



DATABASE

CCGD-ESCC: A Comprehensive Database for Genetic Variants Associated with Esophageal Squamous Cell Carcinoma in Chinese Population



Linna Peng^{1,#,a}, Sijin Cheng^{2,3,#,b}, Yuan Lin^{2,3,#,c}, Qionghua Cui^{1,d},
Yingying Luo^{1,e}, Jiahui Chu^{1,f}, Mingming Shao^{1,g}, Wenyi Fan^{1,h}, Yamei Chen^{1,i},
Ai Lin^{1,j}, Yiyi Xi^{1,k}, Yanxia Sun^{1,l}, Lei Zhang^{4,m}, Chao Zhang^{5,n}, Wen Tan^{1,o},
Ge Gao^{2,3,*p}, Chen Wu^{1,*q}, Dongxin Lin^{1,r}

¹ Department of Etiology and Carcinogenesis, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

² State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Center for Bioinformatics, Peking University, Beijing 100871, China

³ Beijing Advanced Innovation Center for Genomics (ICG), Peking University, Beijing 100871, China

⁴ National Engineering Research Center for Beijing Biochip Technology, Beijing 102206, China

⁵ School of Life Sciences, Peking University, Beijing 100871, China

Received 8 January 2018; revised 20 March 2018; accepted 26 March 2018

Available online 9 September 2018

Handled by Yu Xue

* Corresponding authors.

E-mail: chenwu@cicams.ac.cn (Wu C), gaog@mail.cbi.pku.edu.cn (Gao G).

Equal contribution.

^a ORCID: 0000-0002-4094-8002.

^b ORCID: 0000-0002-2565-7127.

^c ORCID: 0000-0002-9583-9313.

^d ORCID: 0000-0001-5285-5260.

^e ORCID: 0000-0003-4154-4277.

^f ORCID: 0000-0003-2949-1491.

^g ORCID: 0000-0002-4690-0245.

^h ORCID: 0000-0002-7875-7404.

ⁱ ORCID: 0000-0002-9072-5157.

^j ORCID: 0000-0001-7502-7935.

^k ORCID: 0000-0002-1293-0550.

^l ORCID: 0000-0002-2122-4009.

^m ORCID: 0000-0003-4498-8236.

ⁿ ORCID: 0000-0003-4167-4872.

^o ORCID: 0000-0002-7231-7838.

^p ORCID: 0000-0001-6470-8815.

^q ORCID: 0000-0003-4954-1011.

^r ORCID: 0000-0002-8723-8868.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2018.03.005>

1672-0229 © 2018 The Authors. Production and hosting by Elsevier B.V. on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

KEYWORDS

Esophageal cancer;
Germline variants;
Somatic mutations;
Association database;
Chinese population

Abstract Esophageal squamous-cell carcinoma (ESCC) is one of the most lethal malignancies in the world and occurs at particularly higher frequency in China. While several genome-wide association studies (GWAS) of **germline variants** and whole-genome or whole-exome sequencing studies of **somatic mutations** in ESCC have been published, there is no comprehensive database publically available for this cancer. Here, we developed the Chinese Cancer Genomic Database-Esophageal Squamous Cell Carcinoma (CCGD-ESCC) database, which contains the associations of 69,593 single nucleotide polymorphisms (SNPs) with ESCC risk in 2022 cases and 2039 controls, survival time of 1006 ESCC patients (survival GWAS) and gene expression (expression quantitative trait loci, eQTL) in 94 ESCC patients. Moreover, this database also provides the associations between 8833 somatic mutations and survival time in 675 ESCC patients. Our user-friendly database is a resource useful for biologists and oncologists not only in identifying the associations of genetic variants or somatic mutations with the development and progression of ESCC but also in studying the underlying mechanisms for tumorigenesis of the cancer. CCGD-ESCC is freely accessible at <http://db.cbi.pku.edu.cn/ccgd/ESCCdb>.

Introduction

The studies of cancer genetics and genomics powered by rapidly developing high-throughput biochip and sequencing technologies have promoted our understanding of the roles of genetic variants in carcinogenesis and cancer progression. However, most literature only reports the most significant association results due to the space limitation of the paper. Although many well-known genomic databases are available, the information on esophageal squamous-cell carcinoma (ESCC) is scarce due to the limited number of studies and its relatively low prevalence in Caucasian populations [1]. However, ESCC is the fourth most common leading cause of cancer-related deaths and kills ~250,000 people in China annually [2]. Therefore, there is an increasing demand to establish a comprehensive genetic and genomic database for ESCC, which is currently absent.

