

RESEARCH

Open Access



Global comparative transcriptomes uncover novel and population-specific gene expression in esophageal squamous cell carcinoma

Amal Alotaibi^{1†}, Veerendra P. Gadekar^{2†}, Pranav Swaroop Gundla², Sumana Mandarathi^{2,8}, Nidhi Jayendra², Asna Tungekar², B. V. Lavanya², Ashok Kumar Bhagavath³, Mary Anne Wong Cordero¹, Janne Pitkaniemi^{4,5}, Shaik Kalimulla Niazi⁶, Raghavendra Upadhy⁷, Asmatanzeem Bepari^{1*} and Prashantha Hebbar^{2,7,8*}

Abstract

Background Esophageal squamous cell carcinoma (ESCC) has a poor prognosis and is one of the deadliest gastrointestinal malignancies. Despite numerous transcriptomics studies to understand its molecular basis, the impact of population-specific differences on this disease remains unexplored.

Aims This study aimed to investigate the population-specific differences in gene expression patterns among ESCC samples obtained from six distinct global populations, identify differentially expressed genes (DEGs) and their associated pathways, and identify potential biomarkers for ESCC diagnosis and prognosis. In addition, this study deciphers population specific microbial and chemical risk factors in ESCC.

Methods We compared the gene expression patterns of ESCC samples from six different global populations by analyzing microarray datasets. To identify DEGs, we conducted stringent quality control and employed linear modeling. We cross-compared the resulting DEG lists of each populations along with ESCC ATLAS to identify known and novel DEGs. We performed a survival analysis using The Cancer Genome Atlas Program (TCGA) data to identify potential biomarkers for ESCC diagnosis and prognosis among the novel DEGs. Finally, we performed comparative functional enrichment and toxicogenomic analysis.

Results Here we report 19 genes with distinct expression patterns among populations, indicating population-specific variations in ESCC. Additionally, we discovered 166 novel DEGs, such as ENDOU, SLCO1B3, KCNS3, IFI35, among others. The survival analysis identified three novel genes (CHRM3, CREG2, H2AC6) critical for ESCC survival. Notably, our findings showed that ECM-related gene ontology terms and pathways were significantly enriched among the DEGs in ESCC. We also found population-specific variations in immune response and microbial infection-related pathways which included genes enriched for HPV, Ameobiosis, Leishmaniosis, and Human Cytomegaloviruses. Our toxicogenomic analysis identified tobacco smoking as the primary risk factor and cisplatin as the main drug chemical interacting with the maximum number of DEGs across populations.

[†]Amal Alotaibi and Veerendra P. Gadekar have contributed equally.

*Correspondence:

Asmatanzeem Bepari

ambepari@pnu.edu.sa

Prashantha Hebbar

hebbar.prashantha@gmail.com

Full list of author information is available at the end of the article



Conclusion This study provides new insights into population-specific differences in gene expression patterns and their associated pathways in ESCC. Our findings suggest that changes in extracellular matrix (ECM) organization may be crucial to the development and progression of this cancer, and that environmental and genetic factors play important roles in the disease. The novel DEGs identified may serve as potential biomarkers for diagnosis, prognosis and treatment.

Keywords Novel genes in ESCC, CHRM3, CREG2, H2AC6, ENDOU, FHL1, Esophagus cancer, ESCC, Toxicogenomics

Introduction

Esophageal cancer is the eighth most commonly diagnosed cancer, accounting for 3% of all cancer cases and the sixth most common cause of death from cancer. The GLOBOCAN 2020 survey estimated a global increase in esophageal cancers with 604,000 new cases and 544,000 deaths in 2020. New estimation predicts that new cases will increase by about 1 million by 2040 [1]. Of the two histological types of esophageal cancers, 85% of the cases had esophageal squamous cell carcinoma and 14% had esophageal adenocarcinoma [1].

Striking differences were observed in the incidence, clinicopathological features, treatment efficacy, and overall prognosis of ESCC between geographical populations [2]. The highest incidence of ESCC is seen in Eastern Asia and Southern and Eastern Africa whereas the lowest is observed in Western Africa and Central America regions. Black patients had been diagnosed at the median age of 63 whereas Non-Hispanic Whites and Asian patients had 68 years as the median age at diagnosis. Similarly, survival differences were also observed between the ethnic groups. Black patients had the lowest while Non-Hispanic Whites had the better survival of ESCC [2].

It is important to note that specific risk factors for certain diseases can vary depending on the geographic region due to region-specific lifestyle, cultural, and environmental factors [3]. For example, salted meat consumption, alcohol intake, and smoking behavior are known risk factors for ESCC in Uganda [4]. In India, tobacco use is a common risk factor for ESCC, with approximately 34.6% of adults consuming tobacco through cigarettes or chewing [5]. Some malignancies are prone to the onset or progression because of trace element deficit or excess [6]. Trace element imbalance in the soils of South Africa and West Asia [3] showed high incidence and prevalence for ESCC. Dietary zinc deficiency is known to increase the risk of ESCC [7]. Zinc deficiency or excess has many carcinogenic impacts on cell growth, DNA repair, mutagenesis, apoptosis, DNA synthesis, and differentiation, and the overall balance of cellular antioxidants [6]. Similarly, selenium and zinc serum levels have been linked to the incidence of gastroesophageal cancers in West Asia [8]. In addition, frequent drinking of hot Arabic coffee in the Al-Qaseem region of Saudi Arabia [9], consumption of

hot green tea in East Asia, and betel quid/gutka chewing in regions of South and Southeast Asia [10] are associated with the development of ESCC.

There is a growing understanding of the link between gut microbiota and human health [11]. Maintaining homeostasis and good health depends on the intestinal tract microbiota. Further, several types of cancers were initiated and progressed by the microbiome [12]. An imbalance of some species may contribute to the initiation and progression of tumors by harming DNA structure, generating metabolites that promote tumor growth, and inhibiting the immune response against tumors. About 50% of the ESCC tumors had human papillomavirus (HPV) infection in the southeastern region of Poland [13]. Studies have shown geographical region is an important factor for HPV prevalence in ESCC [14, 15].

Besides regional-specific risk factors, there exists a difference in genetic makeup among the global human populations by polymorphisms and their linkage disequilibrium [16]. Several Mendelian Randomization studies using genetic data have revealed causal associations between environmental factors and disease traits including cancers [17, 18]. Increasing evidence shows that environmental and lifestyle factors influence epigenetic changes. Changes in global epigenetic signatures, together with genetic alterations, are driving events in several types of cancer including ESCC [19, 20]. Hence, alteration in epigenetic signatures by region-specific risk factors and their interaction with population-specific genetic determinants may result in population-specific gene expression dysregulation and gene network perturbation in ESCC.

Although it is well-established that cancer causation involves multiple factors and gene-environment interactions [3, 21], to the best of our knowledge, no studies have attempted a comparative analysis between different populations to highlight the molecular connections with the risk factors. In this study, we conducted a comprehensive analysis of ESCC expression profiles from populations of various geographical regions, including Chinese, Japanese, Taiwanese, African American, Brazilian, and European populations. By examining the extent of shared gene perturbation across these populations, we aim to gain a better understanding of population-specific

risk factors, diagnosis, prognosis, and treatment of ESCC. This systematic analysis will provide valuable insights into the molecular etiology of ESCC and help inform future research efforts in this field.

Materials and methods

Dataset

We collected protein-coding mRNA array expression datasets of human ESCC tissues and cell lines from—(1) Gene Expression Omnibus (NCBI-GEO) [22], (2) EBI ArrayExpress [23] and (3) All of the gene expression (AOE) [24]. We selected only those datasets where the studies compared the normal and tumor conditions. For the selected datasets, we inferred their origin population based on the description in the corresponding manuscript, which includes either the explicit mention of the study population or the location of the sample collection centers. If sample collection centers were spread across countries/continental hospitals, we labeled the population 'Unknown'.

Data quality assessment

We used `arrayQualityMetrics` (v3.48.0) [25] in R statistical programming language (referred to as R in further instances) for the quality check of the datasets and to flag the potential outlier samples. Based on further manual inspection of the flagged outlier samples, we decided whether to include or exclude the samples from the downstream analysis. We disregarded the datasets that contained a potential sample mislabeling, class (control/tumor) imbalance issues or if retaining a dataset would decrease the overall common genes among the dataset in meta-analysis.

Platform-specific background correction and normalization

Given that we selected the datasets generated from different platforms, we applied the platform-specific methods for background correction and normalization. We selected the appropriate normalization method that provided the normal data distribution of the data. For gene mapping of array probes, we used the array design files of platforms or Bioconductor annotation packages including `AnnotationDBI` (v1.58.0) [26] in R. Further multiple probes in each array were combined either by taking the mean or median values of probe intensities to represent the source gene.

Differential gene expression of individual datasets

To discover differentially expressed genes in tumors compared to the normal samples in each dataset separately, we first generated the model matrix specifying the design (tumor ~ normal), followed by fitting

the linear model to the design using `lmFit()` from the `Limma` Bioconductor package in R. Next, to perform empirical Bayesian moderation of the t-statistics we used the `ebayes()` from `Limma` that uses the distribution across genes to calculate a robust test statistic. We considered a gene as differentially expressed in the tumor compared to the normal samples if it showed a change in expression with absolute log₂ base fold change (\log_2FC) > 1.5 and adjusted p-value < 0.05.

Meta-analysis for populations with multiple datasets

We performed the meta-analysis for the multiple datasets available for the same population using the `MetaVolcanoR` package (v1.10.0) available from the Bioconductor in R. The gene expression \log_2FC values, adjusted P-values, and 95% confidence intervals across multiple datasets were used as input data for the Random Effect Model (REM) approach implemented in the `MetaVolcanoR`. The random effect method takes into consideration the diversity present in multiple datasets by incorporating a statistical parameter that represents the inter-study variation, including clustering or dependence within a dataset, as well as varying relationships both within and between clusters [27]. REM summarizes the gene fold change taking into account the mean and variance that depends upon the study-specific estimates of the effect size. The genes consistently perturbed across all the studies were ranked based on the `Topconfacts` approach [28] implemented within the package. `Topconfacts` is a method for ranking by confidence bounds on the log fold change, based on the previously developed TREAT test by McCarthy and Smyth [29].

Cross-comparative analysis of differentially expressed gene list

We compared the DEGs between populations using the `ComplexHeatmap` package for generating Upset plots in R. The common genes across at least two populations were compared against the ESCC ATLAS to identify if they were previously known to be associated with the ESCC etiology and if not, they were regarded as “novel DEGs” (in the rest of the manuscript). Similarly, we compared the DEGs unique among the populations against the ESCC ATLAS. The genes that did not map were discarded. For the downstream Functional Enrichment and Toxicogenomic analysis, we considered the common (including novel DEGs) and population-specific unique genes that mapped to ESCC ATLAS.

Prioritization of novel genes

We used the Genotype-Tissue Expression portal (GTEx) portal bulk tissue-specific expression data or Human Protein Atlas (HPA) RNAseq normal tissue expression (tissue samples from 95 human individuals representing 27 different tissues) of Esophagus data to prioritize novel genes. The novel genes that were upregulated in our list but showed low or no expression in at least 3 different Esophagus tissues (Esophagus—Gastroesophageal Junction, Esophagus—Mucosa, and Esophagus—Muscular) in GTEx or HPA RNAseq, or vice versa were considered more important genes in the context to ESCC etiology.

Survival analysis

To reveal the candidate gene contribution to the patient's survival, we performed overall survival prediction in an independent ESCC RNAseq dataset available at TCGA. We used TCGAbiolinks (v2.20.1) in R to access 90 ESCC tumors and 11 normal samples of the total 559 samples corresponding to Esophageal Cancer. Further, we used the survival package (v 0.4.9) and survminer package (v3.4.0) in R for survival analysis and for generating Kaplan–Meier survival plots respectively. For the genes that showed significant association with the survival of ESCC patients, we performed Cox proportional hazard analyses using the dysregulated and intact candidate genes and compared the distribution of log₂ base counts per million (log₂CPM) values for these genes in the tumor and normal samples using t-test statistics.

Functional and Toxicogenomic enrichment analysis

We performed functional enrichment for DEGs compared to all the expressed genes as background from each population separately using the ClusterProfiler (v4.4.4) [30] package in R. We examined the overrepresentation of GO terms and pathways using, gseGO(), gseKEGG(), gsePathway(), and gseWP() functions respectively. Where, the annotation data was used from the Gene Ontology [31], KEGG [32], Reactome [33], and WikiPathway [34] databases respectively. These functions compute the overrepresentation of terms based on the hypergeometric distribution test; the p-values were adjusted for the multiple testing in each case using the Benjamini-Hochberg (BH) procedure. The redundant enrichment terms were simplified using the simplify() function in the ClusterProfiler package. The GO terms and the Pathways were considered to be overrepresented among differentially expressed genes if the adjusted p-value for the enrichment analysis is < 0.05.

