Heliyon 10 (2024) e38011

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Identification of genome-wide copy number variation-driven subtypes for the treatment and prognostic prediction of esophageal carcinoma

Chao Zhao [a], Hui Han [a], Yushuang Tian [a], Guangjin Qu [a], Yingying Xu [a], Yihan Wang [b,**], Lili Shi [a,*]

[a] Department of Gerontology, First Affiliated Hospital of Harbin Medical University, Harbin, 150001, China
[b] College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, 150081, China

ARTICLE INFO

Keywords:
Esophageal carcinoma
Copy number variation
Classification of subtypes
Prognostic
Biomarkers

ABSTRACT

Background: Esophageal carcinoma (ESCA) is a frequently detected gastrointestinal cancer. Copy number variants (CNVs) have a dramatic impact on the screening, diagnosis and prognostic prediction of cancers. However, the mechanism of action of CNVs on ESCA occurrence and progression remains unclear.

Methods: ESCA samples from The Cancer Genome Atlas (TCGA) were typed by consensus clustering using CNV-associated genes. Weighted Gene Co-Expression Network Analysis (WGCNA) was used to section gene modules closely related to the two clusters, and sub-networks were constructed as hub genes. In addition, seven prognosis-correlated genes were further screened and retained by multivariate Cox regression analysis to develop a prognostic assessment model. The ssGSEA algorithm assessed energy metabolism levels in patients from different clusters and risk groups. Finally, quantitative real-time PCR (qRT-PCR) and live-dead cell staining verified the expression of genes associated with CNV risk scores.

Results: ESCA was classified into two subtypes based on CNV values. Compared with cluster 1, cluster 2 had significantly higher level of immune score and tumor-associated immune cell infiltration as well as a noticeably better overall survival. The three modules most associated with the two clusters were identified by WGCNA, and a prognostic model with a strong prediction performance was constructed with their genes. Glycolysis, lactate metabolism, fatty acid synthesis, glutathione, methionine, and tryptophan metabolic pathway enrichment scores were remarkably higher in patients in cluster 1 and the high-risk group than in cluster 2 and the low-risk group. Knockdown PIK3C2A promoted ESCA cells apoptosis and inhibited cell vibiality.

Conclusion: The current research maybe provides new understanding for the pathogenesis of ESCA based on CNV, providing an effective guidance for its clinical diagnosis and prognostic evaluation.

* Corresponding author. Department of Gerontology, First Affiliated Hospital of Harbin Medical University, Harbin, 150001, China.
** Corresponding author.
E-mail addresses: wangyihan@hrbmu.edu.cn (Y. Wang), shilili@hrbmu.edu.cn (L. Shi).

https://doi.org/10.1016/j.heliyon.2024.e38011
Received 21 June 2024; Received in revised form 11 September 2024; Accepted 16 September 2024
Available online 18 September 2024

1. Introduction

Esophageal carcinoma (ESCA) occurs mainly in the upper gastrointestinal tract, and it ranks seventh among the most common types of cancer [1–3]. For many years, the main approach to treat early-stage ESCC has been surgery. Local therapies including surgery, endoscopic resection, or ablation can be effective for tumors that are classified as very early-stage or high-grade dysplasia. In early-stage cases, surgical intervention can enhance the 5-year survival rate to between 60 % and 85 % [4]. Nevertheless, the insidious nature of early symptoms often leads to patients presenting with metastatic or advanced disease upon diagnosis. Surgery by itself has not produced satisfactory outcomes for patients with locally advanced disease, with median survival times ranging from 12 to 18 months and 5-year survival rate between 15 % and 39 % [5,6]. Additionally, patients undergoing surgery alone frequently experience local or systemic recurrence, with reports indicating that 35 %–50 % experience such recurrence [7]. For this reason, it is necessary to mine accurate predictive ESCA biomarkers for improving the prognosis of ESCA.

Copy number variations (CNVs) are widely present in human genome and include duplications, insertions, and deletions, as well as complex multi-site variation [8]. CNVs are capable of causing the differential expression of genes to varying degrees, and the presence of multiple CNVs in the genome leads to genomic and molecular phenotypic heterogeneity, which causes the development of cancer and other complex diseases [9,10]. Several studies have shown that DNA copy number variation is closely involved in the progression of ESCC and that transcriptional disruption caused by CNV is a potential driver of ESCC development [11–13]. By examining esophageal tissues at different ESCC pathological stages, Liu et al. found that CNV load appeared in esophageal tissues and that CNV accumulated more in the cancer as the lesions deepened [14]. CNV functions crucially in tumor development. To examine the predictive association between CNV and clinical characteristics, 86 % of gene expressions were found to be positively linked with CNV [15] after analysis a total of 816 copies of breast cancer genome and 547 copies of transcriptome data from TCGA. In addition, Shai et al. showed a strong association between CNV and patient survival by collecting an anaplastic oligodendroglioma cohort and performing CNV analysis using single-nucleotide polymorphism arrays [16]. Thus, from this perspective, genomic amplification and deletion-based CNV offer promise for targeted treatment of ESCA.

Currently, there are fewer studies on CNV and ESCA. Therefore, this study identifies subtypes of ESCA using consistent clustering based on CNV data about ESCA in the TCGA database and describes their clinical and immunological characteristics and the relationship with immunotherapeutic response. CNV-related risk models were developed by WGCNA and multivariate Cox analysis, and energy metabolic pathway scores were calculated for different clusters and different risk groups. Finally, the sensitivities of high- and low-risk patients to common chemotherapeutic agents were evaluated. Overall, the current research can improve the present understanding on the mechanism of CNV influence on the development of ESCA and provide more scientific basis for the clinical treatment of ESCA patients.

2. Methods and materials

2.1. Data download and pre-processing

The ESCA data from TCGA (https://portal.gdc.cancer.gov/) database was obtained via UCSC Xena (http://xena.ucsc.edu/). Using the R package TCGAbiolinks, CNV data of ESCA patient tumor tissue samples f were collected rom the TCGA database [17]. The clinical follow-up information of ESCA was downloaded from the TCGA database. We delineated it using GISTIC 2.0 [18] to obtain chromosome data as well as loss or gain status.

2.2. Consensus clustering analysis using ESCA-correlated genes

The “ConsensusClusterPlus” R package [19] was applied to cluster ESCA samples in TCGA. According to the cumulative distribution function (CDF) curve, the optimal clustering number defined and delta area were highly consistent in the cluster without significant changes in the area under the CDF curve.

Sample clustering was assessed by the t-SNE using the “Rtsne” (version0.16) packages for difference analysis between the two subtypes. Additionally, we used Kaplan-Meier curves to show the prognosis of the two clusters.

2.3. Immune cell infiltration analysis

According to the 28 immune cells and their corresponding associated genetic data provided by a previous study [20], relative proportion of these 28 immune cell was calculated by single-sample gene-set enrichment analysis (ssGSEA) with the “GSVA” package [21]. The tumor purity, immune and stromal scores were calculated for the all samples in Cluster1 and Cluster2 using the ESTIMATE algorithm [22]. For a supplement, Microenvironment Cell Populations-counter (MCP-counter) [23] and CIBERSORT [24] were also conducted.

2.4. WGCNA and enrichment analysis

Since the expression patterns of genes implicated in the same biological processes or pathways are similar, the R package WGCNA [25] was used to convert cluster1 and cluster2 into a gene co-expression network. To meet a scale-free distribution criterion, we use the “pickSoftThreshold” function of the WGCNA package to select the best β. Notably, the co-expression network has to conform to a

scale-free network, i.e., the log log(k) of the occurrence of a node with connectivity k is negatively related to the log log(P(k)) of the occurrence possibility of that node and the correlation coefficient has to be greater than 0.85. Next, the "blockwiseModules" function is used for one-step network development and module detection, and modules were sectioned by constructing hierarchical clustering tree using dynamic hybrid cutting technology (parameters: reassignThreshold = 0, mergeCutHeight = 0.25, minModuleSize = 30, pamRespectsDendro = FALSE, numericLabels = TRUE, corType = "pearson", verbose = 3). The Pearson algorithm is used to measure the association between the module signature genes and the two clusters. Notably, the focus was on modules with correlation coefficients more than 0.5 and these genes were selected for subsequent analysis. Finally, we uploaded interested module signature genes to the DAVID database (https://david.ncifcrf.gov/home.jsp), and thus understood the biological processes in which the genes within the module are involved. We also downloaded list of genes contained in lipid metabolism, amino acid metabolism, carbohydrate metabolism, metabolism of other amino acids from the KEGG website (https://www.kegg.jp/), and used GSVA packages [26] to calculate the ssGSEA enrichment scores for these metabolism related pathways within each TCGA-ESCA sample.

2.5. PPI network construction and identification of sub-networks

The PPI network was predicted on STRING (http://string-db.org) [27]. Cytoscape [28] and MCODE [29] were employed for visualizing the PPI network and identifying the most significance modules in the network, respectively.

2.6. Prognostic risk scoring model constructed by clustering using genes correlated with ESCA

Firstly, univariate Cox regression analysis was conducted and LASSO Cox regression analysis was performed based on the prognostic genes in the R package "glmnet" [30]. Following that, independent genes were selected by performing multivariate Cox regression analysis. Risk score was calculating for the remaining genes by the following formula:

$$\text{Risk score} = \sum Exp_i * \beta_i \quad (1)$$

Where Exp_i refers to the expression of the target gene, while β_i is the coefficient of the gene. Finally, the 1-, 3-, and 5-year prediction power of the model was shown by plotting the receiver operating characteristic [31] curves.

2.7. Cell culture and transfection

Human normal esophageal epithelial cell HEEC and human esophageal cancer cell lines OE19 and TE-1 purchased from Bena Biotechnology Co. (Beijing, China) were cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco, USA) with penicillin/streptomycin and fetal bovine serum (FBS, Gibco, USA) at 37 °C with 5 % CO_2.

The target sequences for PIK3C2A siRNA and GPC1 siRNA were: siRNA- PIK3C2A: AGCAAAAGATGTGGACAAAGAAG; siRNA-GPC1: CCCTGACTATTGCCGAAATGTGC.

2.8. Quantitative real-time PCR (qRT-PCR)

Separation of total RNA from HEEC, OE19 and TE-1 cell lines was realized with the use of TRIzol reagent (Thermo Fisher, USA). The cDNA was synthesized from 500 ng of RNA with the use of the HiScript II SuperMix (Vazyme, China). The qRT-PCR added with the SYBR Green Master Mix was performed in ABI 7500 System (Thermo Fisher, USA) the following the amplification condition of 46 cycles at 94 °C for 10 min, at 94 °C for 10 s, and at 60 °C for 45 s. GAPDH acted as the internal reference. In addition, see Table 1 for the primer pair sequences for the targeted genes.

2.9. Calcein AM/PI cytotoxicity assay and cell viability

HEEC, OE19 and TE-1 cell lines were inoculated on cell crawls and washed using PBS. The cells were added with a proper volume of Calcein AM/PI assay solution for 2-h incubation at 37 °C. The staining effect was observed under a fluorescent microscope in the dark.

Differently treated cells were cultured at a density of 1×10^3 cells per well in 96-well plates and then added with CCK-8 solution

Table 1
Sequence of target gene primer pairs for qRT-PCR.

Gene	Forward primer sequence (5′-3′)	Reverse primer sequence (5′-3′)
SDC1	CTGCCGCAAATTGTGGCTAC	TGAGCCGGAGAAGTTGTCAGA
GPC1	GCCCTGACTATTGCCGAAATGTG	GAACTTGTCGGTGATGAGCACC
PIK3C2A	AAATGGGACCAGTAGTTTGCC	GGGTTTGTGCGGTGATTGGTA
CLDN3	AACACCATTATCCGGGACTTCT	GCGGAGTAGACGACCTTGG
EGFR	AGGCACGAGTAACAAGCTCAC	ATGAGGACATAACCAGCCACC
INSR	AAAACGAGGCCCGAAGATTTC	GAGCCCATAGACCCGGAAG
LYN	GCTGGATTTCCTGAAGAGCGATG	CGGTGAATGTAGTTCTTCCGCTC
GAPDH	CTGGGCTACACTGAGCACC	AAGTGGTCGTTGAGGGCAATG

(Beyotime, China) at the indicated time points for another 2-h incubation at 37 °C. Finally, the O.D 450 values of each well was measured with a microplate reader.