The genetic etiology of ESCC remains largely unknown. However, several studies indicate that individuals' genetic makeup alone with some environmental factors, such as alcohol intake and cigarette smoking, play an important role in esophageal carcinogenesis in China [3]. Using genome-wide association studies (GWAS), a powerful and successful tool in identification of common disease alleles, 14 susceptible genes or loci including *PLCE1*, *ADH* and *ALDH2* have been shown to be associated with ESCC risk [4–9]. However, the use of stringent *P* value threshold ($P < 10^{-7}$) might miss some true ESCC-associated loci with moderate effect sizes. Moreover, accumulating evidence has demonstrated that genetic variants associated with complex diseases or phenotypes are often located in non-coding regions and could regulate the gene expression. Therefore, expression quantitative trait loci (eQTL) analysis might be an effective method in identification of SNPs with potential regulatory functions [10–12]. However, the genome-wide eQTL analysis for ESCC based on tumors and the paired normal esophageal tissues has not yet been performed, limiting its power to identify functional genes or variants associated with ESCC.

Currently, surgery, chemotherapy and radiotherapy are commonly used to treat ESCC. Unfortunately, the long-term outcome in patients is still dismal, with 5-year survival rate around 30% [13]. Previous studies have suggested that

germline genetic variability may provide important prognostic information for ESCC [14,15], while recent whole-exome sequencing (WES) or whole-genome sequencing (WGS) studies have identified the landscape of somatic mutations as another important genomic factor in ESCC [16–22]. However, most previous GWAS, WGS and WES analyses suffer from the lack of patients' clinical information such as tumor staging, treatment outcome and survival time. Such lack of the clinical information would hinder the exploration and identification of the genomic changes that are associated with phenotypes and affect treatment outcomes, which are critical in translational and precision medicine.

In the present study, we integrated the genetic and genomic data derived from the studies in our laboratory [6,7,9,14,22] and other groups [16–21] to establish the Chinese Cancer Genomic Database-Esophageal Squamous Cell Carcinoma (CCGD-ESCC) database. Our endeavor is not only to provide comprehensive information on the associations of genetic variants or somatic mutations with ESCC risk and prognosis but also to dissect their potential regulatory functions based on eQTL analysis in ESCC tumor and normal esophageal tissues. The CCGD-ESCC database has an intuitive, well-organized and easy-to-use interface for search and display. We believe that this database can enhance our understanding of genetics, genomics and their biological functions in ESCC and facilitate the generations of new genetic and genomic markers or potential therapeutic targets by integrating multidimensional information.

Data collection and processing

To build the CCGD-ESCC database, we integrated multiple genetic association results and set up four sections. These include (1) GWAS of 16,544 SNPs in 2022 ESCC cases and 2039 controls; (2) survival GWAS of 1652 SNPs in 1006 ESCC patients; (3) eQTLs from 94 ESCC patients with both WGS and RNA-seq data in tumor and paired normal tissues; and (4) 8833 single nucleotide variations (SNVs)/indels in the protein-coding regions from 675 ESCC patients with WGS or WES data and their associations with survival time in ESCC patients (Table 1). The criteria of enrollment of study subjects were described in the previous publications [6,7,9,14,22]. Genetic variants were annotated using the ANNOVAR with

Table 1 Basic data information of four sections in the CCGD-ESCC

Section	Subjects	Platform	No. of SNPs and/or genes included
GWAS	2022 ESCC cases and 2039 controls	Affymetrix GeneChip Human Mapping 6.0 set	16,544 SNPs
Survival GWAS	1006 ESCC patients with survival information	Affymetrix GeneChip Human Mapping 6.0 set	1652 SNPs
eQTL	94 ESCC patients	WGS and RNA-seq	16,984 and 54,445 eQTLs in tumor and the adjacent normal tissues, respectively
Somatic mutation	675 ESCC patients with survival information	WGS (94 patients) and WES (581 patients)	8833 SNVs/indels in 662 genes

Note: GWAS, genome-wide association study; eQTL, expression quantitative trait loci; ESCC, esophageal squamous-cell carcinoma; WGS, whole-genome sequencing; WES, whole-exome sequencing; SNP, single nucleotide polymorphism; SNV, single nucleotide variation.

the reference cluster identifier (rsID) from the NCBI dbSNP build 146 [23].