To infer the Toxicogenomic enrichment of our DEGs, we used the CTDquerier (v2.3.1) [35] package available from the Bioconductor in R. The CTDquerier retrieves

the data from the Comparative Toxicogenomics Database (CTD), along with the evidence for direct gene-risk factor/drug associations. The CTD is a meticulously curated repository of interactions between chemicals, genes, and diseases extracted from scientific literature. In this study, we employed a network-based approach to calculate the inference score for genes and CTD terms relationships, specifically focusing on a chosen disease, such as ESCC. For this purpose, the background gene set comprised genes that were curated for ESCC in the CTD. A detailed information about inference score calculation is given in [36].

Results

Datasets selection

We used 14 published and 2 unpublished (namely, GSE23964 and GSE45168) datasets comprising clinical and expression information upon exhaustive search from publicly available data repositories. Detailed Information about the datasets with biological materials, sample size, microarray platforms, and sample collection centers used in each study is provided in Table 1. Based on our stringent quality check (refer to methods) we removed 4 data sets (GSE63941, GSE9982, GSE45168, GSE32424) (marked with the * symbol in Table 1). Based on the population inferences, the 12 datasets that we included in our study corresponded to either Chinese (5), Taiwanese (1), Japanese (1), African American (1), Brazilian (1), European (1), or unknown (2). An illustration of the workflow used in this study is provided in Additional file 1: Fig. S1.

Distribution of DEGs in ESCC among different populations

In order to identify the DEGs in tumors compared to the normal samples among all the populations, we performed the differential gene expression analysis as discussed in the methods section. The observed and expected quantiles of the 12 datasets are shown in the QQ plot (Additional file 1: Fig. S2), indicates no bias or confounding factors in the results of analyzed datasets. The total number of DEGs in each dataset is shown in Fig. 1A.

We found on average 1155 ± 766 (Mean \pm SD) DEGs in the European, African American, and Taiwanese populations, whereas 75 ± 65 (Mean \pm SD) DEGs in the Chinese, Japanese, and Unknown populations. This was expected given in the latter populations, the data set included a lower number of genes. For the Chinese and unknown populations, we had multiple datasets. The total up- and down-regulated DEGs are shown in Fig. 1B.

Table 1 Datasets used in this study

Study	GSE ID	Biological Material (Sample Size)	Data generation platform	Sample collection center	Population
Hu et al. [92]	GSE20347	Tumor (17) and matched normal (17) esophageal tissue pairs	Affymetrix Human Genome U133A 2.0 Array	Shanxi Cancer Hospital in Taiyuan, Shanxi Province, PR China	Chinese [#]
Fujiwara et al. [Unpublished]	GSE23964	Normal (2) and tumor transfected in series of KYSE cell lines (14)	3D-Gene Human Oligo chip 25 k		Unknown
Chen et al. [93]	GSE70409	17 tumor and adjacent normal esophageal tissues	Phalanx Human OneArray	Kaohsiung Medical University Hospital, Taiwan	Taiwanese (inferred)
Aoyagi et al. [94]	GSE22954	Tumor (90) and normal (10) esophageal tissues	Affymetrix Human Genome U95A and U133 Arrays	Central Hospital or East Hospital at the National Cancer Center, Chiba, Japan	Japanese (inferred)
Yan et al. [95]	GSE33426	Tumor (59) and adjacent normal (12) esophageal tissues. GSE33426 had 18 common samples with GSE29001. They were removed	Affymetrix GeneChip human genome U133A 2.0 array	Shanxi Cancer Hospital and Institute, Taiyuan, Shanxi Province, China	Chinese (Taihang Mountain region) [#]
Wang et al. [96]	GSE26886	Tumor (19) and normal esophageal tissues (9)	Affymetrix Human Genome U133 Plus 2.0 Array	Charité Comprehensive Cancer Center, Berlin, Germany	European (inferred)
Yan et al. [97]	GSE29001	Tumor (21), normal basal esophageal epithelium (12) and normal differentiated esophageal epithelium (12)	Affymetrix Genechip Human Genome U133A 2.0 Array	Shanxi Cancer Hospital and Institute, Taiyuan, Shanxi Province, China	Chinese [#]
Su et al. [37]	GSE23400	Tumor (53) and adjacent normal esophageal tissues (53)	Affymetrix GeneChip Human Genome U133A 2.0 Array	Shanxi Cancer Hospital in Taiyuan, Shanxi Province, People's Republic of China	Chinese (inferred)
Lee et al. [98]	GSE17351	Tumor (5) and adjacent normal esophageal tissues (5)	Affymetrix Genechip U133 plus v2.0 Array	Okayama University Hospital, Kitano Hospital, Japan and the Hospital of the University of Pennsylvania, USA	Unknown/Mixed
Nicolau-Neto et al. [99]	GSE75241	Tumor (15) and adjacent normal esophageal tissues (15)	Affymetrix Human Exon 1.0 ST array	Instituto Nacional de Câncer, Rio de Janeiro, Brazil	Brazilian (inferred)
Yang et al. [100]	GSE44021	Pairs of tumor (113) and normal (113) esophageal tissues	Affymetrix GeneChip Human Genome U133A 2.0 Arrays	Shanxi Cancer Hospital in Taiyuan, Shanxi Province, PR China	Chinese (inferred)
Erikizan et al. [101]	GSE77861	Tumor (7) and adjacent normal esophageal tissues (7)	Affymetrix Human Genome U133 Plus 2.0 Array		African American [#]
Tong et al. [102]	GSE32424 *	Tumor (7) and normal (5) esophageal tissues	illumina Genome Analyzer Iix	Linzhou Cancer Hospital, Henan, China	Chinese (inferred)
Saito et al. [103]	GSE63941 *	22 ESCC cell lines and 4 isolated primary human esophageal fibroblasts	Affymetrix GeneChip Human Genome U133A array		Unknown
Shimokuni et al. [104]	GSE9982 *	20 KYSE human ESCC cell lines and 2 non-cancerous HEEC-1 esophageal epithelial cell lines	Codelink Uniset Human 20 K I Bioarray		Unknown
Zhu et al. [Unpublished]	GSE45168 *	Tumor (5) and adjacent normal (5) esophageal tissues	Agilent-026652 Whole Human Genome Microarray 4 × 44 K v2		Unknown

[#] population information explicitly mentioned in the study; ^{*} population information inferred from the sample collection center described in the study; ^{**} indicates these datasets have been removed based on QC

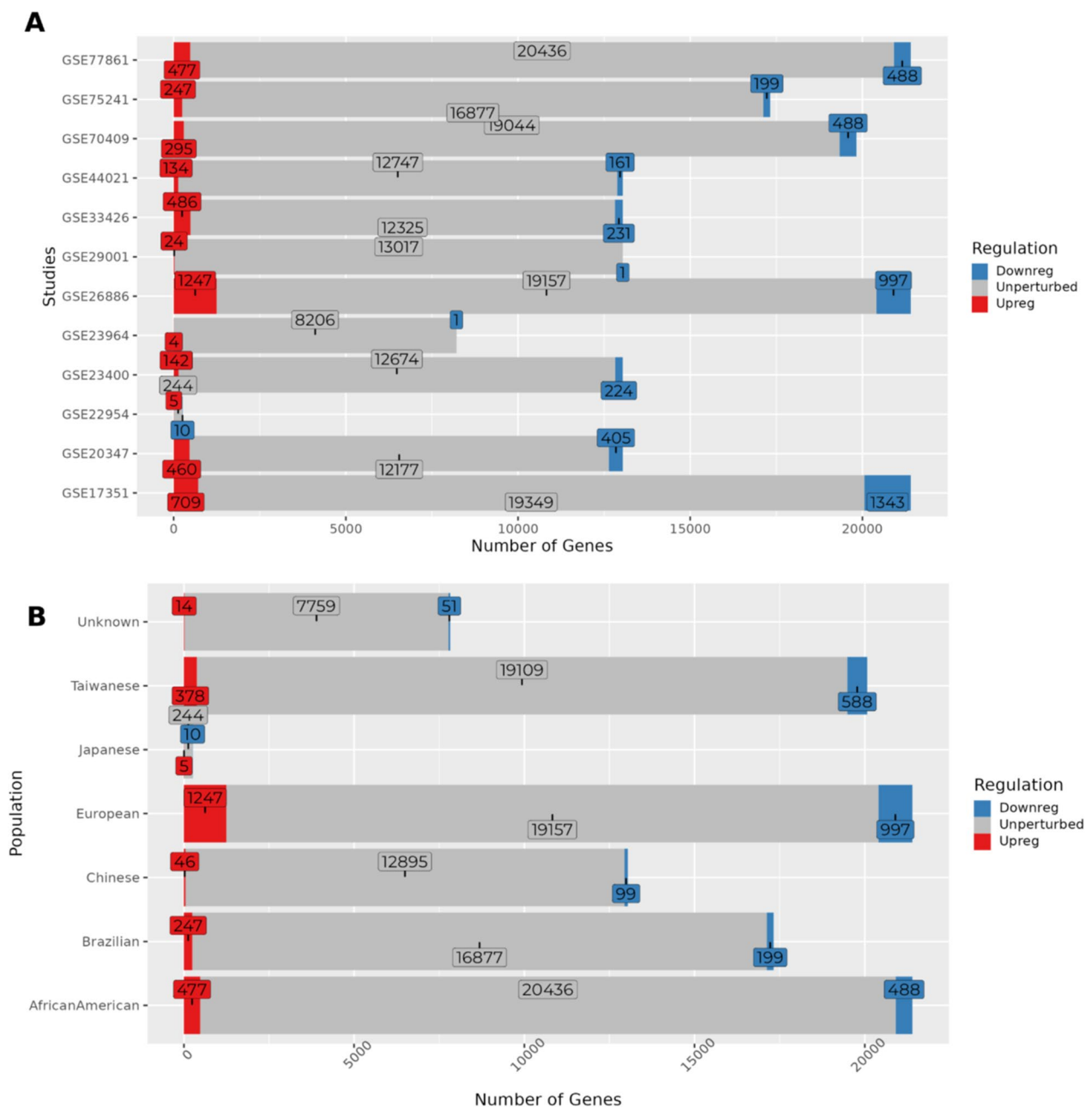


Fig. 1 illustrates **A** Total Up-, Down-regulated and unperturbed genes in the 12 datasets used in this analysis. **B** Population-wise total gene counts (including meta-analysis of 5 Chinese datasets and 2 unknown population datasets). Note: $\log_2FC > 1.5 \mid < -1.5$ and adjusted p-value < 0.05 was used as a threshold for filtering the genes

Comparative analysis of DEGs across populations

We identified 1442 DEGs of which 1423 (Additional file 2: Table S1) were concordant DEGs (expressed commonly up or down-regulated between the population), and the remaining 19 genes were discordant DEGs where the direction of gene expression (up/down) in at least one population is different compared to the rest. Among the 1423 concordant DEGs, we identified 1257 mapped

to ESCC ATLAS. We have summarized the total DEGs across populations upon filtering against ESCC ATLAS in Fig. 2A. The common and unique Up- and Down-regulated genes across populations are shown in Fig. 2B and C respectively. Some of the interesting genes among these include—SASH1 (SAM and SH3 domain containing 1), a potential tumor suppressor with negative regulation of proliferation, and invasion of cancer cells. We

