Identification of the Potential Biomarkers in Barrett'S Esophagus Derived Esophageal Adenocarcinoma

Nan Yi
The people's Hospital of Guangxi Zhuang Autonomous Region
Juan He
Youjiang Medical University For Nationalities
Xike Xie
Youjiang Medical University For Nationalities
Liexin Liang
The people's Hospital of Guangxi Zhuang Autonomous Region
Guowen Zuo
The people's Hospital of Guangxi Zhuang Autonomous Region
Mingyue Xiong
Baise People's Hospital
Yunxiao Liang
The people's Hospital of Guangxi Zhuang Autonomous Region
Tingzhuang Yi (ytz20070101@163.com)
Affiliated Hospital of YouJiang Medical University For Nationalities

Research Article

Keywords: Esophageal adenocarcinoma, Barrett's esophagus, differentially expressed genes, prognosis, bioinformatics analysis

Posted Date: December 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1127858/v1

Abstract

Background

Clinically, almost 50% of esophageal adenocarcinoma (EAC) originates from the progression of Barrett's esophagus (BE). EAC is often diagnosed at late stages and is related to dismal prognosis. However, there are still no effective strategies for stratification and therapy in BE derived EAC (BE-EAC).

Methods

Two public datasets (GSE26886 and GSE37200) are analyzed to identify differentially expressed genes (DEGs) between BE and EAC. A series of bioinformatics analyses are performed to explore potential biomarkers associated with BE-EAC.

Results

27 up- and 104 down-regulated genes are identified between GSE26886 and GSE37200. The GO and KEGG enrichment analysis indicate that the DEGs are highly involved in tumorigenesis. Subsequently, Weighted Gene Co-Expression Network Analysis (WGCNA) is performed to explore the potential genes related to BE-EAC, which are further validated in The Cancer Genome Atlas (TCGA) database. 5 up-regulated genes (*MYO1A*, *ACE2*, *COL1A1*, *LGALS4*, and *ADRA2A*) and 3 down-regulated genes (*AADAC*, *RAB27A*, and *P2RY14*) were found in BE-EAC. Among them, the expression level of *AADAC*, *ACE2* and *ADRA2A* show a significant correlation with patients' survival probability .

Conclusion

MYO1A, *ACE2*, *COL1A1*, *LGALS4*, *ADRA2A, AADAC*, *RAB27A*, and *P2RY14* are potential diagnostic and prognostic biomarkers in BE-EAC.

Background

Esophageal adenocarcinoma (EAC), the predominant subtype in the west, is one of the two main histological types of esophageal cancer, and the incidence of EAC has increased nearly six-fold over the last three decades[1, 2]. Long-term exposure to the acid, bile, and other stomach contents causes great injury of the squamous esophageal epithelium and increases the risk of developing Barrett's esophagus (BE) and later EAC[3, 4]. BE is the only precursor that has been verified involved in the tumorigenesis process of EAC. Individuals with BE are 30-125 times more likely to develop EAC than the general population, and almost 50% of EAC patients are progressed from BE[5]. Patients with BE must undergo regular endoscopic surveillance, as BE surveillance carries an improved prognosis[6]. Given the high cost of endoscopy and many patients developed EAC during endoscopic surveillance, stratification of BE patients is indispensable. Meanwhile, EAC is often diagnosed at late stages and is related to dismal prognosis. Although tremendous strategies have been taken in clinical therapy, including esophagectomy,

chemotherapy, and molecular targeted drugs treatment, the 5-year survival rate of EAC still remains less than 20%[7]. Therefore, it is necessary to explore potential targets for EAC diagnosis and therapy.

Several genes from genome-wide association studies have been identified as potential effectors in the development of BE to EAC. It has been reported that *ELF3*, *KLF5*, *GATA6*, *EHF*, *TTK*, *TPX2*, and *RAD54B* are important genetic regulators that play important roles in the pathogenesis and progression of BE to EAC[1, 8]. Spechler et al. reported that early *CDKN2A* (*P16*) loss or methylation and subsequent loss of *P53* in non-dysplastic BE might contribute to BE-EAC progression[9]. In addition, Dulak et al. claimed that *SMAD4*, *ARID1A*, *PIK3CA*, *SPG20*, *TLR4*, *ELMO1*, and *DOCK2* have significant impact on BE-EAC progression[10]. However, there are still no effective methods for stratification and therapy in BE and EAC.