First section: GWAS

Germline genotype data of 2022 cases and 2039 controls were obtained using Affymetrix GeneChip Human Mapping 6.0 set (Affymetrix). To increase power and coverage for association analyses, we used MACH software to impute untyped SNPs in whole genome using the 1000 Genomes Project Phase 3 ASN (Asian) as the reference panel [24]. We then filtered out SNPs with low imputation quality (estimated R-squared, *i.e.*, $R_{sq} < 0.3$), low minor allele frequency ($MAF < 0.01$) and divergence from Hardy-Weinberg equilibrium ($P < 1.0 \times 10^{-6}$). After the quality control (QC) process, 8,252,518 SNPs on autosomes were used in the final GWAS analysis. For each SNP, we carried out association analysis with ESCC risk using an additive model in a logistic regression (1-degree-of-freedom) adjusted for age, sex, smoking status and drinking status as covariates. The odds ratios (ORs) are presented for the minor allele of each SNP. Consequently, 16,544 SNPs associated with ESCC risk with $P < 1.0 \times 10^{-5}$ were retained in the CCGD-ESCC.

Second section: survival GWAS

We included 1006 ESCC patients with the information on survival outcome available to identify variants associated with length of survival time. The overall survival time of 1006 ESCC patients was calculated from the date of diagnosis to the date of last follow-up or death, with median follow-up time of 101 months. As a result, 8,279,620 SNPs on autosomes passed the same QC criteria as for the GWAS analysis in the first section. Log-rank test was used to calculate the associations between SNPs and survival time in ESCC patients. The additive model was used for survival analysis if the homozygous genotype for the minor allele was $> 1\%$; otherwise, the survival analysis was performed using the dominant model. Consequently, a total of 1652 SNPs with $P < 1.0 \times 10^{-4}$ was retained in the CCGD-ESCC.

Third section: eQTL

We performed eQTL analysis to evaluate the associations between genotypes extracted from WGS data and gene expression based on RNA-seq data in paired tumor and

normal tissues obtained from 94 individuals with ESCC. The sequencing data production and processing were described in detail elsewhere [22]. A total of 15,075,079 SNPs was called using FreeBayes based on the WGS of blood samples from 94 individuals [25]. SNPs were excluded if their minor allele frequency (MAF) is $< 5\%$ or the genotype distributions were deviated from those expected according to the Hardy-Weinberg equilibrium ($P < 1.0 \times 10^{-6}$). We also excluded genes that have absent call in over 90% of the samples. After QC, 6,092,313 SNPs and 18,085 genes on autosomes were selected for final *cis*-eQTL analysis using linear regression with additive model in the R package Matrix eQTL [26]. The effect size was defined as the slope coefficient of linear regression. Only SNPs within a ± 100 -kb window of the transcription start site (TSS) of a given gene on the same chromosome were tested. Consequently, a total of 7,676,942 gene-*cis*-SNP pairs were tested, and we found 16,984 *cis*-eQTLs in tumor tissues and 54,445 *cis*-eQTLs in paired normal tissues, which were significantly associated with gene expression with a false discovery rate (FDR) of < 0.05 .

Fourth section: somatic mutations and their associations with patients' survival

We combined our WGS data in 94 individuals with ESCC [22] and WES data in 581 individuals with ESCC obtained from published studies [16–21] to analyze the effects of somatic mutations on survival time in these patients, whose survival time information is available. After QC, 31,673 SNVs or indels in the protein-coding regions and 2203 genes on autosomes that contained mutations in $> 1\%$ of all samples were analyzed. Kaplan–Meier method was used to estimate the median survival time and log-rank test was used to test the significant difference between patients with or without somatic mutations ($P < 0.05$ as significant threshold). Finally, there are 662 genes retained in the CCGD-ESCC.

