

The Prognostic Value of Plasma Complement Factor B (CFB) In Thyroid Carcinoma

Pu Wu

The First Affiliated Hospital of China Medical University

Wei Sun

The First Affiliated Hospital of China Medical University

Hao Zhang (✉ haozhang@cmu.edu.cn)

The First Affiliated Hospital of China Medical University <https://orcid.org/0000-0002-9938-8433>

Research

Keywords: CFB, ESTIMATE, CIBERSORT, TME, TICs

Posted Date: May 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-510321/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Bioengineered on December 13th, 2021.
See the published version at <https://doi.org/10.1080/21655979.2021.2005745>.

Abstract

Background: Stromal and immune cells are major components of tumor microenvironment (TME) and affect the growth and development of thyroid carcinoma (THCA). However, data on the exact mechanisms that define the relationship between the TME and THCA remains scant.

Results: Here, we calculated stromal and immune cells scores and the proportion of tumor-infiltrating immune cells (TICs) by CIBERSORT and ESTIMATE based on the THCA gene expression data from The Cancer Genome Atlas database (TCGA). In addition, we evaluated differentially expressed genes (DEGs) from high- and low-score groups and then performed Gene Ontology (GO) as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. Besides, we carried out protein-protein interaction (PPI) analysis and COX regression on the DEGs. Our data showed a significant correlation between plasma complement factor B (CFB) and PTC development and prognosis. In addition, Gene Set Enrichment Analysis (GSEA) demonstrated that the CFB was mainly enriched in immune response pathways. Further dissection showed that the expression of CFB was positively correlated with T cells CD8, Macrophages M1, Plasma cells, T cells CD4 memory activated, T cells follicular helper and T cells regulatory (Tregs), whereas negatively correlated with Eosinophils, Macrophages M0, Macrophages M2, Mast cells resting, T cells CD4 memory resting in the TME. Finally, the expression level of CFB was verified by other cohorts from Gene Expression Omnibus (GEO) database.

Conclusions: Taken together, our data demonstrated that the CFB could be a prognostic marker for THCA and its expression influences the infiltration of immune cells.

Background

Thyroid carcinoma (THCA), derived from follicular thyroid cells, is the most common endocrine system malignancy worldwide¹⁻³. The morbidity of THCA has been increasing in the recent decades⁴. Of all the THCA subtypes, papillary thyroid carcinoma (PTC) is the most common, and accounts for approximately 90% of all the THCA cases⁵. However, compared to other forms of cancer, the course of PTC remains indolent and has low mortality⁶. Besides, most patients experience relatively better prognosis following the standard treatment (surgery followed by either radioactive iodine or observation)⁷. Nevertheless, some patients might develop cervical lymph node metastasis, which leads to locoregional recurrence⁸. It is, therefore, crucial to identify patients with aggressive disease in order to administer timely treatment⁹. Thus, understanding the THCA carcinogenesis would help inform the development of effective diagnostic and therapeutic strategies.

Increasing evidences have shown the TME is of great importance in the tumor progression, treatment and clinical outcome^{10,11}. TME is a complex and continuously evolving system. The occurrence, growth and metastasis of tumors is strongly dependent on the internal environment of tumor cells. It not only includes the structure, function and metabolism of tumor tissues, but also relates to the internal environment of tumor cells themselves^{12,13}. Accordingly, the TME significantly influences tumor

therapeutic response and clinical outcomes¹⁴. The TME is mainly composed of stromal cells, soluble molecules and immune cells¹⁵. Both the immune and stromal cells have been proposed to be valuable in tumor diagnosis and evaluation of disease prognosis¹⁶. However, data on the role of the TME stromal cells on tumor progression remains scant. Besides, previous studies have indicated that tumor-infiltrating immune cells (TICs) in TME could serve as prognostic markers in cancer with various bioinformatics tools. The interaction has been valuable for developing innovative THCA-directed immunotherapies. Recent studies show that TME components significantly affect THCA growth and progression¹⁷. Therefore, it is imperative to find out TME-related biomarker to improve the survival and prognosis of THCA.

Here, we obtained 570 transcriptome data and 509 clinical datasets from TCGA database. We then employed ESTIMATE and CIBERSORT algorithms to compute TME-specific immune and stromal scores in THCA. The proportion of the TIC in THCA samples was also calculated from The Cancer Genome Atlas (TCGA) database. We demonstrate that the plasma complement factor B (CFB) is a TME-related biomarker for THCA¹⁸. High CFB expression indicated better clinical prognosis and infiltration of TICs. Therefore, TME-specific CFB gene might provide a potential prognostic marker in patients with THCA.

Methods

Materials

Transcriptome RNA-seq data (FPKM) from 570 THCA cases (58 normal thyroid samples and 512 thyroid tumor samples) and the corresponding clinical data were downloaded from The Cancer Genome Atlas (TCGA) (<https://tcga-data.nci.nih.gov/tcga/>).

ESTIMATE

We used ESTIMATE (Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data) in R language (version 4.0.3) to predict the tumor purity and presence of infiltrating stromal/immune cells in the tumor tissues using gene expression data¹⁹. We generated stromal scores (amount of stroma cells in tumor tissues), immune scores (the infiltration of immune cells in tumor tissues) as well as estimate scores (tumor purity)²⁰.