Therefore, we analyzed two public datasets to identify differentially expressed genes (DEGs) among BE and EAC. Then, Weighted Gene Co-Expression Network Analysis (WGCNA) was performed to explore the potential genes related to BE-EAC. This study aims to screen potential genes that are responsible for BE-EAC progression.

Materials And Methods

Data Retrieving and Processing

The gene expression profiles of GSE26886[11] and GSE37200[12] were downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo). Finally, 20 BE samples and 21 EAC samples in the GSE26886; and 31 BE samples and 15 EAC samples in the GSE37200 were included in this study. EAC data were obtained from The Cancer Genome Atlas (TCGA) database, including 80 EAC samples and 11 normal samples.

Two R packages (GEOquery and limma) were used for the analysis of DEGs. The threshold for the DEGs was set as adjusted-P value <0.05 and $|\log_2$ fold change (FC)$| \geq 1$. Volcano plots and heat maps were drawn using R package “ggplot2” and “complexHeatmap”. Venn diagram was performed using the jvenn tool (http://jvenn.toulouse.inra.fr/app/example.html), and the overlaps represented the intersection between the two datasets.

Gene Ontology (GO) Analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Enrichment Analysis

To identify the function of DEGs, GO and KEGG analysis were performed using Metascape (metascape.org) database. GO is a commonly used bioinformatics tool that supply comprehensive information on gene function of individual genomic products based on defined features and is primarily divided into three parts, molecular function (MF), biological process (BP), and cellular component (CC). KEGG is a database resource for understanding high-level biological functions and utilities. We

determined that results were statistically significant at a level of adjusted-P < 0.05 and false discovery rate (FDR) < 0.05. Then, histograms and chord plots were generated with R package “GOplot”.

Weighted Gene Co-Expression Network Analysis (WGCNA)

GSE26886 was used to detect modules highly correlated with EAC, and WGCNA was performed using R package “WGCNA” and carried out on all genes. The scale-free topology of the networks was assessed for various values of the β shrinkage parameter, and we chose β =8 based on scale-free topology criterion. Finally, the dynamic tree cut algorithm was applied to the dendrogram for module identification with the mini-size of module gene numbers set as 50, and similar modules were merged following a height cutoff of 0.25. In the module-trait analysis, gene-trait significance (GS) value > 0.3 and module membership (MM) value > 0.55 were defined as a threshold[13]. Then, Venn diagram was performed to explore the trait-expression-related genes.

Exploration of Trait-expression-related Genes in TCGA database

Subsequently, the expression levels of trait-expression-related genes were estimated in TCGA database, a receiver operating characteristic (ROC) curve was performed to assess the diagnostic value of the genes, and survival analysis was also performed.

Statistical analysis

Statistical analysis was performed using R software (Version 4.1.0). Statistical comparisons between groups of normalized data were performed using the t-test or Mann-Whitney U-test according to the test condition. A difference with P< 0.05 was considered significant.

Results

Identification of DEGs in the EAC patients

To explore the variation of gene expression during BE to EAC, we downloaded datasets GSE26886 and GSE37200, and the DEGs between BE and EAC were analyzed, respectively (**Figure 1A-D**). Then, we sought for the overlapping DEGs between the two datasets, and found that 27 up- and 104 down-regulated genes were observed (**Figure 1E-F**).