2.10. Analysis of differences in drug reactions

The response of each sample in TCGA-ESCA to the drugs was predicted applying the "pRRophetic" package in R [32] and based on the Genomics of Drug Sensitivity in Cancer (GDSC, https://www.cancerrxgene.org/) database.

The TIDE algorithm was applied for predicting patients' potential clinical responses to immune checkpoint blockade (ICB) therapy [33,34]. A higher TIDE score indicates limited benefits from immune checkpoint inhibitor therapy.

2.11. Statistical analysis

Two-group continuous variables were compared by the Wilcoxon rank-sum test. The association between classified variables was analyzed by Chi-square test. Log-rank test was used for comparing survival differences, and the correlation between survival time and genes were counted using univariate Cox regression. All statistical analyses were carried out by R (version 3.6.0). A $p < 0.05$ showed a statistical significance. The SangerBox platform (http://sangerbox.com/home.html) assisted in data processing.

3. Results

3.1. ESCA classification based on genomic amplification and deletion of CNV values

We calculated CNV values for genomic amplifications and deletions in TCGA-ESCA according to GISTIC and grouped ESCA patients

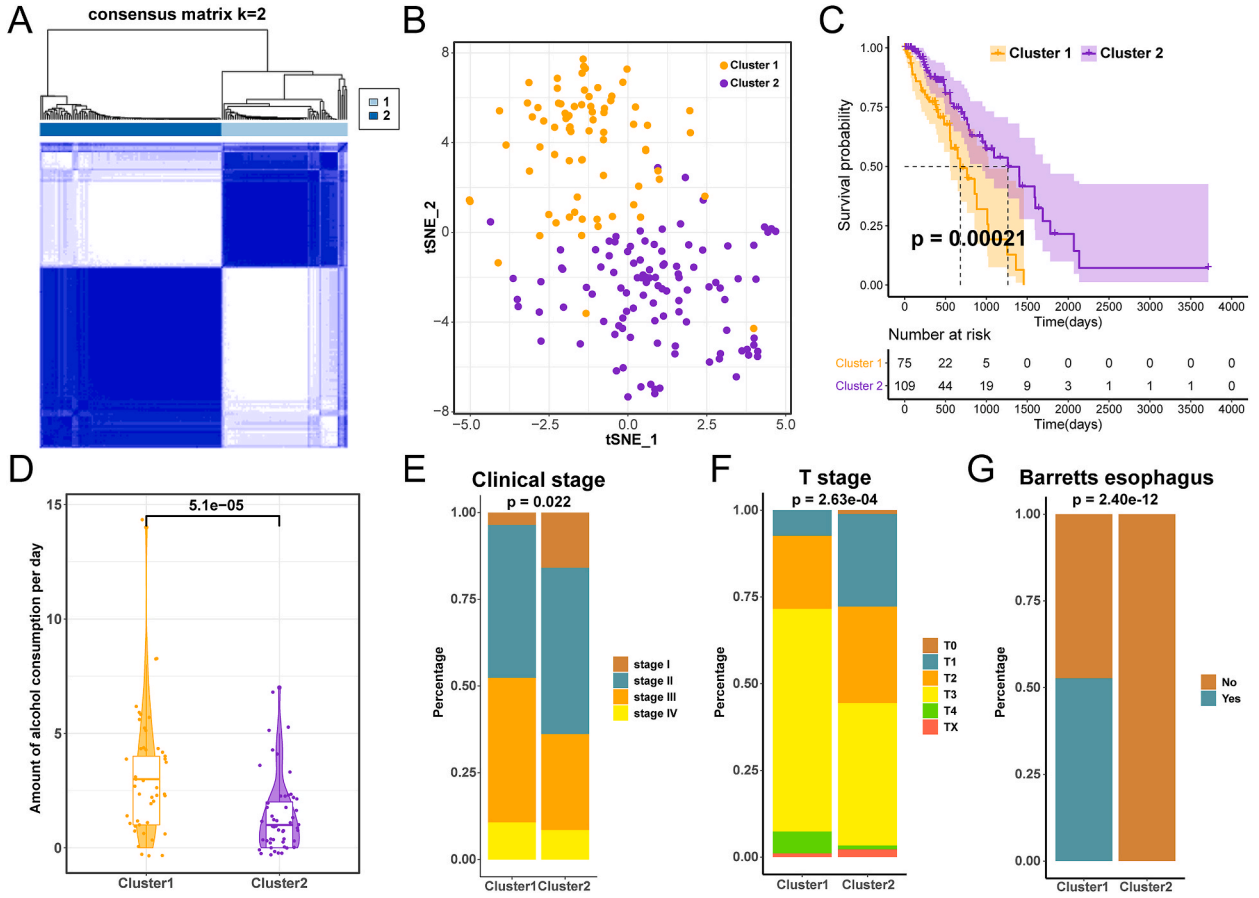


Fig. 1. Identification of ESCA subtypes. (A) Optimal separation of clusters by consensus matrix (k = 2). (B) T-SNE verified the presence of these two clusters in ESCA. The identification of two subgroups without intersection suggests that patterns based on amplification and deletion of ESCA genomic regions can clearly distinguish between cluster 1 and cluster 2. (C) KM survival analysis of cluster 1 and cluster 2. (D) Analysis of the difference in daily alcohol consumption between cluster 1 and cluster 2 groups. (E–G) Comparative analysis of clinical stage, T-stage, and Barrett's esophageal percentage within cluster 1 and cluster 2 groups.

using consistent clustering. Related clustering parameters are shown in Fig. S1. The optimal grouping at k = 2 was determined (Fig. 1A). Thus, we obtain two CNV-based clusters. Using the t-SNE dimensional reduction method, we observed that the patterns of deletion and amplification within genomic regions differed significantly under this grouping (Fig. 1B). Furthermore, observing the two-group difference in CNV values in the amplified and deletion regions, we found that cluster 1 had higher CNV values in the amplified region at 14q21.1, 11q13.3, 3q26.2 and 3q28 (Fig. S2A), while in the deleted region, cluster 1 had lower CNV values at 2q22.1, 9p21.3, 3p14.3, 3p14.2, 3p26.2, 3p25.3 (Fig. S2B).

The shorter survival of cluster1 compared to cluster 2 suggests a poorer prognostic outcome ($p < 0.05$) (Fig. 1C). The proportion of grade, Tobacco smoking history, N stage, and M stage distribution showed no significant difference between the two subtypes (Figs. S2G–F), but there were more patients with daily alcohol consumption, clinical stage, and T stage in cluster 1 (Figure D–F). In addition, compared to cluster 2, cluster 1 is younger in age (Fig. S2C). In particular, Barrett's esophagus is closely related to ESCA. As shown in Fig. 1G, the proportion of Barrett's esophagus was remarkably higher in cluster 1, suggesting that the onset of ESCA in patients in cluster 1 may be related to gastroesophageal reflux disease (GERD) [35].

3.2. Analysis of the degree of immune cell infiltration in two subtypes of ESCA patients

First, differences in the immune microenvironment of the two ESCA subtypes were compared by using ESTIMATE. Relative to cluster1, cluster 2 showed a significant increase in immune score. However, tumor purity and stromal score between the two subtypes of ESCA showed no significant differences (Fig. 2A). Next, we analyzed the level of immune infiltration in the tumor immune microenvironment for both ESCA isoforms using ssGSEA, MCP-counter and CIBERSORT. It was observed that immune cells in 28

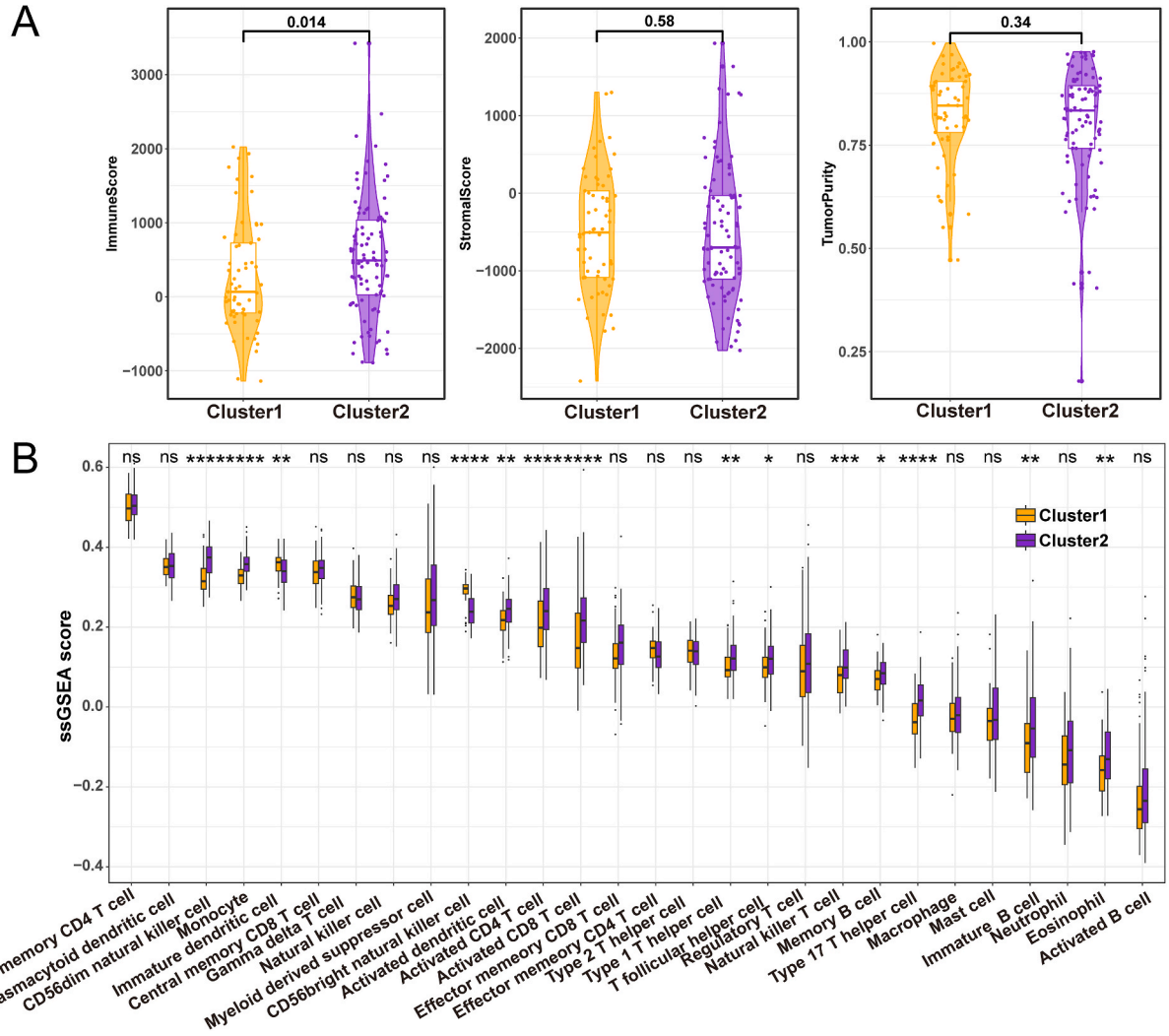


Fig. 2. Relationship between ESCA subtypes and immune microenvironment. (A) The immune score, stromal score, and tumor purity of two ESCA subtypes in TCGA; (B) Evaluation of the enrichment scores of 28 immune cells in two subtypes of ESCA using ssGSEA.

tumors evaluated by ssGSEA, with the exception of immature dendritic cells, showed an enrichment of activity in cluster2, which was remarkably higher than that in cluster1 (Fig. 2B). Additionally, the MCP-counter and CIBERSORT analyses all showed consistent results (Fig. S3). The results in Fig. 2 reflect a large infiltration of immune cells in cluster 2, demonstrating a more active anti-tumor immune microenvironment. This is consistent with the analysis of the prognostic results.

