Identification and functional analysis of novel biomarkers in adenoid cystic carcinoma

Jiang Su[1,#], Wei Zhou[1,#], Huajie Yuan[1,#], Hui Wang[2,#], Hua Zhang[1,*]

[1] Department of Otolaryngology-Head and Neck Surgery, The First Affiliated Hospital of Xin Jiang Medical University, Urumqi, China
[2] Xinjiang Urumqi Eye, Ear, Nose and Throat Hospital, Urumqi, China

# Contributed equally to this work

ARTICLE INFO

ABSTRACT

To explore the key genes associated with the development and progression of adenoid cystic carcinoma (ACC), with the aim of exploring novel biomarkers that can better diagnose ACC, and thus better improve patient prognosis. The GSE59701 and GSE88804 datasets (containing transcriptomic data for a total of 19 normal samples and 20 tumor samples) were downloaded from the Gene Expression Omnibus (GEO) database, combined into one dataset and used to remove batch effects using the SVA algorithm. A total of 711 differentially expressed genes (DEGs) were screened by using the limma package. The metscape database (www. metascape. org) was used for gene ontology (GO) analysis and gene-specific Kyoto Genome Encyclopedia (KEGG) pathway analysis, which showed that the main enriched pathways of DEGs were kinase activity, fertility properties, extracellular matrix structural components, tryptophan metabolism, cancer pathway, PI3K-Akt signaling pathway. The STRING database was used to construct protein-protein interaction (PPI) networks for DEGs, and Cytoscape software was used to visualize the result. Lasso regression and SVM algorithm screened 3 key genes: GABBR1, ITGA9 and MLKL. The results of GSEA on key genes showed that they are mainly enriched in pathways such as cell cycle, and taste transduction mechanisms. CIBERSORT algorithm was used to analyze immune cell infiltration, the "corrplot" package was used to analyze the interaction relationships between immune cells. Spearman correlation analysis demonstrated that GABBR1, ITGA9 and MLKL were all strongly correlated with differentially expressed immune cells. Moreover, correlation analysis of key and differentially regulated genes showed that GABBR1 and MLKL were significantly correlated with MYB and TP53, respectively. In conclusion, GABBR1, ITGA9 and MLKL affect the progression of ACC, where GABBR1 and MLKL may regulate ACC through MYB and TP53, and the relationship between ITGA9 and ECM and PI3K-Akt may have some influence on the development of ACC.

Introduction

Adenoid cystic carcinoma (ACC) is a heterogeneous secretory gland malignancy originating from the intercalated duct reserve cells of the salivary gland (1,2). The prognosis of ACC tends to be unsatisfactory, owing to the characteristics of perineural infiltration, an infiltrating growth pattern and a relatively high probability of distant metastasis (3). Approximately 40-50% of patients experience local regional recurrence and 60% experience distant metastases (4). The research results in proteomics, genomics and clinical data have shown that ACC-I and ACC-II have positive therapeutic implications which the percentage of patients with ACC-I subtype is about 37% and the percentage of patients with ACC-II subtype is about 63% (1).

Currently, it has been demonstrated that transmembrane receptor (NOTCH1), vascular endothelial growth factor (VEGF) and proto-oncogene (MYB), nerve growth factor (NGF) involved in the differentiation, proliferation, invasion and metastasis of ACC cells (5-7). And some therapies have been developed on this basis (1,8,9). Nevertheless, cytotoxic drugs and VEGF inhibitor therapies can produce significant toxicity, and systemic therapy has limitations (10). Therefore, further research on the molecular and pathological mechanisms is critically required.

The dramatic growth of bioinformatics and the rapid refinement of online databases and analytical tools have driven a vast amount of scientific research based on data analysis (11-13). So far, research on ACC has been limited to the identification of key factors that regulate tumor proliferation, recurrence and metastasis, but these findings are far from being relevant to the difficulties encountered in the treatment of ACC (14-16). By reason of the foregoing, our study aimed to further explore the molecular biology and unclear mechanisms of ACC and to provide some theoretical support for the development of novel biomarkers, thus improving patient prognosis.