Generation of differentially expressed genes (DEGs) between the high-score and low-score groups

A total of 570 tumor samples were sorted into high- or low-score based on the median ImmuneScores or StromalScores. Using “limma” package in R language (version 4.0.3), we generated DEGs between the high- and low- immune/stromal score groups. DEGs with a false discovery rate (FDR) <0.05 and fold change ≥ 1 , after transformation (\log_2FC) (high-score group/low-score group) were considered significant. We then constructed Venn plots to screen the intersected DEGs.

GO And KEGG Enrichment Analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using clusterProfiler, org.Hs.eg.db, enrichplot and ggplot2 packages in R language version 4.0.3. Enrichment data with a P-value<0.05 and q-value<0.05 were considered to be statistically significant.

Heatmaps

Heatmaps for the DEGs were generated using heatmap package in R language.

Clinical relevance of the scores

The THCA clinical samples were downloaded from the TCGA database. All the statistical analyses were performed using R language. We used the Wilcoxon rank sum test (for ranked data with two categories) or Kruskal–Wallis test (three or more categories) to evaluate the significance of the clinical stages against pathological stages. The data was compared with the immune scores, stromal scores or ESTIMATE scores.

PPI Network Analysis

The protein-protein interaction (PPI) network was analyzed using STRING database (<https://string-db.org>), followed by reconstruction with Cytoscape version 3.8.2. We used nodes with a highest confidence of interaction score of no less than 0.99 to build the PPI network.

COX Regression Analysis

COX regression analysis was performed by survival package in R language, then we screened out the statistically significant genes by the univariate COX regression analysis (P-value < 0.05).

Gene Set Enrichment Analysis (GSEA)

C2.cp.kegg.v7.1.symbols.gmt [Curated] and c5.all.v7.2.symbols.gmt [Gene ontology] were downloaded from Molecular Signatures Database as the target sets for the execution of GSEA. The GSEA of the whole transcriptome from the tumor sample was performed using gsea-4.1.0 (Number of permutations: 1000; Phenotype labels: h _versus _l). Gene sets with FDR q < 0.05 were considered as significant.

CIBERSORT

CIBERSORT, a deconvolution algorithm, was used to estimate the composition of 21 immune cells in all the tumor samples. We performed 1,000 permutations and retained samples with a p-value < 0.05, which were considered statistically significant. The sum of various immune cells was 1.

GEO database verification

The datasets of PTC patients were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). We selected four mRNA datasets (GSE33630, GSE29265,

GSE35570, GSE27155) of PTC to verify the expression level of CFB in this study.

Results

Analysis Process

The analysis process of this study is shown in Supplementary Fig. 1. The transcriptome RNA-seq data from 570 samples and clinical information of the THCA patients were downloaded from TCGA database and used to estimate immune and stromal components. DEGs were extracted by dividing into high- and low-score groups based on the median stromal and immune scores. DEGs were used to construct PPI network and univariate COX regression analysis. We performed intersection analysis between PPI networks core nodes and the top significant factors gained from the analysis of univariate COX regression. Our data showed the presence of only one intersection gene, the CFB. Subsequently, we evaluated the role of the CFB in the patient survival, as well as its clinical relevance.

Characterization of THCA patients and clinical correlation with stromal or immune scores

A total of 570 THCA samples were obtained from the TCGA database. Stromal scores, immune scores and ESTIMATE scores in the TME were generated based on the ESTIMATE algorithm. Our univariate analysis demonstrated that the clinicopathological characteristics were significantly associated with the stromal scores and immune scores (Fig. 1). Besides, the Stromal score of the M0 stage was higher than that of the M1 stage (Fig. 1B, $p = 0.027$). On the other hand, the immune score differed significantly between the N0 and the N1 stage (Fig. 1C, $p = 0.00014$). The estimate score varied across the different M stages (Fig. 1A, $p = 0.035$). However, there were no statistically significant differences between stromal score, immune score and estimate score in T2 patients from that in T1 or T3 patients. Furthermore, whereas there was significant difference between the scores in stage I and stage II or III, other clinical stages did not show any significant differences. Besides, no statistically significant differences were observed in the stromal score, immune score and estimate score between the patients' age and gender (Fig. 1A, 1B, 1C). However, there were significant differences in the N stage (Fig. 1A, 1B, 1C; $p = 0.0034$, 0.00014 , 0.0003). These data showed that the proportion of immune and stromal cells was associated with the growth and progression of THCA.

DEGs and Functional Enrichment Analysis

To identify the exact alterations in gene expression in the TME, the THCA patients were categorized into high- and low-score groups based on the stromal/immune cells. After performing the differential expression analysis of the groups as categorized by the stromal scores, we obtained a total of 1031 DEGs. Among them, 79 genes were down-regulated while 952 genes were up-regulated (Figure 2A, 2B). Similarly, differential analysis of the high- and low- immune score patients showed 1294 DEGs consisting of 318 downregulated genes and 976 upregulated genes. In addition, the Venn diagrams, drawn from an intersection of the DEGs between stromal and immune score groups, showed a total of 72 downregulated genes and 818 upregulated genes (Figure 2C, 2D). The DEGs might have been responsible for the TME status. All the 890 DEGs were used to conduct GO and KEGG functional enrichment analysis via the

clusterProfiler R package in R language. GO enrichment data showed that the DEGs were significantly mapped into T cell activation and lymphocyte differentiation (Figure 2E, 2F, 2G). In addition, the KEGG enrichment analysis suggested that most of the pathways were associated with immune responses, such as cytokine–cytokine receptor interaction and chemokine signaling pathway (Figure 2H, 2I, 2J).