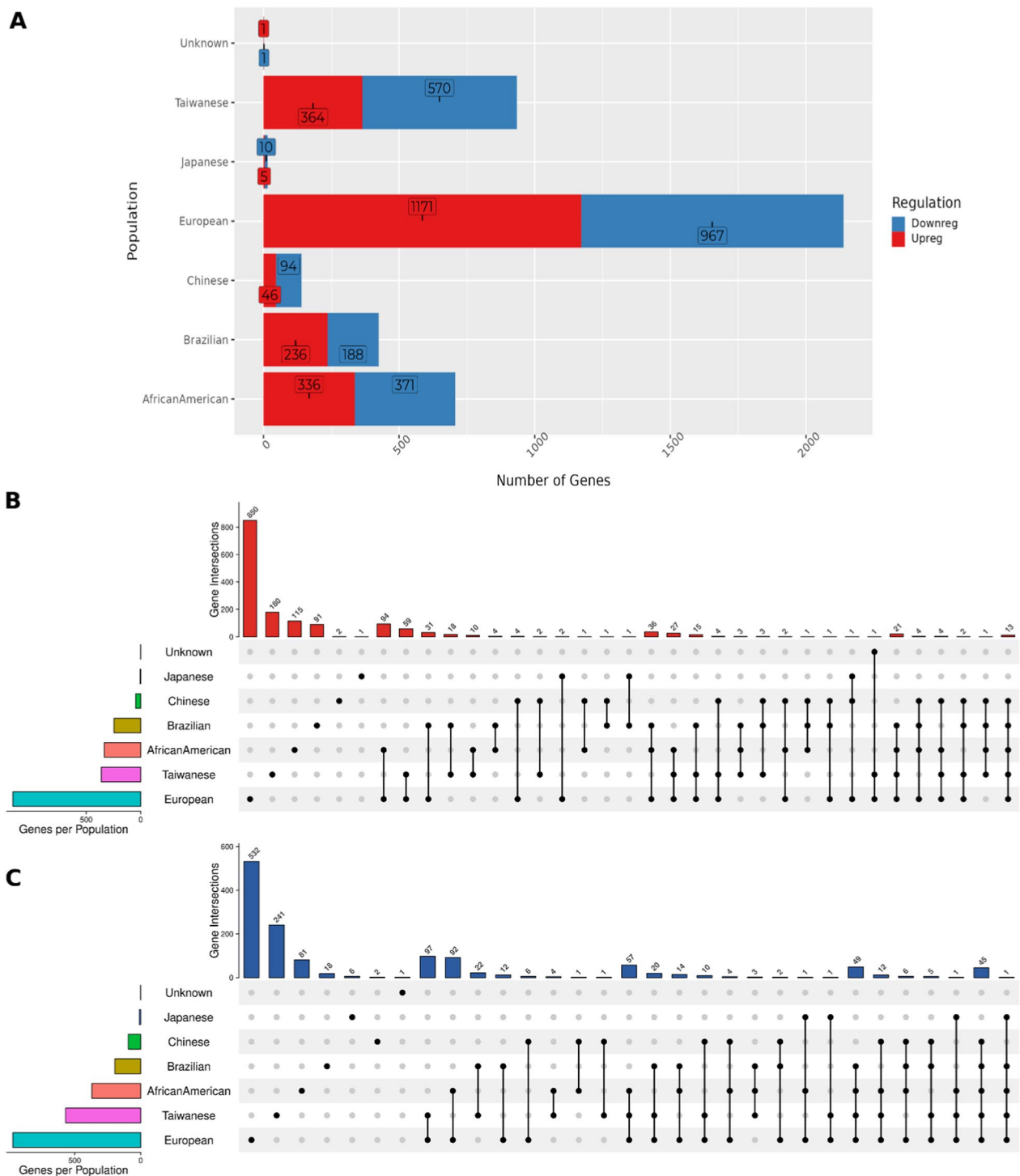


Fig. 2 A Total up- and down-regulated genes in each population intersecting with ESCC ATLAS. Cross-comparison of up (B) and down-regulated (C) genes across populations

identified it is significantly downregulated in African-American, Brazilian, European, Japanese, and Taiwanese populations in our analysis. Additionally, ESCC ATLAS

catalogs a study that suggests SASH1 is downregulated in the Chinese population as well in the onset of ESCC [37].—BLNK (B Cell Linker is a leukocyte protein that

contains SH2 domain) that is known to produce B-cell linker protein crucial for B-cell development [38]. We find this gene is downregulated in 5 populations except in Japanese.

We considered the remaining 166 DEGs out of 1423 concordant DEGs in our analysis as novel genes in the context of ESCC etiology because we did not find any prior evidence for their involvement in ESCC. To investigate the expression of these DEGs in healthy conditions, we compared our list of DEGs with the expression profile of genes in healthy Esophageal-Mucosa using GTEx or HPA RNAseq data. We discovered that 18 DEGs (ENDOU, ANKRD20A5P, CYP4F35P, CYSRT1, EPIST (also known as C5orf66-AS1), LCAL1, LINC02487, LOC100507221, LOC105376081, MUCL3 (also known as DPCR1), SELENOM, TENT5B (also known as FAM46B), TOX2, DSG1 (based on DSG1-AS1 expression), DEGS2, DOUXA2, KLK7, RNF39) exhibited an opposite expression trend in normal Esophagus tissues. This finding provides additional evidence for these genes to play a crucial role in the context of ESCC (Fig. 3).

To highlight a few, here are the novel DEGs that are concordant in at least 4 populations—(1) ENDOU (endonuclease, poly(U) specific), is significantly (fold change and p-value) downregulated in ESCC samples across 5 populations except in Japanese (where it was not included in the array), whereas, it is seen to be overexpressed in healthy Esophageal-Mucosa tissues in GTEx. ENDOU is known to encode for a protein with endoribonuclease activity that binds to the polyuridine-enriched single-stranded RNA. Interestingly, ENDOU in mice is identified as a regulator of activation-induced cell death. (2) KCNS3 (potassium voltage-gated channel modifier subfamily S member 3), is upregulated in at least four populations (see Table 2), but in healthy Esophageal-Mucosa listed in GTEx, it is lowly expressed. Interestingly, the knockdown of KCNS3 inhibits tumor cell proliferation in colon carcinoma and lung adenocarcinoma cell lines [39]. (3) SLC01B3 (solute carrier organic anion transporter family member 1B3), is found to be upregulated in our analysis (see Table 2), however, it is not expressed in healthy Esophageal-Mucosa in GTEx data. Interestingly,

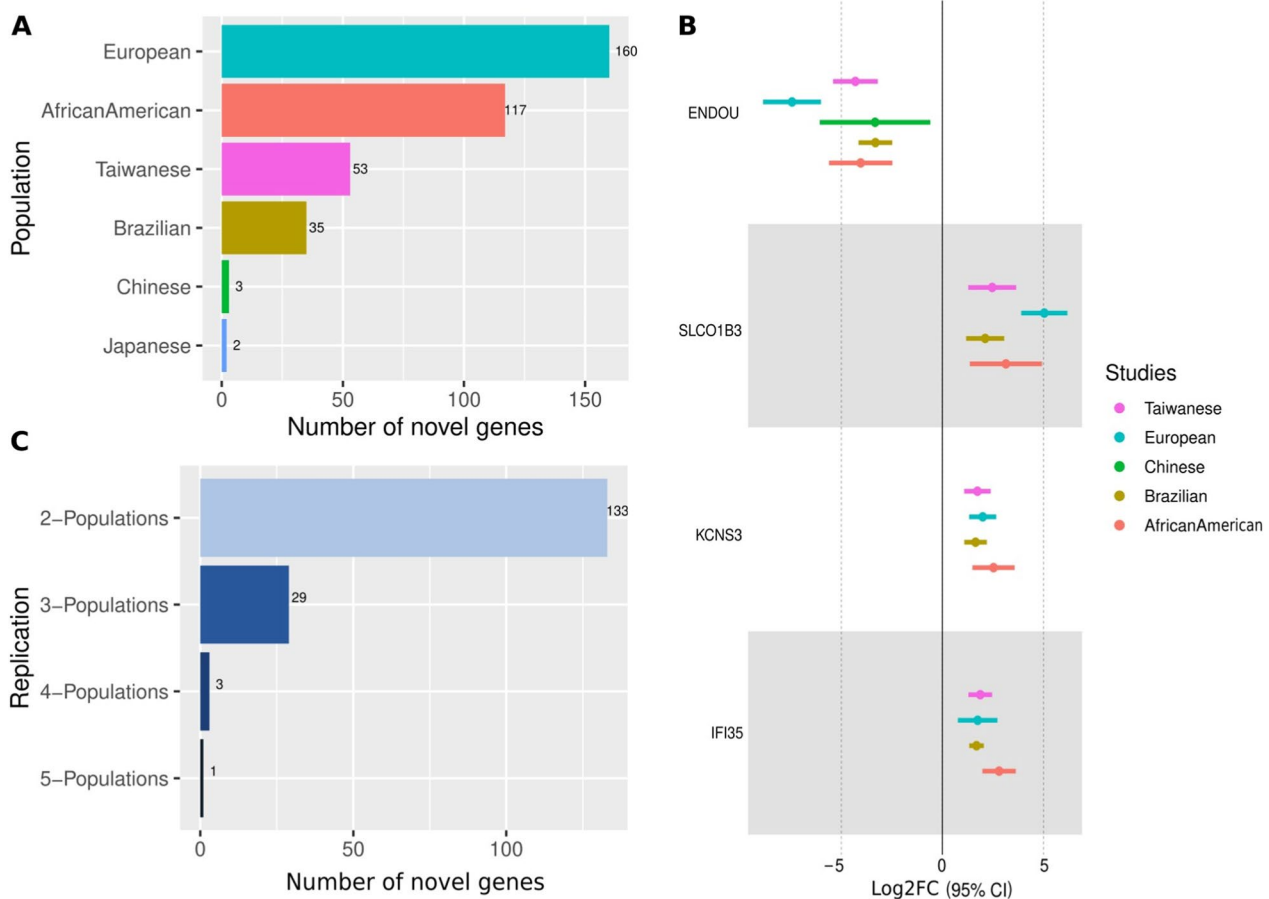


Fig. 3 **A** Total Novel DEGs across populations, **B** Total Novel DEGs common between populations, **C** Expression distribution of 4 Novel DEGs across populations

Table 2 List of Novel DEGs that are concordant in at least 4 populations

Gene	Log2FC	Adj. P-value	Population
ENDOU	-4.04	4.60E-03	African American
ENDOU	-3.32	3.39E-07	Brazilian
ENDOU	-3.33	2.26E-03	Chinese
ENDOU	-7.44	6.73E-09	European
ENDOU	-4.31	3.17E-07	Taiwanese
IFI35	2.8	1.15E-03	African American
IFI35	1.68	3.64E-08	Brazilian
IFI35	1.74	6.53E-03	European
IFI35	1.87	7.53E-06	Taiwanese
KCNS3	2.52	6.41E-03	African American
KCNS3	1.63	2.37E-05	Brazilian
KCNS3	1.99	3.46E-05	European
KCNS3	1.73	1.23E-04	Taiwanese
SLCO1B3	3.14	3.01E-02	African American
SLCO1B3	2.11	5.75E-04	Brazilian
SLCO1B3	5.04	9.99E-08	European
SLCO1B3	2.46	1.94E-03	Taiwanese

We observed consistent downregulation of ENDOU in 5 population and upregulation of IFI35, KCNS3, and SLCO1B3 in 4 population

the overexpression of SLCO1B3 in non-small cell lung cancer cells is known to regulate the epithelial-mesenchymal transition (EMT) related genes [40]. (4) IFI35 (interferon-induced protein 35), is upregulated in four different populations in our analysis (see Table 2), however, it is found to be lowly expressed in healthy Esophageal-Mucosa reported in GTEx. IFI35 is known to be involved in processes such as macrophage activation in immune response and positive regulation of defense response. The list of novel DEGs common in at least two populations is listed in Additional file 2: Table S2.

