in low linkage disequilibrium (LD) with other variants in the region (S2 Fig) including rs10931936 ($r^2 = 0.02$, 1000 Genomes (1KG), Europeans), the lead SNP of a weaker 2q33.1 association (SqCC $P_{\text{meta}} = 1.08$, $P_{\text{meta}} = 1.83 \times 10^{-6}$).

Table 1. Summary of studies included in the aerodigestive SqCC meta-analysis.

| Study                  | Tumor site          | Cases | Controls\(a\) | Array                  | Imputation panel       | Imputation Quality | Number of Variants | Covariates | Ancestry | Publication               |
|-----------------------|---------------------|-------|----------------|------------------------|------------------------|------------------|-------------------|------------|----------|--------------------------|
| Lung cancer OncoArray | Lung                | 7,426 | 55,630         | Illumina OncoArray     | 1000 Genomes v3       | R^2\(>0.3\)       | 10,439,017        | Age, sex,    | European | McKay, Hung et al 2017   |
| Oral and oropharynx cancer OncoArray | Oral and Oropharynx | 5,452 | 5,984          | Illumina OncoArray     | HRC                    | R^2\(>0.3\)       | 7,542,495        | Age, sex,    | European | Lesseur et al 2016        |
|                        |                     | 2,698 |                |                         |                        |                  |                   | PCs         |          |                          |
|                        |                     | 2,414 |                |                         |                        |                  |                   |             |          |                          |
|                        |                     | 340   |                |                         |                        |                  |                   |             |          |                          |
| UADT cancer GWAS       | Larynx              | 693   | 2,847          | Illumina Human-Hap300   | HRC                    | R^2\(>0.3\)       | 8,840,446        | Age, sex,    | European | McKay et al 2011          |
|                        |                     | 316   | 2,847          |                         |                        |                  |                   |             |          |                          |

$R^2$ = imputation quality measure; MAF = minor allele frequency; UADT = upper aerodigestive tract; HRC = Haplotype Reference Consortium panel; \(a\) Overlapping controls N = 2,500 (lung and oral/oropharynx) and N = 2847 (larynx and esophageal).

\(b\) Cases with overlapping oral and oropharyngeal tumors.

https://doi.org/10.1371/journal.pgen.1009254.t001

**Novel loci with pleiotropic aerodigestive SqCC associations**

At 2q33.1, the intronic variant rs56321285[A] within the transmembrane protein 237 (TMEM237) gene was associated with reduced risk of aerodigestive SqCC (OR = 0.90, $P_{\text{meta}} = 6.99 \times 10^{-9}$). This association showed little heterogeneity across cancer sites LuSqCC: OR = 0.92, $P = 2.51 \times 10^{-4}$; OSqCC: OR = 0.89, $P = 2.34 \times 10^{-4}$; LaSqCC: OR = 0.79, $P = 3.83 \times 10^{-3}$; ESqCC: OR = 0.80, $P = 3.56 \times 10^{-2}$ (Fig 2A). rs56321285 is in low linkage disequilibrium (LD) with other variants in the region (S2 Fig) including rs10931936 ($r^2 = 0.02$, 1000 Genomes (1KG), Europeans), the lead SNP of a weaker 2q33.1 association (SqCC $P_{\text{meta}} = 1.08$, $P_{\text{meta}} = 1.83 \times 10^{-6}$).

---

**Fig 1. Manhattan plot of aerodigestive SqCC genome-wide association fixed-effects meta-analysis results.** The y-axis corresponds to $-\log_{10} P$-values, and x-axis to genomic positions. Horizontal red dashed line ($P = 5 \times 10^{-8}$) and black dashed line ($P = 5 \times 10^{-7}$). Highlighted in red are the newly identified pleotropic aerodigestive SqCC loci ($P < 5 \times 10^{-8}$), in blue new loci at $P < 5 \times 10^{-7}$ and in purple previously identified pleotropic loci (at least 2 cancer sites). Loci labeled in black are loci that reached $P < 5 \times 10^{-7}$ but were not pleiotropic associated only in single-cancer analyses.

https://doi.org/10.1371/journal.pgen.1009254.g001
rs10931936 is in LD with nearby CASP8-ALS2CR12 variants which have been previously linked with risk of multiple cancers in Europeans [22], as well as esophageal and lung cancer in Chinese populations [23,24]. CASP8 plays an important role in apoptosis; mutations in this gene have been described in 2% of LuSqCC and 6% of UADT SqCC tumors (S7 Table) [25]. However, the 2q33.1 genome-wide significant SNP (rs56321285) associated with aerodigestive SqCC risk, seems independent from CASP8-ALS2CR12 variants. In eQTL analyses using lung tissues Lung Microarray Study (S3 Table), rs56321285 is a nominally significant cis-eQTL for AOX2P (Laval and Groningen datasets) and CDK15 (Laval and UBC datasets). However, rs56321285 is not a lung or esophageal eQTL in the GTEx catalog [26]. Regulatory annotations from ENCODE [27] and ROADMAP [28] are consistent with rs56321285 mapping to a H3K4me1 enhancer in lung fibroblasts (S5 Table).

The lead SNP (rs12133735) at 1q32.1; the G allele was associated with increased risk of aerodigestive SqCC (OR\textsubscript{meta} = 1.08, P\textsubscript{meta} = 2.16x10\textsuperscript{-7}, Table 2 and Fig 2B), predominantly driven by the LuSqCC (OR = 1.07, P = 4.63x10\textsuperscript{-8}) and OSqCC (OR = 1.14, P = 9.82x10\textsuperscript{-6}) results. rs12133735 is located 3' of MDM4 (S3 Fig) and is a MDM4 eQTL in all datasets from the lung eQTL study (S3 Table and S4 Fig), and in lung and esophageal GTEx tissues [26] (S4 Table). MDM4 is a crucial negative regulator of p53 and its upregulation has been described as a common p53 inactivation mechanism in tumors [29,30]. In contrast, in our analyses rs12133735-G is associated with lower MDM4 expression in lung tissues and increased SqCC risk. However, the regulation of MDM4 expression and interaction with p53 involves complex mechanisms (including alternative splicing) and are reported to differ between normal and cancer tissues [29]. In Europeans, rs12133735[G] is in moderate LD (r\textsuperscript{2} = 0.61, 1KG) with rs4245739 [C] (SqCC OR\textsubscript{meta} = 1.07, P = 7.62x10\textsuperscript{-8}) which has been associated with increased risk of triple