GO and KEGG Pathway Enrichment Analysis

To explore the potential roles of DEGs between BE and EAC, GO and KEGG pathway enrichment analyses were performed. GO analysis showed that the up-regulated genes were mainly involved in biological processes (BP) associated with negative regulation of cell population proliferation, skeletal system development, blood vessel development, positive regulation of programmed cell death, and ossification (**Figure 2A**). In contrast, the down-regulated genes were mainly involved in BP associated with monocarboxylic acid metabolic process, digestion, thrombin-activated receptor signaling pathway,

response to zinc ion, and cellular response to fluid shear stress (**Figure 2B**). These results indicated that the DEGs were highly associated with epithelial-mesenchymal transition (EMT) and nutrition. KEGG analysis indicated that the up-regulated DEGs were primarily enriched in Pertussis, IL-17 signaling pathway, cytokine-cytokine receptor interaction, ECM-receptor interaction, protein digestion and absorption, and Amoebiasis (**Figure 2C**); while the down-regulated genes were enriched in chemical carcinogenesis, Amoebiasis, drug metabolism-other enzymes, steroid hormone biosynthesis, bile secretion, glycerophospholipid metabolism, and inflammatory mediator regulation of trp channels (**Figure 2D**). These results demonstrated that the DEGs were highly involved in tumorigenesis.

Identification of Key Modules by WGCNA

WGCNA analysis provides an overview of the transcriptomic organization, and the relationships between sets of genes with external, biological traits. To identify key modules related to clinical traits, WGCNA was performed by using GSE26886 dataset (**Figure 3A**). The power of β = 8 (scale-free R^2=0.90) was selected as the soft thresholding parameter to construct a scale-free network (**Figure 3B**). Similar module clustering was constructed by using dynamic hybrid cutting (threshold = 0.25). A total of 25 modules were identified (**Figure 3C**). The results in **Figure 3D** showed that the grey module was the highest positive module correlated to EAC (R^2 = 0.86,P = $2e^{-12}$) and was highly negative correlated to BE (R^2 = 0.86,P = $2e^{-12}$). **Figure 3E** showed gene significance for BE and EAC in grey module.

In the module-trait analysis, we intersected the trait-related genes in grey module highly associated with EAC and 131 DEGs generated from expression difference analysis, and finally extracted 27 trait-expression-related genes for the following analysis (**Figure 3F, Table S1-2**). These results showed that the 27 trait-expression-related genes were significantly correlated with the pathogenesis of EAC.

Exploration of Trait-expression-related Genes in TCGA database

Next, further validation and exploration were conducted among the 27 trait-expression-related genes in TCGA database. *MYO1A*, *ACE2*, *COL1A1*, *LGALS4*, and *ADRA2A* were significantly up-regulated in EAC tissue; while *AADAC*, *RAB27A*, and *P2RY14*were abnormally down-expressed in EAC tissue, which indicated that these genes were repeatable in EAC (**Figure 4A**). Subsequently, ROC curves were performed to estimate the diagnostic value in EAC, and the result showed that the genes mentioned above had good diagnostic properties (**Figure 4B**). Later, survival analysis was performed to explore the prognostic value of the 8 genes. Low- *AADAC* and *ACE2* expressionwere significantly correlated with poor progress-free interval (PFI); while low- *ADRA2A* expression was associated with poor overall survival (OS) and disease-specific survival (DSS). These results further illustrated that *AADAC*, *ACE2*,and *ADRA2A* contribute to EAC pathogenesis and progression.

Discussion

Currently, the pathogenesis of BE-EAC is still unclear, and the disease stratification and treatment are also limited. In the present study, we identified 27 up- and 104 down-regulated DEGs in two public datasets,

and the results from GO and KEGG analysis indicated that the DEGs were highly associated with tumorigenesis. Subsequently, 27 trait-expression-related genes highly correlated with EAC were screened out by WGCNA. *MYO1A*, *ACE2*, *COL1A1*, *LGALS4*, *ADRA2A, AADAC*, *RAB27A*, and *P2RY14* were also abnormally regulated in TCGA database and represented good diagnostic properties. Surprisingly, we found that *AADAC*, *ACE2*, and *ADRA2A* were correlated with EAC prognosis.

