

## RESEARCH ARTICLE

# Identification of serum prognostic marker miRNAs and construction of microRNA-mRNA networks of esophageal cancer

Yue Jiang<sup>1‡</sup>, Chengda Zhang<sup>2‡</sup>, Wenbin Shen<sup>3</sup>, Yiming Li<sup>2</sup>, Yun Wang<sup>4</sup>, Jianjun Han<sup>4</sup>, Tao Liu<sup>4</sup>, Li Jia<sup>4</sup>, Fei Gao<sup>4</sup>, Xiaojun Liu<sup>4</sup>, Mi Chen<sup>4</sup>, Guangming Yi<sup>4</sup>, Hongchun Dai<sup>4</sup>, Jun He<sup>4,5\*</sup>

**1** Department of Clinical Medicine, Southwest Medical University, Luzhou, China, **2** Department of Gastroenterology, The Third Hospital of Mian Yang (Sichuan Mental Health Center), Mianyang, China, **3** Department of Oncology, The Third Hospital of Mianyang (Sichuan Mental Health Center), Mianyang, China, **4** Department of Oncology, The First Affiliated Hospital of Chengdu Medical College, Chengdu, China, **5** The Third Hospital of Mianyang (Sichuan Mental Health Center), Mianyang, China

‡ YJ and CZ are co-first authors and contributed equally to this study.

\* [he-001jun@163.com](mailto:he-001jun@163.com)



## Abstract

Esophageal cancer is a common tumor of the digestive system with poor prognosis. This study was to gain a better understanding of the mechanisms involved in esophageal cancer and to identify new prognostic markers. We downloaded the esophageal cancer miRNA expression profile microarray data (GSE113740, GSE112264, GSE122497, GSE113486, and GSE106817) from the GEO database, extracted the esophageal cancer miRNA sequencing data from The Cancer Genome Atlas (TCGA) database, and then used a bioinformatics approach to select common differentially expressed miRNAs (DEMs). Differentially expressed genes (DEGs) were selected by predicting DEM target genes using the miRWalk database and intersecting with differential genes obtained from TCGA database for esophageal cancer. The STRING database was used to obtain protein–protein interaction (PPI) relationships to construct the DEM-DEG network. Furthermore, we selected core genes and core miRNAs associated with esophageal cancer prognosis by performing survival and univariate/multivariate COX analysis on DEMs and DEGs in the network and performed GSEA analysis on core genes alone, and finally the expression of the markers was verified by qPCR in esophageal cancer cell lines Eca109, SKGT-4 and normal esophageal epithelial cells HEEC. Nine DEMs were obtained, of which three were upregulated and six were downregulated, and 326 DEGs were obtained, of which 105 were upregulated and 221 were downregulated. Survival univariate/multivariate COX analysis revealed that five genes, ZBTB16, AQP4, ADCYAP1R1, PDGFD, and VIPR2, and two microRNAs, miR-99a-5p, and miR-508-5p, were related to esophageal cancer prognosis. GSEA analysis showed that the following genes may be involved in esophageal cancer prognosis: ZBTB16 may through the MTOR signaling pathway, AQP4 through the GNRH signaling pathway, ADCYAP1R1 through the PPAR signaling pathway, VIPR2 through the P53 signaling pathway and PDGFD through the PENTOSE-PHOSPHATE signaling pathway.

## OPEN ACCESS

**Citation:** Jiang Y, Zhang C, Shen W, Li Y, Wang Y, Han J, et al. (2021) Identification of serum prognostic marker miRNAs and construction of microRNA-mRNA networks of esophageal cancer. *PLoS ONE* 16(7): e0255479. <https://doi.org/10.1371/journal.pone.0255479>

**Editor:** Suhwan Chang, University of Ulsan College of Medicine, REPUBLIC OF KOREA

**Received:** February 7, 2021

**Accepted:** July 18, 2021

**Published:** July 30, 2021

**Copyright:** © 2021 Jiang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and it's [S1 Checklist](#).

**Funding:** This study was supported by the Health Commission of Mianyang City (grant number 201937).

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Esophageal cancer is a joint malignant tumor of the digestive system, primarily including two types of squamous carcinoma and adenocarcinoma [1, 2]. Worldwide, approximately 500,000 patients are diagnosed with esophageal cancer each year, and their 5-year overall survival rate is generally <15% [3, 4]. Although esophageal cancer treatment has greatly improved the prognosis of patients, the prognosis of patients with esophageal cancer remains poor as endoscopic diagnosis of early stage esophageal cancer is difficult and most patients with esophageal cancer are already in the middle or advanced stages at the time of diagnosis [5–7]. MicroRNAs (miRNAs) are small noncoding RNAs that regulate gene expression and are implicated in the pathogenesis of several cancers [8–10]. It has been shown that measuring circulating miRNA levels in patients' serum may be a simple and non-invasive method for diagnosing certain early-stage cancers [11–13]. In the present study, to explore the mechanism underlying esophageal cancer development and provide novel targets and pathways for esophageal cancer diagnosis and treatment, a bioinformatics approach was used to analyze the esophageal cancer data derived from the public databases The Cancer Genome Atlas (TCGA) [14] and Gene Expression Omnibus (GEO) [15]. The differential genes and differential miRNAs of esophageal cancer were integrated and analyzed to construct a miRNA-mRNA network for esophageal cancer, and survival analysis, univariate/multivariate COX analysis, and GSEA analysis were conducted on the genes and miRNAs in the network to provide research directions for further basic experiments.

## Materials and methods

### Data

The GEO (<https://www.ncbi.nlm.nih.gov/geo/>) [15] database and TCGA (<https://portal.gdc.cancer.gov/>) [14] database provide a large amount of data on patients with esophageal cancer. In the GEO search box, we entered “esophageal cancer serum” and “miRNA” and selected human microarray data with sample size greater than 100 and then selected five datasets, GSE113740 [16], GSE112264 [17], GSE122497 [18], GSE113486 [19], and GSE106817 [20], from the same platform GPL21263 (3D-Gene Human miRNA V21\_1.0.0), as shown in Table 1, including the serum miRNA microarray data from a total of 5869 healthy controls and 769 patients with esophageal cancer (Exclusion of large deviations from the data set GSE124158). RNA sequencing data (miRNA and mRNA) derived from the tumor tissues of patients with esophageal cancer and tissues of healthy controls were downloaded from TCGA database, select TCGA-ESCA's sequencing reads date, including the miRNA sequencing data from 13 healthy tissues and 185 tumor tissues versus the mRNA sequencing data from 11 healthy tissues and 160 tumor tissues.

Human esophageal cancer SKGT-4 and ECA109 cells, human normal esophageal epithelial cells (HEEC cells) were purchased from the Cell Bank of the Chinese Academy of Sciences;

**Table 1. Grouping characteristics of each data set.**

Date Set	Contributors	Samples of EC	Samples of HC	Submission date
GSE113740 [13]	Yamamoto Y, et al	25	969	30-Jul-20
GSE112264 [14]	Urabe F, et al	50	41	Mar 23,2018
GSE122497 [15]	Sudo K, et al	566	2000 (Select the first two thousand)	13-Nov-18
GSE113486 [16]	Usaba W, et al	40	100	20-Apr-18
GSE106817 [17]	Yokoi A, et al	88	2759	13-Nov-17

EC: Esophageal Cancer; HC: healthy control.

<https://doi.org/10.1371/journal.pone.0255479.t001>

fetal bovine serum and RPMI1640 cell culture medium were purchased from Hyclone; RNA extraction kit, Trizol and protein extraction lysis solution were purchased from Biyuntian Biotechnology Company Limited.