### Table 2. Novel genomic regions with pleiotropic aerodigestive SqCC associations.

| Region\textsuperscript{a} | EA/OA\textsuperscript{b} | Gene | EAF | SqCC site | OR | 95%CI | P | OR\textsubscript{meta} | 95%CI\textsubscript{meta} | P\textsubscript{meta}\textsuperscript{c} |
|----------------|----------------|-------|-----|------------|----|-------|---|----------------|----------------|-------------------|
| Genome-wide significant loci |
| 2q33.1 | A/G | TEMEM237 | 0.31 | Lung | 0.92 | 0.88–0.96 | 2.51E-04 | 0.902 | 0.87–0.94 | 6.99E-09 |
| rs56321285 | | Oral/oropharynx | 0.89 | 0.83–0.94 | 2.34E-04 |
| 2:202505545 | | Larynx | 0.79 | 0.67–0.93 | 3.83E-03 |
| | | Esophagus | 0.8 | 0.64–0.98 | 3.56E-02 |
| Suggestive loci |
| 1q32.1 | G/T | MDM4 | 0.35 | Lung | 1.07 | 1.03–1.11 | 4.63E-04 | 1.08 | 1.05–1.12 | 2.16E-07 |
| rs12133735 | | Oral/oropharynx | 1.14 | 1.08–1.21 | 9.82E-06 |
| 1:20456836 | | Larynx | 1.05 | 0.92–1.20 | 5.14E-01 |
| | | Esophagus | 1 | 0.83–1.19 | 9.74E-01 |
| 5q31.2 | G/A | TEMEM173 | 0.28 | Lung | 1.09 | 1.05–1.14 | 5.29E-05 | 1.09 | 1.05–1.13 | 1.74E-07 |
| rs13181561 | | Oral/oropharynx | 1.1 | 1.04–1.17 | 1.74E-03 |
| 5:138850905 | | Larynx | 1.12 | 0.96–1.3 | 1.44E-01 |
| | | Esophagus | 1.06 | 0.87–1.3 | 5.62E-01 |
| 19p13.11 | A/G | ABHD8 | 0.3 | Lung | 1.09 | 1.05–1.14 | 2.05E-05 | 1.09 | 1.05–1.12 | 9.86E-08 |
| rs61494113 | | Oral/oropharynx | 1.09 | 1.03–1.16 | 4.91E-03 |
| 19:17401859 | | Larynx | 1.15 | 1.00–1.31 | 5.22E-02 |
| | | Esophagus | 1.05 | 0.87–1.27 | 5.90E-01 |

\textsuperscript{a} Lead SNP (lowest P, F-E meta-analysis), regions at P<5x10\textsuperscript{-7}.

\textsuperscript{b} EA = Effect allele/OA = other allele; EAF = average effect allele frequency between sites.

\textsuperscript{c} F-E meta-analysis (ASSET) accounting for control overlap between Lung SCC and oral/pharynx analysis and larynx and esophageal analysis.

https://doi.org/10.1371/journal.pgen.1009254.t002
Table 3. Variants within known genomic loci with pleiotropic aerodigestive SqCC associations.

| Region | EA/RA | Gene | EAF | SqCC site  | OR  | 95%CI       | P      | OR_{meta} | 95%CI_{meta} | P_{meta} |
|--------|-------|------|-----|-----------|-----|------------|--------|-----------|-------------|---------|
| 4q23   | T/C   | ADH1B| 0.05| Lung      | 0.92| 0.84–1.01  | 7.35E-02| 0.80      | 0.74–0.86   | 1.89E-09|
| rs1229984 | Oral/oropharynx | 0.58 | 0.50–0.67 | 8.32E-13 |
| 4:100239319 | Larynx | 0.67 | 0.49–0.92 | 1.42E-02 |
|         | Esophagus | 0.28 | 0.15–0.53 | 9.28E-05 |
| 6p21.33 | G/A   | STK19| 0.09| Lung      | 1.29| 1.22–1.37  | 1.41E-16| 1.26      | 1.19–1.32   | 2.42E-19|
| rs389884 | Oral/oropharynx | 1.21 | 1.09–1.35 | 4.41E-04 |
| 6:31940897 | Larynx | 1.21 | 0.95–1.52 | 1.19E-01 |
|         | Esophagus | 1.24 | 0.92–1.68 | 1.65E-01 |
| 6p21.32 | G/A   | HLA-DQA1| 0.41| Lung      | 1.17| 1.12–1.21  | 3.01E-14| 1.16      | 1.12–1.19   | 4.82E-19|
| rs9271611 | Oral/oropharynx | 1.15 | 1.08–1.22 | 1.23E-05 |
| 6:32591609 | Larynx | 1.27 | 1.07–1.49 | 5.14E-03 |
|         | Esophagus | 1.02 | 0.82–1.28 | 8.36E-01 |
| 9p21.3  | T/C   | CDKN2B-AS1| 0.30| Lung      | 1.09| 1.05–1.14  | 1.34E-05| 1.11      | 1.07–1.14   | 5.55E-10|
| rs7857345 | Oral/oropharynx | 1.14 | 1.07–1.21 | 3.73E-05 |
| 9:22087473 | Larynx | 1.18 | 1.03–1.36 | 1.91E-02 |
|         | Esophagus | 1.07 | 0.89–1.29 | 4.63E-01 |
| 13q13.1 | A/G   | BRCA2 | 0.01| Lung      | 2.12| 1.77–2.55  | 1.10E-15| 2        | 1.73–2.32   | 2.30E-21|
| rs11571815 | Oral/oropharynx | 1.67 | 1.28–2.18 | 1.67E-04 |
| 13:32968550 | Larynx | 3.04 | 1.41–6.57 | 4.53E-03 |
|         | Esophagus | 4.73 | 2.14–10.5 | 1.24E-04 |

* Lead SNP (lowest P, F-E meta-analysis), regions at P<5x10^{-7};
* EA = Effect allele/OA = other allele; EAF = average allele frequency between sites;
* F-E meta-analysis (ASSET) accounting for control overlap.

https://doi.org/10.1371/journal.pgen.1009254.t003

Fig 2. Forest plots of ORs for new aerodigestive SqCC-related loci. a) rs56321285 at 2q33.1 (TMEM23). Subset refers to subset based meta-analyses (ASSET). b) rs12133735 at 1q32.1 (near MDM4).  
https://doi.org/10.1371/journal.pgen.1009254.g002
negative breast cancer [31,32] and ovarian cancer [33] (S6 Table). Intriguingly, rs4245739 [C] has also been associated with reduced prostate cancer risk (Europeans) [34] and lower risk of all cancers in Asians [35,36]. A candidate-gene study [37] also described associations between risk of HPV16-associated OSqCC and 1q32.1 MDM4 SNPs including rs11801299 (r² = 0.12, with rs12133735, 1KG, Europeans), which was marginally associated with SqCC risk in our analysis (ORmeta = 0.91, Pmeta = 1.34x10⁻⁵).

The lead variant at 5q31.2 rs13181561[G] (ORmeta = 1.09, Pmeta = 1.74x10⁻⁷, Fig 3A) near TMEM173 (S5 Fig) showed homogenous associations across tumor sites but only significant in LuSqCC and OSqCC. rs13181561 is associated with DNAJC18 and SPATA24 gene expression in lung tissues (Laval, Groningen and UBC; S3 Table), and of DNAJC18 in esophagus (GTEx, S4 Table). rs13181561 overlaps with an enhancer in esophageal and lung tissues (S5 Table). Additionally, rs13181561[G] is highly correlated with rs7447927[C] (r² = 0.94, 1KG, Europeans), the latter (ORmeta = 1.08, P = 5.45x10⁻⁷) has been previously linked to increased ESqCC risk in Chinese populations [38] (S6 Table).