Previous studies showed that *COL1A1*, *RAB27A*, and *P2RY14* were identified as the potential biomarker for esophageal squamous cell cancer (ESCC) and *RAB27A* associated with immune infiltration in ESCC[14-17]. However, there are no further studies to verify their effects on EAC.

To the best of our knowledge, our study, for the first time, screened out 5 genes related to EAC. *MYO1A* is most highly expressed in the digestive tract, and it is associated with stomach adenocarcinoma and colon cancer[18, 19]. *ACE2*, the receptor of COVID-19, is aberrantly expressed in many tumors[20]. It is reported that *LGALS4*, a β-galactoside binding protein, is correlated with prognosis in urothelial carcinoma of bladder and is also a tumor marker in serum immunoassay determination of colorectal carcinoma[21, 22]. *ADRA2A* is thought to be involved in the progression of multiple cancer and can inhibit the activation of PI3K/Akt/mTOR pathway[23]. *AADAC* is a kind of serine hydrolase widely involved in the hydrolysis of drugs and associated with poor prognosis in stomach adenocarcinoma[24, 25]. More future studies are needed to gain more insights into these genes.

Nevertheless, our study also had several limitations. Firstly, further experiments were required to verify these results. Secondly, the lack of BE cases in TCGA database prevented us from comparing EAC and BE, which might impact the outcomes.

Conclusion

MYO1A, *ACE2*, *COL1A1*, *LGALS4*, *ADRA2A, AADAC*, *RAB27A*, and *P2RY14* could be potential novel diagnostic and prognostic biomarkers in BE-EAC. In addition, *AADAC*, *ACE2*, and *ADRA2A* could contribute to EAC progression. Although further validation is still needed, we provide useful and novel information to explore the potential candidate genes for BE-EAC prognosis and therapeutic options.

Abbreviations

EAC: Esophageal adenocarcinoma

BE: Barrett's esophagus

DEGs: Differentially expressed genes

WGCNA: Weighted Gene Co-Expression Network Analysis

GEO: Gene Expression Omnibus

TCGA: The Cancer Genome Atlas

GO: Gene Ontology

KEGG: Kyoto Encyclopedia of Genes and Genomes

MF: molecular function

BP: biological process

CC: cellular component

Declarations

Acknowledgements

The authors appreciate study investigators and staff who participated in this study.

Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author information

Nan Yi, Juan He, and Xike Xie contributed equally to this paper.

Affiliations

Department of Gastroenterology, The people's Hospital of Guangxi Zhuang Autonomous Region, Nanning, Guangxi, 530021, P. R. China

Nan Yi, Liexin Liang, Guowen Zuo, Yunxiao Liang

Youjiang Medical University For Nationalities, Baise, Guangxi, 533000, P. R. China

Juan He, Xike Xie

Department of Hematology, Baise People's Hospital, Baise, Guangxi, 533000, P. R. China

Mingyue Xiong

Department of Oncology, Affiliated Hospital of YouJiang Medical University For Nationalities, Baise, Guangxi, 533000, P. R. China

Tingzhuang Yi

Contributions

Nan Yi, Juan He, and Xike Xie contributed equally to this paper. Nan Yi, Juan He, and Xike Xie analyzed the study data, helped draft the manuscript, made critical revisions of the manuscript. Liexin Liang, Guowen Zuo, and Mingyue Xiong assisted with data collection and the analysis. Yunxiao Liang and Tingzhuang Yi supervised the research and edited the manuscript. All authors contributed to the article and approved the submitted version.

Corresponding author

Yunxiao Liang[1], Tingzhuang Yi[4]

Ethics declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Availability of Data and Materials

Publicly available datasets were analyzed in this study. This data can be found here: TCGA database and GEO database, accession number: GSE26886 and GSE37200.