Web interface

We designed the CCGD-ESCC to enable users to explore the associations between genetic variants, somatic mutations and genes of interest and risk, survival time or gene expression for ESCC. The overview of the database can be found at the upper part of the homepage. A quick search box is provided in the middle part of the homepage for searching by the SNP identifier, gene symbol or chromosomal region (Figure 1A). The

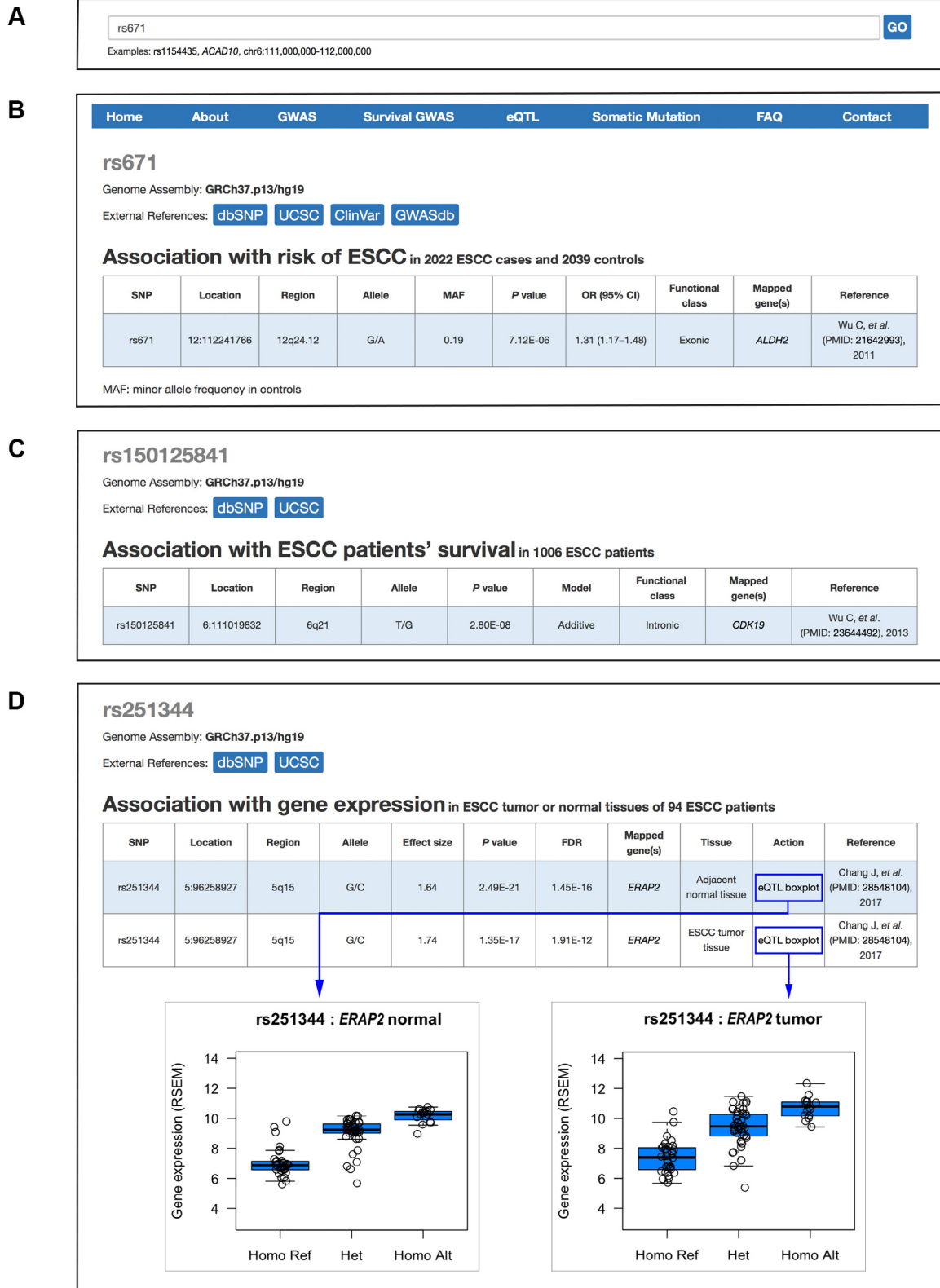


Figure 1 The association results for variants in CCGD-ESCC

A. The search bar of the database; **B.** The association of rs671 with ESCC risk in the GWAS section; **C.** The association of rs150125841 with survival time in ESCC patients in the Survival GWAS section; **D.** The regulatory effect of rs251344 on ERAP2 in both ESCC tumor tissue and paired normal esophageal tissue in the eQTL section. GWAS, genome-wide association study; eQTL, expression quantitative trait loci; ESCC, esophageal squamous-cell carcinoma; MAF, minor allele frequency in controls; OR, odds ratio for the minor allele; RSEM, RNA-seq by expectation maximization.

detailed information of the four sections including GWAS, Survival GWAS, eQTL and Somatic Mutation is shown in tabular format and is further organized into two different web pages: the variant page and the gene page.