Discordant genes are particularly interesting because they point towards potential population-specific expression profiles of important genes in the ESCC etiology. Where up or down-regulation of specific genes could be either deleterious or advantageous for only a subset of the populations. In total, we identified 19 discordant DEGs (see Table 3 and Additional file 2: Table S3) across populations. Some interesting genes include—(1) FHL1 (Four and a half LIM domains protein 1) is upregulated in African American, Brazilian, and European whereas downregulated in Taiwanese. Overexpression of FHL1 is reported to inhibit the cell proliferation, colony formation potential, and expression of CDK4 and Cyclin D1 thereby negatively regulating the Wnt/ β -catenin signaling pathway [41], (2) MYL9 (Myosin light chain 9) is upregulated

Table 3 List of 19 DEGs with discordant expression between populations

Gene	Population (Evidence from ESCC ATLAS)	Log2FC	Adj. P-value	Gene	Population (Evidence from ESCC ATLAS)	Log2FC	Adj. P-value
FHL1	African American	1.68	1.86E-02	TCEAL2	European	1.7	4.91E-02
FHL1	Brazilian	1.55	1.23E-04	TCEAL2	Taiwanese	-2.11	1.04E-02
FHL1	European	1.51	3.16E-02	CNN1	European	2.01	1.20E-02
FHL1	Taiwanese (South Indian, Chinese)	-1.69	1.10E-02	CNN1	Taiwanese (South Indian, Chinese)	-2.95	2.25E-05
TPPP3	African American	3.32	4.64E-02	SYNM	European	1.56	1.83E-03
TPPP3	Chinese	-1.5	5.92E-13	SYNM	Taiwanese (South Indian)	-2.41	5.21E-04
IGSF22	African American	1.81	1.02E-03	PLN	European	2.31	1.10E-02
IGSF22	European	-1.88	3.55E-03	PLN	Taiwanese (South Indian)	-1.91	2.22E-02
KCND1	African American	-1.78	3.12E-02	LYZ	Brazilian (Chinese)	1.68	5.76E-05
KCND1	European	2.01	5.35E-03	LYZ	European	-2.09	2.57E-02
SMPX	African American	-1.84	2.82E-02	ACTG2	Brazilian	1.51	1.90E-02
SMPX	European	1.61	1.78E-02	ACTG2	Taiwanese (South Indian)	-2.5	1.01E-05
TPTE2P1	African American	-1.6	1.69E-02	RBPM52	European	1.82	4.69E-02
TPTE2P1	European	1.96	1.72E-03	RBPM52	Taiwanese	-1.77	1.55E-02
IL36G	African American (South Indian)	2.6	4.20E-02	FGF3	European	2.03	1.50E-02
IL36G	European	-2.77	1.44E-02	FGF3	Taiwanese	-1.7	4.32E-03
PCP4	European	2.16	4.35E-02	KIF26B	European (South Indian)	2.1	3.82E-05
PCP4	Taiwanese	-2.52	1.36E-04	KIF26B	Taiwanese	-1.86	2.35E-05
MYL9	European	1.61	2.10E-02	MYH11	European	1.93	7.33E-03
MYL9	Taiwanese (South Indian)	-1.75	9.96E-03	MYH11	Taiwanese (Chinese)	-2.91	7.20E-05

This table presents a list of 19 differentially expressed genes (DEGs) that exhibit discordant expression patterns between populations. It also includes evidence for the observed gene expression direction from the ESCC ATLAS

in European but downregulated in the Taiwanese population. The overexpression of MYL9 is known to promote cell proliferation, invasion, migration, and angiogenesis [42]. Of these 19 genes, we did not find prior evidence for 6 genes (TPPP3, IGSF22, SMPX, TPTE2P1, RBPMS2, and TCEAL2) for their involvement in ESCC (Additional file 2: Table S4).

Candidate genes associated with ESCC prognosis

In order to identify the candidate genes from our list of 166 novel DEGs in ESCC that could be potential prognostic markers in ESCC, we performed survival analysis (See “Methods”). Interestingly, we identified 3 genes—CHRM3, CREG2, and H2AC6 (alias HIST1H2AC) that were significantly correlated with overall survival in ESCC. The increased expression levels of CHRM3 and CREG2 were significantly associated with lower survival time; this was supported by the Hazard Ratio (HR) of 2.84 (p-value=0.026) and 2.29 (p-value=0.039) respectively. In the case of H2AC6 gene, its downregulation is significantly associated with lower survival time (HR=2.35, p-value=0.023) (Fig. 4). In our analysis we identified CHRM3, CREG2 to be upregulated in African American and Europeans, whereas, H2AC6 was downregulated in these populations. The expression profiles (up/down-regulation) of these three genes in TCGA RNA-seq samples were well correlated with the profiles observed in our datasets. To the best of our knowledge, these genes have

not been discussed in the context of ESCC thus far. Given that they are differentially expressed in ESCC and show a significant impact on survival time, we suggest these three genes as candidate genes for ESCC prognosis.

Gene ontology and pathway enrichment of DEGs

In order to identify the potential functional role of the DEGs in our analysis, we performed functional enrichment analysis using the publicly available GO and pathway annotation data (See “Methods”). Interestingly, we found multiple GO terms and pathways (Additional file 2: Tables S5–S8) related to the Extra Cellular Matrix (ECM) organization, and its intricate molecular cascades were significantly overrepresented in our list of DEGs (Fig. 5). These included—Degradation of the extracellular matrix, collagen metabolic process (activated), cell-substrate junction assembly, epidermal and endodermal cell differentiations, cornified envelope (deactivation), Keratinization (deactivation), Focal adhesion: PI3K-Akt-mTOR signaling pathway (Activated), Integrin binding (Activated), Cell motility, Cell migration, and Cells localization, etc. This is in agreement with previous reports, where the ECM-based communication between cells and their surrounding microenvironment is discussed to affect the genesis and/or development of esophageal tumors [43, 44].

Another important enrichment term discovered was Ossification (Fig. 5A), a condition that involves the

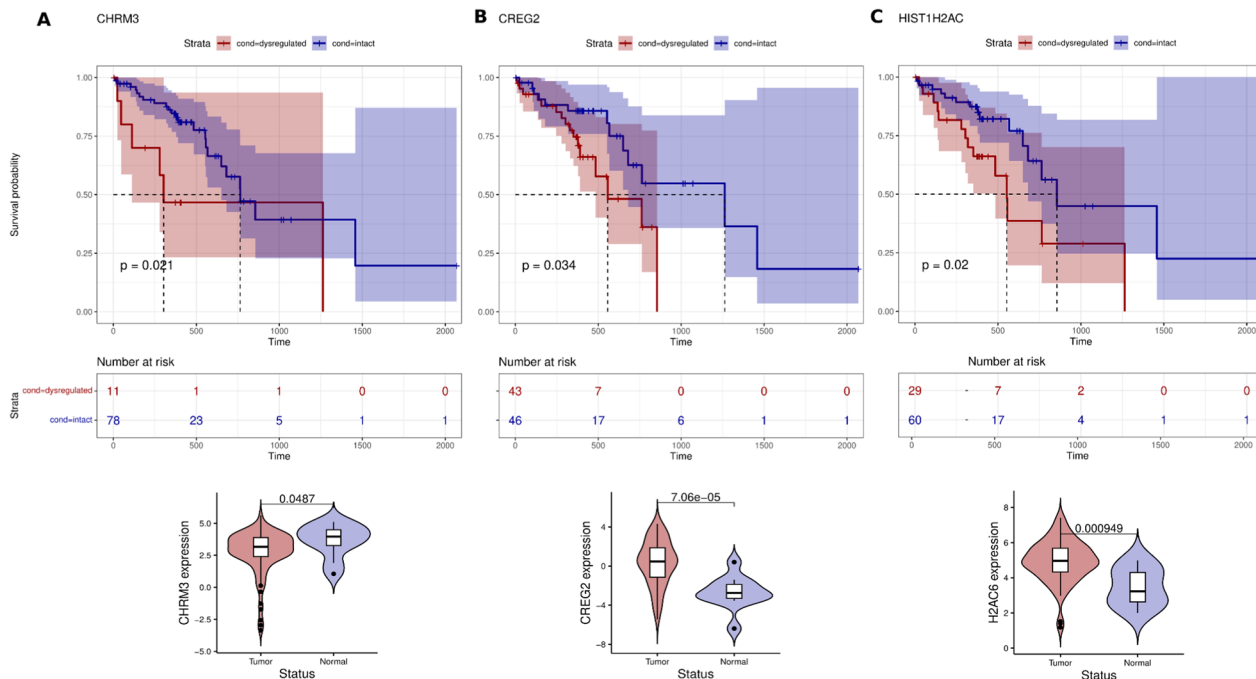
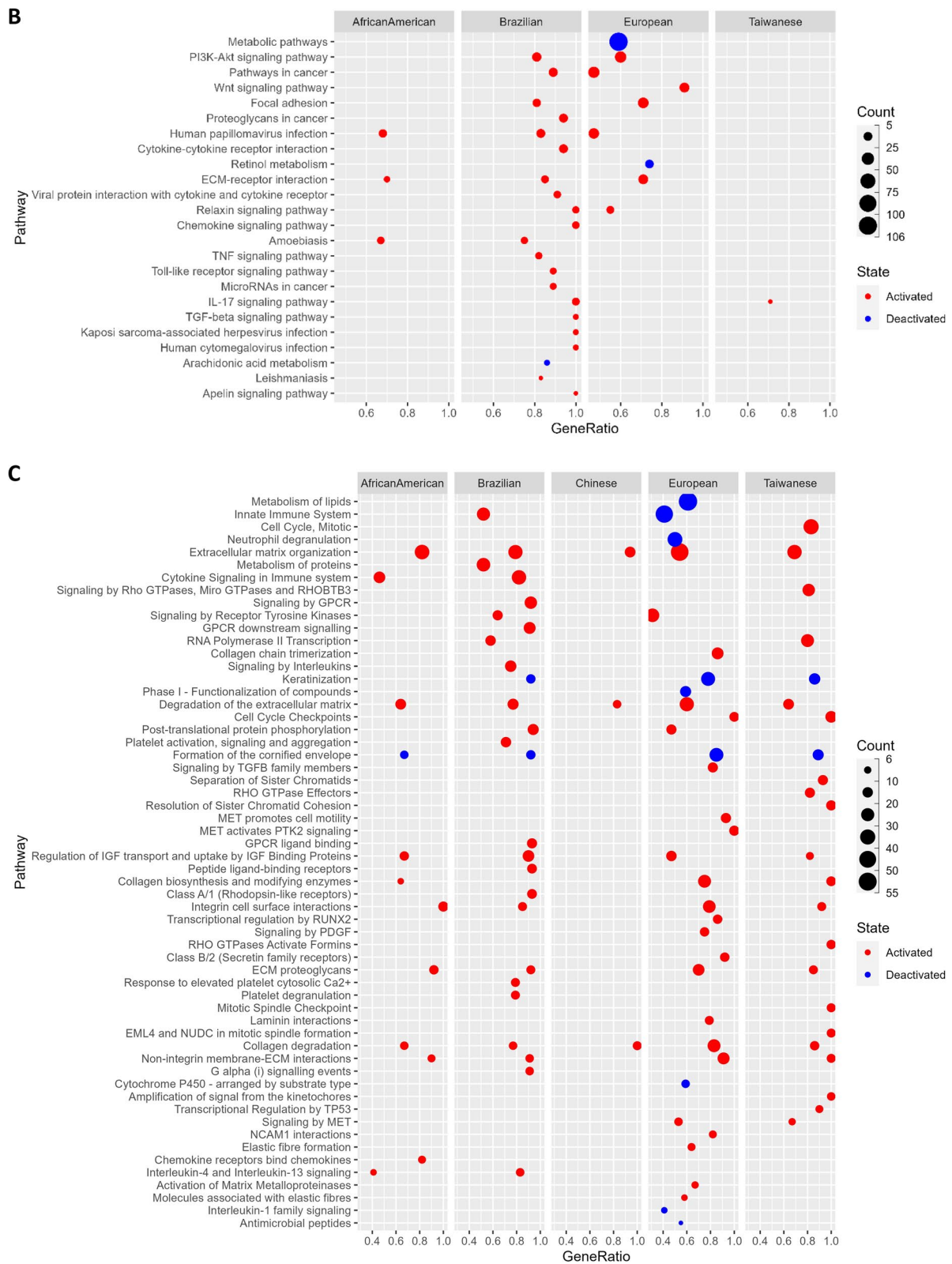


Fig. 4 illustrates DEGs with significant effect on Survival and observed mean differences between normal-tumor samples for the genes **A** CHRM3, **B** CREG2 and **C** HIST1H2AC (alias H2AC6) in ESCC data from TCGA



Fig. 5 Overrepresented **A** GO terms—Biological Process, **B** Pathway terms from KEGG, **C** REACTOME, **D** Wiki pathways, among DEGs