Materials and Methods

Data resource

The mRNA expression profiling datasets of Adenoid cystic carcinoma-related data set GSE59701 and GSE88804 both based on GPL6244 (HuGene-1_0-st) Affymetrix Human Gene 1.0 ST Array (transcript (gene) ver-

sion) were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The GSE59701 dataset contained 12 human salivary glands with adenoid cystic cancer and 12 matched normal adjacent tissues. The GSE88804 dataset contained 13 surgical samples of ACC, 2 ACC xenografts, and 7 normal salivary gland tissues (NSGs). The GSE59701 and GSE88804 datasets were merged and normalized using the "sva" R package, using the normalized data for subsequent analysis.

Recognition and analysis of differentially expressed genes

On the basis of P < 0.05 and $|Log_2FoldChange(FC)| > 1.5$, we first excavated the differentially expressed genes (DEGs) in the merged dataset through 'limma' package. Then, we applied the Metascape database (www.metascape.org) and annotated and visualized the DEGs. Gene ontology (GO) analysis and Kyoto Encyclopedia of Genomic Genes (KEGG) for enrichment pathway analysis of DEGs. The critical value of enrichment was set as Min overlap≥ 3 and P ≤0.01. The STRING database was used to construct protein-protein interaction (PPI) networks for DEGs, and Cytoscape software was used to visualize the result.

Identification and analysis of key genes

Lasso regression algorithm (" glmnet "software package) and SVM feature selection algorithm (" e1071" software package) were used to screen the ACC key genes in 711 DEGs, then the intersection genes of the two algorithms were selected. The differences in the KEGG pathway between the high and low-expression groups of key genes were investigated using GSEA software (https://www.broadinstitute.org/gsea) (substitutions: 1000, type: phenotype). RcisTarget was used to identify enriched transcription factor (TF) binding motifs and candidate TFS. Motifs were annotated to TFs based on pathway enrichment analysis, and those with NES ≥3 were retained.

Immune infiltration analysis

RNA-seq data from the normal and tumor groups were analyzed using the CIBERSORT algorithm to calculate the relative proportion of immune infiltrating cells. The interaction between immune cells was analyzed using the ""corrplot"" package. The relative content of immune cells was plotted using the "vioplot" package, and Spearman correlation analysis was performed between the level of immune cell infiltration and gene expression. P < 0.05 was considered to be statistically different.

Key genes and disease regulatory genes in ACC

Pearson correlation analysis was conducted to determine the relationship between key genes and differentially expressed disease regulatory genes in adenoid cystic carcinoma.

Results

Recognition and analysis of differentially expressed genes

We used the rectified data for differential expression analysis and obtained 711 DEGs, including 372 up-regulated genes and 339 down-regulated genes (Figure 1). Then, we executed a functional enrichment analysis of the role of 711 DEGs in ACC. As is demonstrated in Figure 2A-B, the DEGs mainly focused on the kinase activity, developmental growth, extracellular matrix structural constituent, Tryptophan metabolism, Pathways in cancer, and PI3K-Akt signaling pathway. Furthermore, the PPI network of 711 DEGs was painted through the STRING database (Figure 2C).

Identification of key genes

Aiming to identify the biomarkers of ACC, we performed feature screening by LASSO regression, and the regression coefficient path diagram demonstrated the corresponding coefficients for each gene, and the optimal threshold of 10 gene coefficients was selected based on the cross-validation curve (Figure 3a-b). Simultaneously,
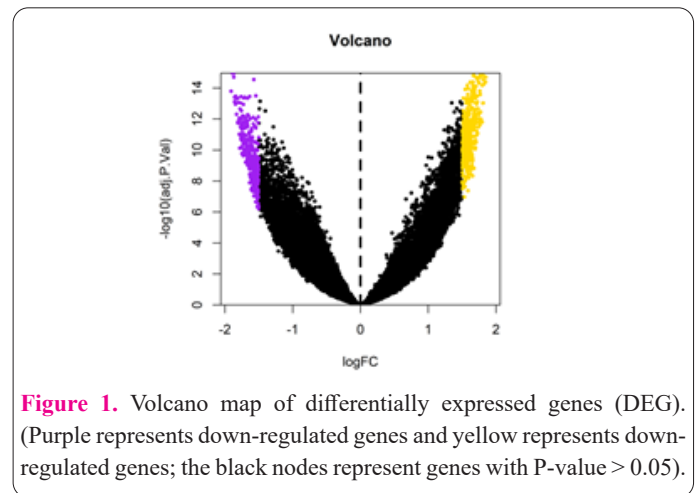