References

1. Ma S, Zhou B, Yang Q, Pan Y, Yang W, Freedland SJ, Ding LW, Freeman MR, Breunig JJ, Bhowmick NA, et al: **A Transcriptional Regulatory Loop of Master Regulator Transcription Factors, PPARG, and Fatty Acid Synthesis Promotes Esophageal Adenocarcinoma.** *Cancer Res* 2021, **81:**1216-1229.
2. Deshpande NP, Riordan SM, Castano-Rodriguez N, Wilkins MR, Kaakoush NO: **Signatures within the esophageal microbiome are associated with host genetics, age, and disease.** *Microbiome* 2018, **6:**227.
3. Hvid-Jensen F, Pedersen L, Drewes AM, Sorensen HT, Funch-Jensen P: **Incidence of adenocarcinoma among patients with Barrett's esophagus.** *N Engl J Med* 2011, **365:**1375-1383.
4. Kaakoush NO, Castano-Rodriguez N, Man SM, Mitchell HM: **Is Campylobacter to esophageal adenocarcinoma as Helicobacter is to gastric adenocarcinoma?** *Trends Microbiol* 2015, **23:**455-462.

5. Desai TK, Krishnan K, Samala N, Singh J, Cluley J, Perla S, Howden CW: **The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: a meta-analysis.** *Gut* 2012, **61:**970-976.
6. El-Serag HB, Naik AD, Duan Z, Shakhatreh M, Helm A, Pathak A, Hinojosa-Lindsey M, Hou J, Nguyen T, Chen J, Kramer JR: **Surveillance endoscopy is associated with improved outcomes of oesophageal adenocarcinoma detected in patients with Barrett's oesophagus.** *Gut* 2016, **65:**1252-1260.
7. Lv J, Guo L, Wang JH, Yan YZ, Zhang J, Wang YY, Yu Y, Huang YF, Zhao HP: **Biomarker identification and trans-regulatory network analyses in esophageal adenocarcinoma and Barrett's esophagus.** *World J Gastroenterol* 2019, **25:**233-244.
8. Kumar S, Buon L, Talluri S, Roncador M, Liao C, Zhao J, Shi J, Chakraborty C, Gonzalez G, Tai YT, et al: **Integrated genomics and comprehensive validation reveal drivers of genomic evolution in esophageal adenocarcinoma.** *Commun Biol* 2021, **4:**617.
9. Spechler SJ, Fitzgerald RC, Prasad GA, Wang KK: **History, molecular mechanisms, and endoscopic treatment of Barrett's esophagus.** *Gastroenterology* 2010, **138:**854-869.
10. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, Bandla S, Imamura Y, Schumacher SE, Shefler E, et al: **Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity.** *Nat Genet* 2013, **45:**478-486.
11. Wang Q, Ma C, Kemmner W: **Wdr66 is a novel marker for risk stratification and involved in epithelial-mesenchymal transition of esophageal squamous cell carcinoma.** *Bmc Cancer* 2013, **13:**137.
12. Myers AL, Lin L, Nancarrow DJ, Wang Z, Ferrer-Torres D, Thomas DG, Orringer MB, Lin J, Reddy RM, Beer DG, Chang AC: **IGFBP2 modulates the chemoresistant phenotype in esophageal adenocarcinoma.** *Oncotarget* 2015, **6:**25897-25916.
13. Wu C, Zhou Y, Wang M, Dai G, Liu X, Lai L, Tang S: **Bioinformatics Analysis Explores Potential Hub Genes in Nonalcoholic Fatty Liver Disease.** *Front Genet* 2021, **12:**772487.
14. Wang Z, Chen M, Qiu Y, Yang Y, Huang Y, Li X, Zhang W: **Identification of potential biomarkers associated with immune infiltration in the esophageal carcinoma tumor microenvironment.** *Biosci Rep* 2021, **41**.
15. Lv J, Guo L, Wang JH, Yan YZ, Zhang J, Wang YY, Yu Y, Huang YF, Zhao HP: **Biomarker identification and trans-regulatory network analyses in esophageal adenocarcinoma and Barrett's esophagus.** *World J Gastroenterol* 2019, **25:**233-244.
16. Yu F, Wu W, Liang M, Huang Y, Chen C: **Prognostic Significance of Rab27A and Rab27B Expression in Esophageal Squamous Cell Cancer.** *Cancer Manag Res* 2020, **12:**6353-6361.
17. Chen FF, Zhang SR, Peng H, Chen YZ, Cui XB: **Integrative genomics analysis of hub genes and their relationship with prognosis and signaling pathways in esophageal squamous cell carcinoma.** *Mol Med Rep* 2019, **20:**3649-3660.
18. Bai Y, Wei C, Zhong Y, Zhang Y, Long J, Huang S, Xie F, Tian Y, Wang X, Zhao H: **Development and Validation of a Prognostic Nomogram for Gastric Cancer Based on DNA Methylation-Driven Differentially Expressed Genes.** *Int J Biol Sci* 2020, **16:**1153-1165.