### Difference analysis

The data in the five datasets were divided into cancer and healthy control groups, and the results were analyzed differentially and visualized using the GEO2R tool based on the limma.R package (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>) [21], setting padj = 0.05 and log2FC = 1. The results were incorporated using the RobustRankAggreg.R [22] package, and an R package for the meta-analysis of multiple sets of GSE data was used to integrate the differential results and select differentially expressed miRNAs with  $p < 0.05$  (DEMs1). Differential analysis of data obtained from TCGA database was performed using the Edge.R [23] package, and to ensure the accuracy of the results, read counts were transformed into counts per million index to eliminate the effect of sequencing depth. Differential miRNAs with Q values  $< 0.05$ ,  $\log_2FC > 1$  were selected as DEMs2 and Q values  $< 0.05$ ,  $\log_2FC > 2$  were used as initial differentially expressed genes (DEGs). DEMs1 and DEMs2 were selected for intersection to obtain differentially expressed miRNAs (DEMs).

### miRNA-mRNA and PPI network analysis

The miRWalk database [24] and the STRING database [25] were used for the construction of the miRNA-mRNA network in esophageal cancer. A total of 326 DEGs were obtained by predicting the target genes of DEMs using the miRWalk website and taking intersections with the abovementioned DEGs. The STRING database contains several protein-protein interaction relationships, and the protein-protein association (PPI) of 326 DEGs was analyzed using the STRING database. The DEGs with a combined score  $> 0.7$  (139 genes in total) were selected and outlined with their upstream miRNAs to create an esophageal cancer miRNA-mRNA network using the Cytoscape [26] software to visualize the results.

### Univariate/Multivariate COX analysis

Univariate / Multivariate COX analysis and survival analysis often establishes a link with disease prognosis and is widely used in oncology research. We used genome-wide clinical data downloaded from TCGA for 160 patients with esophageal cancer to initially screen 139 genes in the network using univariate COX analysis, before further screening esophageal cancer prognostic association DEGs using multivariate COX analysis and constructing a prognostic model for esophageal cancer to test the prognostic model effect using survival curves and ROC curves. Only DEGs that met  $p < 0.05$  for both univariate COX and multivariate COX analyses could be considered as prognostic-associated DEGs for esophageal cancer. Multivariate COX analysis was performed online for DEMs using the OncomiR database [27], and then  $p < 0.05$  Kaplan-Meier survival curves were plotted for DEMs based on the high and low risk values obtained from the COX analysis.

### GSEA analysis of core genes

Gene set enrichment analysis (GSEA) analysis can be used for the pathway enrichment analysis of individual genes, which can better reveal the upregulation and downregulation relationships of individual gene-enriched pathways in disease compared with the common KEGG analysis. The 160 esophageal cancer samples were divided into gene-high and gene-low expression groups, and all esophageal cancer genes were analyzed for enrichment in GSEA software

[28], and some pathways with p-values <0.05 were selected for visualisation. The pathways that were predominantly enriched for genes were examined in preparation for the subsequent basic experiments.

## Cell culture and qPCR

Human esophageal cancer SKGT-4, ECA109 cell lines and HEEC cell lines were cultured in sterile RPMI1640 medium containing 10% fetal bovine serum and 1% antibacterial drugs (penicillin, streptomycin) in a sterile cell culture incubator at 37°C and 5% CO<sub>2</sub>.

RT-qPCR is a sensitive technique for quantifying specific RNA targets. Human esophageal cancer SKGT-4 and ECA109 cells at logarithmic growth stage were collected, with HEEC cells as control. RNA was extracted using the RNA extraction kit, strictly according to the instructions. RNA concentration was detected by qPCR. cDNA was synthesized by reverse transcription and PCR reactions (PCR reaction conditions: 95°C30s,95°C10s,60°C30s) were performed to detect the expression levels of miR-99a-5p, miR-508-5p, ZBTB16, AQP4, ADCYAP1R1, PDGFD and VIPR2. The relative expression levels of the genes were calculated as  $2^{-\Delta\Delta C_t}$ .

## Results and discussion

### Differentially expressed miRNAs (DEMs)

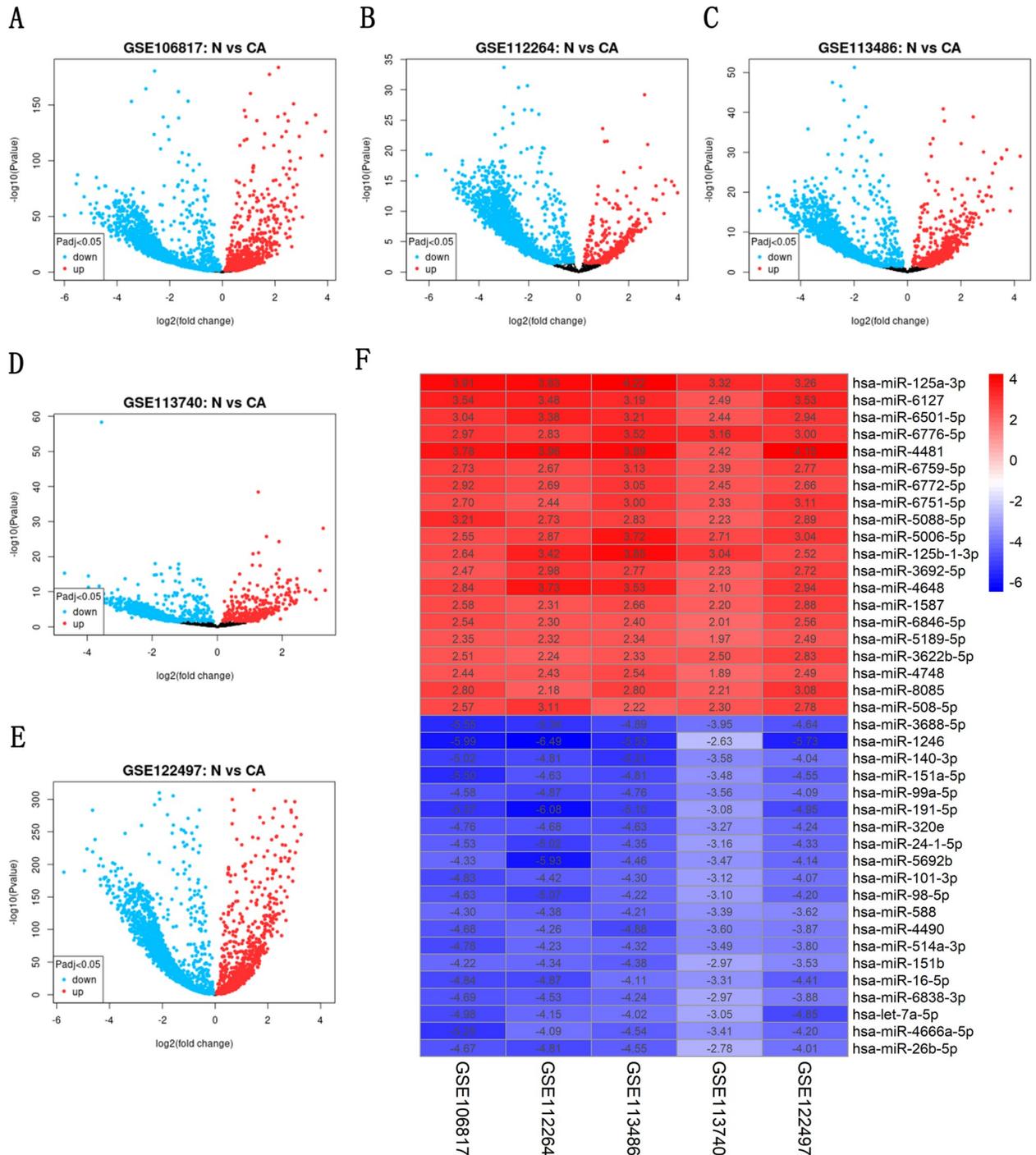
To obtain differential miRNAs for esophageal cancer, we conducted the differential analysis of sequencing data using the GEO2R tool to obtain differential volcano plots (Fig 1A–1E). The results of all differential analyses were analyzed using the RobustRankAggreg.R package, and miRNAs with p values < 0.05 were selected as DEMs1, after which the results were visualized (Fig 1F). Differential analysis of miRNA sequencing data derived from 13 healthy tissues and 185 tumor tissues in TCGA database was performed using the Edge.R package, and a differential volcano plot was obtained (Fig 2A). The differential miRNAs with p < 0.05 and log<sub>2</sub>FC > 1 were selected as DEMs2, and the final DEMs were obtained by taking the intersection of DEMs1 and DEMs2 (Fig 2B and 2C). A total of 169 DEMs1 and 141 DEMs2 were obtained, resulting in 9 DEMs, of which 3 were upregulated and 6 were downregulated.