PPI Network Analysis and Univariate COX Regression Analysis

To further explore the relationship between the DEGs, we used STRING online tool and Cytoscape software (version 3.8.2) to perform a PPI network (Figure 3A, 3B). The top 30 genes with maximum number of nodes in the PPI network were presented as bar plots (Figure 3C). Univariate COX regression analysis of the 890 immune-related DEGs showed that 4 genes were potential prognostic or risk factors for THCA (Figure 3D). Furthermore, an intersection analysis of the top 100 genes in PPI network and the 4 prognostic genes showed CFB as the only one core target gene (Figure 3E).

CFB Expression Levels, Survival and Clinical Correlation Analysis

To evaluate the role of CFB expression in clinical disease manifestations, we profiled the expression of CFB in normal and tumor tissues using “limma” package in R language (Wilcoxon rank sum test, $P < 0.001$). The CFB expression in normal tissues was significantly lower than in tumor tissues (Figure 4A). Furthermore, the CFB expression in paired tumor tissues was significantly higher than the expression level in adjacent normal tissues (Figure 4B). To perform survival analysis, all the THCA cases were divided into CFB high and CFB low-expression groups based on the CFB median expression value. The survival analysis indicated that THCA patients with higher CFB expression had a longer survival compared to those with lower CFB expression (Figure 4C). Besides, our clinical correlation analysis of CFB showed that, except for the low CFB expression in stage III compared to stage I or II, there was no significant difference in the other clinical stages (Figure 4D). In addition, the expression level in N1 was higher than that in N0 (Figure 4G). M0 and M1 did not show any significant differences in the CFB protein expression (Figure 4F). Furthermore, our data showed that there was significant difference between the T2 and T3 or T4 groups (Figure 4E).

GSEA

To determine significantly enriched signaling pathways, we performed GSEA in the high-expression and low-expression groups and then used FDR q-value, nominal p-value and FWER p-value for data interpretation. Our KEGG enrichment data showed that the CFB gene was highly expressed in signaling pathways such as antigen processing and presentation, autoimmune thyroid disease, cell adhesion molecules cams, intestinal immune network for IGA production, leishmania infection, natural killer cell mediated cytotoxicity, *Escherichia coli* infection, systemic lupus erythematosus, type 1 diabetes mellitus or viral myocarditis (Figure 5A). On the other hand, the GO enrichment analysis showed upregulated CFB expression in pathways such as adaptive immune response, interferon gamma mediated signaling pathway, positive regulation of T cell mediated cytotoxicity, positive regulation of T cell mediated immunity, regulation of leukocyte mediated cytotoxicity, regulation of syncytium formation by plasma

membrane fusion, regulation of T cell mediated cytotoxicity, regulation of T cell mediated immunity, T cell activation involved in immune response or T cell mediated immunity (Figure 5B). These results indicated that CFB may be a potential indicator for the TME status.

Tumor-infiltrating immune subsets in the TME

CIBERSORT algorithm be used to calculate the proportion of the tumor-infiltrating immune cells in the TME among the THCA cases. Our data suggested that the proportion of the 21 immune cells in every case, as well as the sum of various immune cells was 1 (Figure 6A). Besides, the correlation analyses indicated that CD8 T cells displayed a strong positive correlation with Plasma cells (Cor=0.57) (Figure 6B). On the contrary, plasma cells were negatively correlated with Macrophages M0 (Cor=-0.58).

Effect of the expression of CFB on TIC levels and the Correlation between the CFB expression and the TICs

To explore the potential functions of CFB in the TME, we classified the patients into high- and low-expression groups based on the expression of CFB. Our analysis showed that a total of 10 TICs were perturbed by the expression of CFB (Figure 7A). B cells naïve ($p=0.034$), plasma cells ($p<0.001$), T cells CD4 memory activated ($p<0.001$), T cells follicular helper ($p=0.004$), T cells regulatory ($p=0.019$) and Macrophages M1 ($p<0.001$) were positively correlated with high CFB expression. However, Macrophages M0 ($p<0.001$), Macrophages M0 ($p=0.002$), Mast cells resting ($p=0.003$) and Eosinophils ($p=0.030$) were significantly upregulated in the low-expression group.

In addition, the data indicated a correlation between the CFB expression and TICs in the THCA patients (Figure 7B). The CFB gene was positively correlated with T cells CD8, Macrophages M1, Plasma cells, T cells CD4 memory activated, T cells follicular helper or T cells regulatory (Tregs), but negatively correlated with Eosinophils, Macrophages M0, Macrophages M2, Mast cells resting or T cells CD4 memory resting.

Consequently, to identify the most closely related immune cells, we performed the intersection analysis based on both the difference analysis and correlation tests with Venn diagram. Our data showed positive correlation between the Plasma cells, T cells CD4 memory activated, T cells follicular helper, T cells regulatory (Tregs), Macrophages M0, Macrophages M1, Macrophages M2, Mast cells resting or Eosinophils with CFB expression (Figure 7C).

GEO database verification

We downloaded four datasets from GEO to further verify the accuracy of the previous results from TCGA analysis. The expression of CFB was significantly higher in PTC compared with than that in normal thyroid samples in GSE33630, GSE29265, GSE35570 and GSE27155 (Figure 8, $P < 0.05$). The results of GEO datasets were essentially in agreement with the findings of the TCGA database. The survival information of CFB in PTC could unfortunately not be obtained from GEO datasets.