The variant page

To acquire the information of a SNP, users can search the database by providing its rsID from dbSNP (build 146). The variant page provides the basic information about the SNP, including its chromosomal location, genomic region, reference allele, alternative allele, as well as the hyperlinks to dbSNP, UCSC, ClinVar, and GWASdb. The GWAS section shows the association results with ESCC risk including MAF in controls, OR, 95% confidence interval (CI), *P* value, and its functional annotations by ANNOVAR (Figure 1B). In the Survival GWAS section, users can obtain the association results of SNPs and patients' overall survival, including *P* value and the functional annotations by ANNOVAR (Figure 1C). The eQTL section provides the genes affected by SNPs with effect size and *P* value obtained in ESCC tumor tissues or normal esophageal tissues (Figure 1D). Users can obtain the boxplot figure of eQTL after clicking on the eQTL results. In addition, users can also reach our related publications in PubMed by clicking the links for more information.

The gene page

To explore a specific gene region, users can search the database using gene symbol. The gene page shows the primary information of the gene, including the full gene name, Entrez ID, brief gene description by NCBI, as well as hyperlinks to GeneCards, HUGO, COSMIC, Ensembl, UCSC and GWASdb. We provide the association results of all SNPs located in and around the gene annotated by ANNOVAR in the GWAS, Survival GWAS and eQTL sections (Figure 2A). The gene page and the variant page are also linked with each other. Users can hyperlink respective variant pages by clicking the SNP column on the gene page and *vice versa*. In addition, the gene page also provides the results of Somatic Mutation section, which shows all somatic mutations in the protein-coding region of the query gene with genomic location, reference allele, mutant allele, amino acid change, mutation type and the association of somatic mutations with ESCC survival time (Figure 2B).

Other pages

Users can also search using a particular chromosomal region (written as "chrN: start-end") to obtain a table of all SNPs within the specified genome region. The corresponding variant page of a SNP is shown when the SNP is clicked. To improve the usability, we provide an About page with detailed description of the database, a FAQ page that gives answers to some common questions and a Contact page for help and feedback.

Discussion

We established the CCGD-ESCC to share data on the associations of genetic variants and somatic mutations with risk or survival of ESCC for cancer researchers worldwide. To the best of our knowledge, the CCGD-ESCC is the first comprehensive database that systematically integrated association results of genetic variants and somatic mutations at the genome-wide level. Another important strength of this database is that it contains patients' survival data and the correlations between the genetic variants or somatic mutations and survival time of patients with ESCC, which provide cancer researchers and clinicians with a precious resource.

The majority of genetic variants potentially associated with ESCC that are identified by GWAS are located in non-coding sequences, suggesting that their functions, if any, may be exerted through regulating the expression of genes producing proteins or non-coding RNAs [27]. Thus, integration of GWAS and eQTL data is essential for elucidating the underlying mechanisms for genetic variants in diseases [28]. The currently available databases, such as the GTEx Project and the NHGRI GWAS Catalog, provide only one-sided association data of genetic variants for gene expression or for cancer risk. The Cancer Genome Atlas (TCGA), which provides the greatest amount of cancer omics data, mainly focuses on somatic mutations not germline variations. Conversely, our CCGD-ESCC database provides integrative data, from which users can explore the functional variants in both ESCC tumor tissues and normal esophageal tissues. More information given in our database, such as systematic analysis of the *cis*-regulatory variants in matched tumor and normal samples, might help users to identify the variants specific for the development of ESCC and perhaps other squamous cell carcinomas as well. However, CCGD-ESCC only provides significant associations that are above the corresponding thresholds as defined in the four sections. Users can send us emails to access to the associations with the SNPs of their interest. We hope that the CCGD-ESCC database can be a one-stop portal for seeking potential functional SNPs and related genes for research on ESCC and other cancers. While the database currently covers ESCC only, we are actively working to update and extend the portal, aiming to generate a comprehensive genomic database for cancers in Chinese populations.