### Differentially expressed genes (DEGs)

To obtain differential genes for esophageal cancer, we used the Edge.R package to perform the differential analysis of mRNA sequencing data derived from 11 healthy and 160 tumor tissues downloaded from TCGA database to obtain a differential volcano plot (Fig 2D), and the gene with p < 0.05 and log<sub>2</sub>FC > 2 was selected as the initial differential gene (DEGs1). The miR-Walk database was used to predict the target genes of DEMs, all possible target genes were obtained, and the final DEGs were obtained by intersecting DEGs1 with the target genes (Fig 2E and 2F). A total of 326 DEGs were obtained, of which 105 were upregulated and 221 were downregulated.

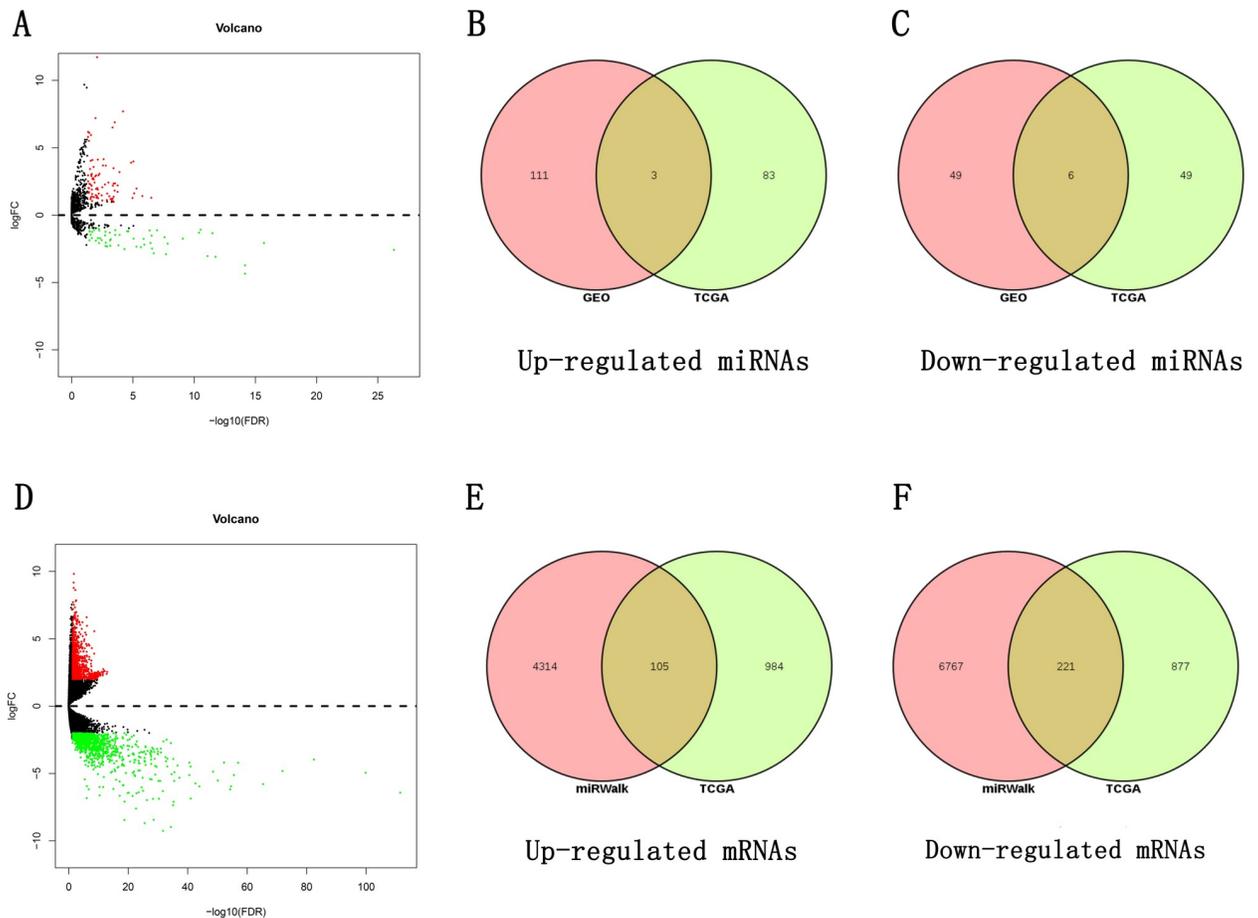
### miRNA-mRNA and PPI networks

To construct the miRNA-mRNA network for esophageal cancer and discover key markers of esophageal cancer from multiple perspectives, we submitted the DEGs to the STRING database for protein-protein interaction (PPI) analysis, identified the gene associations from the protein level, selected genes with stronger associations, and visualized them using the Cytoscape software (Fig 3).



**Fig 1. Volcano plots of differentially expressed genes in GSE113740, GSE112264, GSE122497, GSE113486, and GSE10681 and log<sub>2</sub>FC heatmap of each expression microarray.** (A-E) show the volcano plots of differentially expressed genes in GSE10681, GSE112264, GSE113486, GSE113740, and GSE122497, respectively. The horizontal coordinates represent log<sub>2</sub> (fold change), blue represents downregulated genes, and red represents upregulated genes. The vertical coordinate represents -log<sub>10</sub> (p value). N is the normal control group; CA is the cancer group. (F) shows the log<sub>2</sub>FC heat map of GSE113740, GSE112264, GSE122497, GSE113486, and GSE10681. The numbers in the plot show the log<sub>2</sub>FC for each expression microarray, with red indicating upregulation, and blue indicating downregulation.