3.3. Identification of gene modules of two ESCA subtypes based on CNV typing

To develop a co-expression network, we analyzed the related gene modules based on CNV typing by using WGCNA. The network best fits the characteristics of a scale-free network when the lowest soft threshold $\beta = 4$ (Fig. 3A). Closely related genes were selected as modules using hierarchical clustering and formed a dendrogram of genes (Fig. 3B). Next, we classified three modules that were closely correlated with two subtypes of ESCA (Fig. 3C). Among them, MEmidnightblue and MEpink had the strongest correlation with cluster 1. MEturquoise also had a high correlation with cluster 2.

We compared the three modules in relation to biological processes (Fig. 3D). The results show that MEmidnightblue is mainly associated with DNA damage stimulation, growth regulation, and regulation of embryonic development. MEpink is mainly related to metabolism, DNA replication, and epithelial cell proliferation. Moreover, MEturquoise is mainly associated with cell migration, lipid metabolism, and cell proliferation. In particular, MEturquoise is also related to the positive regulation of T cell proliferation and differentiation.

PPI network analysis and visualization of genes within the three modules MEmidnightblue, MEpink, and MEturquoise were performed and the most tightly connected subnetworks were obtained (Fig. 4A–C). Differentially expressed genes were compared between high and low tumor staging and between high and low grade. As shown in Fig. 4D and E, we obtained multiple differentially expressed genes, such as SDC1, GPC1, EGFR, CLDN3, INSR, PIK3C2A, LYN, HRAS, PCYT1A, and EPHX1.

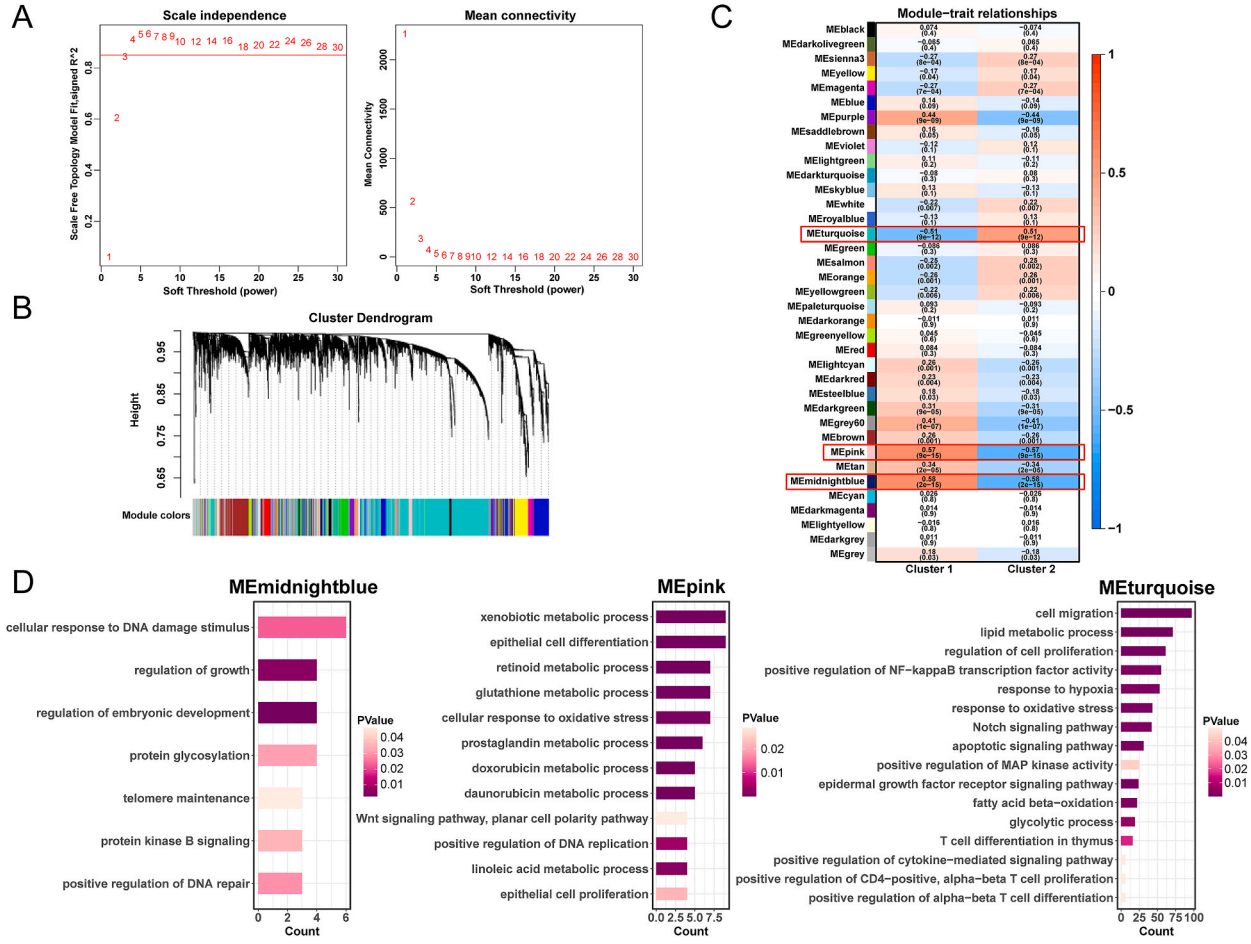


Fig. 3. Identification of gene modules of two ESCA subtypes based on CNV typing. (A) Selection of the optimal soft threshold; (B) Cluster dendrograms obtained based on WGCNA analysis; (C) Correlation between two subtypes of CNV typing and co-expressed gene modules; (D) Correlation between MEmidnightblue, MEpink, and MEturquoise and biological processes.

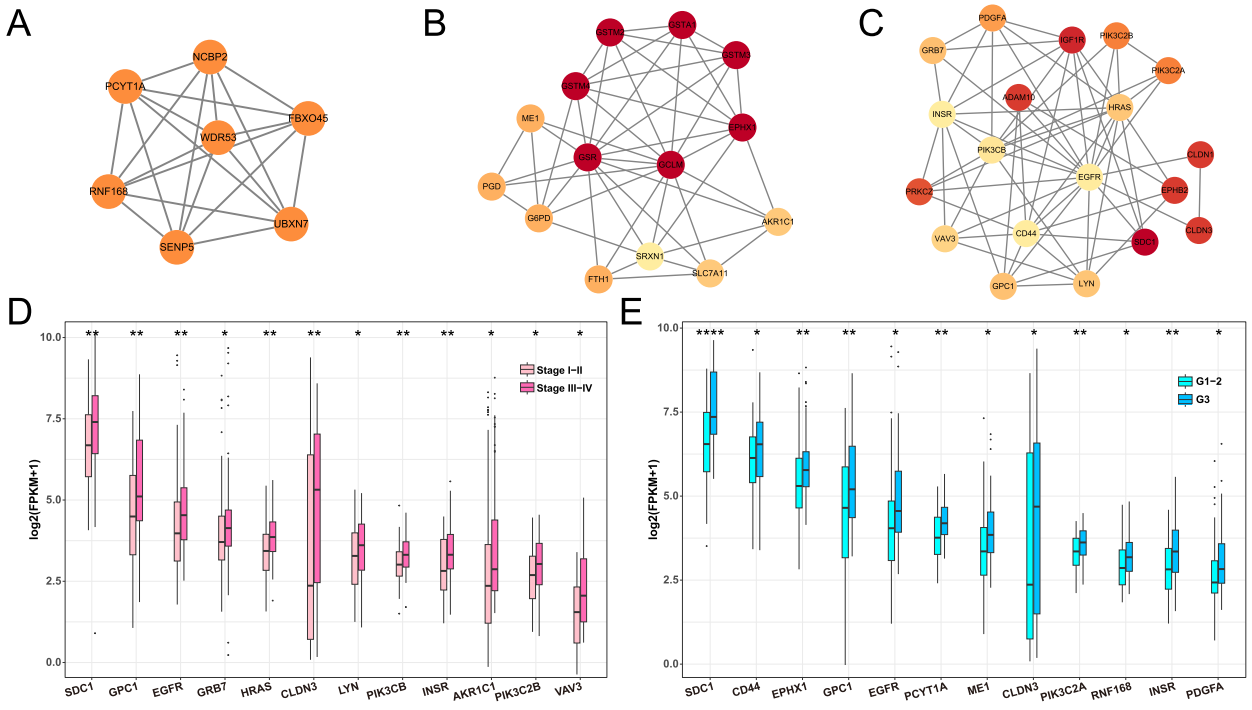


Fig. 4. Identification of key genes within the module. (A–C) PPI network diagram of MEmidnightblue, MEpink, and MEturquoise modular genes; (D) Hub gene expression between high and low tumor stages; (E) Hub gene expression between high and low tumor grades.

3.4. Established and assessment of a relevant prognostic model according to the CNV typing

To screen the hub genes in the modules, seven genes with prognostic relevance, including SDC1, GPC1, PIK3C2A, CLDN3, EGFR, INSR, and LYN, were screened from the three modules by multivariate Cox regression analysis. The formula was applied to calculate risk score for ESCA patients, as follows: Risk score = (0.8980*SDC1) + (−1.1230*GPC1) + (1.3897*PIK3C2A) + (−1.4157*CLDN3) + (0. 6079*EGFR) + (0.7930*INSR) + (0.9503*LYN).

As shown in Fig. 5A, overall survival for high-risk ESCA patients was more unfavorable in comparison to the low-risk group. In addition, Univariate Cox regression analysis demonstrated that stage (HR = 2.43, p = 1.49e-06), and risk score (HR = 1.108, p = 2.91e-09) predicted worse overall survival (Fig. 5B). Meanwhile, multivariate Cox regression analysis showed that risk score (HR = 1.148, p = 3.04e-10) and stage (HR = 2.184, p = 8.41e-05) independently influenced the prognosis of ESCA patients (Fig. 5C). We plotted ROC curves in combination with risk scores and other clinical characteristics to assess survival at 1, 3, and 5 years prognosis for ESCA patients, respectively. The area under the curve for the risk score was the highest in all cases (AUC = 0.673, 0.773, and 0.873, Fig. 5D–F). Our findings suggest that risk scores are more accurate than other clinical characteristics in assessing the prognostic survival of ESCA patients.

Based on the median risk score, ESCA patients were classified into different subgroups in terms of age, gender, grading and staging and into low-risk and high-risk groups. In each subgroup, the prognostic outcome of high-risk patients is worse (Fig. 5G–J). These results are illustrating that the risk scores we calculated are highly independent and are not susceptible to other clinical factors. To enhance the quantification of risk assessment and survival probabilities for ESCA patients, we developed a nomogram aimed at estimating OS at 1, 3, and 5 year(s), utilizing the risk score along with clinicopathological features. As illustrated in Fig. S4A, the risk score significantly influenced the OS of COAD patients. The calibration curves associated with this nomogram indicated a strong alignment between observed and predicted values (Fig. S4B). To assess the model's reliability, decision curve analysis (DCA) was generated, confirming that both the nomogram and risk score were most effective in forecasting prognosis (Fig. S4C).