Another suggestive SqCC association was detected at rs61494113[A] within 19p13.11; (ORmeta = 1.09, Pmeta = 9.9x10⁻⁸); showed similar odds ratios acrossSqCCs sites albeit with limited power for larynx and esophageal SqCCs (Table 1 and Fig 3B). Lung eQTL analysis showed rs61494113 as a significant eQTL for OCEL1 (Laval and Groening, S3 Table). However, the GTEx catalog shows rs61494113 as an esophageal ABHD8 eQTL and a BABAM1 splice-QTL (lung and esophagus) but not for OCEL1 (S4 Table). rs61494113 maps within H3K4me1 histone and DNase marks in normal lung tissues and in lung carcinoma cells (S5 Table). rs61494113[A] is in complete LD with rs56069439[A] (r² = 1, 1KG, Europeans), also associated with a SqCC risk (S1 Table and S6 Fig) and previously linked with increased risk of ER-negative breast [39] and ovarian [39,40] cancers (S6 Table). The 19p13.11 region of LD contains multiple genes involved in DNA damage repair including BABAM1 and BRCA1-interacting protein [41] and ANKLE1 [42].

**Known risk loci with pleiotropic aerodigestive SqCC associations**

Chromosome 6 showed a large aerodigestive SqCC association signal overlapping the human leukocyte antigen (HLA), region previously identified in the LuSqCC [16] and oral/pharyngeal SqCC [17] cancer analyses. rs389884 near STK19 was the top pleotropic SNP at 6p21.33 (SqCC ORmeta = 1.26; P = 2.4x10⁻¹⁹, Table 3 and S2 Table). 6p21.33 SNPs are in moderate LD (rs389884 and rs115785414, r² > 0.4, 1KG, Europeans) with variants at 6p22.1, suggesting a common haplotype. We also detected associations at 6p21.32; rs9271611 near HLA-DQA1 (class II) is not correlated (S7 Fig) with 6p21.33 SNPs, pairwise LD between rs9267123 and rs9271611 r² = 0.09 (1 KG, Europeans). This second association reduced aerodigestive SqCC risk (ORmeta = 0.85; P = 1x10⁻¹⁷) mainly for OSqCC and LuSqCC (Table 3). These observations are in concert with our previous findings [17,43]; of at least two haplotypes within the HLA region with different effects on cancer risk. Genetic variants at 9p21.33 have also been associated with multiple malignancies including lung adenocarcinoma [16] and OSqCC [17].

rs7857345 (9p21.33 lead SNP mapped to CDKN2B-AS1, S8 Fig) was associated with a slight increase in SqCC risk (ORmeta = 1.11, Pmeta = 5.55x10⁻¹⁰, Table 3). rs7857345 is in LD with rs61271866 (r² = 0.51) previously associated with ESqCC in Chinese populations [38]. However, rs7857345 is not in LD with rs885518 (r² = 0.006, 1KG, Europeans), suggesting that this is a different signal to that previously reported for lung adenocarcinoma risk in Europeans [16].

Expectedly, two other previously reported rare variants at 4q23 (rs1229984, ADH1B) and 13q13.1 (rs11571815, BRCA2) also displayed pleiotropy in this analyses [10–12] [44] (Fig 1, Table 3). Of note, we did not observe pleiotropy for variants at 15q25.1, a known locus related
to lung cancer [5] and smoking behavior [45]. The lead SNP in this region rs55781567 (CHRNA5) was prominent for SqCC (OR<sub>meta</sub> = 1.19; P<sub>meta</sub> = 1.74x10<sup>-29</sup>), but this result was primarily driven by the LuSqCC (OR = 1.3; P = 4.6x10<sup>-41</sup>) with no effect in any other SqCC site (OSqCC P = 0.26; LaSqCC P = 0.68 and ESqCC P = 0.7, S1 Table and S9 Fig). Variants at 15q25 have been interrogated before in relation to upper aerodigestive cancer risk (including samples used in this study) [46]; significant associations were found only in women and unrelated to smoking behavior suggesting that 15q25.1 SNPs relate differently to LuSqCC compared to oral/oropharyngeal SqCCs. Importantly, none of the published HNC GWAS in Chinese or Europeans have reported associations with 15q25 variants. Thus, while the relation between this locus with smoking behavior and lung cancer is unequivocal; to date there is no evidence of a clear link between 15q25 and head and neck cancer. Future studies including more cases and stratified analyses should examine this further.

Aerodigestive SqCCs risk genes and pathways
To gain further functional insight into aerodigestive SqCC genetic susceptibility, we used the results of the F-E meta-analyses to map risk variants to genes (FUMA) and to perform a genome-wide gene-based association analysis (GWGAS) using MAGMA. The SNP to gene analyses highlighted 182 genes within 21 genomic regions (S8 Table) and the gene-based analysis identified 51 significant genes related to aerodigestive SqCC (Bonferroni correction P<2.67x10<sup>-6</sup>, S9 Table). Next, we overlapped the results from the two analyses and obtained a list of 48 SqCC-related genes (S10 Table), which includes TMEM237 (2q33.1), MDM4 (1q32.1), AC138517.1 (5q31.2) and BABAM1 (19p13.11) located within the pleotropic SqCC risk regions identified in the meta-analyses. Expectedly, the HLA region in chromosome 6 had the highest number of genes mapped (24), however and interestingly the most prominent signals were located within 6p22 and included >10 histone genes. Gene-set enrichment analyses using
these SqCC-related genes and canonical pathways resulted in 63 significant gene sets (S11 Table) most of which mapped to DNA damage pathways (telomere, checkpoint, oxidative stress and strand break response) as well as epigenetic regulation pathways related to histones and DNA methylation.

**Discussion**

This study identified one novel genome-wide significant loci associated with aerodigestive SqCC risk (2q33.1). Four other loci (1q32.1, 5q31.2 and 19p13.11) showed suggestive associations with SqCC. Amongst known SqCC loci, four showed evidence of pleiotropy across cancer sites. Our results demonstrate the power of cross-cancer analyses of histologically-related tumors to identify genetic risk loci.

It is notable that many of the detected associations are plausibly related to cancer risk. Our results from 2q33.1 and 5q31.2 combined with previous evidence [23,24,38] indicate that these loci relate to SqCC risk across distinct ancestries (Asians and Europeans). Moreover, the signal at 2q33.1 was also proximal to CASP8-ALS2CR12, a region previously associated with other cancers [47,48]. Likewise, the 1q32.1 and 19p13.11 genomic regions implicate genes like MDM4 and BABAM1, which have been previously associated with risk of other epithelial malignancies including breast, prostate and ovary [31,36,39,49]. Lastly, these observations not only make our findings more plausible but also expand our understanding of cross-cancer genetic susceptibility and complex biology behind these associations.