<https://doi.org/10.1371/journal.pone.0255479.g001>

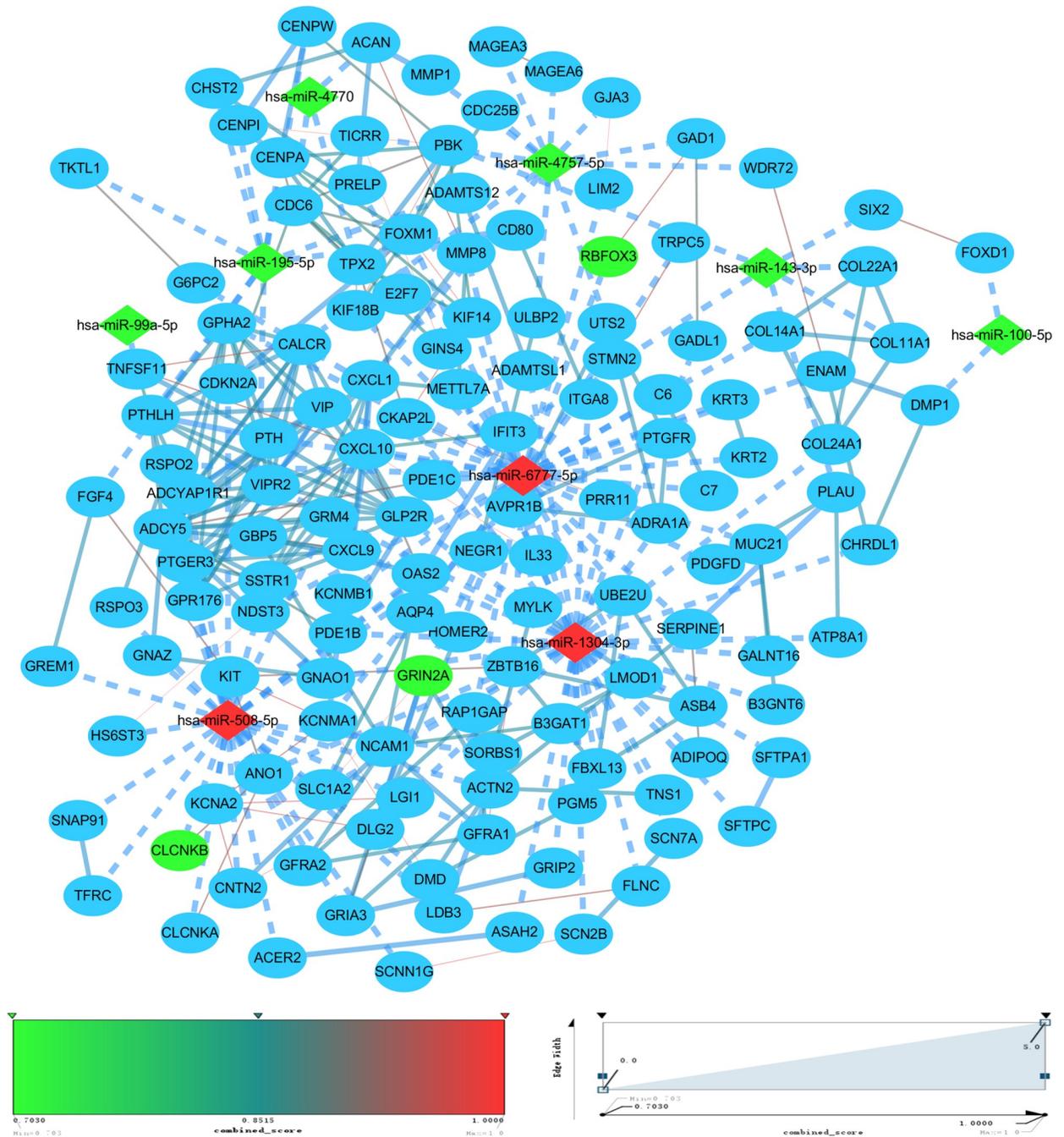


**Fig 2. Differentially expressed miRNAs (DEMs) and differentially expressed genes (DEGs).** (A) and (D) indicate the differential volcano plots of differentially expressed miRNAs in esophageal cancer tissues versus normal tissues in TCGA database and differential volcano plots of differentially expressed genes in esophageal cancer tissues versus normal tissues in TCGA database, respectively, where red indicates upregulation, and green indicates downregulation. (B) and (C) indicate the intersecting Venn diagrams of upregulated DEM1 and upregulated DEM2 and the intersecting Venn diagrams of downregulated DEM1 and downregulated DEM2, respectively. (E) and (F) indicate the Wayne diagrams of intersection of DEMs predicted upregulated target genes and TCGA database differentially expressed upregulated genes and the Wayne diagrams of intersection of DEMs predicted downregulated target genes and TCGA database differentially expressed downregulated genes, respectively.

<https://doi.org/10.1371/journal.pone.0255479.g002>

### Univariate / Multivariate COX analysis

To obtain further core genes and core miRNAs associated with the prognosis of esophageal cancer, this study used clinical data from 11 healthy patients and 160 esophageal cancer patients in the TCGA database to perform univariate/multivariate COX analysis on 139 DEGs in the network, and univariate COX analysis yielded 22 prognosis-associated genes. Multivariate COX analysis yielded SIX2, ADIPOQ, ZBTB16, AQP4, ADCYAP1R1, PDGFD and VIPR2 seven (of which  $p < 0.05$  for a total of five genes ZBTB16, AQP4, ADCYAP1R1, PDGFD and VIPR2) genes associated with prognosis in esophageal cancer, seven-gene prognostic model survival  $p = 0.01252$ , area under the ROC curve AUC = 0.634 (Fig 4C–4F). OncomiR, a powerful database for studying cancer miRNAs online, was used to perform multivariate COX survival analysis of DEMs using 184 esophageal cancer clinical data in OncomiR, and miRNAs with  $p < 0.05$  were selected as the last associated with esophageal cancer prognosis of de novo markers (Fig 4A and 4B).

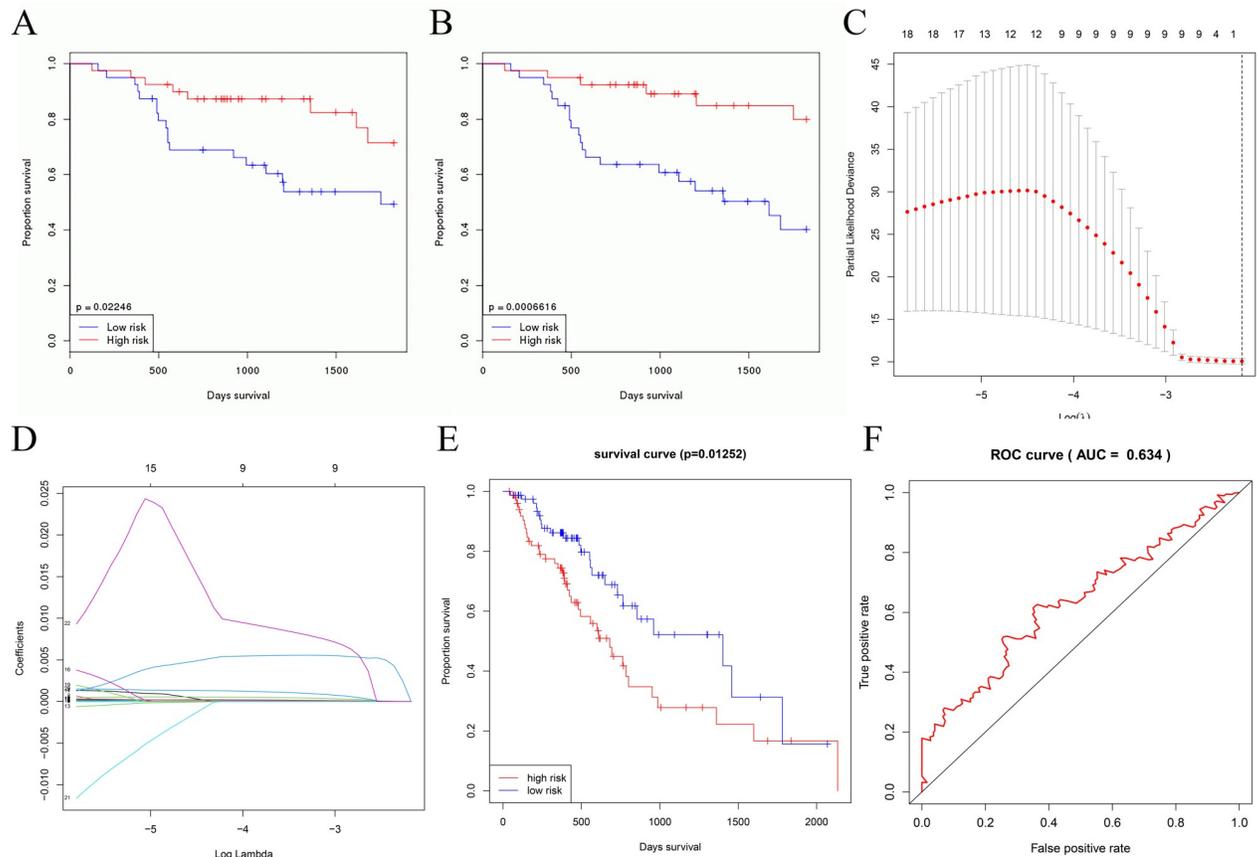


**Fig 3. Diagram of miRNA-mRNA network in esophageal cancer and PPI network of DEGs.** The above miRNA-mRNA network is based on the miRWalk database, where miRNAs are represented by diamonds, and genes are represented by circles. PPI linkage of DEGs was obtained by STRING database screening, where red indicates upregulation in esophageal cancer, green indicates downregulation, marquee dash-type lines indicate miRNA-gene interactions, and solid lines indicate gene-to-gene interactions. The color and thickness of the lines in the PPI network of DEG are drawn according to the combined score from the SRTING database.