Discussion

Previous studies showed an association between TME and the tumor development^{21,22}. Besides, TME was shown to promote tumor inflammation and angiogenesis^{23,24}. The prognosis of cancer patients depends on the purity of tumor TME, which is dependent on the individual components²⁵. Stromal and immune cells are two main types of tumor TME components and have been proposed to aid in tumor diagnosis, prognosis and survival analysis²⁶⁻²⁸.

In this study, we intend to explore TME-related gene signature associated with THCA patient survival and outcomes. We performed ESTIMATE to acquire the stromal and immune scores in TME, then identified TME-related genes from TCGA database to determine the correlation with clinical information. The data showed that the genes were significantly associated with clinical stage and TNM stages. We then employed GO and KEGG enrichment analysis to evaluate the potential roles of 2325 DEGs in the high- or low-score groups of stromal and immune cells. The GO analysis showed that most of the DEGs regulated biological processes, such as T cell activation and lymphocyte differentiation in the TME. On the other hand, the KEGG analysis suggested that most of the DEGs triggered immune responses. In addition, to understand the interactions of the DEGs, we constructed the PPI network. The COX regression analysis was performed to understand predictive values for prognosis in the THCA patients. We obtained 4 related genes that were thought to have significant predictive values. The genes could be considered as potential biomarkers for THCA. The intersection of the 4 genes with the top 100 genes in the PPI network nodes resulted into the identification of plasma Complement Factor B (CFB). Unlike in normal tissues, the CFB gene was significantly upregulated in the TME of THCA patients. Besides, the high expression of the CFB gene was associated with longer survival and considerably better prognosis. On the other hand, to further dissect the functions of the CFB in the TME, we performed GSEA analysis. The 10 most enriched signaling pathways in the high or low expression phenotypes of CFB were determined. The data showed that CFB could be a potential marker in the TME status. Additionally, 9 kinds of TICs were associated with CFB expression. The TICs included plasma cells, T cells CD4 memory activated, T cells follicular helper, T cells regulatory (Tregs), Macrophages M0, Macrophages M1, Macrophages M2, Mast cells resting and Eosinophils. Other reports have shown high expression of CFB in pancreatic cancer and cutaneous squamous cell carcinoma, thus confirming the association between the CFB expression and carcinogenesis^{31,32}.

To date, the correlation between THCA and the CFB expression remains unclear. Here, we evaluated TME and TICs to identify the possible role of the CFB in the growth and progression of THCA^{33,34}. However, previous studies have reported gene signatures that define the growth and progression of THCA. In this study, we focused on the gene signatures in complex TME, which affected the progression and prognosis of THCA. The expression of CFB was closely associated with the prognosis of THCA. Thus, CFB might be used as a potential marker for the prognosis of THCA.

However, it was difficult to precisely determine the mechanisms of CFB on TME or TICs in THCA patients. Besides, clinical information from the TCGA database was incomplete, thus we did not perform

comprehensive prognostic evaluation.

Conclusions

Taken together, our data demonstrated that the expression of CFB is strongly correlated with the prognosis of THCA, thus could be used as a potential prognostic and therapeutic response marker for patients with THCA.

Declarations

Acknowledgements

We acknowledge the Cancer Genome Atlas (TCGA) database and the Gene Expression Omnibus (GEO) database for uploading their meaning datasets.

Authors' contributions

WP performed the data analysis and wrote the manuscript. SW and ZH reviewed and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (81902726), China Postdoctoral Science Foundation (2018M641739).

Availability of data and materials

Gene expression profiles and clinical data of THCA in this study are available from the public database (TCGA, <https://portal.gdc.cancer.gov/> and GEO, <https://www.ncbi.nlm.nih.gov/geo/>).

Ethics approval and consent to participate

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests" in this section.

References

1. Bray F, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68:394–424. doi:10.3322/caac.21492.

2. Scoazec JY, Couvelard A, Réseau T. [Classification of pancreatic neuroendocrine tumours: Changes made in the 2017 WHO classification of tumours of endocrine organs and perspectives for the future]. *Ann Pathol.* 2017;37:444–56. doi:10.1016/j.annpat.2017.10.003.
3. Ito Y, Nikiforov YE, Schlumberger M, Vigneri R. Increasing incidence of thyroid cancer: controversies explored. *Nat Rev Endocrinol.* 2013;9:178–84. doi:10.1038/nrendo.2012.257.
4. Seib CD, Sosa JA. Evolving Understanding of the Epidemiology of Thyroid Cancer. *Endocrinol Metab Clin North Am.* 2019;48:23–35. doi:10.1016/j.ecl.2018.10.002.
5. Siegel R, Naishadham D, Jemal A, Cancer statistics. 2013. *CA Cancer J Clin* **63**, 11–30, doi:10.3322/caac.21166 (2013).
6. Scharpf J. Achieving Active Surveillance for Thyroid Cancer-Not a Euphemism for Watching a Ticking Time Bomb. *JAMA Otolaryngol Head Neck Surg.* 2017;143:1021–2. doi:10.1001/jamaoto.2017.1453.
7. Vaccarella S, et al. Worldwide Thyroid-Cancer Epidemic? The Increasing Impact of Overdiagnosis. *N Engl J Med.* 2016;375:614–7. doi:10.1056/NEJMp1604412.
8. Haugen BR, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* **26**, 1-133, doi:10.1089/thy.2015.0020 (2016).
9. Raue F, Frank-Raue K. Thyroid Cancer: Risk-Stratified Management and Individualized Therapy. *Clin Cancer Res.* 2016;22:5012–21. doi:10.1158/1078-0432.CCR-16-0484.
10. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med.* 2013;19:1423–37. doi:10.1038/nm.3394.
11. Fidler IJ. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nat Rev Cancer.* 2003;3:453–8. doi:10.1038/nrc1098.
12. Fan T, et al. Immune profile of the tumor microenvironment and the identification of a four-gene signature for lung adenocarcinoma. *Aging* 12, doi:10.18632/aging.202269 (2020).
13. Privat-Maldonado A, Bengtson C, Razzokov J, Smits E, Bogaerts A. Modifying the Tumour Microenvironment: Challenges and Future Perspectives for Anticancer Plasma Treatments. *Cancers (Basel)* 11, doi:10.3390/cancers11121920 (2019).
14. Şenbabaoğlu Y, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol.* 2016;17:231. doi:10.1186/s13059-016-1092-z.
15. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell.* 2012;21:309–22. doi:10.1016/j.ccr.2012.02.022.
16. Jiang Y, Han QJ, Zhang J. Hepatocellular carcinoma: Mechanisms of progression and immunotherapy. *World J Gastroenterol.* 2019;25:3151–67. doi:10.3748/wjg.v25.i25.3151.