19. McIntosh BB, Ostap EM: **Myosin-I molecular motors at a glance.** *J Cell Sci* 2016, **129:**2689-2695.
20. Chai P, Yu J, Ge S, Jia R, Fan X: **Genetic alteration, RNA expression, and DNA methylation profiling of coronavirus disease 2019 (COVID-19) receptor ACE2 in malignancies: a pan-cancer analysis.** *J Hematol Oncol* 2020, **13:**43.
21. Ferlizza E, Solmi R, Miglio R, Nardi E, Mattei G, Sgarzi M, Lauriola M: **Colorectal cancer screening: Assessment of CEACAM6, LGALS4, TSPAN8 and COL1A2 as blood markers in faecal immunochemical test negative subjects.** *J Adv Res* 2020, **24:**99-107.
22. Ding Y, Cao Q, Wang C, Duan H, Shen H: **LGALS4 as a Prognostic Factor in Urothelial Carcinoma of Bladder Affects Cell Functions.** *Technol Cancer Res Treat* 2019, **18:**1078144249.
23. Wang W, Guo X, Dan H: **alpha2A-Adrenergic Receptor Inhibits the Progression of Cervical Cancer Through Blocking PI3K/AKT/mTOR Pathway.** *Onco Targets Ther* 2020, **13:**10535-10546.
24. Zou Q, Lv Y, Gan Z, Liao S, Liang Z: **Identification and Validation of a Malignant Cell Subset Marker-Based Polygenic Risk Score in Stomach Adenocarcinoma Through Integrated Analysis of Bulk and Single-Cell RNA Sequencing Data.** *Front Cell Dev Biol* 2021, **9:**720649.
25. Wu C, Wu Z, Tian B: **Five gene signatures were identified in the prediction of overall survival in resectable pancreatic cancer.** *Bmc Surg* 2020, **20:**207.