3.5. Experimental verification of the reliability of the risk score model

As importance of method validation in the analysis of biomarkers [36], we examined the expression of SDC1, GPC1, PIK3C2A, CLDN3, EGFR, INSR, and LYN in HEEC, OE19, and TE-1 cell lines by qRT-PCR to further verify the reliability of the risk score model. The expressions of SDC1, PIK3C2A, EGFR, INSR and LYN were upregulated in OE19 and TE-1 cell lines (Fig. 6A–G). In contrast, GPC1 and CLDN3 expression was elevated in normal esophageal epithelial cells HEEC. We then selected the two molecules that contribute most to the risk model, PIK3C2A and GPC1, and suppressed them in esophageal cancer cell lines and normal esophageal cells, respectively. The results of live-dead cell staining demonstrated that the percentage of apoptosis in OE19 and TE-1 cell lines was elevated after inhibiting the expression of PIK3C2A (Fig. 7A–B, 7D-E), whereas the percentage of apoptosis in normal esophageal

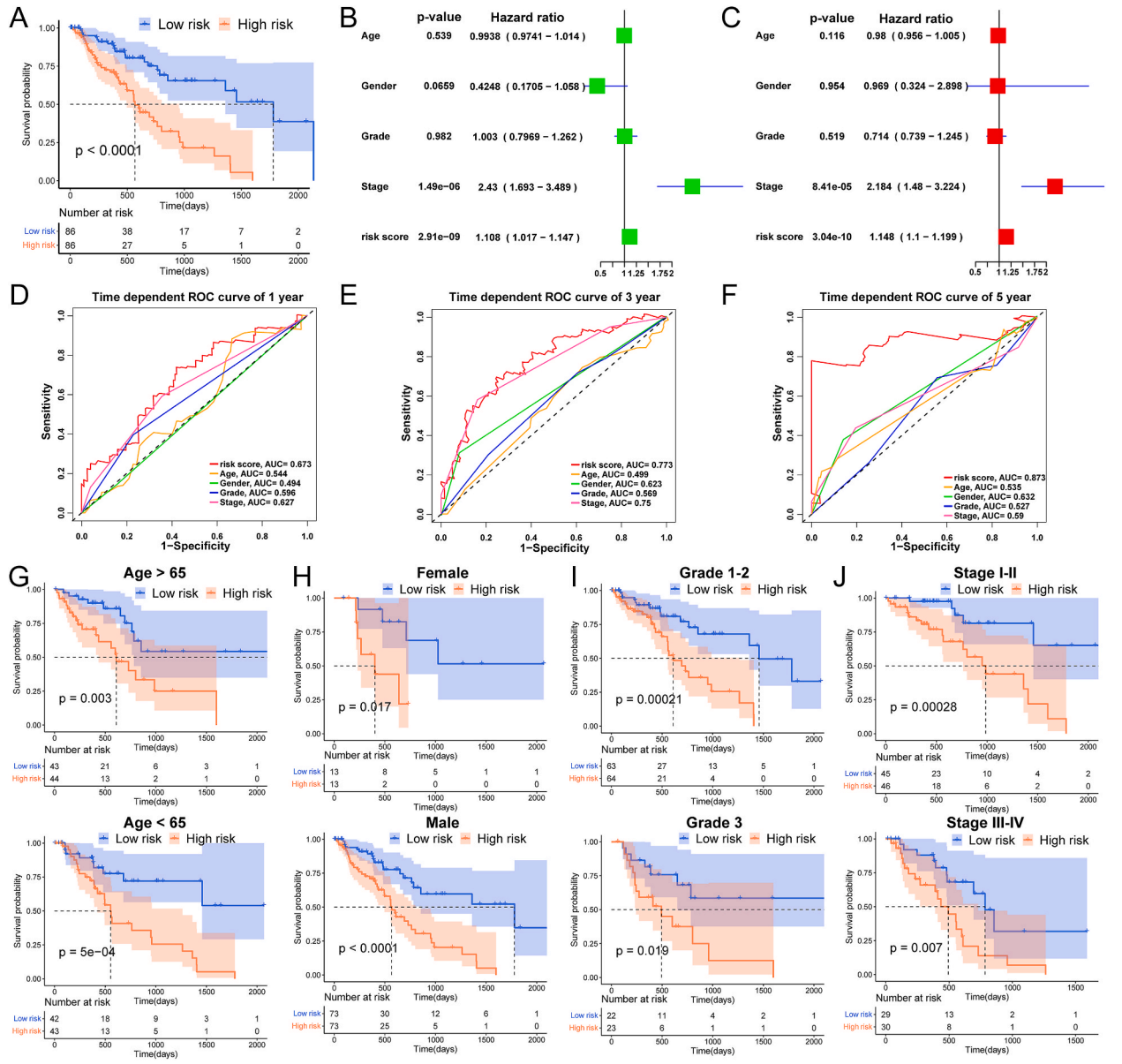


Fig. 5. Construction of a prognostic model for ESCA subtypes based on CNV typing. (A) Kaplan-Meier curves of risk score; (B–C) Univariate and multivariate Cox analysis of risk scores and clinical factors; (D–F) ROC curves of risk score and clinical factors for 1,3,5-year survival assessment, respectively; (G–J) Kaplan-Meier curves of risk scores at different ages, genders, grades and stages.

cancer cells HEEC was decreased after inhibiting the expression of GPC1 (Fig. 7C and F). The results for CCK8 were also consistent with live-dead cell staining. As shown in Fig. 7G–I, inhibition of PIK3C2A expression suppressed cell viability, whereas inhibition of GPC1 expression promoted cell viability. These findings facilitate the understanding of two prognostic markers, GPC1 and PIK3C2A.

3.6. Analysis of differences in energy metabolism

To further understand the relationship of energy metabolism on ESCA progression and survival, we analyzed differences in lipid metabolism, carbohydrate metabolism, and amino acid metabolism between different subtypes of ESCA and different risk groups. It was observed that cluster 1 was more active in glycolysis, lactate metabolism, and pentose phosphate-related pathways, and less active in the Citrate (TCA) cycle (Fig. 8A). In terms of lipid metabolism, cluster 1 was more active in the fatty acid synthesis pathway and less active in fatty acid metabolism (Fig. 8B). In addition, in terms of pathway scores related to amino acid metabolism, we found that cluster 1 was more capable of metabolizing glutathione, cysteine, methionine, and tryptophan, and less capable of arginine synthesis, as well as the degradation of valine, leucine, and isoleucine (Fig. 8C). Interestingly, on energy metabolism-related pathway scores, we

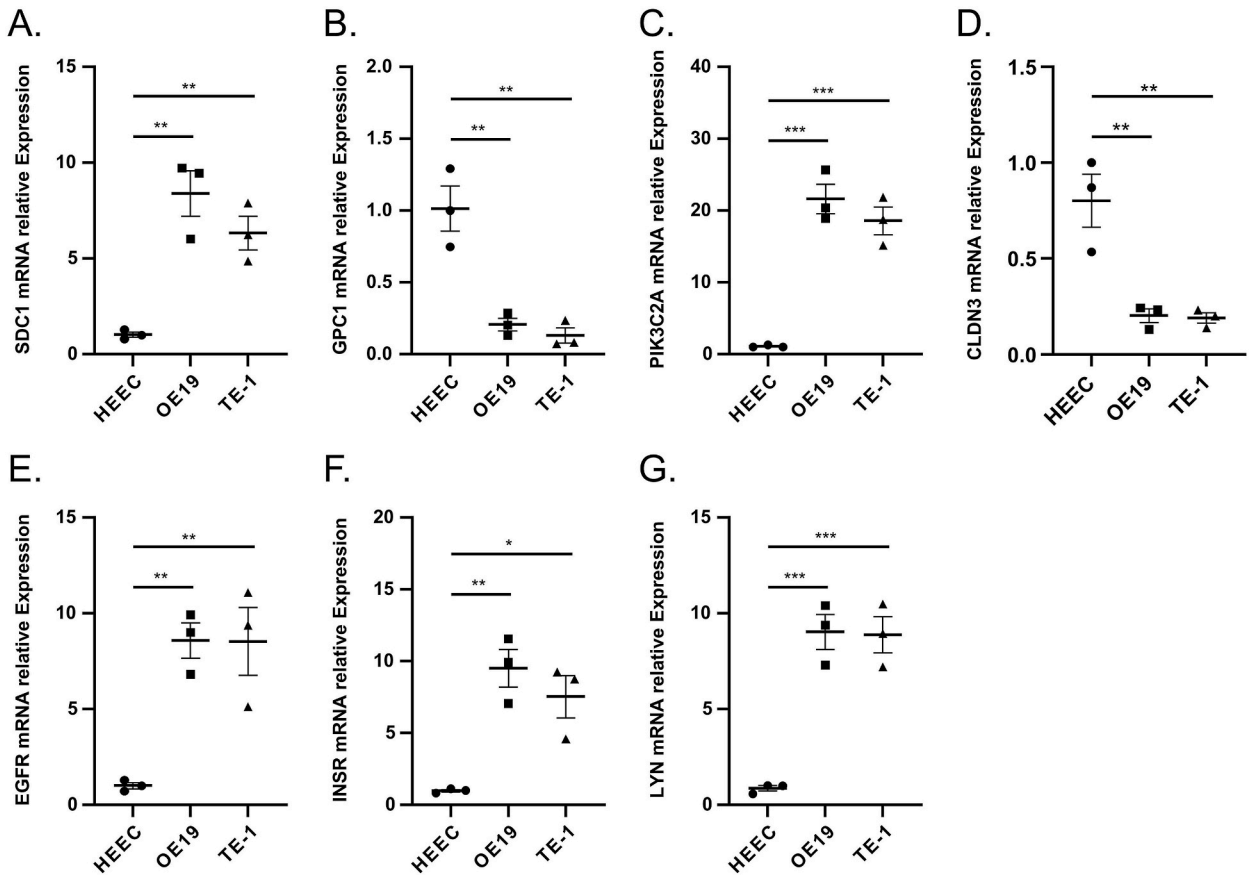


Fig. 6. qT-PCR validation of the reliability of the risk score model. (A–G) mRNA expression of SDC1, GPC1, PIK3C2A, CLDN3, EGFR, INSR, and LYN in normal esophageal epithelial cells HECC as well as in the esophageal cancer cell lines OE19 and TE-1. N = 3, * ≤0.05, ** ≤0.01, *** ≤0.001, **** ≤0.0001. The results are presented as mean ± SEM.

found similarities between different risk groups for ESCA and different subtypes of ESCA. The ESCA high-risk group showed initially higher scores on glycolysis, lactate metabolism and fatty acid synthesis, and significantly higher ability to metabolize glutathione, cysteine, methionine and tryptophan than the low-risk group. Furthermore, the high-risk patients had a weaker capacity for TCA cycling, fatty acid degradation, and arginine synthesis compared with the low-risk patients (Fig. 9).

To further assess the correlation between risk scores and energy metabolism enrichment scores in ESCA, we screened for KEGG pathways by screening for $p < 0.05$ and obtained the same results as above. The risk score was significantly positively related to glycolysis and lactate metabolism and significantly negatively related to the TCA cycle (Fig. 10A). In terms of lipid metabolism, the risk score was positively associated with fatty acid synthesis and negatively related to fatty acid degradation (Fig. 10B). Similarly, risk scores were positively correlated with glutathione, tryptophan, cysteine and methionine metabolism and negatively correlated with arginine biosynthesis (Fig. 10C).

3.7. Evaluation of differences in drug response

Cisplatin, Docetaxel, Paclitaxel, Gemcitabine, and Methotrexate are common chemotherapeutic drugs used in the treatment of ESCA [37–39]. For this purpose, we calculated the sensitivities of high-risk patients and low-risk patients to these five common chemotherapeutic agents. As shown in Fig. 11A, the IC_{50} values for Cisplatin, Docetaxel, and Paclitaxel were lower in the low-risk group compared to the high-risk score group. This indicates that low-risk ESCA patients are more sensitive to these three chemotherapeutic agents. However, Gemcitabine and Methotrexate showed no significant differences between the two groups.

The results of the TIDE score showed that high-risk patients had lower TIDE scores, indicating that high-risk ESCA patients had better immunotherapy outcomes and might be better treated with immune checkpoint inhibitors (Fig. 11B). Specifically, as shown in Fig. 11C, it was observed that multiple immune checkpoint genes, such as PDCD1, PD-1, HAVCR2, LILRB4, and TIGIT, were highly expressed in the high-risk group of patients. The above results suggested that ESCA patients in the high-risk group are more suitable for treatment with immune checkpoint inhibitors.