17. Jolly LA, et al. Fibroblast-Mediated Collagen Remodeling Within the Tumor Microenvironment Facilitates Progression of Thyroid Cancers Driven by BrafV600E and Pten Loss. *Cancer Res.* 2016;76:1804–13. doi:10.1158/0008-5472.CAN-15-2351.
18. Yoshihara K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612. doi:10.1038/ncomms3612.
19. Chakraborty H, Hossain A. R package to estimate intracluster correlation coefficient with confidence interval for binary data. *Comput Methods Programs Biomed.* 2018;155:85–92. doi:10.1016/j.cmpb.2017.10.023.
20. Yang SYC, et al. Landscape of genomic alterations in high-grade serous ovarian cancer from exceptional long- and short-term survivors. *Genome Med.* 2018;10:81. doi:10.1186/s13073-018-0590-x.
21. Bi KW, Wei XG, Qin XX, Li B. BTK Has Potential to Be a Prognostic Factor for Lung Adenocarcinoma and an Indicator for Tumor Microenvironment Remodeling: A Study Based on TCGA Data Mining. *Front Oncol.* 2020;10:424. doi:10.3389/fonc.2020.00424.
22. Zhang S, et al. PD-L1 and CD4 are independent prognostic factors for overall survival in endometrial carcinomas. *BMC Cancer.* 2020;20:127. doi:10.1186/s12885-020-6545-9.
23. van den Meiracker AH, Danser AH, Sleijfer S, Kappers MH, Re. Hypertension as a biomarker of efficacy in patients with metastatic renal cell carcinoma treated with sunitinib. *J Natl Cancer Inst* **103**, 1557; author reply 1558, doi:10.1093/jnci/djr328 (2011).
24. Cao J, et al. Screening and Identifying Immune-Related Cells and Genes in the Tumor Microenvironment of Bladder Urothelial Carcinoma: Based on TCGA Database and Bioinformatics. *Front Oncol.* 2019;9:1533. doi:10.3389/fonc.2019.01533.
25. Siegel RL, Miller KD, Jemal A, Cancer Statistics. 2017. *CA Cancer J Clin* **67**, 7–30, doi:10.3322/caac.21387 (2017).
26. Wang J, et al. CD52 Is a Prognostic Biomarker and Associated With Tumor Microenvironment in Breast Cancer. *Front Genet.* 2020;11:578002. doi:10.3389/fgene.2020.578002.
27. Porcelli T, Sessa F, Luongo C, Salvatore D. Local ablative therapy of oligoproggressive TKI-treated thyroid cancer. *J Endocrinol Invest.* 2019;42:871–9. doi:10.1007/s40618-019-1001-x.
28. Xu ZY, et al. Analysis of prognostic genes in the tumor microenvironment of lung adenocarcinoma. *PeerJ.* 2020;8:e9530. doi:10.7717/peerj.9530.
29. Ricklin D, Hajishengallis G, Yang K, Lambris JD. Complement: a key system for immune surveillance and homeostasis. *Nat Immunol.* 2010;11:785–97. doi:10.1038/ni.1923.
30. Rutkowski MJ, Sughrue ME, Kane AJ, Mills SA, Parsa AT. Cancer and the complement cascade. *Mol Cancer Res.* 2010;8:1453–65. doi:10.1158/1541-7786.MCR-10-0225.
31. Riihila P, et al. Complement Component C3 and Complement Factor B Promote Growth of Cutaneous Squamous Cell Carcinoma. *Am J Pathol.* 2017;187:1186–97. doi:10.1016/j.ajpath.2017.01.006.

32. Kim SH, et al. Prognostic potential of the preoperative plasma complement factor B in resected pancreatic cancer: A pilot study. *Cancer Biomark*. 2019;24:335–42. doi:10.3233/CBM-181847.
33. Cayrol F, Sterle HA, Díaz Flaqué MC, Arcos B, M. L. & Cremaschi GA. Non-genomic Actions of Thyroid Hormones Regulate the Growth and Angiogenesis of T Cell Lymphomas. *Front Endocrinol (Lausanne)*. 2019;10:63. doi:10.3389/fendo.2019.00063.
34. Ferrari SM, et al. Immune and Inflammatory Cells in Thyroid Cancer Microenvironment. *Int J Mol Sci* 20, doi:10.3390/ijms20184413 (2019).