<https://doi.org/10.1371/journal.pone.0255479.g003>

### GSEA analysis of core genes

The expression values of five genes, ZBTB16, AQP4, ADCYAP1R1, PDGFD and VIPR2, were sorted and divided into gene-high and low expression groups for patients with esophageal



**Fig 4. Prognostic risk models for ZBTB16, AQP4, ADCYAP1R1, PDGFD, VIPR2 and survival curves for miR-99a-5p and miR-508-5p.** (A) and (B) indicate the survival curves of miR-99a-5p and miR-508-5p, respectively. The higher the expression of miR-99a-5p and miR-508-5p in esophageal cancer, the better the prognosis of patients. (C-F) indicate the cvfit plot, lambda plot, survival curve and ROC curve of the LASSO regression of the genetic prognostic model, respectively.

<https://doi.org/10.1371/journal.pone.0255479.g004>

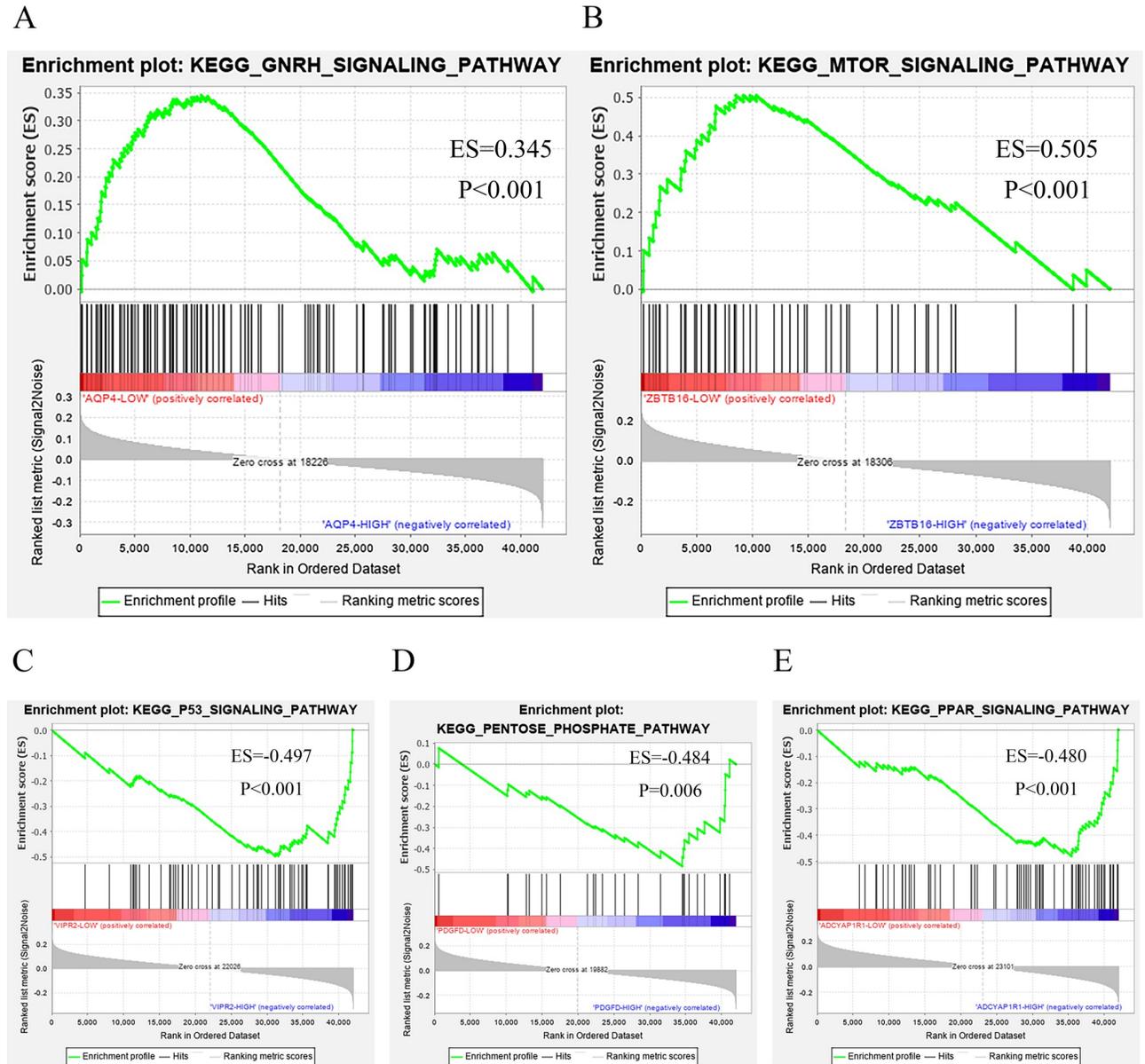
cancer, and then the core genes were sequenced using the esophageal cancer RNA downloaded from the TCGA database, and after counts per million (CPM) quantification, the GSEA software was used to The high- and low-expression groups were subjected to GSEA pathway enrichment analysis, and the top-ranked pathways with  $p$  value  $< 0.05$  were selected for visualization, respectively (Fig 5A–5E).

### qPCR

The results showed that the relative expression levels of miR-508-5p, AQP4, ADCYAP1R1 and VIPR2 were significantly up-regulated in human esophageal cancer SKGT-4 and ECA109 cells compared with HEEC cells, and the differences were all statistically significant (Fig 6B, 6D, 6F and 6G). The relative expression levels of miR-99a-5p were significantly down-regulated in human esophageal cancer SKGT-4 cells compared with HEEC cells, and the differences were statistically significant (Fig 6A).