The top associations displayed homogenous effect direction across SqCC sites and stronger associations in the F-E meta-analyses compared to the subset-based meta-analyses. This could relate to the shared histology and risk factors of aerodigestive SqCCs. Nonetheless, we cannot rule out heterogeneous SqCC associations that we were not able to detect in our data. However, and not surprisingly, for these loci effect sizes were small (range of OR\textsubscript{meta} 0.89–1.09) which limited association detection in the smaller single cancer-analysis using commonly applied GWAS \textit{P}-values thresholds. In contrast, most of the known loci that exhibited pleiotropy in our analysis have larger effects sizes, particularly true for the less common variants within BRCA2 and CHEK2.

Our study has several major strengths. Firstly, we leveraged available European GWAS data sets to perform a large-scale meta-analysis of aerodigestive SqCC risk. Secondly, we analyzed tissue-specific gene expression data from multiple studies and integrated these data with publicly available information on epigenetic regulatory profiles of relevant tissues to aerodigestive SqCCs. Thirdly, for the newly discovered loci we also integrated our results with existent data from genetic susceptibility studies in other populations as well as available tumor repository information. However, this study has a number of limitations. The sample sizes for laryngeal and esophageal SqCC were very limited; this impacted our power to identify more signals at the GWAS threshold. The described associations (particularly those at \( P > 5 \times 10^{-8} \)) could be spurious due to the high testing burden and lack of replication; other studies should examine these regions further to replicate these results. Our criteria to identify pleiotropic loci tried to capture robust loci across multiple aerodigestive SqCC while accommodating for the sample size imbalances across tumor sites. We recognize that this approach did not fully account for multiple testing and could have missed some pleiotropic regions. Pleiotropic studies are limited by sample size of existent GWAS data, as well as the frequency of variants in these regions. Future studies should investigate this further using larger samples, different methodology, and if possible, including SqCCs from other sites (e.g. cervical, anal and bladder). Also, our analyses were restricted to individuals of European ancestry, performing a similar analysis including other genetic backgrounds offers the potential to pinpoint loci that exert effects across
ethnicities. In summary, we provide evidence for one new locus (2q33.1) influencing aerodigestive SqCC risk, and highlight loci for future investigation. Future work should investigate the biological mechanisms underscoring these associations to unearth shared and divergent molecular features of these histologically similar tumors.

Methods

Ethics statement

Informed written consent was obtained from all participants, and all contributing studies have been approved by the IARC Institutional Review Board (IRB; references: 14–03, 13–17, 07–02) which requires to obtain local ethics committees approvals prior to their enrolment and evaluation.

Study population

This meta-analysis includes data from three previous studies of lung squamous cell[16], oral/pharyngeal[17] and upper aerodigestive tract (UADT) cancers[8], totaling 13,887 cases and 61,961 non-overlapping controls. The characteristics and references for each study are summarized in Table 1. The SqCC cases comprise 7,426 LuSqCC, 5,452 OSqCC, part of the OncoArray Consortium (http://epi.grants.cancer.gov/oncoarray/) [16,17], and additional 693 LaSqCC and 316 ESqCC previously included in a upper aerodigestive cancer GWAS[8]. Controls partially overlapped (N = 2,500) between the LuSqCC and OSqCC analyses, and completely overlap (N = 2,847) between the ESqCC and LaSqCC analyses. For this analysis, GWAS summary-statistics for single-site SqCCs were derived using only individuals of European ancestry across multiple epidemiological studies from Europe, North and South America.

Genotyping and imputation

For each of the studies, genomic DNA samples were previously isolated from blood or buccal cells. Genotyping for the lung and oral/pharyngeal cancers OncoArray Consortium[18] studies, was performed at the Center for Inherited Disease Research (CIDR) using the Illumina OncoArray custom designed for cancer studies. Genotype calls were made in GenomeStudio software (Illumina) using a standardized cluster file for OncoArray studies. The esophageal and larynx cancer cases and controls from the upper aerodigestive tract GWA study[8] were genotyped using the Illumina Sentrix HumanHap300 BeadChip at the Centre d’Etude du Polymorphisme Humain (CEPH) and the Centre National Genotypage (CNG Evry France) as previously described[8]. Genotype data have been deposited dbGaP (https://www.ncbi.nlm.nih.gov/gap/) accession number phs001202.v1.p1 for the oral and pharyngeal study[17] and for the lung data[16] accession numbers phs001273.v3.p2 and phs000876.v2.p1. The lung cancer GWA study[16] was imputed using the 1000 genomes reference panel (phase3) (http://phase3browser.1000genomes.org/index.html/) and the oral/pharyngeal cancer, larynx and esophageal cancer GWAS were imputed using the Haplotype Reference Consortium Panel[50] (http://www.haplotype-reference-consortium.org/) using the University of Michigan Imputation Server [51] (https://imputationserver.sph.umich.edu/). Only variants with imputation quality of R2 > 0.3 were used in the meta-analysis.

Summary association statistics and meta-analyses

Cancer risk association results from two previous OncoArray Consortia studies (LuSqCC[16] and OSqCC[17]) and the esophageal and laryngeal analyses ORs, P-values and standard errors
for each SNP for each cancer site were obtained using logistic regression with a log additive models adjusted for age, sex and principal components using plink2[52] (https://www.cog-genomics.org/plink2/) and R[53] (http://www.r-project.org/). Summary statistics for the lung SqCC data are deposited in dbGaP (phs001273.v3.p2). The oral and pharyngeal GWAS summary statistics by cancer site and world region have been deposited in the IEU Open GWAS platform (https://gwas.mrcieu.ac.uk/) under the GWAS IDs: ieu-b-89, ieu-b-90, ieu-b-94, ieu-b-96, ieu-b-93, ieu-b-97, ieu-b-91, ieu-b-95 and 98. Meta-analyses were performed using a fixed-effects (F-E) and subset-based meta-analysis using the ASSET software tool [19] (https://dceg.cancer.gov/tools/analysis/asset/). ASSET allows exploration of all possible subsets of studies to identify the strongest association signal, while accounting for subset search multiple testing, and adjust standard errors to account for overlapping controls between analyses; partial overlap (N = 2,500) between the LuSqCC and OSqCC and complete overlap between the ESqCC and LaSqCC analyses. Meta-analysis for a SNP was performed when at least three cancer sites had association results. P-values from both analyses were two-sided. Meta-analyses results for fixed-effects and subset-based were considered noteworthy if these reach $P < 5 \times 10^{-7}$.

Loci were considered as new if these had not been previously reported in the single SqCC cancers analysis ($P > 5 \times 10^{-7}$ for any single site). Loci with previously reported LuSqCC or OSqCC were characterized as pleiotropic if: 1) $P_{\text{meta}} < 5 \times 10^{-7}$; 2) two single cancer association results at $P < 5 \times 10^{-4}$ and consistent effect direction across all cancer sites. All analyses were performed using the R statistical environment version 3.4.3[53]. Linkage disequilibrium (LD) calculations ($R^2$) were performed using the LDlink[54] tool and the 1000 Genomes Project European ancestry populations. Regional association plots were generated using stand-alone LocusZoom v1.4[55] (https://github.com/statgen/locuszoom-standalone/). Forest plots of association results were produce using the metafor R package[56].