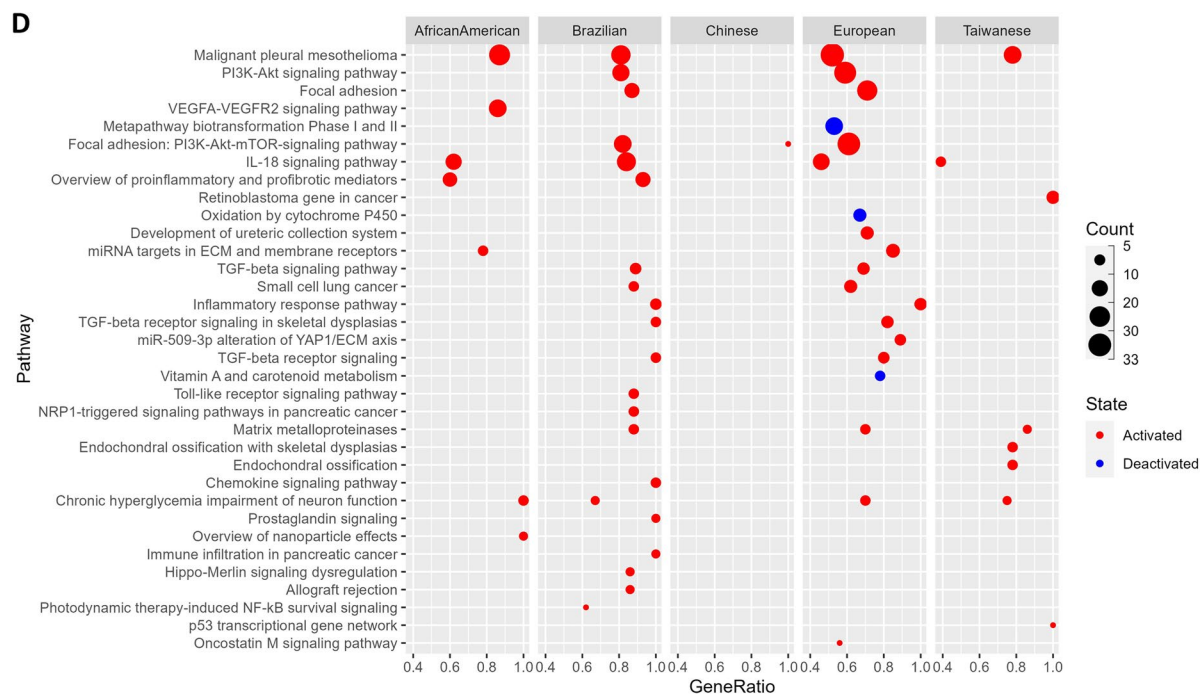


Fig. 5 continued

muscles becoming bruised and forming bone-like structures. Several ossification-related cases were reported in ESCC [45]. Ossifying tumors may arise in the skeleton, the viscera, or the soft tissues and are classified according to the tissue of origin and histological characteristics. We also identified several inflammatory response-related interleukin activations pathways to be overrepresented in our DEGs; these included—Interleukin (IL) -4 and IL-13 signaling, IL-17 signaling pathway, and IL-18 signaling pathway (Fig. 5B–D). The ESCC patients indeed were reported to show high IL-7R expression that potentially contributes to the ESCC progression by promoting the development of various malignant phenotypes [46]. We also identified the signaling pathways such as VEGFA-VEGFR2 (activated) signaling that regulates the cell migration, proliferation, and survival in the formation of new blood vessels to be overrepresented in our DEGs list in context to ESCC. In fact, blocking the VEGFA1-VEGFR2 signaling is a suggested therapy with potential benefits for patients with aggressive EC [47]. Similarly, the transforming growth factor beta (TGF-Beta) signaling pathway (activated) is overrepresented by our DEGs which involves many cellular processes, including cell growth, cell differentiation, cell migration, apoptosis, cellular homeostasis, and other cellular functions. Interestingly, one of the previous reports suggested that the inhibition of this pathway prevents ESCC-induced neoangiogenesis [48, 49].

We also identified the overrepresentation of Toll-like receptor signaling pathways that play crucial roles in the innate immune system by recognizing pathogen-associated molecular patterns derived from various microbes [50], which potentially involves the genes linked to Human Papilloma Virus which is suggested to contribute to ESCC in a high-risk population [51], such as African American based on our analysis. Other microbial infections include Amoebiasis in African American and Brazilian populations, Leishmaniasis, and Human Cytomegalovirus infection in Brazilian populations (Fig. 5B). Literature evidences support these plausible microbial infections in ESCC. For example (i) Infection of several species of *Leishmania* was observed with Squamous Cell Carcinoma in Brazil and Iran [52–54], and a long term infection of it suspected to induce DNA methylation alterations triggering tumorigenesis [52], (ii) CMV infection associated gastrointestinal tract lesion- esophagitis in a Japanese patient observed with a moderately differentiated squamous cell carcinoma [55] and reactivation of CMV infection among esophageal cancer patients undergoing chemotherapy in Japan [56], (iii) Cervical cancer patients in India [57] and colon cancer patients in India and Japan [58, 59] found with Amoebiasis. Further, *Entamoeba histolytica* infects epithelial cells expressing EhADH adhesin together with the EhCP112 cysteine protease and damage epithelium [60]. The genes

Table 4 List of microbial infection terms seen activated across populations

Microbial infection terms	Normalized Enrichment score	Adjusted P-value	Genes	Population
Human papillomavirus infection	0.49	7.1E-04	COL1A2, COL4A1, COL4A2, COL6A3, CREB5, FN1, FZD2, HEY1, ISG15, LAMC2, PTGS2, SPP1, TNC	African American
Human papillomavirus infection	0.56	5.1E-05	AKT3, CCNA1, COL1A1, COL1A2, FN1, HEY1, ITGA5, ITGA6, LAMB3, LAMC1, LAMC2, PTGS2, SPP1, THBS1, TNC	Brazilian
Human papillomavirus infection	0.40	1.28E-05	CCNA1, COL1A2, COL4A1, COL4A2, COL4A6, COL6A3, COL9A3, FN1, FZD2, HES4, HEY1, ISG15, ITGA3, ITGA5, LAMB3, LAMC1, LAMC2, PTGS2, RELN, SPP1, THBS1, THBS2, THBS4, TNC	European
Amoebiasis	0.56	4.7E-04	COL1A2, COL3A1, COL4A1, COL4A2, CXCL1, CXCL8, FN1, IL6, LAMC2, PLCB4	African American
Amoebiasis	0.49	0.047	COL1A1, COL1A2, COL3A1, CXCL1, FN1, IL1B, LAMB3, LAMC1, LAMC2	Brazilian
Human cytomegalovirus infection	0.68	0.030	AKT3, CCL2, CCR1, GNA12, IL1B, PTGS2	Brazilian
Leishmaniasis	0.79	5.2E-03	C3, FCGR2A, FCGR3A, IL1B, PTGS2	Brazilian

overrepresented for microbial infections among respective populations is listed in Table 4.

Toxicogenomic risk factors, and drug chemicals enrichment of DEGs

In order to investigate if our list of DEGs in ESCC was linked to any potential adverse effects that could result from factors such as exposure to environmental or chemical toxins we inferred their toxicogenomic enrichment from the Comparative Toxicogenomics Database (See Methods). We identified 11 genes such as PTGS2, SLC39A6, TAGLN, HMGN5, KRT17, FAT1, SERPINB3, ALDH2, ANXA1, SALL4, and TPM1 with evidence of direct association with the risk factors such as 4-Nitroquinoline-1-oxide, nitroso benzylmethylamine, Tobacco Smoke Pollution, and Zinc, and drug chemicals such as Cisplatin, diallyl trisulfide, Docetaxel, Fluorouracil, Mitomycin and Vinorelbine (Table 5). Of these interesting genes, SLC39A6 is identified as a therapeutic target in previous reports [61] for Ladiratuzumab Vedotin (a Zinc transporter ZIP6 binding agent) drug to treat triple-negative breast cancer and angiosarcoma. Polymorphisms in ALDH2 increased the risk of esophageal cancer with exposure to ethanol and cigarette smoking [62]. Sall4 is an essential regulator in cisplatin-induced apoptosis, and knockdown of Sall4 may restore cisplatin sensitivity in acquired resistant cells [63]. Similarly, FAT1 downregulation enhanced cisplatin resistance and stemness in ESCC [64]. Dietary Zinc known to modulate PTGS2 expression [65]. TPM1 is known to be downregulated by overexpression of miR-21 in inflammatory esophagus and tongue of Zinc deficient rat [66].

If we considered all the potential associations ignoring evidence for the “direct” associations, we found

that ~60% of all the DEGs were linked with tobacco smoke across all the populations, however, this was relatively lower in the Japanese population (~40%) (Fig. 6A). Similarly, ~25% and ~10% of DEGs were associated with Zinc and Nitroso Benzylmethylamine across populations, with an exception being Japanese where we found ~30% of DEGs were associated with Nitroso Benzylmethylamine. Nitroso Benzylmethylamine is a known carcinogen and mutagen, with the potential to cause cancer and genetic mutations in the cells. The Japanese population showed a higher fraction of DEGs that were associated with Nitroso Benzylmethylamine because of the fact that certain traditional Japanese foods, such as pickled vegetables and fermented fish products, contain high levels of Nitroso Benzylmethylamine [67–69].

We additionally retrieved the gene-drug associations from the CTD. We found the highest fraction of our DEG list was associated with Cisplatin across all populations, followed by Fluorouracil. Both of these chemotherapeutic agents are commonly used for the treatment of ESCC [70, 71]. We once again observed that the DEGs particularly in the Japanese population showed different trends of association with the drugs compared to all other populations (Fig. 6A). We found a relatively low fraction of DEGs associated with Cisplatin, but a higher fraction with Fluorouracil, diallyl trisulfide, and docetaxel. This could be because of the fact that only fluorouracil-based regimens showed higher survival incidences and lower hematologic toxic effects than cisplatin plus fluorouracil-based regimen, in the JCOG9205 trial that was conducted by the Japan Clinical Oncology Group [72]. Docetaxel is another chemotherapy drug that was investigated in a phase II clinical trial, where the tolerability of docetaxel as a

Table 5 List of DEGs directly interacting with risk factors or drug chemicals inferred from the comparative toxicogenomic database

Gene	Evidence type	Inference network	Inference score	Reference count	Population
PTGS2	Marker/mechanism	4-Nitroquinoline-1-oxide Cisplatin diallyl trisulfide Docetaxel Fluorouracil Mitomycin nitrosobenzylmethylamine Tobacco Smoke Pollution Vinorelbine Zinc	9.41	14	African American, Brazilian, European
SLC39A6	Therapeutic (LADIRATUZUMAB VEDOTIN-Zinc transporter ZIP6 binding agent) for tripple -ve breast cancer and angiosarcoma	Tobacco Smoke Pollution Zinc	4.9	7	African American
TAGLN	Marker/mechanism	Cisplatin Fluorouracil Tobacco Smoke Pollution	5.5	5	African American
HMGN5	Marker/mechanism	Cisplatin Fluorouracil Tobacco Smoke Pollution	8.74	5	African American, European
KRT17	Marker/mechanism	Cisplatin Fluorouracil nitrosobenzylmethylamine Tobacco Smoke Pollution	10.22	6	African American, Brazilian, European, Chinese
FAT1	Marker/mechanism	Mitomycin nitrosobenzylmethylamine Tobacco Smoke Pollution	10.22	6	Brazilian, European
SERPINB3	Marker/mechanism	nitrosobenzylmethylamine Tobacco Smoke Pollution Zinc	9.69	8	Chinese, European, Taiwanese
ALDH2	Marker/mechanism	Cisplatin nitrosobenzylmethylamine Tobacco Smoke Pollution Zinc	7.51	11	European
ANXA1	Marker/mechanism	Cisplatin diallyl trisulfide Fluorouracil nitrosobenzylmethylamine Tobacco Smoke Pollution Zinc	13.15	11	European
SALL4	Marker/mechanism	Cisplatin Tobacco Smoke Pollution	4.95	5	European
TPM1	Marker/mechanism	Cisplatin Fluorouracil Tobacco Smoke Pollution Zinc	8.18	9	European

single agent in Japanese patients with metastatic esophageal cancer was analyzed [73]. The results showed that the Docetaxel was fairly well tolerated and there were no treatment-related deaths. A higher fraction of DEGs particularly from the Japanese population was also associated with the dietary supplement Diallyl trisulfide (DATS) which is a characteristic flavor component of the essential oil prepared from garlic (*Allium sativum* L.) [74] (Fig. 6A). The antitumor activities of DATs are widely investigated across different types of cancers [75, 76]. The genes inferred for “direct” and “indirect” association with toxicogenomic terms in each population are provided in Additional file 2: Table S9.

Discussion

We conducted a study where we analyzed microarray datasets of ESCC from six distinct global populations and compared the gene expression patterns. Our analysis revealed that while the majority of the DEGs were consistent across all populations, we identified a small subset of discordant DEGs (19 genes). These genes can

provide valuable insights into the genetic and molecular mechanisms underlying the disease, and their discordant expression patterns suggest that population-specific environmental factors, genetic signatures, epigenetic variation and Cis- and Trans-Acting Expression Quantitative Trait Loci [77–79] may contribute to differences in disease progression, drug response, and overall prognosis.