Figure 1. Volcano map of differentially expressed genes (DEG). (Purple represents down-regulated genes and yellow represents down-regulated genes; the black nodes represent genes with P-value > 0.05).
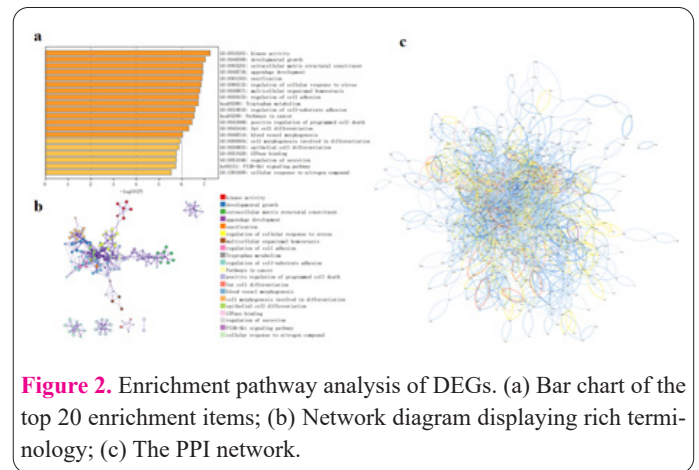

Figure 2. Enrichment pathway analysis of DEGs. (a) Bar chart of the top 20 enrichment items; (b) Network diagram displaying rich terminology; (c) The PPI network.
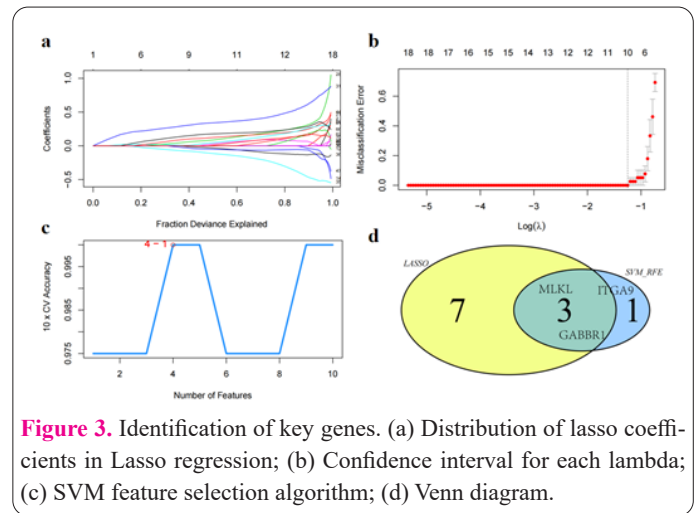

Figure 3. Identification of key genes. (a) Distribution of lasso coefficients in Lasso regression; (b) Confidence interval for each lambda; (c) SVM feature selection algorithm; (d) Venn diagram.

we used the SVM-RFE algorithm to evaluate key genes (Figure 3c). As shown in Figure 3d, we plotted Venn diagrams to screen for co-genes to both algorithms, the GAB-BR1, ITGA9 and MLKL will be used as key genes for subsequent study.

GSEA of key genes

GSEA for analysis of KEGG pathway enriched by high and low expression groups of key genes. As shown in Figure 4, high expression of GABBR1 is mainly enriched in the KEGG CELL CYCLE, KEGG SPLICEOSOME and KEGG TASTE TRANSDUCTION; high expression of ITGA9 is mainly enriched in the KEGG TASTE TRANS-DUCTION, KEGG CELL CYCLE and KEGG ADHE-RENS JUNCTION; high expression of MLKL is mainly enriched in the KEGG GLYCEROLIPID METABOLISM, KEGG INTESTINAL IMMUNE NETWORK FOR IGA PRODUCTION and KEGG ARGININE AND PROLINE METABOLISM.

Immune infiltration analysis

The immune microenvironment is made up of immune cells, extracellular matrix, various growth factors, inflammatory factors and specific physicochemical characteristics, which have an essential influence on the diagnosis and clinical treatment sensitivity of tumors. Figure 5 shows the T cells CD8, NK cells activated, Macrophages M0 and T cells CD4 native (P=0.052) were significantly increased in the tumor group, compared with the normal group.