Figures

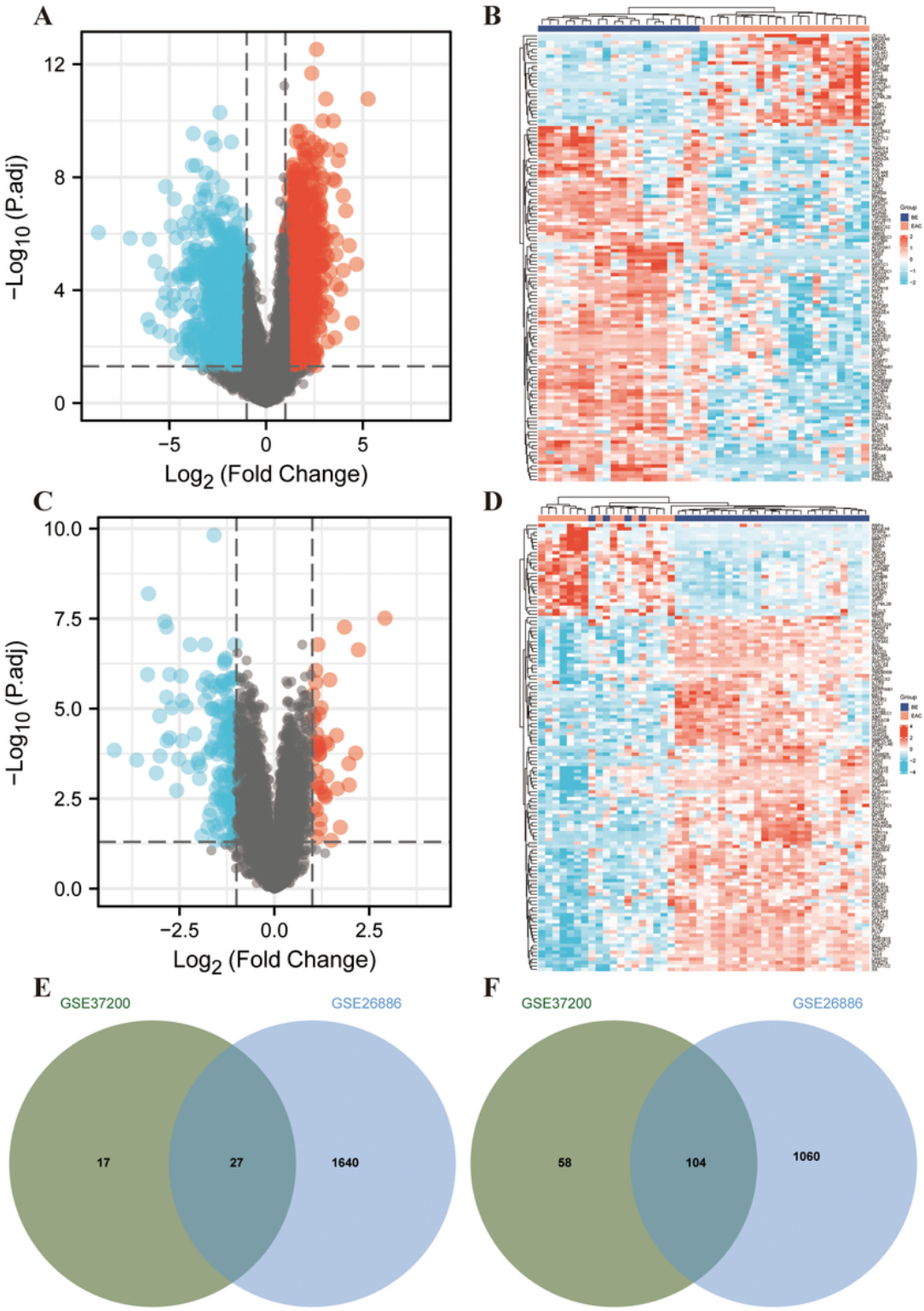

Figure 1

Identification of differentially expressed genes (DEGs) between BE and EAC. (A) The volcano plot of DEGs in GSE26886; (B) The Heatmap of DEGs in GSE26886; (C) The volcano plot of DEGs in GSE37200; (D) The Heatmap of DEGs in GSE37200; (E-F) Venn diagrams displayed the overlapping DEGs of up-(E) and down-(F) regulated genes between BE and EAC.

Figure 2

GO and KEGG pathway enrichment analysis. (A-B) GO analysis of up-(A) and down-(B) regulated DEGs; (C-D) KEGG analysis of up-(C) and down-(D) regulated DEGs.

Figure 3

WGCNA to identify trait-related modules and genes. (A) Sample dendrogram and trait heat map; (B) Calculating soft-thresholding power; Left: scale-free fit indices using different soft-thresholding powers; Right: mean connectivity using different soft-thresholding powers; (C) The dendrogram clustered by Dynamic Tree Cut algorithm; (D) The heatmap profiling the correlations between module eigengenes and the clinical characteristics; (E) Scatter plot of gene significance in grey module; Left: BE; Right: EAC; (F) Venn diagrams displayed the overlapping genes between trait-related genes in grey module and DEGs.