Figures

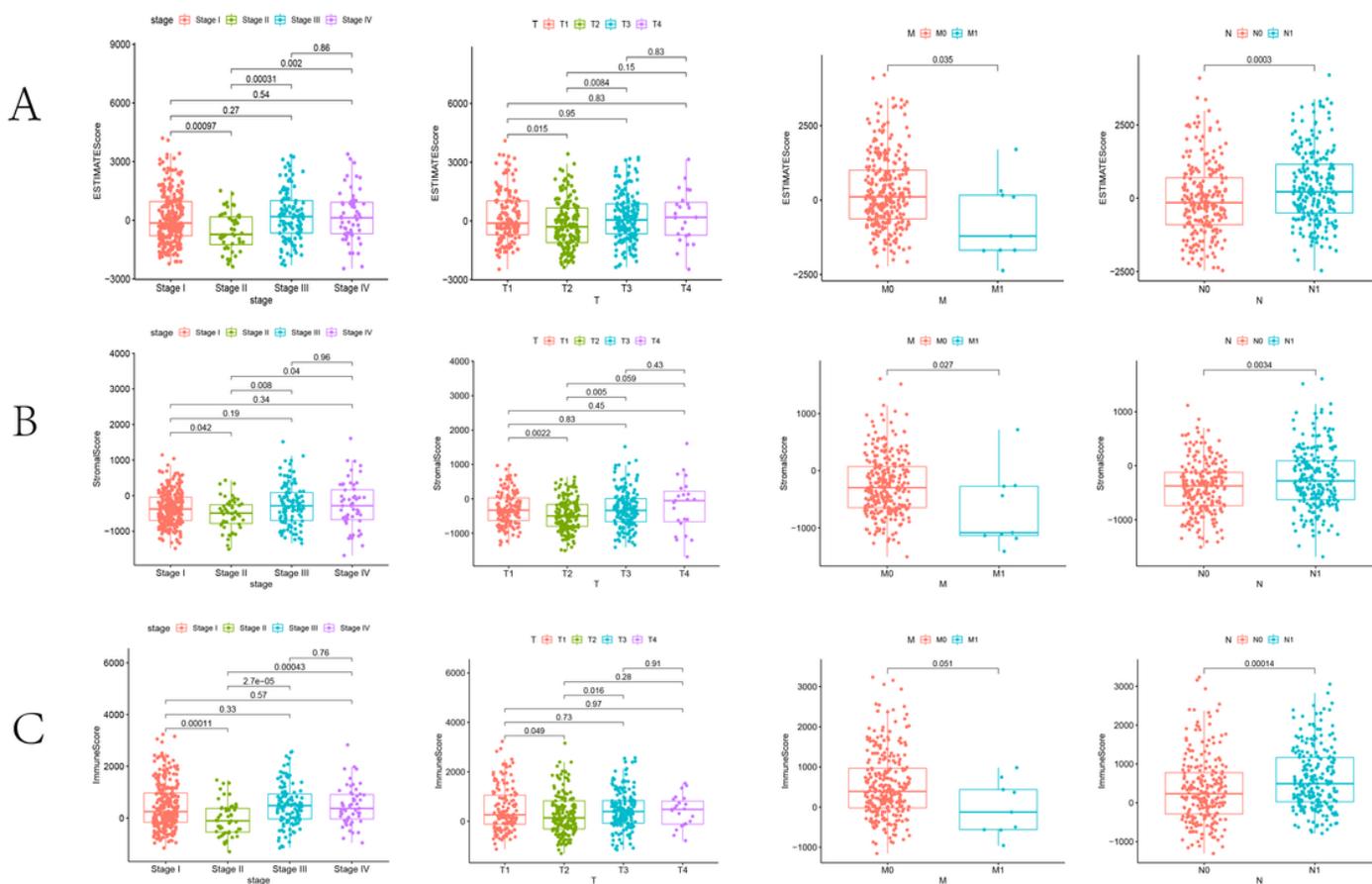


Figure 1

Correlation of StromaScore and ImmuneScore with clinicopathological staging characteristics. (A-C) Distribution of estimate scores, stromal scores and immune scores in tumor TNM stage.