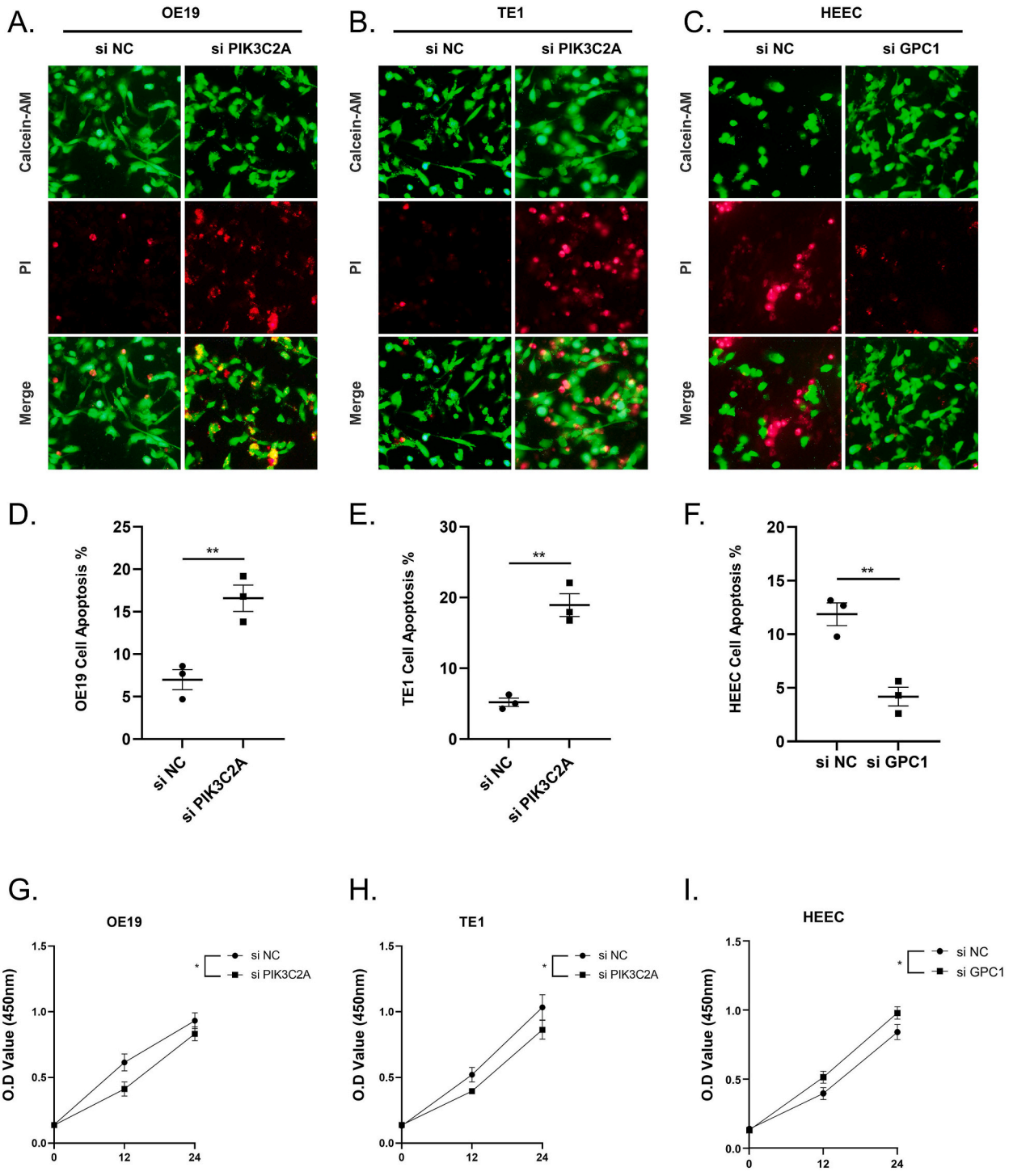
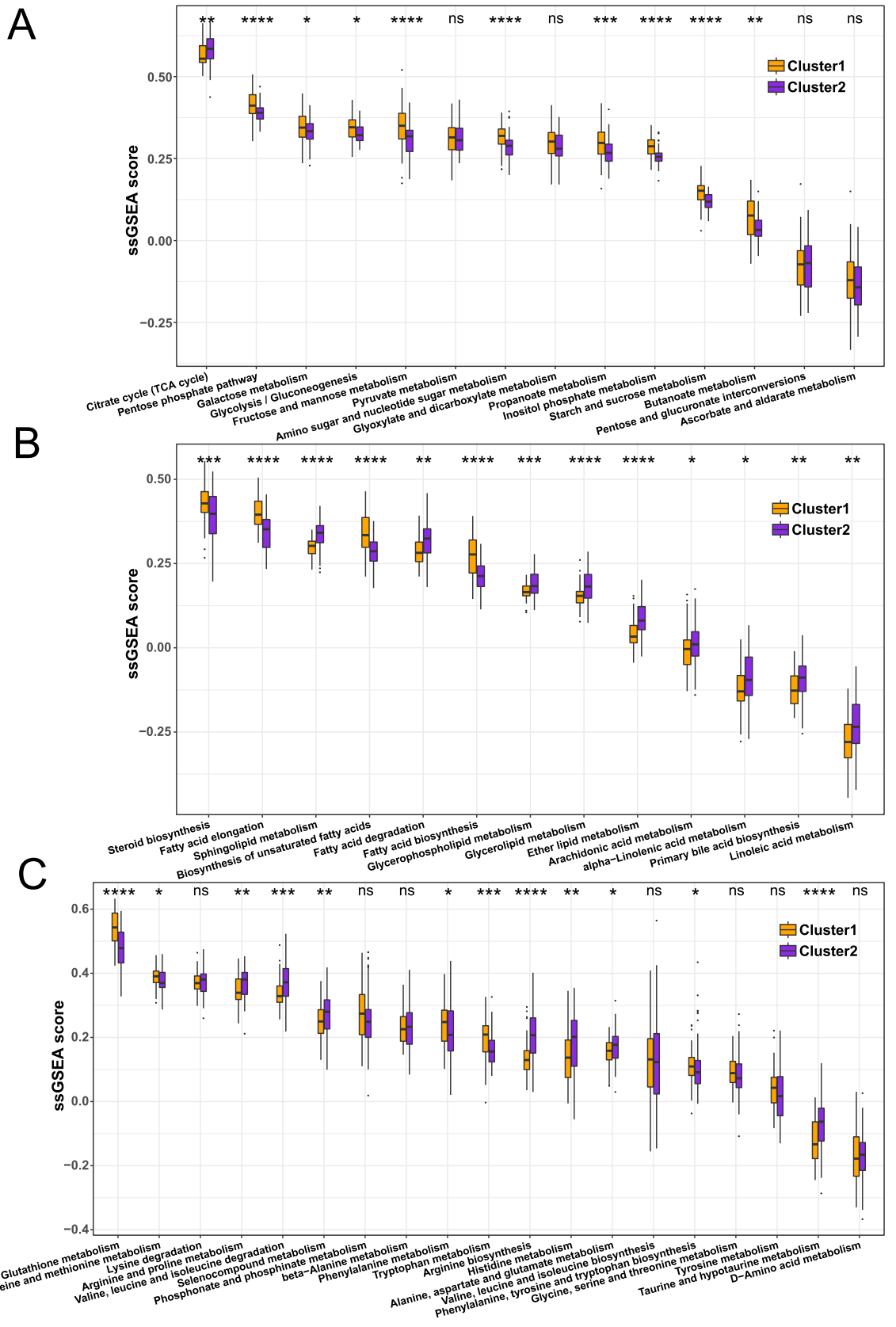


Fig. 7. Alterations in viability of HECC, OE19, and TE1 cell lines after inhibition of PIK3C2A or GPC1. [16] Representative live-dead cell staining images of PIK3C2A inhibition in OE19 and TE-1 cell lines; (C) Representative live-dead cell staining images of GPC1 inhibition in HEEC cell; (D–F) Statistical results on the percentage of apoptotic cells; (G–I) Alterations in cell viability following inhibition of PIK3C2A as well as GPC1 expression in OE19, TE1 and HEEC. N = 3, * ≤0.05, ** ≤0.01, *** ≤0.001, **** ≤0.0001. The results are presented as mean ± SEM.



(caption on next page)

Fig. 8. Differences in energy metabolic pathway enrichment scores between ESCA subtypes. Comparison of Cluster 1 and Cluster 2 enrichment scores in pathways related to carbohydrate metabolism (A), lipid metabolism (B), and amino acid metabolism (C).

4. Discussion

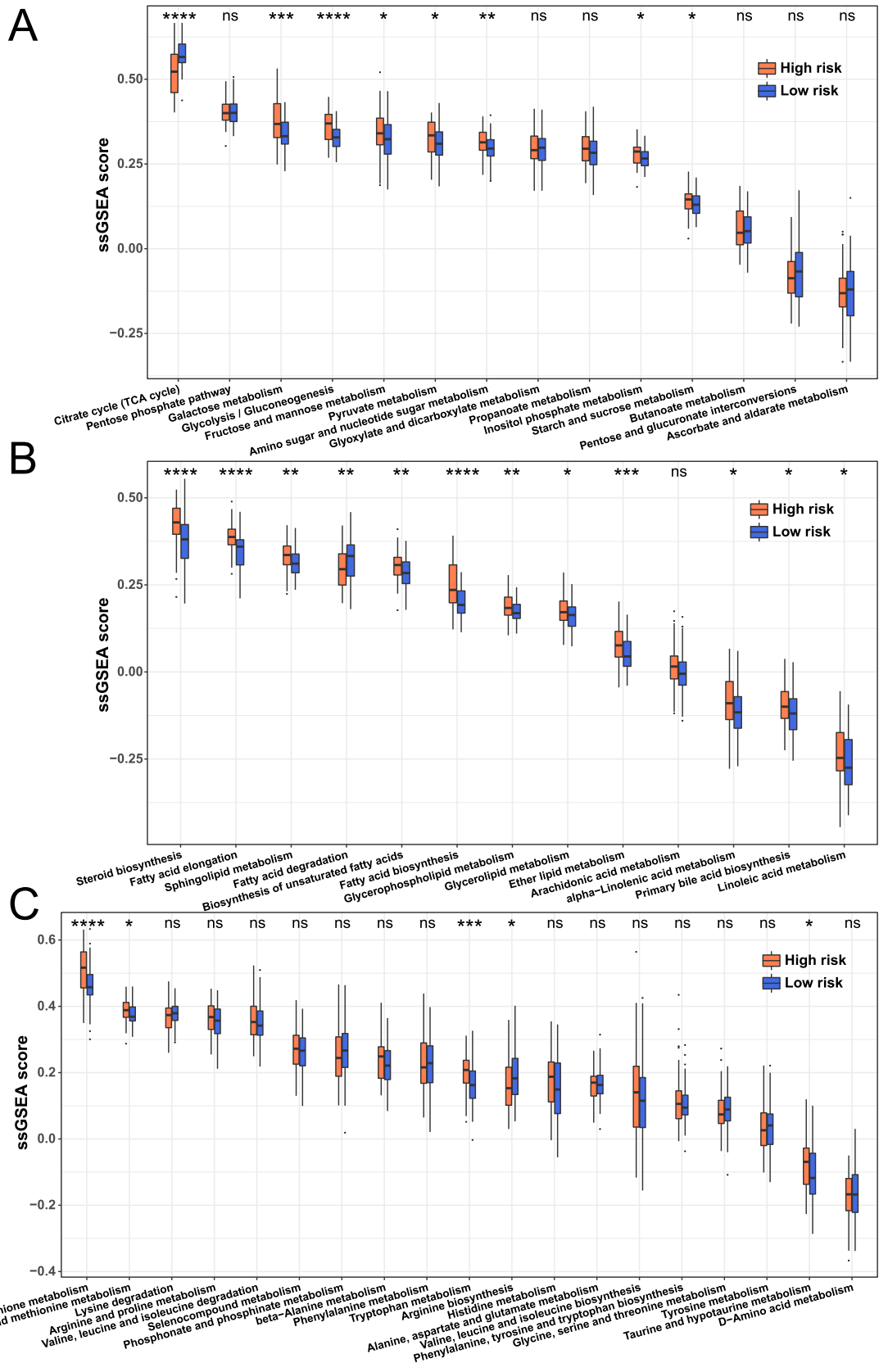
Accurate cancer typing and classification reliably predicts disease progression and is essential for improving diagnosis and predicting treatment response [40,41]. Genetic variation captured from CNV data is considered to be the most stable, and it can contribute to the generation of powerful biomarkers [42,43]. Recent studies indicated that CNVs play a functional role in a variety of tumorigenesis and suggest that typing of cancers based on CNV can provide new biomarkers as a diagnostic tool for cancer [31,39,44]. This study comprehensively investigated the effects of CNV values in genomic amplification and deletion based regions on ESCA staging, prognosis, tumor microenvironment, and immunotherapy response by bioinformatics analysis.