### Significance and impacts

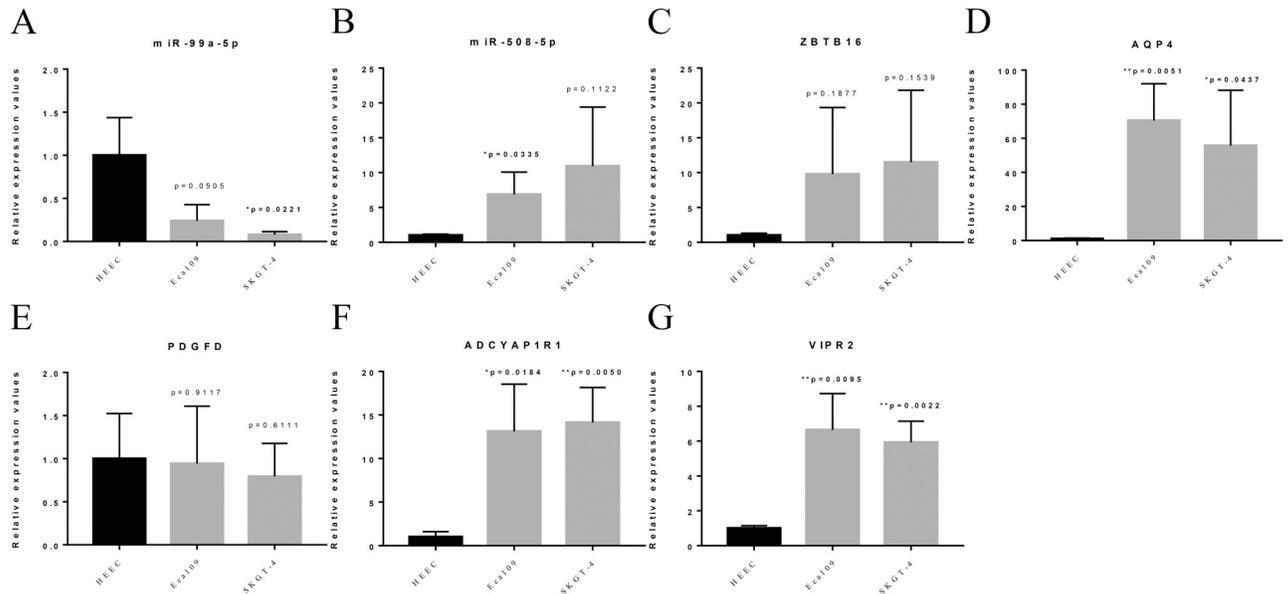
The method of analyzing sequencing and microarray data through bioinformatics approaches is widely applied in the medical field, including the field of research on tumor mechanisms



**Fig 5. GSEA enrichment curves for ZBTB16, AQP4, ADCYAP1R1, PDGFD and VIPR2.** (A-E) GSEA enrichment curves for AQP4, ZBTB16, VIPR2, PDGFD and ADCYAP1R1 based on gene-high and low expression groups using GSEA software, respectively. Where ES denotes Enrichment Score and p denotes Nominal p-value. (A) AQP4 probably through the GNRH signaling pathway, (B) ZBTB16 probably through the MTOR signaling pathway, (C) VIPR2 probably through the P53 signaling pathway, (D) PDGFD probably through the PENTOSE- PHOSPHATE signaling pathway and (E) ADCYAP1R1 may play a role in the prognosis of esophageal cancer through the PPAR signaling pathway.

<https://doi.org/10.1371/journal.pone.0255479.g005>

[29–31]. Bioinformatics can use an existing massive database to make predictions about unknown problems, which is a good solution to the challenge of how to screen out the core genes that are closely associated with diseases from a large number of genes. Esophageal cancer is one of the common tumors with poor prognosis, and according to current statistical results, China has the highest incidence of esophageal cancer in the world [32, 33]. To better screen people at high risk of esophageal cancer at an early stage, to improve the prognosis of esophageal cancer and to provide new directions for research on esophageal cancer, in the present study, we used data from the GEO database and TCGA database to analyze differentially



**Fig 6. Relative expression of markers.** (A-G) denote the expression of miR-99a-5p, miR-508-5p, ZBTB16, AQP4, PDGFD, ADCYAP1R1 and VIPR2, respectively. qPCR results showed that: (A) miR-99a-5p was significantly down-regulated in human esophageal cancer SKGT-4 cells compared with HEEC cells, and the difference was statistically significant. The differences were statistically significant. (B) The relative expression levels of miR-508-5p were significantly up-regulated in human esophageal cancer ECA109 cells compared with HEEC cells, and the differences were statistically significant. The relative expression levels of AQP4, ADCYAP1R1 and VIPR2 were significantly upregulated in human esophageal cancer SKGT-4 and ECA109 cells, and the differences were statistically significant.

<https://doi.org/10.1371/journal.pone.0255479.g006>

expressed miRNAs in esophageal cancer and obtained a total of nine DEMs, of which three were upregulated and six were downregulated. We used sequencing data obtained from TCGA database to analyze DEGs in esophageal cancer and combined them with DEMs. The miRNA-mRNA network of esophageal cancer was constructed by combining the sequencing data with DEMs, which provides a new direction for research on esophageal cancer. To further understand the new biomarkers that are closely related to esophageal cancer prognosis, we also performed the univariate/multivariate COX analysis of miRNAs and mRNAs in the network and found five core genes, ZBTB16, AQP4, ADCYAP1R1, PDGFD and VIPR2, that may be associated with esophageal cancer prognosis and two core miRNAs, miR-99a-5p, and miR-508-5p. Finally, we conducted KEGG enrichment analysis on each of the five core genes using data obtained from TCGA database and GSEA software, expecting to identify the pathways acted upon by the core genes for subsequent basic experiments.

Zinc Finger And BTB Domain Containing 16 (ZBTB16) is a member of the Krueppel C2H2 type zinc finger protein family, which is associated with skeletal defects, genital hypoplasia, mental retardation and acute promyelocytic leukaemia [34–36]. Recent studies have demonstrated that ZBTB16 can act as an oncogene in breast cancer by upregulating ZBTB28 and antagonizing BCL6 [37]. Aquaporin 4 (AQP4), a member of the family of membrane protein water channel proteins, has been suggested to promote tumor progression, invasion and metastasis [38, 39] and is a promising target in oncology research. In lung squamous cell carcinoma, AQP4 is transcribed at low levels [40]. However, the role and mechanisms of ZBTB16 and AQP4 in esophageal cancer have not yet been investigated. Adenylate Cyclase Activating Polypeptide 1 (Pituitary) Receptor Type I (ADCYAP1R1) encodes a type I adenylate cyclase activating polypeptide receptor, which is associated with diseases such as post-traumatic stress disorder and regulatory spasticity [14, 15]. ADCYAP1R1 is differentially expressed in a variety

of tumors, such as gliomas, breast and prostate cancers, and gastric cancer [41–43]. This also includes esophageal cancer, where Zhang, Yuefeng et al. found low expression of ADCYAP1R1 in esophageal cancer by bioinformatics analysis consistent with our experimental results [44], but the exact prognostic mechanism of ADCYAP1R1 affecting esophageal cancer has not been investigated. PDGFD (platelet-derived growth factor D) is a protein-coding gene, Yang, Xiao et al. demonstrated through bioinformatics that PDGFD is associated with ovarian cancer [45]. Although PDGFD has not been shown to be associated with prognosis in esophageal cancer, the association between platelet-derived growth factor-BB expression and prognosis in human esophageal squamous cell carcinoma has been studied [46]. VIPR2 encodes the receptor for vasoactive intestinal peptide (a small neuropeptide), which has been associated with diseases such as schizophrenia [47]. VIPR2 is under-expressed in ESCC, which is consistent with our experimental results [48].

## Conclusion

In conclusion, our study screened prognostic markers for esophageal cancer, constructed a miRNA-mRNA network for esophageal cancer, and predicted possible convergence pathways for core genes, laying the foundation for subsequent basic research on esophageal cancer. However, there are still shortcomings and limitations in our study. 1) We cross-analyzed the obtained serum differential DEMs with the tumor tissue DEMs in the TCGA database, which is a case of excluding miRNAs secreted by tumor tissue into serum, which may lead to the omission of some key miRNAs. To facilitate peer research on circulating miRNAs in esophageal cancer serum, we have published the results of the GEO analysis as a supplementary paper. (2) Our proposal of circulating miRNAs in serum as a possible new diagnostic marker is based on speculation from previous studies and theories, and the exact method of how to extract the corresponding miRNAs from serum needs to continue to be investigated. (3) Our experiments were validated using only one experimental method, qPCR, and two esophageal cancer cell lines, which likely led to biased results. The expression of each gene also appeared to be overexpressed. (4) The pathways and miRNA-mRNA networks were obtained by bioinformatics analysis and should be further validated by experiments.