Immune cell infiltration and key genes

To further explore the potential molecular mechanisms of key genes affecting the progression of adenoid cystic carcinoma, we analyzed the relationship between key genes and immune cell infiltration. Spearman correlation analysis showed that GABBR1, ITGA9 and MLKL were strongly correlated with differentially immune cells (Figure 6). Among them, GABBR1 positively correlated with T cells CD4 native (correlation coefficient = 0.44) and negatively correlated with Monocytes (correlation coefficient = -0.5); ITGA9 was positively correlated with NK cells activated (correlation coefficient = 0.45) and negatively correlated with Macrophages M1 (correlation
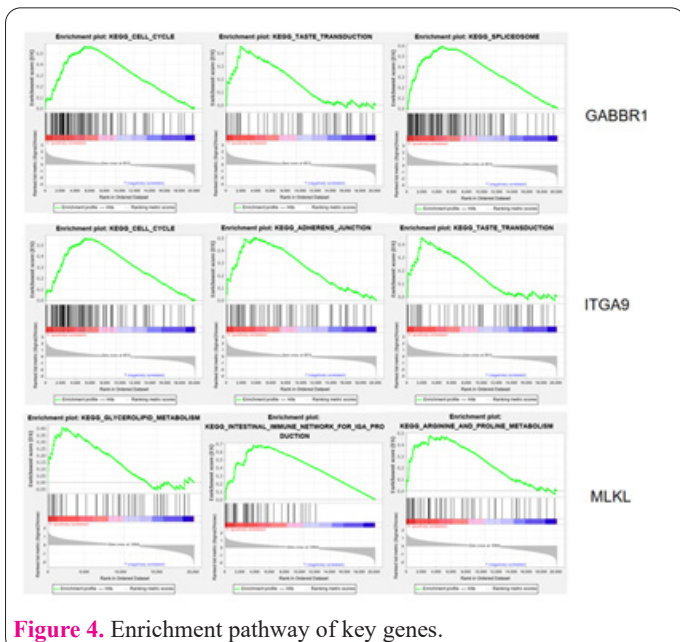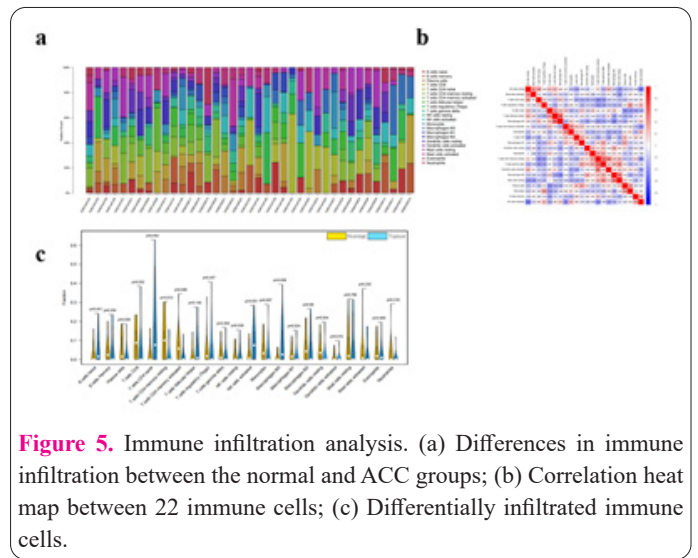

Figure 5. Immune infiltration analysis. (a) Differences in immune infiltration between the normal and ACC groups; (b) Correlation heat map between 22 immune cells; (c) Differentially infiltrated immune cells.
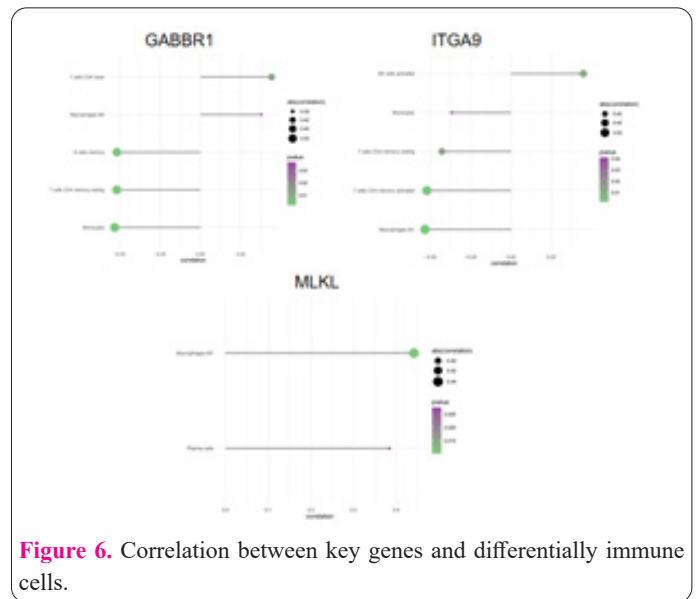