Further characterization of these population-specific gene expression patterns could potentially help identify potential therapeutic targets to certain populations and enable the development of personalized treatment strategies that account for genetic and molecular variations across populations. Additionally, this information could aid in the development of biomarkers to predict drug response or disease prognosis in specific populations, which can ultimately lead to more effective and targeted treatments for patients with ESCC.

We additionally report 166 novel DEGs in ESCC, which is a significant finding. These DEGs have not been previously associated with ESCC and may play a role in the development and progression of the disease. The list of

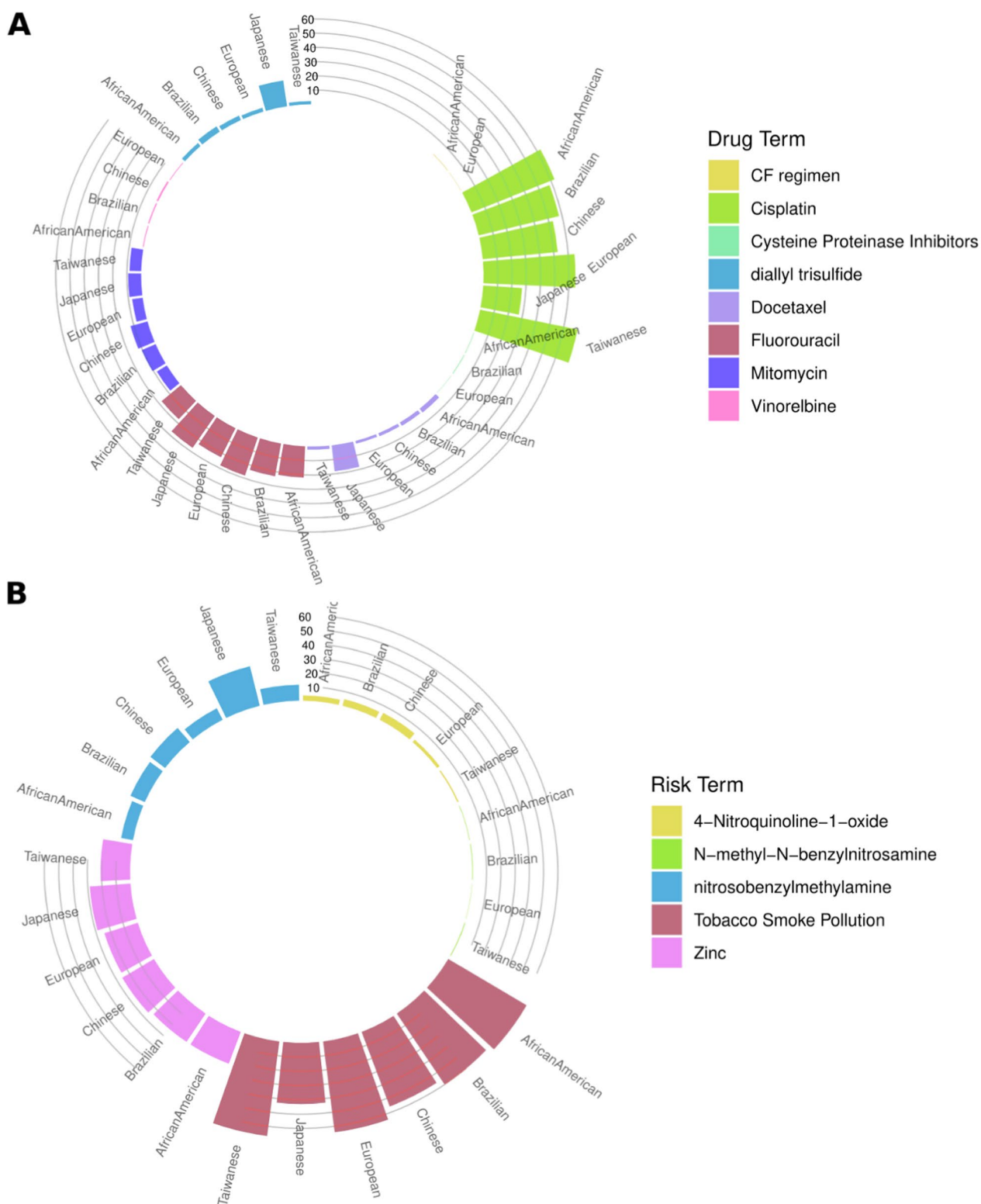


Fig. 6 Overrepresented **A** drug chemicals and **B** risk factors interaction among the DEGs inferred from Comparative Toxicogenomic Database

identified genes includes various novel genes as well as some that have previously been implicated in cancer. For example, the upregulation of KLK7, a gene encoding kallikrein-related peptidase 7, has been linked to the promotion of cell proliferation and invasion in several types of cancer [80, 81]. Similarly, the upregulation of DSG1, a gene encoding desmoglein-1, has been implicated in the development of several types of cancer, including head and neck squamous cell carcinoma. Serine aspartate repeat containing protein D (SdrD) of *Staphylococcus aureus* has been found to directly interact with DSG1 of human squamous cells [82]. It is important to note that DSG2 which is known to upregulate in ESCC has been identified to be a substrate for *Helicobacter Pylori* HtrA receptor in epithelial cells [83]. Similarly, ENDOU is substrate for Nsp15 of Nido family viruses, which encodes for a protein with endoribonuclease activity that binds to the polyuridine-enriched single-stranded RNA [62]. NSP15 has been reported to be involved in the viral replication process and in the evasion of the host immune system [84]. Although there is limited information on the role of ENDOU in cancer, it has been shown to play a role in other cellular processes such as DNA damage response, stress response, and cell death. ENDOU regulates c-Myc expression by regulating the AICD of B cells [85]. The downregulation of ENDOU in ESCC samples plausibly promote tumorigenesis or progression of the disease. However, further studies are needed to elucidate the underlying mechanisms by which ENDOU may contribute to ESCC development and progression. The fact that ENDOU is overexpressed in healthy esophageal mucosa tissues suggests that it may play a role in maintaining the normal function of the esophagus. Understanding the functional role of ENDOU in the esophagus and how it is altered in the context of ESCC could provide valuable insights into the development of novel diagnostic and therapeutic approaches for this disease.

The identification of candidate genes for prognosis in ESCC is an important step towards the development of new biomarkers for early detection and treatment of this cancer. In this study, we identified three novel genes—CHRM3, CREG2, and H2AC6—that showed a significant association with overall survival in ESCC. CHRM3 (cholinergic receptor muscarinic 3) is a novel gene in ESCC, but its expression has been associated with poor prognosis in endometrial carcinoma [86]. Interestingly, CHRM3 has been linked to a well-reported gene KLF4 in ESCC via CHRM3-AS2. A previous report showed that silencing of CHRM3-AS2 expression inhibited cell viability, colony formation, migration, and invasion and promoted apoptosis effects by targeting miR-370-5p/KLF4 in Glioma [87]. This finding suggests a potential role of CHRM3 in tumorigenesis and progression in ESCC. CREG2 (cellular

repressor of E1A stimulated genes 2) is highly expressed in malignant gastric cancer tissues and is positively correlated with tumor clinical stage, tumor metastasis, and stages of tumor infiltration. In our study, we found that CREG2 was highly expressed in ESCC, suggesting its potential as a prognostic marker in ESCC. However, GTEx data suggests that CREG2 is not highly expressed in esophageal tissues, indicating that its expression may be tissue-specific [88]. H2AC6 (H2A Clustered Histone 6; alias HIST1H2AC) is moderately expressed in esophageal tissues, but in our study, we found that it was downregulated in ESCC. HIST1H2AC has been shown to be progressively downregulated in HPV-positive neoplastic keratinocytes derived from uterine cervical preneoplastic lesions at different levels of malignancy [89]. This finding suggests that the downregulation of HIST1H2AC may also play a role in the development and progression of ESCC. Overall, these findings provide a foundation for further investigations into the underlying mechanisms of the disease.

Based on our functional enrichment analysis, we identified the Extracellular Matrix related terms were particularly overrepresented in the DEGs across populations. The extracellular matrix is a complex network of proteins and other molecules that provide structural support to cells and play important roles in cell signaling, adhesion, and migration. Alterations in ECM organization and function have been implicated in the development and progression of many types of cancer [90]. Hence, the overrepresentation of ECM-related GO terms and pathway annotations among DEGs in ESCC suggests that alterations in ECM organization may play a significant role in the development and progression of this cancer. ECM changes may contribute to tumor cell invasion and metastasis, as well as alterations in cell signaling pathways that promote tumor growth and survival. Understanding the role of the ECM in cancer development and progression, and targeting ECM-related pathways and molecules may be a promising strategy for developing new treatments that can improve patient outcomes in ESCC.

ESCC is a type of cancer that is commonly associated with chronic inflammation, a characteristic shared by many other types of cancer. As expected, our analysis revealed an overrepresentation of gene ontology terms related to the immune response and signaling pathways among the DEGs in ESCC. Interestingly, we also observed population-specific differences in overrepresented terms. For instance, while the humoral immune response and antimicrobial humoral response were found to be deactivated in Europeans, adaptive immunity was activated and innate immunity was deactivated in Brazilians. The respective human microbiomes may influence

these differences in immune response in their geographic regions.

Our analysis also revealed an overrepresentation of pathways related to virus infections among the DEGs, which was not surprising given the well-established links between certain microbial species such as HPV [91], CMV [55, 56] (Table 4), and ESCC. In addition, we also observed an overrepresentation of KEGG pathways related to protozoans such as Amoebiasis and Leishmaniasis in our analysis, which is an interesting finding because both of these infections are prevalent disease throughout tropical and subtropical regions of the world. In this line we also identified genes associated to CMV and Leishmaniasis were enriched in Brazilian samples, highlighting potential population and environment-specific differences in the etiology of ESCC.

Our investigation revealed a consistent link between tobacco smoking and ESCC across various populations. The smoke from tobacco contains harmful chemicals such as nitrosamines and polycyclic aromatic hydrocarbons, which can cause DNA damage and increase cancer risk. ESCC patients who smoke also tend to have more advanced disease stages and a higher likelihood of recurrence and mortality than non-smokers. Moreover, smoking may interfere with the effectiveness of various ESCC treatments such as chemotherapy, radiation therapy, and surgery. Our study also identified Cisplatin and Fluorouracil as commonly used drugs for treating ESCC across different populations. We further examined genes associated with the response to these drugs. Interestingly, we observed a higher fraction of DEGs linked to the dietary supplement Diallyl trisulfide, a component of garlic oil, in the Japanese population. This compound has been extensively studied in Japan for its anti-tumor properties.

Conclusion

In conclusion, this study provides an analysis of DEGs across 12 datasets of ESCC in different populations. By filtering against ESCC ATLAS, we identified 1442 DEGs, including 1423 concordant DEGs and 19 discordant DEGs. Among the 1423 concordant DEGs, we found 1257 mapped to ESCC ATLAS, and some interesting genes identified included SASH1, which was downregulated in multiple populations, and BLNK, which was downregulated in all populations except the Japanese. We also identified 166 novel DEGs, of which 19 showed the exact opposite expression trend in healthy esophageal-mucosa, indicating their importance in ESCC. This study highlights the differences in DEG expression across different populations and genomic landscaping of microbial connections including Nido family viruses, HPV, *Entamoeba histolytica*, *Lieshmania*, and *staphylococcus*

aureus provides novel insights into the coinfection in etiology of ESCC. Further research could investigate the functional roles of these DEGs in the pathogenesis of ESCC, potentially leading to the development of precise targeted therapies for this disease.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13027-023-00525-8>.

Additional file 1 Fig. 1 Workflow followed in the data analysis. **Fig. 2** QQ-plot showing observed versus theoretical quantiles of expression in all the 12 ESCC 'normal vs tumor' data-sets

Additional file 2 Table 1 List of identified DEGs (1432) in ESCC. **Table 2** List of Novel DEGs (166) identified in ESCC. **Table 3** List of Discordant DEGs (19) identified in ESCC. **Table 4** List of Novel discordant DEGs (6) found in our ESCC analysis. **Table 5** List of overrepresented Gene Ontology (GO) terms in ESCC. **Table 6** List of overrepresented KEGG pathways in ESCC. **Table 7** List of overrepresented REACTOME pathways in ESCC. **Table 8** List of overrepresented Wiki pathways in ESCC. **Table 9** List of genes found connected to Toxicogenomics terms in ESCC

Acknowledgements

We thank the Deanship of Scientific Research, Princess Nourah bint Abdulrahman University for funding this research work

Author contributions

PH, A.A, and A.B designed the study. V.P.G, P.S.G, and PH performed data analysis and interpretation. V.P.G, and PH wrote the manuscript. A.A, A.B, R.U and J.P were involved in discussions and critical review of the manuscript. S.M, N.J, A.T, L.B.V, A.K.B, M.A.W.C and S.K.N were responsible for data curation from literature. A.A and A.B proofread the manuscript. All authors have reviewed and approved the final manuscript.