## Supporting information

**S1 Checklist.**  
(DOCX)

## Author Contributions

**Conceptualization:** Yue Jiang, Chengda Zhang, Jun He.

**Data curation:** Yue Jiang, Jun He.

**Investigation:** Yue Jiang, Chengda Zhang, Wenbin Shen, Yun Wang.

**Methodology:** Chengda Zhang, Wenbin Shen.

**Resources:** Wenbin Shen, Yiming Li, Yun Wang, Jianjun Han, Tao Liu, Li Jia, Fei Gao, Xiaojun Liu, Mi Chen, Guangming Yi, Hongchun Dai.

**Writing – original draft:** Yue Jiang, Chengda Zhang, Jun He.

**Writing – review & editing:** Jun He.

## References

1. Daly J, Fry W, Little A, Winchester D, McKee R, Stewart A, et al. Esophageal cancer: results of an American College of Surgeons Patient Care Evaluation Study. *Journal of the American College of Surgeon*. 2000; 190(5):562–72; discussion 72–3. [https://doi.org/10.1016/s1072-7515\(00\)00238-6](https://doi.org/10.1016/s1072-7515(00)00238-6) PMID: 10801023.
2. Meng J, Tong Q, Liu X, Yu Z, Zhang J, Gao B. Epigallocatechin-3-gallate inhibits growth and induces apoptosis in esophageal cancer cells through the demethylation and reactivation of the p16 gene. *Oncology letters*. 2017; 14(1):1152–6. <https://doi.org/10.3892/ol.2017.6248> PMID: 28693288.
3. Bray F, Ferlay J, Soerjomataram I, Siegel R, Torre L, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2018; 68(6):394–424. <https://doi.org/10.3322/caac.21492> PMID: 30207593.
4. Zhai R, Chen F, Liu G, Su L, Kulke M, Asomaning K, et al. Interactions among genetic variants in apoptosis pathway genes, reflux symptoms, body mass index, and smoking indicate two distinct etiologic patterns of esophageal adenocarcinoma. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2010; 28(14):2445–51. <https://doi.org/10.1200/JCO.2009.26.2790> PMID: 20385987.
5. Fitzmaurice C, Allen C, Barber R, Barregard L, Bhutta Z, Brenner H, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. *JAMA oncology*. 2017; 3(4):524–48. <https://doi.org/10.1001/jamaoncol.2016.5688> PMID: 27918777.
6. Lu D, Liu X, Feng S, Dong X, Shi X, Ren P, et al. The Current Situation of Esophageal Cancer Staging and Perioperative Strategies Determination in Central and Southern China: A Cross Sectional Survey. *Frontiers in oncology*. 2019; 9:1098. <https://doi.org/10.3389/fonc.2019.01098> PMID: 31696056
7. Grotenhuis BA, Van Heijl M, Zehetner J, Moons J, Wijnhoven B, van Berge Henegouwen M, et al. Surgical management of submucosal esophageal cancer: extended or regional lymphadenectomy? *Annals of surgery*. 2010; 252(5):823–30. <https://doi.org/10.1097/SLA.0b013e3181fcd730> PMID: 21037438
8. Ramsingh G, Koboldt DC, Trissal M, Chiappinelli KB, Wylie T, Koul S, et al. Complete characterization of the microRNAome in a patient with acute myeloid leukemia. *Blood*. 2010; 116(24):5316–26. Epub 2010/09/30. <https://doi.org/10.1182/blood-2010-05-285395> PMID: 20876853; PubMed Central PMCID: PMC3012545.
9. Sandiford O, Moore C, Du J, Boulad M, Gergues M, Eltouky H, et al. Human Aging and Cancer: Role of miRNA in Tumor Microenvironment. *Advances in experimental medicine*. 2018; 1056:137–52. [https://doi.org/10.1007/978-3-319-74470-4\\_9](https://doi.org/10.1007/978-3-319-74470-4_9) PMID: 29754179.
10. Tutar Y. miRNA and cancer; computational and experimental approaches. *Current pharmaceutical biotechnology*. 2014; 15(5):429. <https://doi.org/10.2174/138920101505140828161335> PMID: 25189575.
11. Madhavan D, Cuk K, Burwinkel B, Yang R. Cancer diagnosis and prognosis decoded by blood-based circulating microRNA signatures. *Frontiers in genetics*. 2013; 4:116. <https://doi.org/10.3389/fgene.2013.00116> PMID: 23802013
12. Yu H, Guan Z, Cuk K, Zhang Y, Brenner H. Circulating microRNA biomarkers for lung cancer detection in East Asian populations. *Cancers*. 2019; 11(3):415. <https://doi.org/10.3390/cancers11030415> PMID: 30909610
13. Guo X, Liu P, Yang X, Wang K, Wang Q, Guo Q, et al. A multiple amplification strategy for nucleic acid detection based on host–guest interaction between the  $\beta$ -cyclodextrin polymer and pyrene. *Analyst*. 2015; 140(6):2016–22. <https://doi.org/10.1039/c4an02402h> PMID: 25672284
14. Hutter C, Zenklusen J. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell*. 2018; 173(2):283–5. <https://doi.org/10.1016/j.cell.2018.03.042> PMID: 29625045.
15. Barrett T, Wilhite S, Ledoux P, Evangelista C, Kim I, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic acids research*. 2013; 41:D991–5. <https://doi.org/10.1093/nar/gks1193> PMID: 23193258.
16. Yamamoto Y, Kondo S, Matsuzaki J, Esaki M, Okusaka T, Shimada K, et al. Highly Sensitive Circulating MicroRNA Panel for Accurate Detection of Hepatocellular Carcinoma in Patients With Liver Disease. *Hepatology communications*. 2020; 4(2):284–97. <https://doi.org/10.1002/hep4.1451> PMID: 32025611.
17. Urabe F, Matsuzaki J, Yamamoto Y, Kimura T, Hara T, Ichikawa M, et al. Large-scale Circulating microRNA Profiling for the Liquid Biopsy of Prostate Cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2019; 25(10):3016–25. <https://doi.org/10.1158/1078-0432.CCR-18-2849> PMID: 30808771.