The samples were classified into two clusters using consistency clustering according to the amplification and deletion of CNV values in the TCGA-ESCA genome region. The two clusters have significantly different patterns of amplification and deletion within genomic regions, and survival analyses reflect differences in survival times in different pattern states (Fig. 1B and C). There is strong evidence that differences in CNVs based on amplification and deletion within genomic regions are distinctive features of ESCA.

Aneuploidy is the abnormal acquisition or deletion of chromosomes in cells with critical functions in the malignancy of cancer cells, cancer metastasis and tumor recurrence [45,46]. Spurr et al. suggested the use of tumor aneuploidy as a therapeutic target and biomarker, making it possible to personalize therapy for cancer patients treated with radiotherapy and immune checkpoint blockade. In the present study, we found higher CNV values in the amplified regions of chromosome 3q26.2 and 3q28 in cluster 1 (Fig. S2A). Chromosome arm 3q has been shown to be a critical region for genomic amplification in various cancers, including squamous lung cancer, and cervical cancer [47,48]. In particular, Bochen et al. found that the 3q26-encoded genes SOX2 and SEC62 are overexpressed in a variety of cancers [49]. In addition, cluster 1 occurs with significant deletions at 3p14.3, 3p14.2, 3p26.2, 3p25.3 and 9p21.3. Among them, deletion of 3p is considered to be one of the most widely detected genetic alterations in ESCC, and one or more tumor-suppressive genes may be present in the region where it is frequently missing [50]. By using RealSeqS technology, Douville et al. demonstrated that aneuploidy is a tool capable of assessing the progression of Barrett's esophagus and that aneuploidy progression is associated with deletion of chromosome 9p [51]. Our study also found a significant deletion of 9p in cluster 1 and a higher proportion of Barrett's esophagus in cluster 1. This suggests that in cluster 1, deletion of 9p is closely involved in the development of Barrett's esophagus. Taken together, changes in these gene regions may lead to dysregulation of gene expression, which may affect the onset and development of ESCA. Subsequently, we identified three modules with relevance to CNV subtypes and identified hub genes by constructing co-expression networks. Using multivariate Cox regression analysis, seven genes, including SDC1, GPC1, PIK3C2A, CLDN3, EGFR, INSR, and LYN, were screened for prognostic association and were not influenced by other clinical factors, and a risk model was constructed. GPC1 was specifically accumulated in cancer-cell-derived exosomes which circulate in the blood [52]. This finding indicates that we could distinguish early stages of ESCA patients through testing the levels of GPC1 in the serum. CLDN3 belonging to claudin families, functioning crucially in maintaining the physical barrier function between cells and participating in molecular cell transmission [53]. Based on this background, Che et al. found that overexpressed CLDN3 could repress the epithelial-mesenchymal transition (EMT) and invasion of lung squamous cell carcinoma cells. Lyn is reported to involve in inflammation and tumorigenesis and development such as cancer proliferation, migration as well as apoptosis [54]. EGFR is considered a driving factor for tumor oncogenesis [55] and is increasingly recognized as a biomarker of tumor resistance, as its amplification or secondary mutations have been found to occur under drug pressure [56]. To our knowledge, GPC1, CLDN3, LYN and EGFR are related to the occurrence of ESCA or to be significant predictors of overall survival in ESCA patients [57–60]. Although remained three genes, SDC1, PIK3C2A, and INSR, have not been previously detected to be implicated in ESCA prognosis, several studies have demonstrated that these genes can predict prognosis in a variety of cancers. It has been shown that PIK3C2A promotes tumor survival [61]. Similarly, Qin et al. showed that aberrant expression level of PIK3C2A can affect the biological activities of renal clear cell carcinoma [62]. This is the same as our findings, where we found that PIK3C2A was high-expressed when in ESCA cells and its cell viability was significantly reduced by inhibiting PIK3C2A expression (Fig. 6; Fig. 7). Liao et al. found that SDC1 has different roles in different types of cancers, and it was shown to be a marker of epithelial-to-mesenchymal transition that can be used to predict tumor prognosis [63]. In addition, INSR is overexpressed in various cancers and affects tumor invasiveness through regulating E-cadherin glycosylation [64]. These results indicated the potential of these 7 genes to serve as new biomarkers in ESCA diagnosis and treatment.

This study verified the function of CNV-related genes in ESCA for the first time. High-risk and low-risk ESCA patients were classified by the risk score value. The KM curve demonstrated considerably lower overall survival for high-risk patients (Fig. 5). The ROC curve confirmed risk score as an independent factor affecting ESCA prognosis, with stronger diagnostic impact than other clinical factors. The AUC value of the risk score developed by Zhang et al. developed was higher than gender (0.497), age (0.595), stage (0.481). In the present study, our prognostic models based on CNV constructed for the relevant genes had the highest AUC values for both risk scores, 0.673, 0.773, and 0.873, respectively. This does not seem to be a poor predictive power of our prognostic models compared with previous studies.

Cellular metabolism is recognized as a complex process where metabolites can be generated through interrelated pathways. Importantly, metabolic abnormalities are recognized as markers of cancer [65]. Glucose, fatty acids and some amino acids (including glutamate and glycine) have been shown to be associated with metabolic pathways in esophageal cancer [66–68]. These pathways may be involved in energy metabolism and the tricarboxylic acid cycle in ESCA, and the identification of relevant metabolic markers has also been used to differentiate between patients with ESCA and healthy populations [66–69]. Cancer tissues are able to synthesize the



(caption on next page)

Fig. 9. Differences in energy metabolism-related pathway enrichment scores between ESCA risk groups. Comparison of enrichment scores in pathways related to carbohydrate metabolism (A), lipid metabolism (B), and amino acid metabolism (C) between high- and low-risk groups.

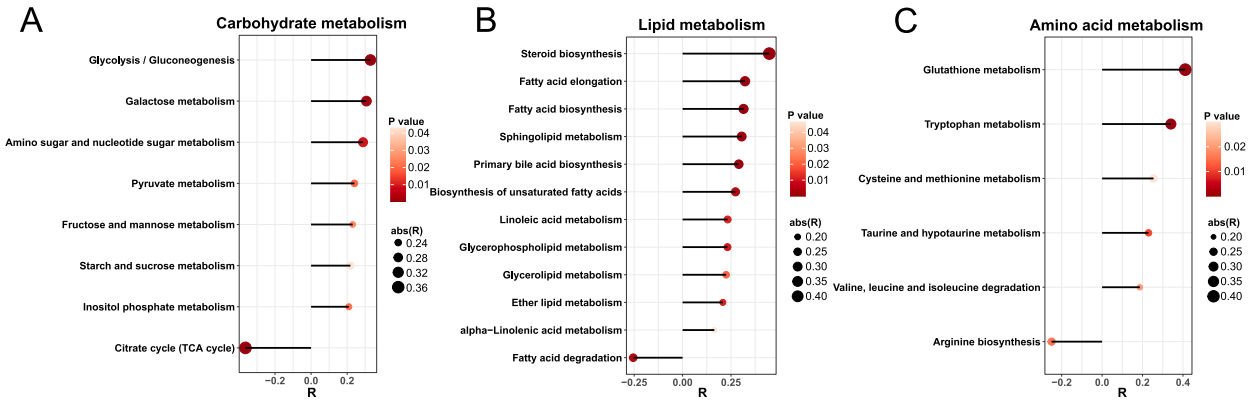


Fig. 10. Correlation analysis between risk scores and energy metabolism-related pathway enrichment scores. Correlation of risk scores with carbohydrate metabolism (A), lipid metabolism (B), and amino acid metabolism (C) pathway enrichment scores, respectively. The horizontal coordinate represents the correlation coefficient, the circle size represents the absolute value of the correlation coefficient, and the color represents the p-value.

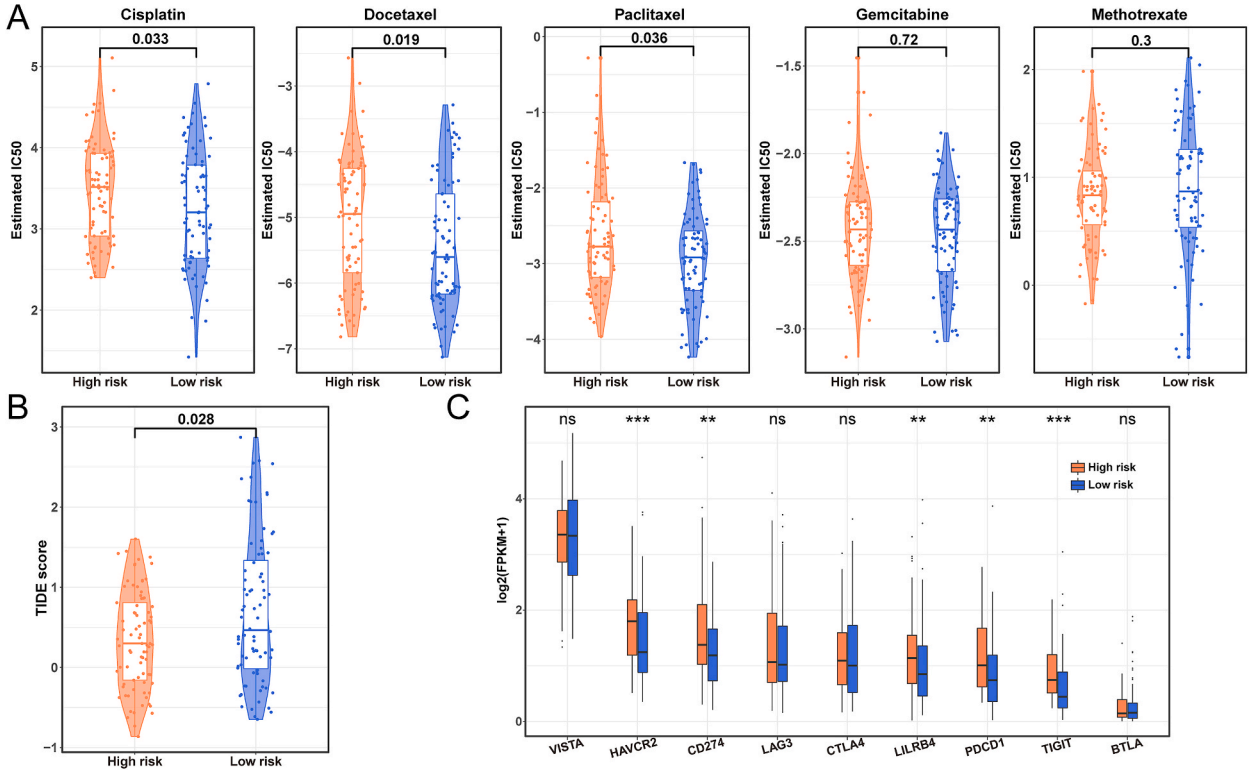


Fig. 11. Differences in drug response between high- and low-risk groups of ESCA patients. (A) Analysis of the differences in sensitivity to common chemotherapeutic drugs between ESCA patients in high and low risk groups; (B) Comparison of differences in TIDE scores; (C) Analysis of differences in immune checkpoint gene expression within ESCA patients in high and low risk groups.

energy-carrying molecule adenosine triphosphate (ATP) and a variety of key precursor molecules such as proteins, lipids, and nucleic acids through glycolysis [70,71]. This study found that cluster 1 and ESCA high-risk groups were predominantly enriched in the glycolytic and pentose phosphate (PPP) pathways. Notably, the glycolytic pathway is able to promote tumor formation and proliferation by consuming only low levels of oxygen and keep acidic environment [72,73]. Based on this finding, a new study idea for targeting the bicarbonate transporter SLC4A4, which could lead to acidosis in the tumor environment, could be a therapeutic target for

treating patients suffering from ESCA [74]. In addition, the PPP pathway is capable of providing precursors for nucleotide and amino acid biosynthesis, as well as promoting the nicotinamide adenine dinucleotide phosphate (NADPH) production [75,76]. This demonstrates that CNV may regulate the energy metabolic pathways of tumor cells, which in turn would promote the growth and proliferation of ESCA cells.