Authors' contributions

CW, GG, and DL conceptualized and supervised this study. CW, GG, LP, SC, and Y Lin contributed to the study design. LP, QC, Y Luo, JC, MS, WF, YC, AL, YX, YS, LZ, CZ, and WT performed genotyping assays, imputation analysis and association analysis. SC, LP, and Y Lin prepared the CCGD-ESCC web server. DL, CW, GG, LP, SC, and Y Lin prepared the manuscript. All authors read and approved the final manuscript.

A

Home	About	GWAS	Survival GWAS	eQTL	Somatic Mutation	FAQ	Contact
------	-------	------	---------------	------	------------------	-----	---------

ACAD10
acyl-CoA dehydrogenase family member 10

External References: [GeneCards](#) [HUGO](#) [COSMIC](#) [Ensembl](#) [UCSC](#) [GWASdb](#)

Entrez ID: **80724**

NCBI Description of ACAD10

This gene encodes a member of the acyl-CoA dehydrogenase family of enzymes (ACADs), which participate in the beta-oxidation of fatty acids in mitochondria. The encoded enzyme contains a hydrolase domain at the N-terminal portion, a serine/threonine protein kinase catalytic domain in the central region, and a conserved ACAD domain at the C-terminus. Several alternatively spliced transcript variants of this gene have been described, but the full-length nature of some of these variants has not been determined. [provided by RefSeq, Nov 2008]

Genome Assembly: **GRCh37.p13/hg19**

Association with risk of ESCC in 2022 ESCC cases and 2039 controls

SNP	Location	Region	Allele	MAF	P value	OR (95% CI)	Functional class	Mapped gene(s)	Reference
rs11066015	12:112168009	12q24.12	G/A	0.19	1.87E-06	1.33 (1.19-1.5)	Intronic	ACAD10	Wu C, et al. (PMID: 21642993), 2011

MAF: minor allele frequency in controls

Association with gene expression in ESCC tumor or normal tissues of 94 ESCC patients

SNP	Location	Region	Allele	Effect size	P value	FDR	Mapped gene(s)	Tissue	Action	Reference
rs3803171	12:112104748	12q24.12	A/C	0.14	1.92E-05	5.83E-03	ACAD10	Adjacent normal tissue	eQTL boxplot	Chang J, et al. (PMID: 28548104), 2017
rs141088484	12:112079747	12q24.12	TG/T	-0.14	1.95E-05	5.89E-03	ACAD10	Adjacent normal tissue	eQTL boxplot	Chang J, et al. (PMID: 28548104), 2017
rs74411430	12:112075997	12q24.12	G/A	0.14	2.91E-05	7.99E-03	ACAD10	Adjacent normal tissue	eQTL boxplot	Chang J, et al. (PMID: 28548104), 2017

B

Somatic mutations in the protein coding region of 675 ESCC patients

Location	Allele	Amino acid change	Mutation type
12:112143731	G>C	D176H	Missense_Mutation
12:112147468	C>T	L224F	Missense_Mutation
12:112153649	A>T	D292V	Missense_Mutation
12:112153654	G>A	G294R	Missense_Mutation
12:112185114	G>A	A804T	Missense_Mutation
12:112187059	C>T	I909I	Silent
12:112187150	G>A	-	Splice_Site

Association of somatic mutations in the coding region and ESCC patients' survival

Study	No. of patients with SNVs/indels	No. of patients without SNVs/indels	P value
Cancer Genome Atlas Research Network. (PMID: 28052061), 2017	2	88	2.12E-03

Figure 2 The association results for genes in CCGD-ESCC

A. Summary of basic information of the gene and hyperlinks of external public resources, as well as association data of *ACAD10* from GWAS, Survival GWAS or eQTL sections. Results of the association data are tailored since the actual web page is too large to fit. **B.** The results of Somatic Mutation section are displayed for the genes with somatic mutations significantly affecting ESCC patients' survival. MAF, minor allele frequency in controls; OR, odds ratio for the minor allele; eQTL, expression quantitative trait loci; SNV, single nucleotide variation.

Competing interests

The authors have declared no competing interests.

Acknowledgments

We would like to thank Yun Li and Mengjie Chen at University of North Carolina and bioinformatic core facility at Peking University for their help in imputation analysis and helpful suggestions of this study. Part of the analysis is performed on the Computing Platform of the Center for Life Sciences of Peking University. This work was supported by CAMS Innovation Fund for Medical Sciences (Grant No. 2016-I2M-4-002 awarded to CW) and the National Key R&D Program (Grant No. 2016YFC1302701 to CW and Grant No. 2016YFC0901603 to GG).