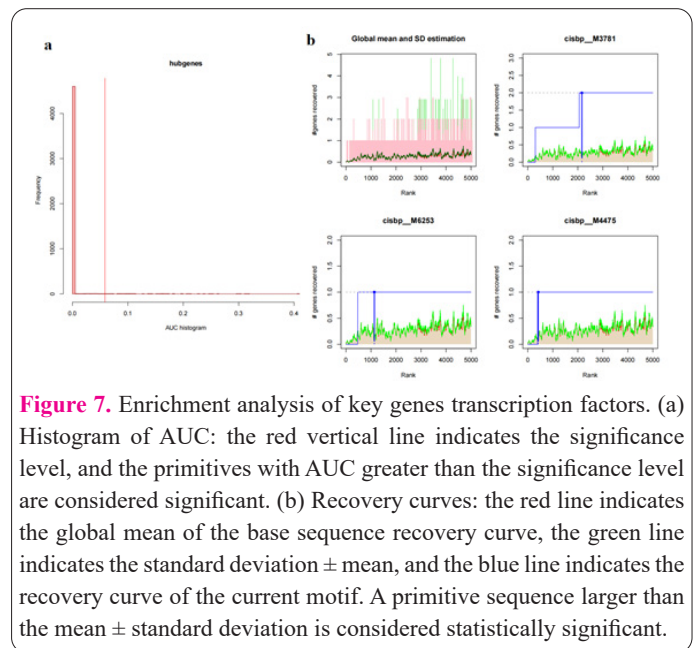**Figure 6.** Correlation between key genes and differentially immune cells.

Figure 7. Enrichment analysis of key genes transcription factors. (a) Histogram of AUC: the red vertical line indicates the significance level, and the primitives with AUC greater than the significance level are considered significant. (b) Recovery curves: the red line indicates the global mean of the base sequence recovery curve, the green line indicates the standard deviation ± mean, and the blue line indicates the recovery curve of the current motif. A primitive sequence larger than the mean ± standard deviation is considered statistically significant.

coefficient = -0.5); MLKL was positively correlated with Macrophages M1 (correlation coefficient = 0.44).

Enrichment analysis of key genes transcription factors

RcisTarge was able to identify over-expressed transcription factors (TF) in the gene list combined with motifs

Figure 4. Enrichment pathway of key genes.

and perform enrichment analysis. As shown in Figure 7, the highest NES of Motif-TF annotated as cisbp__M0564 was 16.3, and 2 key genes were enriched in this motif. It indicates that the transcription-binding structural domain was the main regulator of the key genes.

Correlation analysis of key genes and disease regulatory genes

Disease regulatory genes are involved in disease occurrence and progression. Thus, we identified disease regulatory genes such as BRCA1, CDH1, CTNNB1, MET, MUC1, MYB, NFIB, and TP53 as significantly different in the normal and tumor groups (Figure 8a). Correlation analysis of key genes and differentially disease regulatory genes showed that GABBR1 was significantly positively correlated with MYB (correlation coefficient = 0.9); MLKL was significantly negatively correlated with TP53 (correlation coefficient = -0.89) (Figure 8b). The above results suggest that GABBR1 and MLKL may be involved in the regulation of ACC by MYB and TP53, respectively.

Discussion

As a result of the high risk of invasion of adjacent soft tissues, bones and nerves, ACC is considered to be a highly heterogeneous malignancy that is almost impossible to cure (2,17,18). Currently, various types of treatments such as surgical resection, radiotherapy and chemotherapeutic drugs are used to control the disease but do not prevent its recurrence (3,19). Identifying potential ACC biomarkers and investigating their role in the origin, progression and metastasis of the disease is critical to advancing the status of ACC research and treatment. Previously, several reports stated that MYB as a key transcriptional regulator affecting clinical outcomes in patients with ACC (20-22), and also found that NOTCH1 and RUNX1 regulate a series of key ACC-related genes (23).