The TIDE score reflects the sensitivity to immunotherapy in the high- and low-risk groups in ESCA. The low TIDE score in the high-risk group suggests that high-risk ESCA patients could benefit more from immunotherapy. Immune checkpoint inhibitors (ICIs) are regarded as the most effective treatments to improve the prognosis of ESCA patients. Among them, anti-PD-1 and anti-PD-L1 antibodies have been shown to have a better therapeutic effect on cancer [77]. Programmed cell death protein 1 (PDCD1, PD-1) blocks T-cell co-stimulatory signaling by binding to the ligands PD-L1 (CD274) and PD-L2 (PDCD1LG2), and the strategy of blocking these T cell-inhibitory molecules has been proved to successfully release the patient's own T-cell immunity, which in turn fights tumor antigens in cancer [78]. The current research observed that the high-risk score was highly expressed in various immune checkpoint genes, including PDCD1 and CD274. These results not only suggest that CNV may influence ESCA development by modulating immune components in the tumor microenvironment, but also confirm that patients in the high-risk group of ESCA were more suitable for ICIs treatment to be able to achieve a better therapeutic outcome.

However, our study has some limitations. First, the current results were all derived from bioinformatics analyses based on public databases, and the clinical information was prone to bias. Our future study will combine clinical data from multiple centers and data from different databases to increase sample diversity and accuracy. Second, the results of the bioinformatics analysis of this study were validated by in vivo experiments to verify the function of specific genes. Finally, the specific molecular mechanisms of the model in ESCA need to be investigated. Large-scale clinical trials should be performed to de-humanize the correlation between ESCA prognosis and CNVs.

5. Conclusion

In summary, this study classified ESCA into two subtypes based on the CNV values of genomic amplification and deletion regions. Significant differences in overall survival, immune cell infiltration, associated biological processes and suitability for immunotherapy in different subtypes of patients were observed. A risk score model was established to identify patients at high risk of ESCA and to explore the enrichment of energy metabolism pathways in different clusters and different risk groups. The TIDE algorithm and drug sensitivity were combined to assess patients at high risk of ESCA as more suitable for treatment with ICIs. Our study provided new insights into the pathogenesis of cancer based on CNV, providing a guidance for the clinical diagnosis and prognostic treatment of ESCA.

Funding

This work was financially supported by National Natural Science Foundation of China (32200539).

Data availability statement

The datasets generated during and analyzed during the present research are available from the corresponding author upon reasonable request.

Ethics approval and consent to participate

Not applicable.

CRediT authorship contribution statement

Chao Zhao: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Data curation, Conceptualization. **Hui Han:** Visualization, Software, Resources, Project administration, Methodology. **Yushuang Tian:** Software, Methodology, Investigation, Formal analysis, Data curation. **Guangjin Qu:** Visualization, Supervision, Software, Resources. **Yingying Xu:** Visualization, Supervision, Resources, Methodology, Formal analysis. **Yihan Wang:** Writing – original draft, Visualization, Validation, Project administration, Methodology, Funding acquisition. **Lili Shi:** Writing – review & editing, Visualization, Validation, Supervision, Software, Methodology, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e38011.

Abbreviations

ROC	receiver operating characteristic analysis
ssGSEA	single-sample gene set enrichment analysis
TCGA	The Cancer Genome Atlas
WGCNA	Weighted Gene Co-Expression Network Analysis
AUC	area under ROC curve
TIDE	The Tumor Immune Dysfunction and Exclusion
ESCA	esophageal carcinoma
MCODE	Molecular Complex Detection
ESCC	esophageal squamous cell carcinoma
LASSO	least absolute shrinkage and selection operator
FDR	false discovery rate
GRED	gastroesophageal reflux disease
INSR	Insulin receptor
OS	overall survival
HR	hazard ratio
PPI	protein-protein interaction network
CDF	cumulative distribution function
TIDE	Tumor Immune Dysfunction and Exclusion
STRING	Search Tool for the Retrieval of Interacting Genes

References

[1] H. Luo, Y. Sun, L. Wang, H. Liu, R. Zhao, M. Song, et al., Targeting endoplasmic reticulum associated degradation pathway combined with radiotherapy enhances the immunogenicity of esophageal cancer cells, Cancer Biol. Ther. 24 (1) (2023) 2166763.

[2] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA A Cancer J. Clin. 71 (3) (2021) 209–249.

[3] L. Qi, Z. Li, Exploring the interplay between iron metabolism imbalance and esophageal cancer, Oncologie 26 (4) (2024) 509–523.

[4] P.S. Plum, A.H. Hölscher, K. Pacheco Godoy, H. Schmidt, F. Berlth, S.H. Chon, et al., Prognosis of patients with superficial T1 esophageal cancer who underwent endoscopic resection before esophagectomy-A propensity score-matched comparison, Surg. Endosc. 32 (9) (2018) 3972–3980.

[5] V. Oppedijk, A. van der Gaast, J.J. van Lanschot, P. van Hagen, R. van Os, C.M. van Rij, et al., Patterns of recurrence after surgery alone versus preoperative chemoradiotherapy and surgery in the CROSS trials, J. Clin. Oncol. : official journal of the American Society of Clinical Oncology 32 (5) (2014) 385–391.

[6] Y. Liu, Perioperative immunotherapy for esophageal squamous cell carcinoma: now and future, World J. Gastroenterol. 29 (34) (2023) 5020–5037.

[7] G. Chen, Z. Wang, X.Y. Liu, F.Y. Liu, Recurrence pattern of squamous cell carcinoma in the middle thoracic esophagus after modified Ivor-Lewis esophagectomy, World J. Surg. 31 (5) (2007) 1107–1114.

[8] B. Trost, S. Walker, Z. Wang, B. Thiruvahindrapuram, J.R. MacDonald, W.W.L. Sung, et al., A comprehensive workflow for read depth-based identification of copy-number variation from whole-genome sequence data, Am. J. Hum. Genet. 102 (1) (2018) 142–155.

[9] M. Qiu, W. Xia, R. Chen, S. Wang, Y. Xu, Z. Ma, et al., The circular RNA circPRKCI promotes tumor growth in lung adenocarcinoma, Cancer Res. 78 (11) (2018) 2839–2851.

[10] X.-J. Zhou, Y.-W. Tian, R.-H. Li, The whole-genome survey of Acer griseum, its polymorphic simple sequence repeats development and application, Biocell 47 (8) (2023) 1907–19013.

[11] Y. Sun, N. Shi, H. Lu, J. Zhang, Y. Ma, Y. Qiao, et al., ABCC4 copy number variation is associated with susceptibility to esophageal squamous cell carcinoma, Carcinogenesis 35 (9) (2014) 1941–1950.

[12] L. Hu, Y. Wu, X. Guan, Y. Liang, X. Yao, D. Tan, et al., Germline copy number loss of UGT2B28 and gain of PLEC contribute to increased human esophageal squamous cell carcinoma risk in Southwest China, Am. J. Cancer Res. 5 (10) (2015) 3056–3071.

[13] Y.B. Chen, W.H. Jia, A comprehensive genomic characterization of esophageal squamous cell carcinoma: from prognostic analysis to in vivo assay, Chin. J. Cancer 35 (1) (2016) 76.

[14] Z. Liu, R. Su, A. Ahsan, C. Liu, X. Liao, D. Tian, et al., Esophageal squamous cancer from 4NQO-induced mice model: CNV alterations, Int. J. Mol. Sci. 23 (22) (2022).

[15] W. Ahmed, M.F.A. Malik, M. Saeed, F. Haq, Copy number profiling of Oncotype DX genes reveals association with survival of breast cancer patients, Mol. Biol. Rep. 45 (6) (2018) 2185–2192.

[16] S. Rosenberg, F. Ducray, A. Alentorn, C. Dehais, N. Elarouci, A. Kamoun, et al., Machine learning for better prognostic stratification and driver gene identification using somatic copy number variations in anaplastic oligodendroglioma, Oncol. 23 (12) (2018) 1500–1510.

[17] A. Colaprico, T.C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, et al., TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data, Nucleic Acids Res. 44 (8) (2016) e71.

[18] C.H. Mermel, S.E. Schumacher, B. Hill, M.L. Meyerson, R. Beroukhim, G. Getz, GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers, Genome Biol. 12 (4) (2011) R41.

[19] M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, Bioinformatics 26 (12) (2010) 1572–1573.

[20] P. Charoentong, F. Finotello, M. Angelova, C. Mayer, M. Efremova, D. Rieder, et al., Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade, Cell Rep. 18 (1) (2017) 248–262.

[21] S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7.

[22] K. Yoshihara, M. Shahmoradgoli, E. Martinez, R. Vegesna, H. Kim, W. Torres-Garcia, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (2013) 2612.

[23] E. Becht, N.A. Giraldo, L. Lacroix, B. Buttard, N. Elarouci, F. Petitprez, et al., Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression, Genome Biol. 17 (1) (2016) 218.

[24] B. Chen, M.S. Khodadoust, C.L. Liu, A.M. Newman, A.A. Alizadeh, Profiling tumor infiltrating immune cells with CIBERSORT, Methods Mol. Biol. 1711 (2018) 243–259.