### Lung and esophageal cis-eQTLs

To investigate the association between lead SCC variants and mRNA expression, we used three lung eQTL data sets from the Microarray eQTL study [57], lung tissues for eQTL analysis were obtained from patients who underwent lung surgery at three academic sites: Laval University, the University of British Columbia (UBC) and the University of Groningen. Whole-genome gene expression profiling in the lung was performed on a custom Affymetrix array and is available through GEO (https://www.ncbi.nlm.nih.gov/geo/) accession number GSE23546. Genotyping was carried out on the Illunima Human 1M-Duo BeadChip array, data is accessible in dbGaP (phs001745.v1.p1). Genotypes and gene expression levels were available for 408 (Laval University), 342 (Groningen) and 287 (UBC) patients. Microarray and genotypes preprocessing, quality control and eQTL mapping have been described previously[58]. We also investigated top aerodigestive SqCC associations in the public GTEx catalog (V8)[26] for lung and esophageal tissue eQTLs and sQTLs, summary statistics based on RNAseq and genotypes analyses obtained via the GTEx data portal (https://www.gtexportal.org).

### Functional genomic annotation and gene-based analyses

To functionally annotate newly identified aerodigestive SqCC regions, we leveraged multiple resources: the Encyclopedia of DNA Elements (ENCODEx27] (https://www.encodeproject. org/) and ROADMAP Epigenomics[28] (http://www.roadmapepigenomics.org/) catalogs to obtain epi/genomic regulatory annotations (chromatin states, histones, enhancers, promoters and transcription binding sites) for lung and esophageal tissues and cell-types obtained through HaploReg 4.1 using the HaploR R package[59]); the NHGRI-EBI GWAS Catalog.
for previously reported disease/phenotype associations and the COSMIC catalogue (v90, https://cancer.sanger.ac.uk/cosmic) for cancer somatic mutation information. To provide additional insight into functional and biological mechanisms underlying aerodigestive SqCC genetic susceptibility, we performed a genome-wide gene-based association analysis (GWGAS) of the SqCC meta-analyses results using MAGMA (Multi-marker Analysis of GenoMic Annotation)[20]. We also used the Functional Mapping and Annotation (FUMA, https://fuma.ctglab.nl/)[21] which maps individually significant SNPs to genes. We selected overlapping genes from the MAGMA (Bonferroni-corrected \( P \)-value \(<2.7x10^{-6}\)) and FUMA results were used to assemble a list of genes implicated in aerodigestive SqCC genetic risk. This gene list was used to perform a gene-set analysis for curated canonical biological pathways (containing between 10 and 500 genes) from MSigDB collections[61]; including GO[62], KEGG[63], REACTOME[64] and BIOCARTA[61]. Pathway analyses were performed using MAGMA default settings of 10,000 permutations and applied a Bonferroni correction.

**Supporting information**

**S1 Fig. SqCC F-E meta-analyses.** Quantile-quantile plot of the p-values for ASSET F-E meta-analyses results including lung, oral/oropharyngeal, larynx and esophageal SqCCs. (corrected \( \lambda = 1.006 \)).

**S2 Fig. Regional association plot at 2q33.1.** Chromosome positions (x-axis) and \(-\log_{10} P\)-value (y-axis) SqCC meta-analysis at 2q33.1. Genetic variants colored red according to their LD with rs56321285 (2q33.1 lead SNP) and colored in blue according to LD values with second lead SNP rs1830298. rs56321285 and rs1830298 \( r^2 = 0.02 \).

**S3 Fig. Regional association plot at 1q32.1.** Chromosome positions (x-axis) and \(-\log_{10} P\)-value (y-axis) of SqCC F-E meta-analysis at 1q32.1. Genetic variants are colored according to their LD with the rs12133735 (red) and with rs4245739 (blue) a variant previously associated with cancer risk; rs12133735 and rs4245739 \( r^2 = 0.63 \).

**S4 Fig. rs12133735 MDM4 lung eQTL.** Boxplots for rs12133735 and MDM4 gene expression in 3 datasets from the Microarray eQTL study, from left to right: Laval University, University of British Columbia (UBC) and 3. University of Groningen.

**S5 Fig. Regional association plot at 5q31.2.** Chromosome positions (x-axis) and \(-\log_{10} P\)-value (y-axis) SqCC F-E meta-analysis at 5q31.2. Genetic variants colored according to their LD with the labeled SNP (purple diamond). rs13181561 and rs7447927 \( r^2 = 0.94 \).

**S6 Fig. Regional association plot at 19p13.11.** Chromosome positions (x-axis) and \(-\log_{10} P\)-value (y-axis) SqCC F-E meta-analysis at 19p13.11. Genotyped and imputed variants colored according to their LD with the labeled SNP (purple diamond). rs61494113 and rs56069439 \( r^2 = 1 \).

**S7 Fig. Regional association plot at 6p22.1-6p21.33.** Chromosome positions (x-axis) and \(-\log_{10} P\)-value (y-axis) SqCC F-E meta-analysis at 6p22.1-6p21.33. Variants colored according
to their LD with SNP rs9267123 (lead variant at 6p21.33). rs3116813 (6p22.1) is in moderate LD with rs9267123 ($r^2 = 0.5$). rs1049213 at 6p21.33 is not correlated with rs9267123 ($r^2 = 0.01$).

**S8 Fig. Regional association plot at 9p21.3.** Chromosome positions (x-axis) and -log10 P-value (y-axis) SqCC meta-analysis at 9p21.3. Variants colored according to their LD with SNP rs7857345 (9p21.3 lead variant).

(TIF)

**S9 Fig. Regional association plot at 15p25.1.** Regional association plot at 15q25 Chromosome positions (x-axis) and -log10 P-value (y-axis). A. aerodigestive SqCC P-values; B. Lung SqCC P-values; C. Oral and oropharyngeal cancer SqCC P-values. Genetic variants colored red according to their LD with rs55781567 (lowest P-value at 15q5 in the meta-analysis).

(TIF)

**S1 Table. Results with $P<5\times 10^{-5}$ aerodigestive SqCC meta-analyses.** All variants with $P<5\times 10^{-5}$ in the fixed-effects (F-E) ASSET meta-analyses of aerodigestive SqCC. Results for each SqCC site are also shown.

(XLSX)

**S2 Table. Pleiotropic aerodigestive SqCC risk variants.** 108 variants with $P_{\text{meta}}<5\times 10^{-7}$; in the fixed-effects (F-E) meta-analyses; two single-cancer analyses at $P<5\times 10^{-4}$ and consistent effect direction across cancer sites.

(CSV)

**S3 Table. Lung Cis-eQTLs for aerodigestive SqCC loci.** Cis-eQTLs for novel SqCC loci in the lung Microarray eQTL study datasets.

(XLSX)

**S4 Table. Cis-eQTLs and cis-sQTLs for new SqCC loci.** Lung and esophageal Cis-eQTLs and cis-sQTLs in the GTEx catalog V8 for new SqCC loci.

(XLSX)

**S5 Table. Chromatin states and histone marks in lung and esophageal tissues or cells for new SqCC loci.** Chromatin and histone annotations for new SqCC loci from the Roadmap and ENCODE projects.