In this study, GSE59701 and GSE88804 from Gene Expression Omnibus (GEO) database were used for gene expression analysis. Firstly, 711 DEGs were identified, which were mainly enriched in kinase activity, fertility properties, extracellular matrix structural components, tryptophan metabolism, cancer pathway and PI3K-Akt signaling pathway. Similarly, the DEG of ACC was found to be enriched in the cancer pathway and PI3K-Akt signaling pathway in a study by Tang et al. (24) on ACC. Furthermore, we obtained GABBR1, ITGA9 and MLKL as ACC key genes by using LASSO and SVM algorithm screening. Subsequently, we further explored the biological pathways and molecular mechanisms of GABBR1, ITGA9 and MLKL through pathway enrichment, immune infiltration analysis and association with disease genes.

Wu et al. (25) suggested that ITGA9 is a potential biomarker and therapeutic target in different tumors. ITGA9 regulates the PI3K-Akt pathway through a reaction between the cell membrane and the extracellular matrix (ECM). Moreover, cell-ECM interactions play a vital role in ACC biological processes, such as cell adhesion, cell contact and gene expression leading to cell migration, proliferation and differentiation. ECM components accumulated in the cell interstitial space give rise to typical morphological structures in the ACC (26). Consequently, based on the results of this study and previous reports, it is reasonable to speculate that the relationship between ITGA9
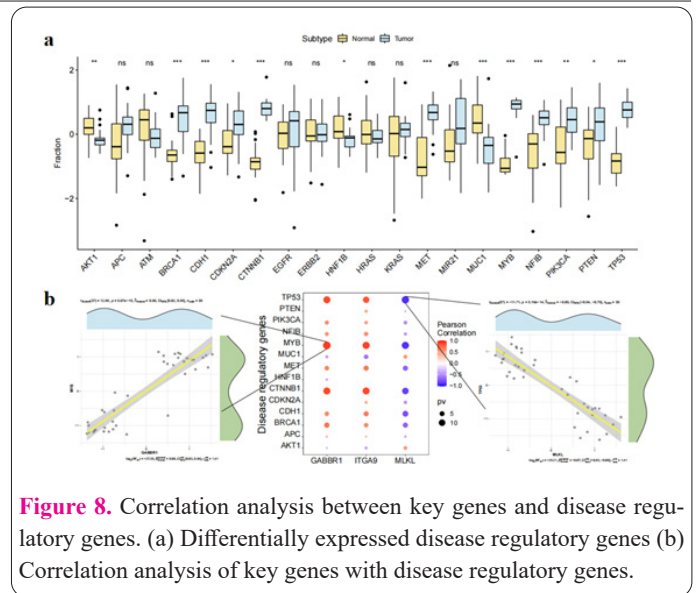

Figure 8. Correlation analysis between key genes and disease regulatory genes. (a) Differentially expressed disease regulatory genes (b) Correlation analysis of key genes with disease regulatory genes.

and ECM as well as PI3K-Akt may have a non-negligible impact on the development of ACC, suggesting that continued exploration of the relationship between the three in future studies is essential for a closer understanding of the development of ACC.

Correlation analysis of key genes and differentially expressed disease genes in ACC revealed that GABBR1 was significantly and positively correlated with MYB. MYB is an oncogene located on chromosome 6. Several reports stated that MYB fusions with the NFIB gene on chromosome 9 can increase MYB oncogene expression (27,28), and MYB-NFIB fusions often occur in ACC. It demonstrates that the activated Myb transcription factor encoded by MYB is the driver oncogene in tumors, and MYB activation is thought to underlie the development of ACC (29-31). We, therefore, hypothesized that overexpression of GABBR1 could facilitate similarly high expression of MYB, thus being a driving factor in the development of ACC. MLKL is a mixed-spectrum kinase structural domain-like protein that may influence cancer development and metastasis through necroptosis-dependent and independent functions (32). Nevertheless, there are no reports investigating the effects of MLKL in ACC. Somatic mutations in the TP53 are the most commonly occurring genetic variant in cancers and a potential prognostic biomarker, as well as a target for drug intervention. In animal models, defective TP53 can induce cancer progression, while restoration of TP53 function would enable the regression of formed tumors (33,34). This study demonstrated that significant negative correlation between MLKL and TP53, thus we hypothesized that MLKL could facilitate the progression of ACC.