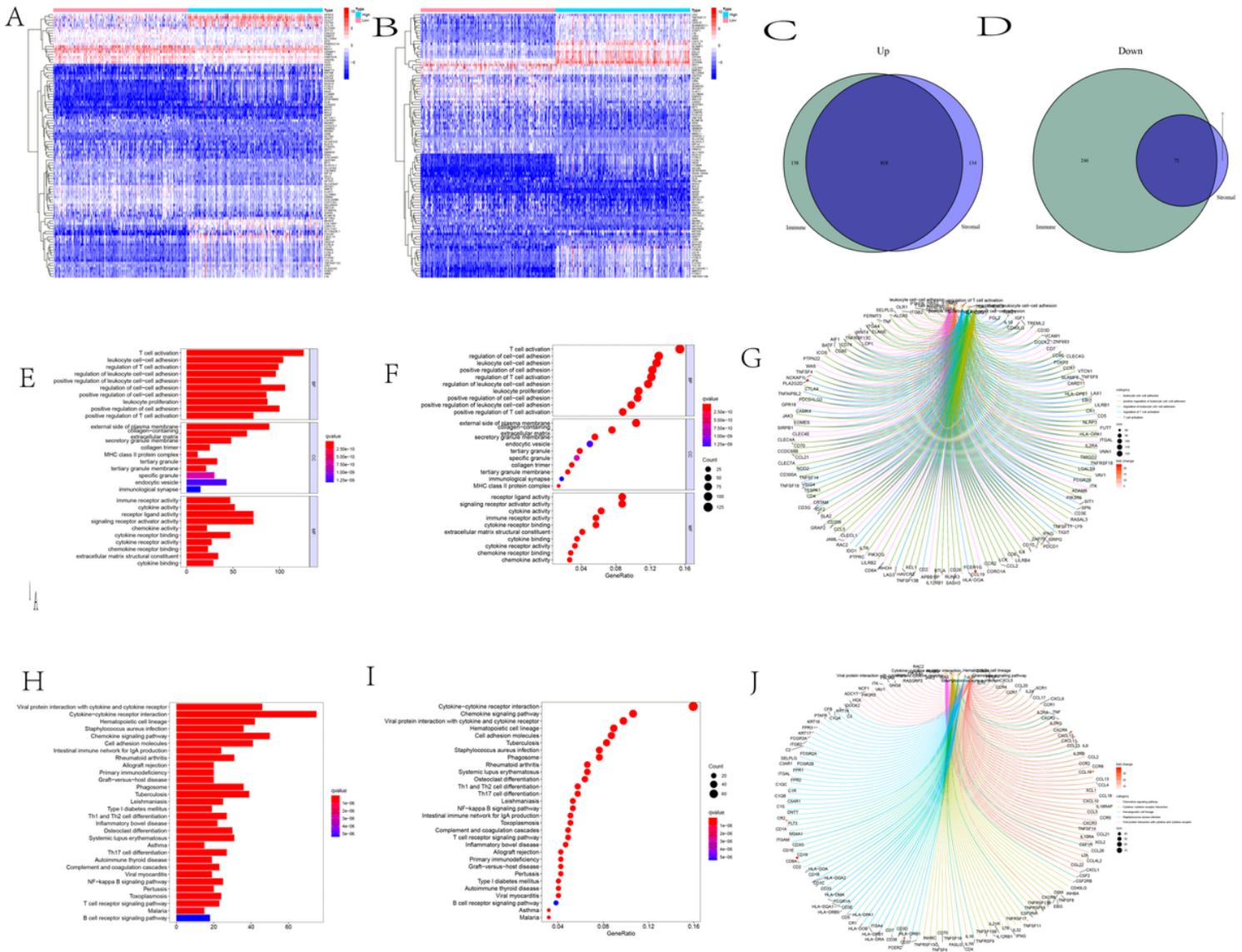


Figure 2

High score groups and low score groups and enrichment analysis of GO and KEGG for DEGs. (A, B) Heatmaps for DEGs in stromal scores and immune scores. Venn diagrams drawn from up-regulated (C) and down-regulated (D) DEGs in stromal scores and immune scores. (E, F, G) GO term enrichment analysis results of DEGs. (H, I, J) KEGG term enrichment analysis results of DEGs.