(XLSX)

**S6 Table. Summary of reported cancer risk associations within the newly SqCC risk loci.** NHGRI-EBI Catalog (v1.0 e98 2020-02-08) reported cancer risk associations for lead SNP (or proxies $r^2>0.6$) within the new SqCC loci.

(XLSX)

**S7 Table. Aerodigestive loci genes with somatic mutations.** Genes within SqCC new loci with somatic mutations in the COSMIC catalogue. (release v90, 5th September 2019).

(XLSX)

**S8 Table. Significant results from the gene-based aerodigestive SqCC associations.** Analyses performed with MAGMA with 18669 protein-coding genes.

(XLSX)

**S9 Table. Aerodigestive SqCC results from the FUMA SNPs to genes mapping.**

(XLSX)
S10 Table. Genes overlapping between FUMA and MAGMA analyses.
(XLSX)

S11 Table. Aerodigestive SqCC results from gene set enrichment analyses.
(XLSX)

Acknowledgments

We acknowledge all the participants involved in this research and the funders and support.

The authors would like to thank the staff at the Respiratory Health Network Tissue Bank of
the FRQS for their valuable assistance with the lung eQTL data set at Laval University. The
lung eQTL study at Laval University was supported by the Fondation de l’Institut Universitaire
de Cardiologie et de Pneumologie de Québec, the Respiratory Health Network of the FRQS
and the Canadian Institutes of Health Research (MOP-123369). Y. Bossé holds a Canada
Research Chair in Genomics of Heart and Lung Diseases.

We thank the ARCADE study investigators and team including: Pagona Lagiou, Tatiana V.
Macfarlane, Franco Merletti, Jerry Polesel, Kristina Kjaerheim, Max Robinson, Wolfgang
Ahrens, Lorenzo Simonato, Ariana Znaor, Xavier Castellsague (deceased June 2016), David I.
Conway, Ivana Holcátová, Claire M. Healy and Peter Thomson. We thank L. Fernandez for
her contribution to the IARC ORC multicenter study. We are also grateful to S. Koifman for
his contribution to the IARC Latin America multicenter study (S. Koifman passed away in
May 2014).

Where authors are identified as personnel of the International Agency for Research on Can-
cer / World Health Organization, the authors alone are responsible for the views expressed
in this article and they do not necessarily represent the decisions, policy or views of the Interna-
tional Agency for Research on Cancer / World Health Organization.

Author Contributions

Conceptualization: Corina Lesseur, Rayjean J. Hung, Paul Brennan.

Data curation: Xiangjun Xiao.

Formal analysis: Corina Lesseur, Aida Ferreiro-Iglesias, Yohan Bosse´, Valerie Gaborieau.

Funding acquisition: Christopher I. Amos, Rayjean J. Hung, Paul Brennan.

Project administration: Corina Lesseur, Aida Ferreiro-Iglesias.

Resources: James D. McKay, Mattias Johansson, Maria Teresa Landi, David C. Christiani,
Neil C. Caporaso, Stig E. Bojesen, Christopher I. Amos, Sanjay Shete, Geoffrey Liu, Gadi
Rennert, Demetrius Albanes, Melinda C. Aldrich, Adonina Tardon, Chu Chen, Liloglou
Triantafillos, John K. Field, Marion Dawn Teare, Lambertus A. Kiemeney, Brenda Dier-
gaarde, Robert L. Ferris, Shanbeh Zienolddiny, Stephen Lam, Andrew F. Olshan, Mark C.
Weissler, Martin Lacko, Angela Risch, Heike Bickeböller, Andy R. Ness, Steve Thomas,
Loic Le Marchand, Matthew B. Schabath, Victor Wünsch-Filho, Eloiza H. Tajara, Angeline
S. Andrew, Gary M. Clifford, Philip Lazarus, Kjell Grankvist, Mikael Johansson, Susanne
Arnold, Olle Melander, Hans Brunström, Stefania Boccia, Gabriella Cadoni, Wim Timens,
Ma’en Obeidat, Richard S. Houlston, Rayjean J. Hung, Paul Brennan.

Supervision: Rayjean J. Hung, Paul Brennan.

Writing – original draft: Corina Lesseur, Aida Ferreiro-Iglesias, Rayjean J. Hung, Paul
Brennan.
Writing – review & editing: James D. McKay, Mattias Johansson, Maria Teresa Landi, David C. Christiani, Neil C. Caporaso, Stig E. Bojesen, Christopher I. Amos, Sanjay Shete, Geoffrey Liu, Gadi Rennert, Demetrios Alvanes, Melinda C. Aldrich, Adonina Tardon, Andrew F. Olshan, Mark C. Weissler, Martin Lacko, Angela Risch, Heike Bickeboller, Andy R. Ness, Steve Thomas, Loic Le Marchand, Matthew B. Schabath, Victor Wünsch-Filho, Eloiza H. Tajara, Angeline S. Andrew, Gary M. Clifford, Philip Lazarus, Kjell Grankvist, Mikael Johansson, Susanne Arnold, Olle Melander, Hans Brunström, Stefania Boccia, Gabriella Cadoni, Wim Timens, Ma’en Obeidat, Richard S. Houlston, Rayjean J. Hung, Paul Brennan.