In summary, GABBR1, ITGA9 and MLKL may have an impact on ACC through different biological pathways, respectively. Therefore, they can be used as novel biomarkers for diagnosing the disease. Among them, GABBR1 promotes the progression of ACC by promoting the expression of oncogene MYB; the relationship between ITGA9 and ECM and PI3K-Akt may have an impact on the development of ACC; and MLKL may play a role in promoting tumor progression in the process of ACC by inhibiting the expression of TP53.

This study had several limitations. The interaction studies and identification of pivotal genes at the protein level for DEG are incomplete and require further analysis.

Meanwhile, there is a lack of validation of the expression profile of key genes on clinical samples. Hereafter, more emphasis should be placed on the relationship between identified key genes and related signaling pathways, and the specific molecular mechanisms of action should be studied in depth to provide a solid theoretical basis and experimental support for the clinical diagnosis and subsequent treatment of ACC.

Conflict of Interests
The authors declared no conflict of interest.

References

1. de Sousa LG, Neto FL, Lin J, Ferrarotto R. Treatment of Recurrent or Metastatic Adenoid Cystic Carcinoma. Curr Oncol Rep 2022; 24(5): 621-631.

2. Ellington CL, Goodman M, Kono SA, et al. Adenoid cystic carcinoma of the head and neck: Incidence and survival trends based on 1973-2007 Surveillance, Epidemiology, and End Results data. Cancer-Am Cancer Soc 2012; 118(18): 4444-4451.

3. Vander PV, Balm AJ, Hilgers FJ, et al. Prognostic factors for long term results of the treatment of patients with malignant submandibular gland tumors. Cancer-Am Cancer Soc 1999; 85(10): 2255-2264.

4. Lorini L, Ardighieri L, Bozzola A, et al. Prognosis and management of recurrent and/or metastatic head and neck adenoid cystic carcinoma. Oral Oncol 2021; 115: 105213.

5. Drier Y, Cotton MJ, Williamson KE, et al. An oncogenic MYB feedback loop drives alternate cell fates in adenoid cystic carcinoma. Nat Genet 2016; 48(3): 265-272.

6. Ferrarotto R, Mitani Y, Diao L, et al. Activating NOTCH1 Mutations Define a Distinct Subgroup of Patients With Adenoid Cystic Carcinoma Who Have Poor Prognosis, Propensity to Bone and Liver Metastasis, and Potential Responsiveness to Notch1 Inhibitors. J Clin Oncol 2017; 35(3): 352-360.

7. Hao L, Xiao-lin N, Qi C, Yi-ping Y, Jia-quan L, Yan-ning L. Nerve growth factor and vascular endothelial growth factor: retrospective analysis of 63 patients with salivary adenoid cystic carcinoma. Int J Oral Sci 2010; 2(1): 35-44.

8. de Sousa LG, Jovanovic K, Ferrarotto R. Metastatic Adenoid Cystic Carcinoma: Genomic Landscape and Emerging Treatments. Curr Treat Option On 2022; 23(8): 1135-1150.

9. Sahara S, Herzog AE, Nor JE. Systemic therapies for salivary gland adenoid cystic carcinoma. Am J Cancer Res 2021; 11(9): 4092-4110.

10. Laurie SA, Licitra L. Systemic therapy in the palliative management of advanced salivary gland cancers. J Clin Oncol 2006; 24(17): 2673-2678.

11. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 2013; 41(Database issue): D991-D995.

12. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000; 28(1): 27-30.

13. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019; 47(D1): D607-D613.

14. Chen W, Liu BY, Zhang X, et al. Identification of differentially expressed genes in salivary adenoid cystic carcinoma cells associated with metastasis. Arch Med Sci 2016; 12(4): 881-888.

15. Morris L, Chandramohan R, West L, et al. The Molecular Landscape of Recurrent and Metastatic Head and Neck Cancers: Insights From a Precision Oncology Sequencing Platform. Jama Oncol 2017; 3(2): 244-255.