[25] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559.
[26] Z. Song, J. Yu, M. Wang, W. Shen, C. Wang, T. Lu, et al., CHDTEPDB: transcriptome expression profile database and interactive analysis platform for congenital heart disease, Congenit. Heart Dis. 18 (6) (2023) 693–701.
[27] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, et al., STRING v9.1: protein-protein interaction networks, with increased coverage and integration, Nucleic Acids Res. 41 (D1) (2012) D808–D815.
[28] M.E. Smoot, K. Ono, J. Ruscheinski, P.L. Wang, T. Ideker, Cytoscape 2.8: new features for data integration and network visualization, Bioinformatics 27 (3) (2011) 431–432.
[29] W.P. Bandettini, P. Kellman, C. Mancini, O.J. Booker, S. Vasu, S.W. Leung, et al., MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study, J. Cardiovasc. Magn. Reson. 14 (1) (2012) 83.
[30] N. Simon, J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for cox's proportional hazards model via coordinate descent, J. Stat. Software 39 (5) (2011) 1–13.
[31] Cancer Genome Atlas Research N, Analysis Working Group, U. Asan, B.C.C. Agency, Women's H. Brigham, I. Broad, et al., Integrated genomic characterization of oesophageal carcinoma, Nature 541 (7636) (2017) 169–175.
[32] P. Geeleher, N. Cox, R.S. Huang, pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels, PLoS One 9 (9) (2014) e107468.
[33] P. Jiang, S. Gu, D. Pan, J. Fu, A. Sahu, X. Hu, et al., Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response, Nat. Med. 24 (10) (2018) 1550–1558.
[34] W. Yu, F. Liu, Q. Lei, P. Wu, L. Yang, Y. Zhang, Identification of key pathways and genes related to immunotherapy resistance of LUAD based on WGCNA analysis, Front. Oncol. 11 (2021) 814014.
[35] S. Blaine-Sauer, T.L. Samuels, K. Yan, N. Johnston, The protease inhibitor amprenavir protects against pepsin-induced esophageal epithelial barrier disruption and cancer-associated changes, Int. J. Mol. Sci. 24 (7) (2023).
[36] B. Seyfinejad, A. Jouyban, Importance of method validation in the analysis of biomarker, Curr. Pharmaceut. Anal. 18 (6) (2022) 567–569.
[37] K. Nakamura, K. Kato, H. Igaki, Y. Ito, J. Mizusawa, N. Ando, et al., Three-arm phase III trial comparing cisplatin plus 5-FU (CF) versus docetaxel, cisplatin plus 5-FU (DCF) versus radiotherapy with CF (CF-RT) as preoperative therapy for locally advanced esophageal cancer (JCOG1109, NExT study), Jpn. J. Clin. Oncol. 43 (7) (2013) 752–755.
[38] H. Wu, S. Chen, J. Yu, Y. Li, X.Y. Zhang, L. Yang, et al., Single-cell transcriptome analyses reveal molecular signals to intrinsic and acquired Paclitaxel resistance in esophageal squamous cancer cells, Cancer Lett. 420 (2018) 156–167.
[39] Y. Zhang, L. Chen, G.Q. Hu, N. Zhang, X.D. Zhu, K.Y. Yang, et al., Gemcitabine and cisplatin induction chemotherapy in nasopharyngeal carcinoma, N. Engl. J. Med. 381 (12) (2019) 1124–1135.
[40] C. Ji, Y. He, Y. Wang, Identification of necroptosis subtypes and development of necroptosis-related risk score model for in ovarian cancer, Front. Genet. 13 (2022) 1043870.
[41] C.-C. Sim, E.U.-H. Sim, C.-W. Lee, K. Narayanan, Multigenic prognosis assessment model for nasopharyngeal carcinoma via a modified meta-analysis approach, Oncologie 25 (4) (2023) 355–365.
[42] S. Rajpal, A. Rajpal, M. Agarwal, V. Kumar, A. Abraham, D. Khanna, et al., XAI-CNVMarker: explainable AI-based copy number variant biomarker discovery for breast cancer subtypes, Biomed. Signal Process Control 84 (2023).
[43] M. Treccani, E. Locatelli, C. Patuzzo, G. Malerba, A broad overview of genotype imputation: standard guidelines, approaches, and future investigations in genomic association studies, Biocell 47 (6) (2023) 1225–1241.
[44] W. Liu, J.M. Snell, W.R. Jeck, K.A. Hoadley, M.D. Wilkerson, J.S. Parker, et al., Subtyping sub-Saharan esophageal squamous cell carcinoma by comprehensive molecular analysis, JCI Insight 1 (16) (2016).
[45] U. Ben-David, A. Amon, Context is everything: aneuploidy in cancer, Nat. Rev. Genet. 21 (1) (2019) 44–62.
[46] M. Li, F. Gao, X. Ren, G. Dong, H. Chen, A.Y. Lin, et al., Nonhematogenic circulating aneuploid cells confer inferior prognosis and therapeutic resistance in gliomas, Cancer Sci. 113 (10) (2022) 3535–3546.
[47] J. Qian, Y. Zou, J. Wang, B. Zhang, P.P. Massion, Global gene expression profiling reveals a suppressed immune response pathway associated with 3q amplification in squamous carcinoma of the lung, Genom Data 5 (2015) 272–274.
[48] M. Linxweiler, F. Bochen, B. Schick, S. Wemmert, B. Al Kadah, M. Greiner, et al., Identification of SEC62 as a potential marker for 3q amplification and cellular migration in dysplastic cervical lesions, BMC Cancer 16 (1) (2016) 676.
[49] F. Bochen, H. Adisurya, S. Wemmert, C. Lerner, M. Greiner, R. Zimmermann, et al., Effect of 3q oncogenes SEC62 and SOX2 on lymphatic metastasis and clinical outcome of head and neck squamous cell carcinomas, Oncotarget 8 (3) (2017) 4922–4934.
[50] Y. Li, C.L. Zhu, C.J. Nie, J.C. Li, T.T. Zeng, J. Zhou, et al., Investigation of tumor suppressing function of CACNA2D3 in esophageal squamous cell carcinoma, PLoS One 8 (4) (2013) e60027.
[51] C. Douville, H.R. Moinova, P.N. Thota, N.J. Shaheen, P.G. Iyer, M.I. Canto, et al., Massively parallel sequencing of esophageal brushings enables an aneuploidy-based classification of patients with Barrett's esophagus, Gastroenterology 160 (6) (2021) 2043–20454 e2.
[52] S.A. Melo, L.B. Luecke, C. Kahlert, A.F. Fernandez, S.T. Gammon, J. Kaye, et al., Glypican-1 identifies cancer exosomes and detects early pancreatic cancer, Nature 523 (7559) (2015) 177–182.
[53] M.J. Orea, J.C. Angulo, A. González-Corpas, D. Echegaray, M. Marvá, M.V.T. Lobo, et al., Claudin-3 loss of expression is a prognostic marker in castration-resistant prostate cancer, Int. J. Mol. Sci. 24 (1) (2023).
[54] Y. Sun, Y. Yang, Y. Zhao, X. Li, Y. Zhang, Z. Liu, The role of the tyrosine kinase Lyn in allergy and cancer, Mol. Immunol. 131 (2021) 121–126.
[55] X. Li, L. Zhao, C. Chen, J. Nie, B. Jiao, Can EGFR be a therapeutic target in breast cancer? Biochim. Biophys. Acta Rev. Canc 1877 (5) (2022) 188789.
[56] P.A. Jänne, C. Baik, W.C. Su, M.L. Johnson, H. Hayashi, M. Nishio, et al., Efficacy and safety of patritumab deruxtecan (HER3-DXd) in EGFR inhibitor-resistant, EGFR-mutated non-small cell lung cancer, Cancer Discov. 12 (1) (2022) 74–89.
[57] E. Munekage, S. Serada, S. Tsujii, K. Yokota, K. Kiuchi, K. Tominaga, et al., A glypican-1-targeted antibody-drug conjugate exhibits potent tumor growth inhibition in glypican-1-positive pancreatic cancer and esophageal squamous cell carcinoma, Neoplasia 23 (9) (2021) 939–950.
[58] M. Ma, Y. Chen, X. Chong, F. Jiang, J. Gao, L. Shen, et al., Integrative analysis of genomic, epigenomic and transcriptomic data identified molecular subtypes of esophageal carcinoma, Aging (Albany NY) 13 (5) (2021) 6999–7019.
[59] D.L.Y.N. Liu, A key gene from bioinformatics analysis, contributes to development and progression of esophageal adenocarcinoma, Med Sci Monit Basic Res 21 (2015) 253–261.
[60] L. Spohn, C. Fichter, M. Werner, S. Lassmann, Subcellular localization of EGFR in esophageal carcinoma cell lines, J Cell Commun Signal 10 (1) (2016) 41–47.
[61] S.K. Ng, S.Y. Neo, Y.W. Yap, R.K. Karuturi, E.S. Loh, K.H. Liau, et al., Ablation of phosphoinositide-3-kinase class II alpha suppresses hepatoma cell proliferation, Biochem. Biophys. Res. Commun. 387 (2) (2009) 310–315.
[62] C. Qin, S. Liu, S. Zhou, Q. Wang, X. Xia, J. Hu, et al., PIK3C2A is a prognostic biomarker that is linked to immune infiltrates in kidney renal clear cell carcinoma, Front. Immunol. 14 (2023) 1114572.
[63] S. Liao, C. Liu, G. Zhu, K. Wang, Y. Yang, C. Wang, Relationship between SDC1 and cadherin signalling activation in cancer, Pathol. Res. Pract. 216 (1) (2020) 152756.
[64] I. Donner, T. Kiviluoto, A. Ristimaki, L.A. Aaltonen, P. Vahteristo, Exome sequencing reveals three novel candidate predisposition genes for diffuse gastric cancer, Fam. Cancer 14 (2) (2015) 241–246.
[65] U.E. Martinez-Outschoorn, M. Peiris-Pagés, R.G. Pestell, F. Sotgia, M.P. Lisanti, Cancer metabolism: a therapeutic perspective, Nat. Rev. Clin. Oncol. 14 (2) (2017) 113.

[66] J.H. Liang, Y. Lin, T. Ouyang, W. Tang, Y. Huang, W. Ye, et al., Nuclear magnetic resonance-based metabolomics and metabolic pathway networks from patient-matched esophageal carcinoma, adjacent noncancerous tissues and urine, World J. Gastroenterol. 25 (25) (2019) 3218–3230.
[67] J. Xu, Y. Chen, R. Zhang, Y. Song, J. Cao, N. Bi, et al., Global and targeted metabolomics of esophageal squamous cell carcinoma discovers potential diagnostic and therapeutic biomarkers, Mol. Cell. Proteomics 12 (5) (2013) 1306–1318.
[68] M. Munemoto, K.I. Mukaisho, T. Miyashita, K. Oyama, Y. Haba, K. Okamoto, et al., Roles of the hexosamine biosynthetic pathway and pentose phosphate pathway in bile acid-induced cancer development, Cancer Sci. 110 (8) (2019) 2408–2420.
[69] J. Xu, Y. Chen, R. Zhang, J. He, Y. Song, J. Wang, et al., Global metabolomics reveals potential urinary biomarkers of esophageal squamous cell carcinoma for diagnosis and staging, Sci. Rep. 6 (2016) 35010.
[70] W.H. Koppenol, P.L. Bounds, C.V. Dang, Otto Warburg's contributions to current concepts of cancer metabolism, Nat. Rev. Cancer 11 (5) (2011) 325–337.
[71] M.G. Vander Heiden, L.C. Cantley, C.B. Thompson, Understanding the Warburg effect: the metabolic requirements of cell proliferation, Science 324 (5930) (2009) 1029–1033.
[72] R.A. Cairns, I.S. Harris, T.W. Mak, Regulation of cancer cell metabolism, Nat. Rev. Cancer 11 (2) (2011) 85–95.
[73] A.F. Abdel-Wahab, W. Mahmoud, R.M. Al-Harizy, Targeting glucose metabolism to suppress cancer progression: prospective of anti-glycolytic cancer therapy, Pharmacol. Res. 150 (2019) 104511.
[74] F. Cappellesso, M.P. Orban, N. Shirgaonkar, E. Berardi, J. Serneels, M.A. Neveu, et al., Targeting the bicarbonate transporter SLC4A4 overcomes immunosuppression and immunotherapy resistance in pancreatic cancer, Nat. Can. (Ott.) 3 (12) (2022) 1464–1483.
[75] A. Stincone, A. Prigione, T. Cramer, M.M. Wamelink, K. Campbell, E. Cheung, et al., The return of metabolism: biochemistry and physiology of the pentose phosphate pathway, Biol. Rev. Camb. Phil. Soc. 90 (3) (2015) 927–963.
[76] E. Tsouko, A.S. Khan, M.A. White, J.J. Han, Y. Shi, F.A. Merchant, et al., Regulation of the pentose phosphate pathway by an androgen receptor-mTOR-mediated mechanism and its role in prostate cancer cell growth, Oncogenesis 3 (5) (2014) e103.
[77] Y. Nagata, S. Yamamoto, K. Kato, Immune checkpoint inhibitors in esophageal cancer: clinical development and perspectives, Hum. Vaccines Immunother. 18 (6) (2022) 2143177.
[78] S. Kannan, G.M. O'Connor, E.Y. Bakker, Molecular mechanisms of PD-1 and PD-L1 activity on a pan-cancer basis: a bioinformatic exploratory study, Int. J. Mol. Sci. 22 (11) (2021).