18. Sudo K, Kato K, Matsuzaki J, Boku N, Abe S, Saito Y, et al. Development and Validation of an Esophageal Squamous Cell Carcinoma Detection Model by Large-Scale MicroRNA Profiling. *JAMA network open*. 2019; 2(5):e194573. <https://doi.org/10.1001/jamanetworkopen.2019.4573> PMID: 31125107.
19. Usuba W, Urabe F, Yamamoto Y, Matsuzaki J, Sasaki H, Ichikawa M, et al. Circulating miRNA panels for specific and early detection in bladder cancer. *Cancer science*. 2019; 110(1):408–19. <https://doi.org/10.1111/cas.13856> PMID: 30382619.
20. Yokoi A, Matsuzaki J, Yamamoto Y, Yoneoka Y, Takahashi K, Shimizu H, et al. Integrated extracellular microRNA profiling for ovarian cancer screening. *Nature communications*. 2018; 9(1):4319. <https://doi.org/10.1038/s41467-018-06434-4> PMID: 30333487.
21. Davis S, Meltzer P. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*. 2007; 23(14):1846–7. <https://doi.org/10.1093/bioinformatics/btm254> PMID: 17496320.
22. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. 2012; 28(4):573–80. <https://doi.org/10.1093/bioinformatics/btr709> PMID: 22247279.
23. Robinson M, McCarthy D, Smyth G. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616> PMID: 19910308.
24. Ru Y, Kechris K, Tabakoff B, Hoffman P, Radcliffe R, Bowler R, et al. The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. *Nucleic acids research*. 2014; 42(17):e133. <https://doi.org/10.1093/nar/gku631> PMID: 25063298.
25. Szklarczyk D, Gable A, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*. 2019; 47:D607–D13. <https://doi.org/10.1093/nar/gky1131> PMID: 30476243.
26. Shannon P, Markiel A, Ozier O, Baliga N, Wang J, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 2003; 13(11):2498–504. <https://doi.org/10.1101/gr.1239303> PMID: 14597658.
27. Sarver A, Sarver A, Yuan C, Subramanian S. OMCD: OncomiR Cancer Database. *BMC cancer*. 2018; 18(1):1223. <https://doi.org/10.1186/s12885-018-5085-z> PMID: 30522456.
28. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov J. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics*. 2007; 23(23):3251–3. <https://doi.org/10.1093/bioinformatics/btm369> PMID: 17644558.
29. Kristensen L, Ebbesen K, Sokol M, Jakobsen T, Korsgaard U, Eriksen A, et al. Spatial expression analyses of the putative oncogene ciRS-7 in cancer reshape the microRNA sponge theory. *Nature communications*. 2020; 11(1):4551. <https://doi.org/10.1038/s41467-020-18355-2> PMID: 32917870.
30. Sheffield N, Pierron G, Klughammer J, Datlinger P, Schönegger A, Schuster M, et al. DNA methylation heterogeneity defines a disease spectrum in Ewing sarcoma. *Nature medicine*. 2017; 23(3):386–95. <https://doi.org/10.1038/nm.4273> PMID: 28134926.
31. Yang Y, Chen L, Gu J, Zhang H, Yuan J, Lian Q, et al. Recurrently deregulated lncRNAs in hepatocellular carcinoma. *Nature communications*. 2017; 8:14421. <https://doi.org/10.1038/ncomms14421> PMID: 28194035.
32. Klima T, Christ A, Marana I, Kalbermatter S, Uthoff H, Burri E, et al. Sodium chloride vs. sodium bicarbonate for the prevention of contrast medium-induced nephropathy: a randomized controlled trial. *European heart journal*. 2012; 33(16):2071–9. <https://doi.org/10.1093/eurheartj/ehr501> PMID: 22267245.
33. Yu C, Tang H, Guo Y, Bian Z, Yang L, Chen Y, et al. Hot Tea Consumption and Its Interactions With Alcohol and Tobacco Use on the Risk for Esophageal Cancer: A Population-Based Cohort Study. *Annals of internal medicine*. 2018; 168(7):489–97. <https://doi.org/10.7326/M17-2000> PMID: 29404576.
34. Fischer S, Kohlhase J, Böhm D, Schweiger B, Hoffmann D, Heitmann M, et al. Biallelic loss of function of the promyelocytic leukaemia zinc finger (PLZF) gene causes severe skeletal defects and genital hypoplasia. *Journal of medical genetics*. 2008; 45(11):731–7. <https://doi.org/10.1136/jmg.2008.059451> PMID: 18611983
35. Plaisier C, Bennett B, He A, Guan B, Lusi A, Reue K, et al. Zbtb16 has a role in brown adipocyte bioenergetics. *Nutrition & diabetes*. 2012; 2(9):e46–e.
36. Mao A-P, Ishizuka IE, Kasal DN, Mandal M, Bendelac A. A shared Runx1-bound Zbtb16 enhancer directs innate and innate-like lymphoid lineage development. *Nature communications*. 2017; 8(1):1–14. <https://doi.org/10.1038/s41467-016-0009-6> PMID: 28232747
37. He J, Wu M, Xiong L, Gong Y, Yu R, Peng W, et al. BTB/POZ zinc finger protein ZBTB16 inhibits breast cancer proliferation and metastasis through upregulating ZBTB28 and antagonizing BCL6/ZBTB27. *Clinical epigenetics*. 2020; 12(1):1–15. <https://doi.org/10.1186/s13148-020-00867-9> PMID: 32517789

38. Wu X, Wu G, Zhang H, Peng X, Huang B, Huang M, et al. MiR-196b Promotes the Invasion and Migration of Lung Adenocarcinoma Cells by Targeting AQP4. *Technology in Cancer Research & Treatment*. 2021; 20:1533033820985868.
39. Li Y, Sun S, Han X. Down-regulation of AQP4 inhibits proliferation, migration and invasion of human breast cancer cells. *Folia Biol (Praha)*. 2016; 62(3):131–7. PMID: [27516192](https://pubmed.ncbi.nlm.nih.gov/27516192/)
40. Warth A, Muley T, Meister M, Herpel E, Pathil A, Hoffmann H, et al. Loss of aquaporin-4 expression and putative function in non-small cell lung cancer. *BMC cancer*. 2011; 11(1):1–9. <https://doi.org/10.1186/1471-2407-11-161> PMID: [21548930](https://pubmed.ncbi.nlm.nih.gov/21548930/)
41. Hu G, Wang R, Wei B, Wang L, Yang Q, Kong D, et al. Prognostic markers identification in glioma by gene expression profile analysis. *Journal of Computational Biology*. 2020; 27(1):81–90. <https://doi.org/10.1089/cmb.2019.0217> PMID: [31433208](https://pubmed.ncbi.nlm.nih.gov/31433208/)
42. Sun C, Gu Y, Chen G, Du Y. Bioinformatics Analysis of Stromal Molecular Signatures Associated with Breast and Prostate Cancer. *Journal of Computational Biology*. 2019; 26(10):1130–9. <https://doi.org/10.1089/cmb.2019.0045> PMID: [31180245](https://pubmed.ncbi.nlm.nih.gov/31180245/)
43. Fu T, Ji X, Bu Z, Zhang J, Wu X, Zong X, et al. Identification of key long non-coding RNAs in gastric adenocarcinoma. *Cancer Biomarkers*. 2020;(Preprint):1–13. <https://doi.org/10.3233/CBM-192389> PMID: [32176636](https://pubmed.ncbi.nlm.nih.gov/32176636/)
44. Zhang Y, Xu Y, Li Z, Zhu Y, Wen S, Wang M, et al. Identification of the key transcription factors in esophageal squamous cell carcinoma. *Journal of thoracic disease*. 2018; 10(1):148. <https://doi.org/10.21037/jtd.2017.12.27> PMID: [29600044](https://pubmed.ncbi.nlm.nih.gov/29600044/)
45. Yang X, Zhu S, Li L, Zhang L, Xian S, Wang Y, et al. Identification of differentially expressed genes and signaling pathways in ovarian cancer by integrated bioinformatics analysis. *OncoTargets and therapy*. 2018; 11:1457. <https://doi.org/10.2147/OTT.S152238> PMID: [29588600](https://pubmed.ncbi.nlm.nih.gov/29588600/)
46. Matsumoto S, Yamada Y, Narikiyo M, Ueno M, Tamaki H, Miki K, et al. Prognostic significance of platelet-derived growth factor-BB expression in human esophageal squamous cell carcinomas. *Anticancer research*. 2007; 27(4B):2409–14. PMID: [17695532](https://pubmed.ncbi.nlm.nih.gov/17695532/)
47. Vacic V, McCarthy S, Malhotra D, Murray F, Chou H-H, Peoples A, et al. Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature*. 2011; 471(7339):499–503. <https://doi.org/10.1038/nature09884> PMID: [21346763](https://pubmed.ncbi.nlm.nih.gov/21346763/)
48. Shi Z-Z, Shang L, Jiang Y-Y, Shi F, Xu X, Wang M-R, et al. Identification of genomic biomarkers associated with the clinicopathological parameters and prognosis of esophageal squamous cell carcinoma. *Cancer Biomarkers*. 2015; 15(6):755–61. <https://doi.org/10.3233/CBM-150517> PMID: [26406417](https://pubmed.ncbi.nlm.nih.gov/26406417/)