Funding

This research project was funded by the Deanship of Scientific Research, Princess Nourah bint Abdulrahman University, through the Program of Research Project Funding After Publication, Grant No (43-PRFA-P-8).

Availability of data and materials

The data generated in this study are available within the article and its supplementary data files.

Declarations

Competing interests

Authors do not have any competing interests.

Author details

¹Basic Science Department, College of Medicine, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. ²Mbiomics LLC, 16192 Coastal Highway, Lewes, DE 19958, USA. ³Department of Cellular and Molecular Biology, University of Texas Health Science Center, Tyler, TX, USA. ⁴Finnish Cancer Registry, Unioninkatu 22, 00130 Helsinki, Finland. ⁵Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland. ⁶Department of Preparatory Health Sciences, Riyadh Elm University, Riyadh, Saudi Arabia. ⁷Manipal Center for Biotherapeutics Research, Manipal Academy of Higher Education, Manipal, India. ⁸Meta Biosciences Pvt Ltd, Manipal-GOK Biocubator, Manipal, India.

Received: 6 April 2023 Accepted: 21 August 2023

Published online: 28 August 2023

References

- Morgan E, Soerjomataram I, Runggay H, Coleman HG, Thrift AP, Vignat J, et al. The global landscape of esophageal squamous cell carcinoma and esophageal adenocarcinoma incidence and mortality in 2020 and projections to 2040: new estimates from GLOBOCAN 2020. *Gastroenterology*. 2022;163:649–658.e2.
- Chen Z, Ren Y, Du XL, Yang J, Shen Y, Li S, et al. Incidence and survival differences in esophageal cancer among ethnic groups in the United States. *Oncotarget*. 2017;8:47037–51.
- Tarazi M, Chidambaram S, Markar SR. Risk factors of esophageal squamous cell carcinoma beyond alcohol and smoking. *Cancers (Basel)*. 2021;13:1009.
- Lin S, Wang X, Huang C, Liu X, Zhao J, Yu ITS, et al. Consumption of salted meat and its interactions with alcohol drinking and tobacco smoking on esophageal squamous-cell carcinoma. *Int J Cancer*. 2015;137:582–9.
- Mangalaparthi KK, Patel K, Khan AA, Manoharan M, Karunakaran C, Murugan S, et al. Mutational landscape of esophageal squamous cell carcinoma in an Indian cohort. *Front Oncol*. 2020;10:1457.
- Dar NA, Mir MM, Salam I, Malik MA, Gulzar GM, Yattoo GN, et al. Association between copper excess, zinc deficiency, and TP53 mutations in esophageal squamous cell carcinoma from Kashmir Valley, India—a high risk area. *Nutr Cancer*. 2008;60:585–91.
- Taccioli C, Chen H, Jiang Y, Liu XP, Huang K, Smalley KJ, et al. Dietary zinc deficiency fuels esophageal cancer development by inducing a distinct inflammatory signature. *Oncogene*. 2012;31:4550–8.
- Hashemi SM, Mashhadi M, Moghaddam AA, Yousefi J, Mofrad AD, Sadeghi M, et al. The relationship between serum selenium and zinc with gastroesophageal cancers in the Southeast of Iran. *Indian J Med Paediatr Oncol*. 2017;38:169–72.
- Amer MH. Epidemiologic aspects of esophageal cancer in Saudi Arabian Patients. *Ann Saudi Med*. 1985;5:69–77.
- Domper Arnal MJ, Ferrández Arenas Á, Lanás AA. Esophageal cancer: Risk factors, screening and endoscopic treatment in Western and Eastern countries. *World J Gastroenterol*. 2015;21:7933–43.
- Reitano E, deAngelis N, Gavriilidis P, Gaiani F, Memeo R, Inchingolo R, et al. Oral bacterial microbiota in digestive cancer patients: a systematic review. *Microorganisms*. 2021;9:2585.
- Yang W, Chen C-H, Jia M, Xing X, Gao L, Tsai H-T, et al. Tumor-associated microbiota in esophageal squamous cell carcinoma. *Front Cell Dev Biol*. 2021;9: 641270.
- Dąbrowski A, Kwaśniewski W, Skoczylas T, Bednarek W, Kuźma D, Goździcka-Józefiak A. Incidence of human papilloma virus in esophageal squamous cell carcinoma in patients from the Lublin region. *World J Gastroenterol*. 2012;18:5739–44.
- Petrick JL, Wyss AB, Butler AM, Cummings C, Sun X, Poole C, et al. Prevalence of human papillomavirus among oesophageal squamous cell carcinoma cases: systematic review and meta-analysis. *Br J Cancer*. 2014;110:2369–77.
- Syrjänen K. Geographic origin is a significant determinant of human papillomavirus prevalence in oesophageal squamous cell carcinoma: systematic review and meta-analysis. *Scand J Infect Dis*. 2013;45:1–18.
- 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061–73.
- Park HA, Neumeyer S, Michailidou K, Bolla MK, Wang Q, Dennis J, et al. Mendelian randomisation study of smoking exposure in relation to breast cancer risk. *Br J Cancer*. 2021;125:1135–45.
- Tang H, Yang D, Han C, Mu P. Smoking, DNA methylation, and breast cancer: a mendelian randomization study. *Front Oncol*. 2021;11: 745918.
- Alegria-Torres JA, Baccarelli A, Bollati V. Epigenetics and lifestyle. *Epigenomics*. 2011;3:267–77.
- Talukdar FR, Soares Lima SC, Khoueiry R, Laskar RS, Cuenin C, Sorroche BP, et al. Genome-wide DNA methylation profiling of esophageal squamous cell carcinoma from global high-incidence regions identifies crucial genes and potential cancer markers. *Cancer Res*. 2021;81:2612–24.
- Abnet CC, Arnold M, Wei W-Q. Epidemiology of esophageal squamous cell carcinoma. *Gastroenterology*. 2018;154:360–73.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41:D991–995.
- Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, et al. ArrayExpress update - from bulk to single-cell expression data. *Nucleic Acids Res*. 2019;47:D711–5.
- Bono H. All of gene expression (AOE): an integrated index for public gene expression databases. *PLoS ONE*. 2020;15: e0227076.
- Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics*. 2009;25:415–6.
- Pagès H, Carlson M, Falcon S, Li N. AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor. R package version 1580. 2022; <https://bioconductor.org/packages/AnnotationDbi>
- DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials*. 2007;28:105–14.
- Harrison PF, Pattison AD, Powell DR, Beilharz TH. Topconfcats: a package for confident effect sizes in differential expression analysis provides a more biologically useful ranked gene list. *Genome Biol*. 2019;20:67.
- McCarthy DJ, Smyth GK. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics*. 2009;25:765–71.
- Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)*. 2021;2:100141.
- Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. *Genetics*. 2023;224:iyad031.
- Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*. 2023;51:D587–92.
- Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res*. 2022;50:D687–92.
- Martens M, Ammar A, Riutta A, Waagmeester A, Slenfer DN, Hanspers K, et al. WikiPathways: connecting communities. *Nucleic Acids Res*. 2021;49:D613–21.
- Hernandez-Ferrer C, Gonzalez JR. CTDquerier: a bioconductor R package for Comparative Toxicogenomics Database™ data extraction, visualization and enrichment of environmental and toxicological studies. *Bioinformatics*. 2018;34:3235–7.
- King BL, Davis AP, Rosenstein MC, Wiegiers TC, Mattingly CJ. Ranking transitive chemical-disease inferences using local network topology in the comparative toxicogenomics database. *PLoS ONE*. 2012;7: e46524.
- Su H, Hu N, Yang HH, Wang C, Takikita M, Wang Q-H, et al. Global gene expression profiling and validation in esophageal squamous cell carcinoma and its association with clinical phenotypes. *Clin Cancer Res*. 2011;17:2955–66.
- Lagresle-Peyrou C, Millili M, Luce S, Boned A, Sadek H, Rouiller J, et al. The BLNK adaptor protein has a nonredundant role in human B-cell differentiation. *J Allergy Clin Immunol*. 2014;134:145–54.
- Lee J-H, Park J-W, Byun JK, Kim HK, Ryu PD, Lee SY, et al. Silencing of voltage-gated potassium channel KV9.3 inhibits proliferation in human colon and lung carcinoma cells. *Oncotarget*. 2015;6:8132–43.
- Hase H, Aoki M, Matsumoto K, Nakai S, Nagata T, Takeda A, et al. Cancer type-SLCO1B3 promotes epithelial-mesenchymal transition resulting in the tumour progression of non-small cell lung cancer. *Oncol Rep*. 2021;45:309–16.
- Liu Y, Wang C, Cheng P, Zhang S, Zhou W, Xu Y, et al. FHL1 inhibits the progression of colorectal cancer by regulating the Wnt/ β -catenin signaling pathway. *J Cancer*. 2021;12:5345–54.
- Feng M, Dong N, Zhou X, Ma L, Xiang R. Myosin light chain 9 promotes the proliferation, invasion, migration and angiogenesis of colorectal cancer cells by binding to Yes-associated protein 1 and regulating Hippo signaling. *Bioengineered*. 2022;13:96–106.
- Palumbo A, Meireles Da Costa N, Pontes B, Leite de Oliveira F, Lohan Codeço M, Ribeiro-Pinto LF, et al. Esophageal cancer development: crucial clues arising from the extracellular matrix. *Cells*. 2020;9:455.
- Wang X, Peng Y, Xie M, Gao Z, Yin L, Pu Y, et al. Identification of extracellular matrix protein 1 as a potential plasma biomarker of ESCC by proteomic analysis using iTRAQ and 2D-LC-MS/MS. *Proteomics Clin Appl*. 2017;11.