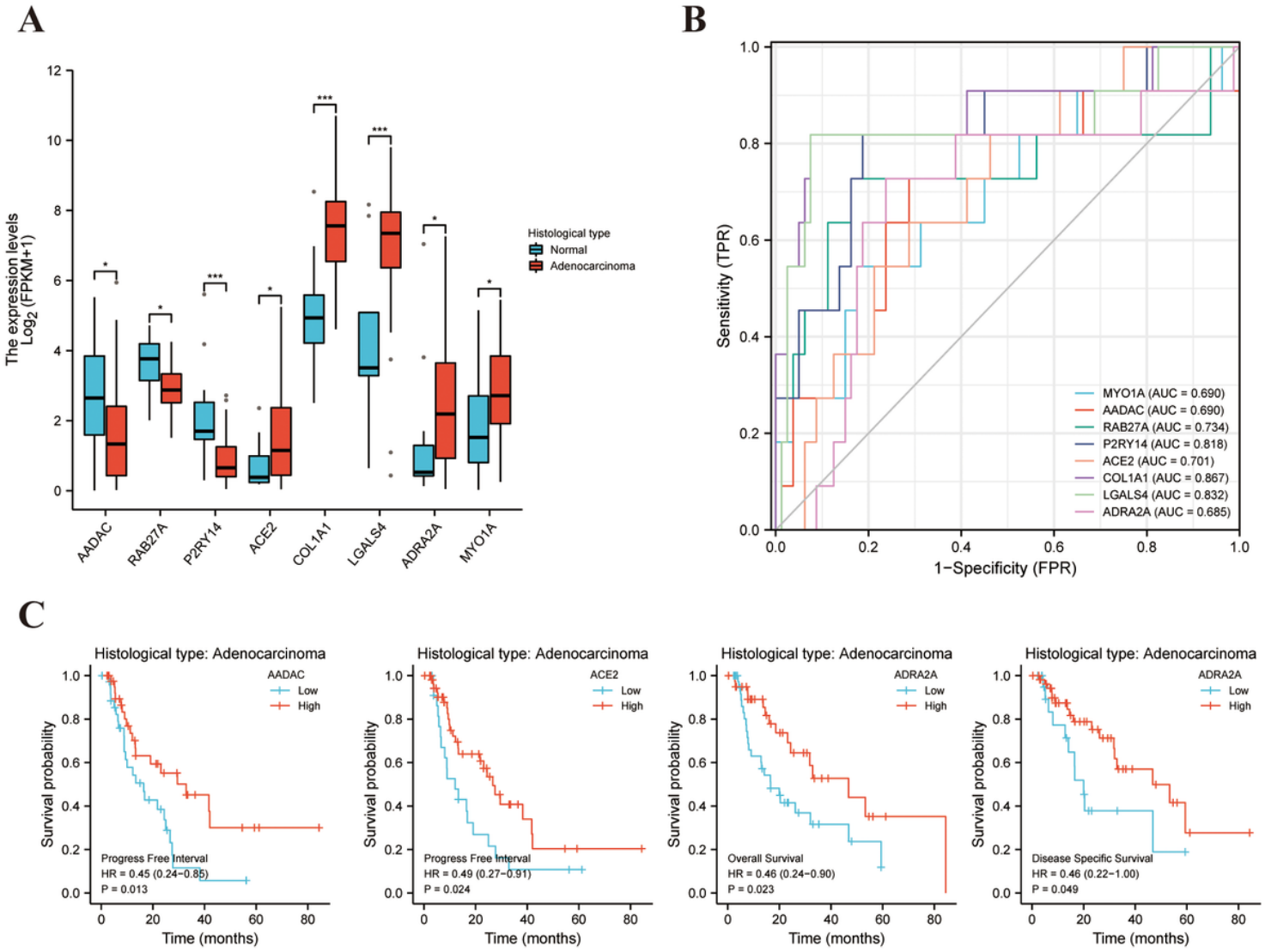

Figure 4

Further validation in TCGA database (*p < 0.05, **p<0.01, ***p<0.001). (A) Box plot assessing the expression of 8 genes in TCGA database; (B) ROC curve for the 8 genes; (C) Survival plots of AADAC, ACE2, and ADRA2A in overall survival, disease specific survival, and progress free survival.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xlsx
- TalbeS2.xlsx