16. Ross JS, Gay LM, Wang K, et al. Comprehensive genomic profiles of metastatic and relapsed salivary gland carcinomas are associated with tumor type and reveal new routes to targeted therapies. Ann Oncol 2017; 28(10): 2539-2546.

17. Spiro RH, Huvos AG, Strong EW. Adenoid cystic carcinoma of salivary origin. A clinicopathologic study of 242 cases. Am J Surg 1974; 128(4): 512-520.

18. Kokemueller H, Eckardt A, Brachvogel P, Hausamen JE. Adenoid cystic carcinoma of the head and neck--a 20 years experience. Int J Oral Max Surg 2004; 33(1): 25-31.

19. Lloyd S, Yu JB, Wilson LD, Decker RH. Determinants and patterns of survival in adenoid cystic carcinoma of the head and neck, including an analysis of adjuvant radiation therapy. Am J Clin Oncol-Canc 2011; 34(1): 76-81.

20. Ho AS, Kannan K, Roy DM, et al. The mutational landscape of adenoid cystic carcinoma. Nat Genet 2013; 45(7): 791-798.

21. Liu X, Chen D, Lao X, Liang Y. The value of MYB as a prognostic marker for adenoid cystic carcinoma: Meta-analysis. Head Neck-J Sci Spec 2019; 41(5): 1517-1524.

22. Wagner VP, Bingle CD, Bingle L. MYB-NFIB fusion transcript in adenoid cystic carcinoma: Current state of knowledge and future directions. Crit Rev Oncol Hemat 2022; 176: 103745.

23. Quintana AM, Liu F, O'Rourke JP, Ness SA. Identification and regulation of c-Myb target genes in MCF-7 cells. Bmc Cancer 2011; 11: 30.

24. Tang YF, Wu WJ, Zhang JY, Zhang J. Reconstruction and analysis of the aberrant lncRNA-miRNA-mRNA network based on competitive endogenous RNA in adenoid cystic carcinoma of the salivary gland. Transl Cancer Res 2021; 10(12): 5133-5149.

25. Wu Y, Chen J, Tan F, Wang B, Xu W, Yuan C. ITGA9: Potential Biomarkers and Therapeutic Targets in Different Tumors. Curr Pharm Design 2022; 28(17): 1412-1418.

26. Liu Z, Gao J, Yang Y, Zhao H, Ma C, Yu T. Potential targets identified in adenoid cystic carcinoma point out new directions for further research. Am J Transl Res 2021; 13(3): 1085-1108.

27. Andersson MK, Afshari MK, Andren Y, Wick MJ, Stenman G. Targeting the Oncogenic Transcriptional Regulator MYB in Adenoid Cystic Carcinoma by Inhibition of IGF1R/AKT Signaling. Jnci-J Natl Cancer I 2017; 109(9):

28. Persson M, Andren Y, Mark J, Horlings HM, Persson F, Stenman G. Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. P Natl Acad Sci Usa 2009; 106(44): 18740-18744.

29. Mitani Y, Li J, Rao PH, et al. Comprehensive analysis of the MYB-NFIB gene fusion in salivary adenoid cystic carcinoma: Incidence, variability, and clinicopathologic significance. Clin Cancer Res 2010; 16(19): 4722-4731.

30. Brayer KJ, Frerich CA, Kang H, Ness SA. Recurrent Fusions in MYB and MYBL1 Define a Common, Transcription Factor-Driven Oncogenic Pathway in Salivary Gland Adenoid Cystic Carcinoma. Cancer Discov 2016; 6(2): 176-187.

31. Mitani Y, Liu B, Rao PH, et al. Novel MYBL1 Gene Rearrangements with Recurrent MYBL1-NFIB Fusions in Salivary Adenoid Cystic Carcinomas Lacking t(6;9) Translocations. Clin Cancer Res 2016; 22(3): 725-733.

32. Martens S, Bridelance J, Roelandt R, Vandenabeele P, Takahashi N. MLKL in cancer: more than a necroptosis regulator. Cell Death Differ 2021; 28(6): 1757-1772.

33. Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use. Csh Perspect Biol 2010; 2(1): a1008.

34. Aubrey BJ, Strasser A, Kelly GL. Tumor-Suppressor Functions of the TP53 Pathway. Csh Perspect Med 2016; 6(5): a026062.