References
1. Yan W, Wistuba II, Emmert-Buck MR, Erickson HS. Squamous Cell Carcinoma—Similarities and Differences among Anatomical Sites. Am J Cancer Res. 2011; 1(3):275–300. Epub 2011/09/23. https://doi.org/10.1158/1538-7445.am2011-275 PMID: 21938273.
2. Dotto GP, Rustgi AK. Squamous Cell Cancers: A Unified Perspective on Biology and Genetics. Cancer Cell. 2016; 29(5):622–37. Epub 2016/05/12. https://doi.org/10.1016/j.ccell.2016.04.004 PMID: 27165741.
3. Campbell JD, Yau C, Bowlby R, Liu Y, Brennan K, Fan H, et al. Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas. Cell Rep. 2018; 23(1):194–212 e6. Epub 2018/04/05. https://doi.org/10.1016/j.celrep.2018.03.063 PMID: 29617660.
4. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell. 2018; 173(2):291–304.e6. Epub 2018/04/07. https://doi.org/10.1016/j.cell.2018.03.022 PMID: 29625048.
5. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nature genetics. 2008; 40(5):616–22. Epub 2008/04/04. https://doi.org/10.1038/ng.109 PMID: 18385676.
6. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature. 2008; 452(7187):633–7. Epub 2008/04/04. https://doi.org/10.1038/nature06885 PMID: 18385738.
7. Truong T, Hung RJ, Amos CI, Wu X, Bickeboller H, Rosenberger A, et al. Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. Journal of the National Cancer Institute. 2010; 102(13):959–71. Epub 2010/06/16. https://doi.org/10.1093/jnci/djv037 PMID: 20548021.
8. McKay JD, Truong T, Gaborieau V, Chabrier A, Chuang S-C, Byrnes G, et al. A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. PLoS genetics. 2011; 7(3):e1001333. https://doi.org/10.1371/journal.pgen.1001333 PMID: 21437268.
9. Akbari MR, Malekzadeh R, Nasrollahzadeh D, Amanian D, Islami F, Li S, et al. Germline BRCA2 mutations and the risk of esophageal squamous cell carcinoma. Oncogene. 2008; 27(9):1290–6. Epub 2008/04/08. https://doi.org/10.1038/sj.onc.1210739 PMID: 17724471.
10. Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nature genetics. 2014; 46(7):736–41. Epub 2014/06/02. https://doi.org/10.1038/ng.3002 PMID: 24880342.
11. Delahaye-Sourdeix M, Anantharaman D, Timofeeva MN, Gaborieau V, Chabrier A, Vallee MP, et al. A rare truncating BRCA2 variant and genetic susceptibility to upper aerodigestive tract cancer. J Natl Cancer Inst. 2015; 107(5). Epub 2015/04/04. https://doi.org/10.1093/jnci/djv037 PMID: 25938448.
12. Brennan P, McKay J, Moore L, Zaridze D, Mukeria A, Szeshenya-Dabrowska N, et al. Uncommon CHEK2 missense variant and reduced risk of tobacco-related cancers: case–control study. Human molecular genetics. 2007; 16(15):1794–801. https://doi.org/10.1093/hmg/ddm127 PMID: 17517688.
13. Cybulski C, Masojc B, Oszutowska D, Jaworowska E, Grodzki T, Waloszczyk P, et al. Constitutional CHEK2 mutations are associated with a decreased risk of lung and laryngeal cancers. Carcinogenesis. 2008; 29(4):762–5. Epub 2008/02/19. https://doi.org/10.1093/carcin/bgn044 PMID: 18281249.
14. Jiang X, Finucane HK, Schumacher FR, Schmit SL, Tyner JP, Han Y, et al. Shared heritability and functional enrichment across six solid cancers. Nature communications. 2019; 10(1):431. Epub 2019/01/27. https://doi.org/10.1038/s41467-018-08054-4 PMID: 30683880.
15. de Martel C, Plummer M, Vignat J, Franceschi S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. Int J Cancer. 2017; 141(4):664–70. Epub 2017/04/04. https://doi.org/10.1002/ijc.30716 PMID: 28369882.

16. McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nature genetics. 2017; 49(7):1126–32. Epub 2017/06/13. https://doi.org/10.1038/ng.3892 PMID: 28604730.

17. Lesueur C, Diergaarde B, Olshan AF, Wunsch-Filho V, Ness AR, Liu G, et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. Nature genetics. 2016; 48(12):1544–50. Epub 2016/11/01. https://doi.org/10.1038/ng.3685 PMID: 27749845.

18. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. Cancer Epidemiol Biomarkers Prev. 2017; 26(1):126–35. Epub 2016/10/05. https://doi.org/10.1158/1055-9965.EPI-16-0106 PMID: 27697780.

19. Bhattacharjee S, Rajaraman P, Jacobs KB, Wheeler WA, Melin BS, Hartge P, et al. A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. Am J Hum Genet. 2012; 90(5):821–35. Epub 2012/05/09. https://doi.org/10.1016/j.ajhg.2012.03.015 PMID: 22560090.

20. de Leeuw CA, Mooij JM, Heeskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS computational biology. 2015; 11(4):e1004219. Epub 2015/04/18. https://doi.org/10.1371/journal.pcbi.1004219 PMID: 25885710.

21. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nature communications. 2017; 8(1):1826. Epub 2017/12/01. https://doi.org/10.1038/s41467-017-01261-5 PMID: 29184086.

22. Fehringer G, Kraft P, Pharoah PD, Eeles RA, Chatterjee N, Schumacher FR, et al. Cross-Cancer Genome-Wide Analysis of Lung, Ovary, Breast, Prostate, and Colorectal Cancer Reveals Novel Pleiotropic Associations. Cancer Res. 2016; 76(17):5103–14. Epub 2016/05/20. https://doi.org/10.1158/0008-5472.CAN-15-2980 PMID: 27197191.

23. Abnet CC, Wang Z, Song X, Hu N, Zhou FY, Freedman ND, et al. Genotypic variants at 2q33 and risk of esophageal squamous cell carcinoma in China: a meta-analysis of genome-wide association studies. Human molecular genetics. 2012; 21(9):2132–41. Epub 2012/02/11. https://doi.org/10.1038/hmg/ddo029 PMID: 22333630.

24. Zhao XK, Mao YM, Meng H, Song X, Hu SJ, Lv S, et al. Shared susceptibility loci at 2q33 region for lung and esophageal cancers in high-incidence areas of esophageal cancer in northern China. PLoS One. 2017; 12(5):e0177504. Epub 2017/05/26. https://doi.org/10.1371/journal.pone.0177504 PMID: 28542283.

25. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic acids research. 2019; 47(D1):D941–d7. Epub 2018/10/30. https://doi.org/10.1093/nar/gky1015 PMID: 30371878.

26. The Genotype-Tissue Expression (GTEx) project. Nature genetics. 2013; 45(6):580–5. Epub 2013/05/30. https://doi.org/10.1038/ng.2653 PMID: 23715323.

27. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489(7414):57–74. Epub 2012/09/08. https://doi.org/10.1038/nature11247 PMID: 22955616.

28. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518(7539):317–30. Epub 2015/02/20. https://doi.org/10.1038/nature14248 PMID: 25693563.

29. Haupt S, Mejía-Hernández JO, Vijayakumaran R, Keam SP, Haupt Y. The long and the short of it: the MDM4 tail so far. Journal of molecular cell biology. 2019; 11(3):231–44. Epub 2019/01/29. https://doi.org/10.1038/s41999-019-0041-1 PMID: 30689920.

30. Marine JC, Jochemsen AG. MDMX (MDM4), a Promising Target for p53 Reactivation Therapy and Beyond. Cold Spring Harbor perspectives in medicine. 2016; 6(7). Epub 2016/07/03. https://doi.org/10.1101/cshperspect.a026237 PMID: 27371671.

31. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nature genetics. 2013; 45(4):392–8, 8e1-2. Epub 2013/03/20. https://doi.org/10.1038/ng.2561 PMID: 23357333.

32. Milne RL, Kuchenbaecker KB, Michailidou K, Beesley J, Kar S, Lindstrom S, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nature genetics. 2017; 49(12):1767–78. Epub 2017/10/24. https://doi.org/10.1038/ng.3785 PMID: 29058716.

33. Gansmo LB, Bjornsmot M, Halle MK, Salvesen HB, Dørum A, Birkeland E, et al. The MDM4 SNP34091 (rs4245739) C-allele is associated with increased risk of ovarian-but not endometrial cancer. Tumour
Biol. 2016; 37(8):10697–702. Epub 2016/02/13. https://doi.org/10.1007/s13277-016-4940-2 PMID: 26867771.

34. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nature genetics. 2018; 50(7):928–36. Epub 2018/06/13. https://doi.org/10.1038/s41588-018-0142-8 PMID: 29892016.

35. Zhai Y, Dai Z, He H, Gao F, Yang L, Dong Y, et al. A PRISMA-compliant meta-analysis of MDM4 genetic variants and cancer susceptibility. Oncotarget. 2016; 7(45):73935–44. Epub 2016/10/16. https://doi.org/10.18632/oncotarget.12558 PMID: 27738340.