45. Nakajima Y, Ohta S, Okada T, Miyawaki Y, Hoshino A, Suzuki T, et al. Osteoplastic bone metastasis in esophageal squamous cell cancer: report of a case. *Surg Today*. 2012;42:376–81.
46. Kitamura Y, Koma Y-I, Tanigawa K, Tsukamoto S, Azumi Y, Miyako S, et al. Roles of IL-7R induced by interactions between cancer cells and macrophages in the progression of esophageal squamous cell carcinoma. *Cancers (Basel)*. 2023;15:394.
47. Xu WW, Li B, Lam AKY, Tsoo SW, Law SYK, Chan KW, et al. Targeting VEGFR1- and VEGFR2-expressing non-tumor cells is essential for esophageal cancer therapy. *Oncotarget*. 2015;6:1790–805.
48. Wan CC, Nisar MF, Wu H. Pharmacological activities of natural products through the TGF- β signalling pathway. *Evid Based Complement Alternat Med*. 2022;2022:9823258.
49. Lu Z, Chen Z, Li Y, Wang J, Zhang Z, Che Y, et al. TGF- β -induced NKILA inhibits ESCC cell migration and invasion through NF- κ B/MMP14 signaling. *J Mol Med (Berl)*. 2018;96:301–13.
50. Kawasaki T, Kawai T. Toll-like receptor signaling pathways. *Front Immunol*. 2014;5:461.
51. Xu W, Liu Z, Bao Q, Qian Z. Viruses, other pathogenic microorganisms and esophageal cancer. *Gastrointest Tumors*. 2015;2:2–13.
52. Vega-Benedetti AF, Loi E, Zavattari P. DNA methylation alterations caused by Leishmania infection may generate a microenvironment prone to tumour development. *Front Cell Infect Microbiol*. 2022;12:984134.
53. Quintella LP, Cuzzi T, de Fátima Madeira M, Valette-Rosalino CM, de Matos SM, de Camargo F, Vasconcellos E, et al. Cutaneous leishmaniasis with pseudoepitheliomatous hyperplasia simulating squamous cell carcinoma. *Am J Dermatopathol*. 2011;33:642–4.
54. Khorsandi-Ashtiani M-T, Hasibi M, Yazdani N, Paydarfar JA, Sadri F, Mirashrafi F, et al. Auricular leishmaniasis mimicking squamous cell carcinoma. *J Laryngol Otol*. 2009;123:915–8.
55. Murakami D, Harada H, Yamato M, Amano Y. Cytomegalovirus-associated esophagitis on early esophageal cancer in immunocompetent host: a case report. *Gut Pathog*. 2021;13:24.
56. Kitagawa K, Okada H, Miyazaki S, Funakoshi Y, Sanada Y, Chayahara N, et al. Cytomegalovirus reactivation in esophageal cancer patients receiving chemoradiotherapy: a retrospective analysis. *Cancer Med*. 2021;10:7525–33.
57. Mamilla S, Agarwal V, Indulkar S, Vuta T. Cervical amoebiasis mimicking cancer cervix. *J Obstet Gynaecol India*. 2023;73:285–6.
58. Fernandes H, D'Souza CRS, Swethadri GK, Naik CNR. Ameboma of the colon with amebic liver abscess mimicking metastatic colon cancer. *Indian J Pathol Microbiol*. 2009;52:228–30.
59. Abe T, Kawai N, Yasumaru M, Mizutani M, Akamatsu H, Fujita S, et al. Ameboma mimicking colon cancer. *Gastrointest Endosc*. 2009;69:757–8 (**discussion 758**).
60. Betanzos A, Zanatta D, Bañuelos C, Hernández-Nava E, Cuellar P, Orozco E. Epithelial cells expressing EhADH, an entamoeba histolytica adhesin, exhibit increased tight junction proteins. *Front Cell Infect Microbiol*. 2018;8:340.
61. Koscielny G, An P, Carvalho-Silva D, Cham JA, Fumis L, Gasparyan R, et al. Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res*. 2017;45:D985–94.
62. Cui R, Kamatani Y, Takahashi A, Usami M, Hosono N, Kawaguchi T, et al. Functional variants in ADH1B and ALDH2 coupled with alcohol and smoking synergistically enhance esophageal cancer risk. *Gastroenterology*. 2009;137:1768–75.
63. Jiang G, Liu C-T. Knockdown of SALL4 overcomes cisplatin-resistance through AKT/mTOR signaling in lung cancer cells. *Int J Clin Exp Pathol*. 2018;11:634–41.
64. Zhai Y, Shan C, Zhang H, Kong P, Zhang L, Wang Y, et al. FAT1 down-regulation enhances stemness and cisplatin resistance in esophageal squamous cell carcinoma. *Mol Cell Biochem*. 2022;477:2689–702.
65. Fong LYY, Zhang L, Jiang Y, Farber JL. Dietary zinc modulation of COX-2 expression and lingual and esophageal carcinogenesis in rats. *J Natl Cancer Inst*. 2005;97:40–50.
66. Alder H, Taccioli C, Chen H, Jiang Y, Smalley KJ, Fadda P, et al. Dysregulation of miR-31 and miR-21 induced by zinc deficiency promotes esophageal cancer. *Carcinogenesis*. 2012;33:1736–44.
67. Blot SI, Hoste EA, Vandewoude KH, Colardyn FA. Estimates of attributable mortality of systemic candida infection in the ICU. *J Crit Care*. 2003;18:130–1 (**author reply 131**).
68. Song P, Wu L, Guan W. Dietary nitrates, nitrites, and nitrosamines intake and the risk of gastric cancer: a meta-analysis. *Nutrients*. 2015;7:9872–95.
69. Lilleyman JS, Lennard L, Rees CA, Morgan G, Maddocks JL. Childhood lymphoblastic leukaemia: sex difference in 6-mercaptopurine utilization. *Br J Cancer*. 1984;49:703–7.
70. Xu J, Bai Y, Xu N, Li E, Wang B, Wang J, et al. Tislelizumab plus chemotherapy as first-line treatment for advanced esophageal squamous cell carcinoma and gastric/gastroesophageal junction adenocarcinoma. *Clin Cancer Res*. 2020;26:4542–50.
71. Hiramoto S, Kato K, Shoji H, Okita N, Takashima A, Honma Y, et al. A retrospective analysis of 5-fluorouracil plus cisplatin as first-line chemotherapy in the recent treatment strategy for patients with metastatic or recurrent esophageal squamous cell carcinoma. *Int J Clin Oncol*. 2018;23:466–72.
72. Ohtsu A, Shimada Y, Shirao K, Boku N, Hyodo I, Saito H, et al. Randomized phase III trial of fluorouracil alone versus fluorouracil plus cisplatin versus uracil and tegafur plus mitomycin in patients with unresectable, advanced gastric cancer: the Japan Clinical Oncology Group Study (JCOG9205). *J Clin Oncol*. 2003;21:54–9.
73. Muro K, Hamaguchi T, Ohtsu A, Boku N, Chin K, Hyodo I, et al. A phase II study of single-agent docetaxel in patients with metastatic esophageal cancer. *Ann Oncol*. 2004;15:955–9.
74. Mitra S, Das R, Emran TB, Labib RK, Noor-E-Tabassum N, Islam F, et al. Diallyl disulfide: a bioactive garlic compound with anticancer potential. *Front Pharmacol*. 2022;13:943967.
75. Hosono T, Fukao T, Ogihara J, Ito Y, Shiba H, Seki T, et al. Diallyl trisulfide suppresses the proliferation and induces apoptosis of human colon cancer cells through oxidative modification of beta-tubulin. *J Biol Chem*. 2005;280:41487–93.
76. Xiao D, Singh SV. Diallyl trisulfide, a constituent of processed garlic, inactivates Akt to trigger mitochondrial translocation of BAD and caspase-mediated apoptosis in human prostate cancer cells. *Carcinogenesis*. 2006;27:533–40.
77. Price AL, Patterson N, Hancks DC, Myers S, Reich D, Cheung VG, et al. Effects of cis and trans genetic ancestry on gene expression in African Americans. *PLoS Genet*. 2008;4: e1000294.
78. Davis AR, Kohane IS. Expression differences by continent of origin point to the immortalization process. *Hum Mol Genet*. 2009;18:3864–75.
79. Porcellini D, Westram AM, Pascual M, Gaston KJ, Butlin RK, Snook RR. Gene expression clines reveal local adaptation and associated trade-offs at a continental scale. *Sci Rep*. 2016;6:32975.
80. Walker F, Nicole P, Jallane A, Soosaipillai A, Mosbach V, Oikonomopoulou K, et al. Kallikrein-related peptidase 7 (KLK7) is a proliferative factor that is aberrantly expressed in human colon cancer. *Biol Chem*. 2014;395:1075–86.
81. Gong W, Liu Y, Diamandis EP, Kiechle M, Bronger H, Dorn J, et al. Prognostic value of kallikrein-related peptidase 7 (KLK7) mRNA expression in advanced high-grade serous ovarian cancer. *J Ovarian Res*. 2020;13:125.
82. Askarian F, Ajayi C, Hanssen A-M, van Sorge NM, Pettersen I, Diep DB, et al. The interaction between *Staphylococcus aureus* SdrD and desmoglein 1 is important for adhesion to host cells. *Sci Rep*. 2016;6:22134.
83. Bernegger S, Vidmar R, Fonovic M, Posselt G, Turk B, Wessler S. Identification of Desmoglein-2 as a novel target of *Helicobacter pylori* HtrA in epithelial cells. *Cell Commun Signal*. 2021;19:108.
84. Zheng A, Shi Y, Shen Z, Wang G, Shi J, Xiong Q, et al. Insight into the evolution of nidovirus endoribonuclease based on the finding that nsp15 from porcine Deltacoronavirus functions as a dimer. *J Biol Chem*. 2018;293:12054–67.

85. Poe JC, Kountikov EI, Lykken JM, Natarajan A, Marchuk DA, Tedder TF. EndoU is a novel regulator of AICD during peripheral B cell selection. *J Exp Med*. 2014;211:57–69.
86. Wang Y, Li J, Wen S, Yang X, Zhang Y, Wang Z, et al. CHR3 is a novel prognostic factor of poor prognosis in patients with endometrial carcinoma. *Am J Transl Res*. 2015;7:902–11.
87. Wang D, Chen Q, Liu J, Liao Y, Jiang Q. Silencing of lncRNA CHR3-AS2 Expression Exerts Anti-Tumour Effects Against Glioma via Targeting microRNA-370-5p/KLF4. *Front Oncol*. 2022;12: 856381.
88. Xu L, Wang F, Liu H, Xu X-F, Mo W-H, Xia Y-J, et al. Increased expression of cellular repressor of E1A-stimulated gene (CREG) in gastric cancer patients: a mechanism of proliferation and metastasis in cancer. *Dig Dis Sci*. 2011;56:1645–55.
89. Rotondo JC, Bosi S, Bassi C, Ferracin M, Lanza G, Gafà R, et al. Gene expression changes in progression of cervical neoplasia revealed by microarray analysis of cervical neoplastic keratinocytes. *J Cell Physiol*. 2015;230:806–12.
90. Lu P, Takai K, Weaver VM, Werb Z. Extracellular matrix degradation and remodeling in development and disease. *Cold Spring Harb Perspect Biol*. 2011;3: a005058.
91. Syrjänen KJ. HPV infections and oesophageal cancer. *J Clin Pathol*. 2002;55:721–8.
92. Hu N, Clifford RJ, Yang HH, Wang C, Goldstein AM, Ding T, et al. Genome wide analysis of DNA copy number neutral loss of heterozygosity (CNNLOH) and its relation to gene expression in esophageal squamous cell carcinoma. *BMC Genomics*. 2010;11:576.
93. Chen Y-K, Tung C-W, Lee J-Y, Hung Y-C, Lee C-H, Chou S-H, et al. Plasma matrix metalloproteinase 1 improves the detection and survival prediction of esophageal squamous cell carcinoma. *Sci Rep*. 2016;6:30057.
94. Aoyagi K, Minashi K, Igaki H, Tachimori Y, Nishimura T, Hokamura N, et al. Artificially induced epithelial-mesenchymal transition in surgical subjects: its implications in clinical and basic cancer research. *PLoS ONE*. 2011;6: e18196.
95. Yan W, Shih J, Rodriguez-Canales J, Tangrea MA, Player A, Diao L, et al. Three-dimensional mRNA measurements reveal minimal regional heterogeneity in esophageal squamous cell carcinoma. *Am J Pathol*. 2013;182:529–39.
96. Wang Q, Ma C, Kemmer W. Wdr66 is a novel marker for risk stratification and involved in epithelial-mesenchymal transition of esophageal squamous cell carcinoma. *BMC Cancer*. 2013;13:137.
97. Yan W, Shih JH, Rodriguez-Canales J, Tangrea MA, Ylaja K, Hipp J, et al. Identification of unique expression signatures and therapeutic targets in esophageal squamous cell carcinoma. *BMC Res Notes*. 2012;5:73.
98. Lee JJ, Natsuzaka M, Ohashi S, Wong GS, Takaoka M, Michaylira CZ, et al. Hypoxia activates the cyclooxygenase-2-prostaglandin E synthase axis. *Carcinogenesis*. 2010;31:427–34.
99. Nicolau-Neto P, Da Costa NM, de Souza Santos PT, Gonzaga IM, Ferreira MA, Guaraldi S, et al. Esophageal squamous cell carcinoma transcriptome reveals the effect of FOXM1 on patient outcome through novel PIK3R3 mediated activation of PI3K signaling pathway. *Oncotarget*. 2018;9:16634–47.
100. Yang H, Su H, Hu N, Wang C, Wang L, Giffen C, et al. Integrated analysis of genome-wide miRNAs and targeted gene expression in esophageal squamous cell carcinoma (ESCC) and relation to prognosis. *BMC Cancer*. 2020;20:388.
101. Erkizan HV, Johnson K, Ghimbovski S, Karkera D, Trachiotis G, Adib H, et al. African-American esophageal squamous cell carcinoma expression profile reveals dysregulation of stress response and detox networks. *BMC Cancer*. 2017;17:426.
102. Tong M, Chan KW, Bao JY, Wong KY, Chen J-N, Kwan PS, et al. Rab25 is a tumor suppressor gene with antiangiogenic and anti-invasive activities in esophageal squamous cell carcinoma. *Cancer Res*. 2012;72:6024–35.
103. Saito S, Morishima K, Ui T, Hoshino H, Matsubara D, Ishikawa S, et al. The role of HGF/MET and FGF/FGFR in fibroblast-derived growth stimulation and lapatinib-resistance of esophageal squamous cell carcinoma. *BMC Cancer*. 2015;15:82.
104. Shimokuni T, Tanimoto K, Hiyama K, Otani K, Ohtaki M, Hihara J, et al. Chemosensitivity prediction in esophageal squamous cell carcinoma: novel marker genes and efficacy-prediction formulae using their expression data. *Int J Oncol*. 2006;28:1153–62.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