References

- [1] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015;65:87–108.
- [2] Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin* 2016;66:115–32.
- [3] Xing D, Tan W, Lin D. Genetic polymorphisms and susceptibility to esophageal cancer among Chinese population (a review). *Oncol Rep* 2003;10:1615–23.
- [4] Abnet CC, Freedman ND, Hu N, Wang Z, Yu K, Shu XO, et al. A shared susceptibility locus in *PLCE1* at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat Genet* 2010;42:764–7.
- [5] Wang LD, Zhou FY, Li XM, Sun LD, Song X, Jin Y, et al. Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies susceptibility loci at *PLCE1* and *C20orf54*. *Nat Genet* 2010;42:759–63.
- [6] Wu C, Hu Z, He Z, Jia W, Wang F, Zhou Y, et al. Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. *Nat Genet* 2011;43:679–84.
- [7] Wu C, Kraft P, Zhai K, Chang J, Wang Z, Li Y, et al. Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. *Nat Genet* 2012;44:1090–7.
- [8] 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. *Hum Mol Genet* 2012;21:2132–41.
- [9] 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. *Nat Genet* 2014;46:1001–6.
- [10] Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* 2015;16:197–212.
- [11] Gibson G, Powell JE, Marigorta UM. Expression quantitative trait locus analysis for translational medicine. *Genome Med* 2015;7:60.
- [12] Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci* 2013;368:20120362.
- [13] Liu J, Xie X, Zhou C, Peng S, Rao D, Fu J. Which factors are associated with actual 5-year survival of oesophageal squamous cell carcinoma? *Eur J Cardiothorac Surg* 2012;41:e7–11.
- [14] Wu C, Li D, Jia W, Hu Z, Zhou Y, Yu D, et al. Genome-wide association study identifies common variants in *SLC39A6* associated with length of survival in esophageal squamous-cell carcinoma. *Nat Genet* 2013;45:632–8.
- [15] Cheng X, Wei L, Huang X, Zheng J, Shao M, Feng T, et al. Solute carrier family 39 member 6 gene promotes aggressiveness of esophageal carcinoma cells by increasing intracellular levels of zinc, activating phosphatidylinositol 3-kinase signaling, and up-regulating genes that regulate metastasis. *Gastroenterology* 2017;152:1985–97.e12.
- [16] Song Y, Li L, Ou Y, Gao Z, Li E, Li X, et al. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* 2014;509:91–5.
- [17] Gao YB, Chen ZL, Li JG, Hu XD, Shi XJ, Sun ZM, et al. Genetic landscape of esophageal squamous cell carcinoma. *Nat Genet* 2014;46:1097–102.
- [18] Zhang L, Zhou Y, Cheng C, Cui H, Cheng L, Kong P, et al. Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma. *Am J Hum Genet* 2015;96:597–611.
- [19] Sawada G, Niida A, Uchi R, Hirata H, Shimamura T, Suzuki Y, et al. Genomic landscape of esophageal squamous cell carcinoma in a Japanese population. *Gastroenterology* 2016;150:1171–82.
- [20] Qin HD, Liao XY, Chen YB, Huang SY, Xue WQ, Li FF, et al. Genomic characterization of esophageal squamous cell carcinoma reveals critical genes underlying tumorigenesis and poor prognosis. *Am J Hum Genet* 2016;98:709–27.
- [21] Cancer Genome Atlas Research Network. Integrated genomic characterization of oesophageal carcinoma. *Nature* 2017;541:169–75.
- [22] Chang J, Tan W, Ling Z, Xi R, Shao M, Chen M, et al. Genomic analysis of oesophageal squamous-cell carcinoma identifies alcohol drinking-related mutation signature and genomic alterations. *Nat Commun* 2017;8:15290.
- [23] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
- [24] 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature* 2015;526:68–74.
- [25] Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv* 2012, 1207.3907v2.
- [26] Shabalín AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012;28:1353–8.
- [27] GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;348:648–60.
- [28] Montgomery SB, Dermitzakis ET. From expression QTLs to personalized transcriptomics. *Nat Rev Genet* 2011;12:277–82.