36. Wang MJ, Luo YJ, Shi ZY, Xu XL, Yao GL, Liu RP, et al. The associations between MDM4 gene polymorphisms and cancer risk. Oncotarget. 2016; 7(34):55611–23. Epub 2016/10/16. https://doi.org/10.18632/oncotarget.10877 PMID: 27742919.

37. Yu H, Sturgis EM, Liu Z, Wang LE, Wei Q, Li G. Modifying effect of MDM4 variants on risk of HPV16-associated squamous cell carcinoma of oropharynx. Cancer. 2012; 118(6):1684–92. Epub 2011/08/09. https://doi.org/10.1002/cncr.21823114.

38. Wu C, Wang Z, Song X, Feng XS, Abnet CC, He J, et al. Joint analysis of three genome-wide association studies of esophageal squamous cell carcinoma in Chinese populations. Nature genetics. 2014; 46(9):1001–6. Epub 2014/08/19. https://doi.org/10.1038/ng.3064 PMID: 25129146.

39. Lawrenson K, Kar S, McCue K, Kuchenbaecker K, Michailidou K, Tyrer J, et al. Functional mechanisms underlying pleiotropic risk alleles at the 19p13.1 breast-ovarian cancer susceptibility locus. Nature communications. 2016; 7:12675. Epub 2016/09/08. https://doi.org/10.1038/ncomms12675 PMID: 27601076.

40. Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nature genetics. 2017; 49(5):680–91. Epub 2017/03/28. https://doi.org/10.1038/ng.3826 PMID: 28346442.

41. Vikrant, Sawant UU, Varma AK. Role of MERIT40 in stabilization of BRCA1 complex: a protein-protein interaction study. Biochem Biophys Res Commun. 2014; 446(4):1139–44. Epub 2014/03/29. https://doi.org/10.1016/j.bbrc.2014.03.073 PMID: 24667804.

42. Brachner A, Foisner R. Lamina-associated polypeptide (LAP) 2alpha and other LEM proteins in cancer biology. Adv Exp Med Biol. 2014; 773:143–63. Epub 2014/02/25. https://doi.org/10.1007/978-1-4899-8032-8_7 PMID: 24563347.

43. Ferreira-Iglesias A, Lesueur C, McKay J, Hung RJ, Han Y, Zong X, et al. Fine mapping of MHC region in lung cancer highlights independent susceptibility loci by ethnicity. Nature communications. 2018; 9(1):3927. Epub 2018/09/27. https://doi.org/10.1038/s41598-018-22834-x PMID: 30254314.

44. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nature genetics. 2002; 31(1):55–9. Epub 2002/04/23. https://doi.org/10.1038/ng879 PMID: 11967536.

45. Hancock DB, Regnissin GW, Gaddis NC, Chen X, Saccone NL, Lutz SM, et al. Genome-wide meta-analysis reveals common splice site acceptor variant in CHRNA4 associated with nicotine dependence. Transl Psychiatry. 2015; 5:e651. Epub 2015/10/07. https://doi.org/10.1038/tp.2015.149 PMID: 26440539.

46. Chen D, Truong T, Gaborieau V, Byrnes G, Chabrier A, Chuang S-c, et al. A sex-specific association between a 15q25 variant and upper aerodigestive tract cancers. Cancer Epidemiology Biomarkers & Prevention. 2011; 20(4):658–64. https://doi.org/10.1158/1055-9965.EPI-10-1008 PMID: 21335511.

47. Stacey SN, Hilderson G, Gudjonsson SA, Thorleifsson G, Zink F, Sigurdsson A, et al. New basal cell carcinoma susceptibility loci. Nature communications. 2015; 6:6825. Epub 2015/04/10. https://doi.org/10.1038/ncomms7825 PMID: 25865136.

48. Couch FJ, Kuchenbaecker KB, Michailidou K, Mendoza-Fandino GA, Nord S, Lileyquist J, et al. Identification of four novel susceptibility loci for oestrogen receptor negative breast cancer. Nature communications. 2016; 7:11375. Epub 2016/04/28. https://doi.org/10.1038/ncomms11375 PMID: 27117709.

49. Eeles RA, Olama AA, Benlloch S, Saunders EJ, Leongarmbrt DA, Tymrakiewicz M, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nature genetics. 2013; 45(4):385–91, 91e1-2. Epub 2013/03/29. https://doi.org/10.1038/ng.2560 PMID: 23535732.

50. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature. 2015; 526(7571):68–74. Epub 2015/10/04. https://doi.org/10.1038/ nature15393 PMID: 26432245.

51. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nature genetics. 2016; 48(10):1284–7. Epub 2016/08/30. https://doi.org/10.1038/ng.3656 PMID: 27571263.
52. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015; 4:7. Epub 2015/02/28. https://doi.org/10.1186/s13742-015-0047-8 PMID: 25722852.

53. R Core Development Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.

54. 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(21):3555–7. Epub 2015/07/04. https://doi.org/10.1093/bioinformatics/btv402 PMID: 26139635.

55. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010; 26(18):2336–7. Epub 2010/07/17. https://doi.org/10.1093/bioinformatics/btq419 PMID: 20634204.

56. Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of statistical software. 2010; 36(3).

57. Lamontagne M, Berube JC, Obeidat M, Cho MH, Hobbs BD, Sakornsakolpat P, et al. Leveraging lung tissue transcriptome to uncover candidate causal genes in COPD genetic associations. Human molecular genetics. 2018; 27(10):1819–29. Epub 2018/03/17. https://doi.org/10.1093/hmg/ddy091 PMID: 29547942.

58. Hao K, Bosse Y, Nickle DC, Pare PD, Postma DS, Laviolette M, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. PLoS Genet. 2012; 8(11):e1003029. Epub 2012/12/05. https://doi.org/10.1371/journal.pgen.1003029 PMID: 23209423.

59. Zhbannikov IY, Arbeev K, Ukraintseva S, Yashin AI. haploR: an R package for querying web-based annotation tools. F1000Res. 2017; 6:97. Epub 2017/06/20. https://doi.org/10.12688/f1000research.10742.2 PMID: 28620455.

60. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic acids research. 2017; 45(D1):D896–D901. Epub 2016/12/03. https://doi.org/10.1093/nar/gkw1133 PMID: 27899670.

61. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell systems. 2015; 1(6):417–25. Epub 2016/01/16. https://doi.org/10.1016/j.cels.2015.12.004 PMID: 26771021.

62. Gene Ontology Consortium: going forward. Nucleic acids research. 2015; 43(Database issue):D1049–56. Epub 2014/11/28. https://doi.org/10.1093/nar/gku179 PMID: 25428369.

63. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic acids research. 2016; 44(D1):D457–62. Epub 2015/10/18. https://doi.org/10.1093/nar/gkv1070 PMID: 26476454.

64. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. Nucleic acids research. 2020; 48(D1):D498–d503. Epub 2019/11/07. https://doi.org/10.1093/nar/gkz1031 PMID: 31691815.