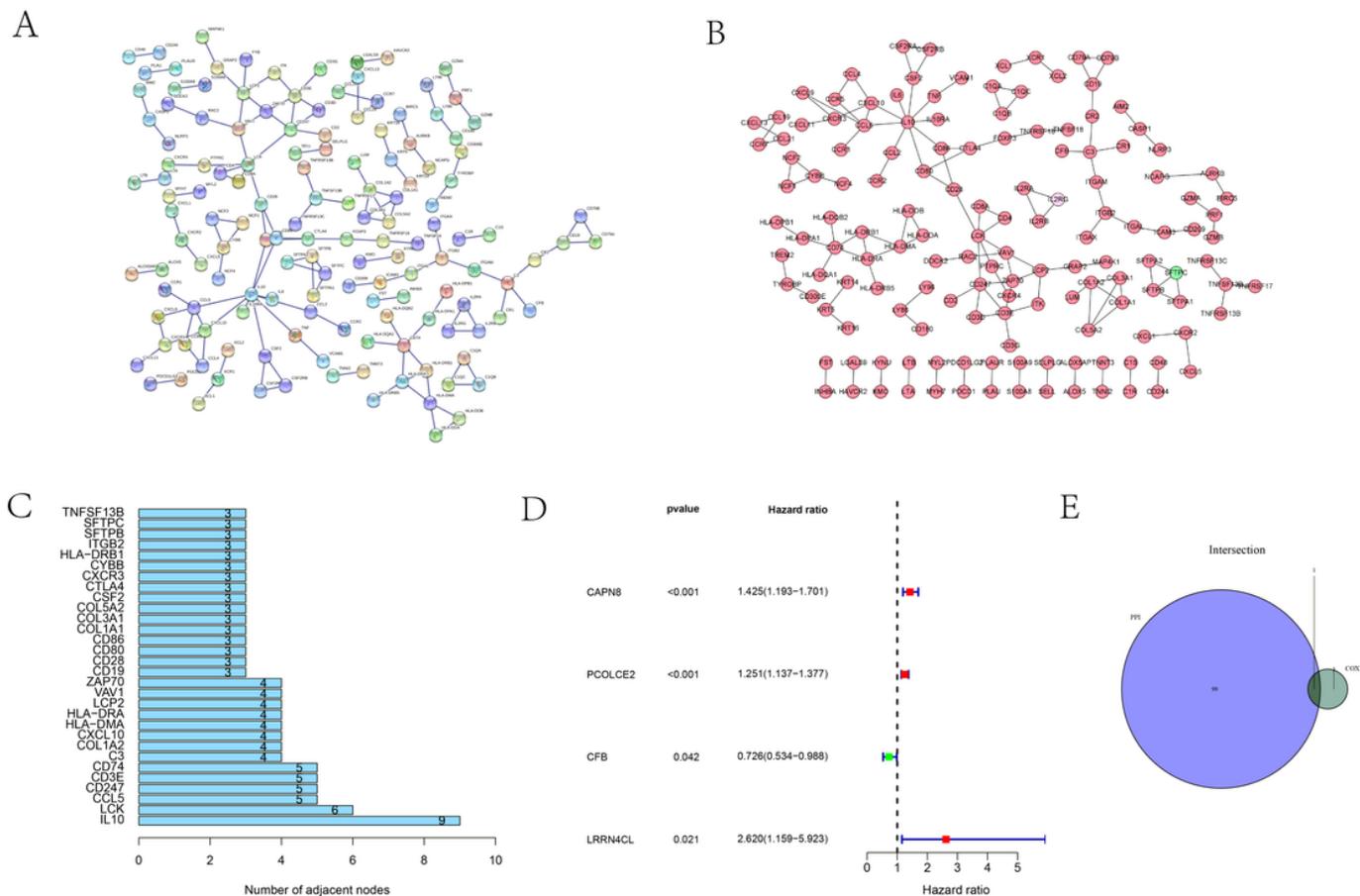


Figure 3

Protein-protein interaction network and univariate Cox regression analysis. (A) PPI network of the DEGs based on STRING with interaction confidence value > 0.99. (B) Visualized PPI analysis of DEGs based on Cytoscape. (C) The top 30 genes with maximum ordered nodes. (D) Univariate Cox regression analysis of DEGs, listing the genes with $p < 0.05$ in the forest plot. (E) Venn diagram showing the key gene (CFB) in PPI and Univariate Cox regression analysis.

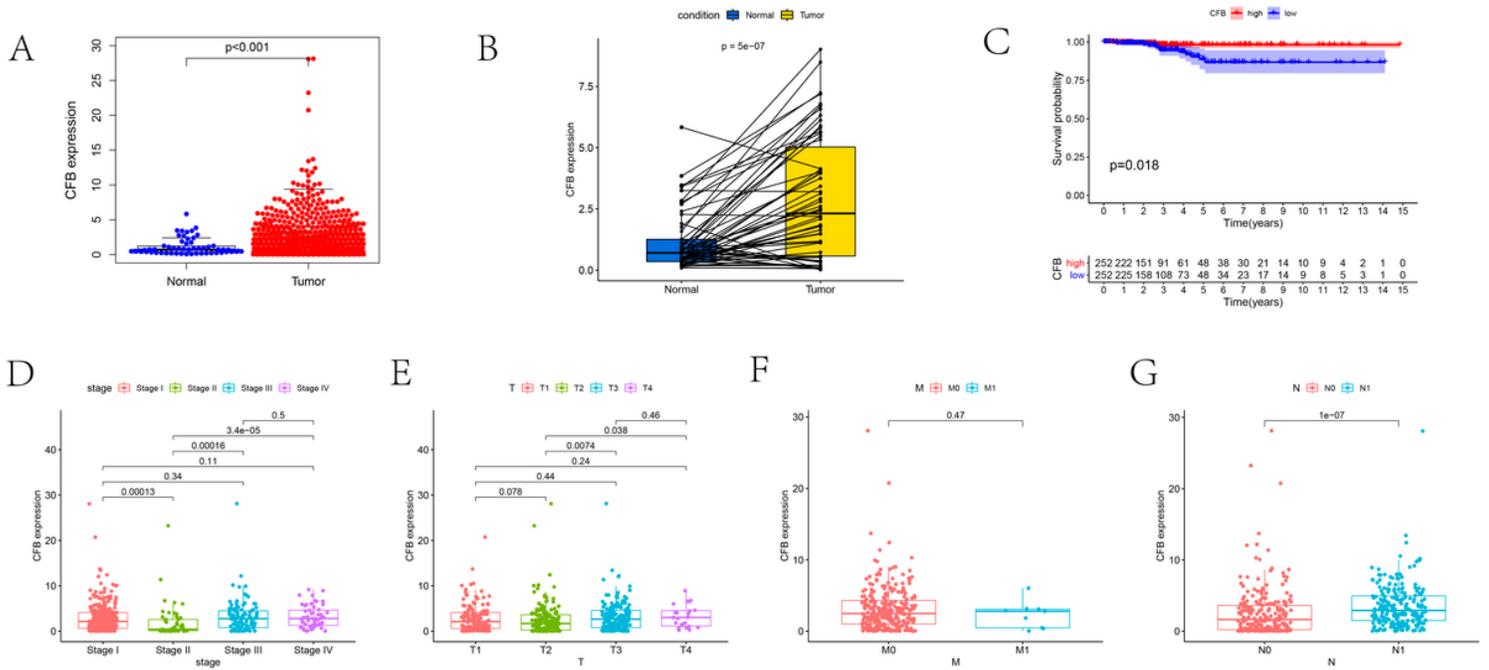
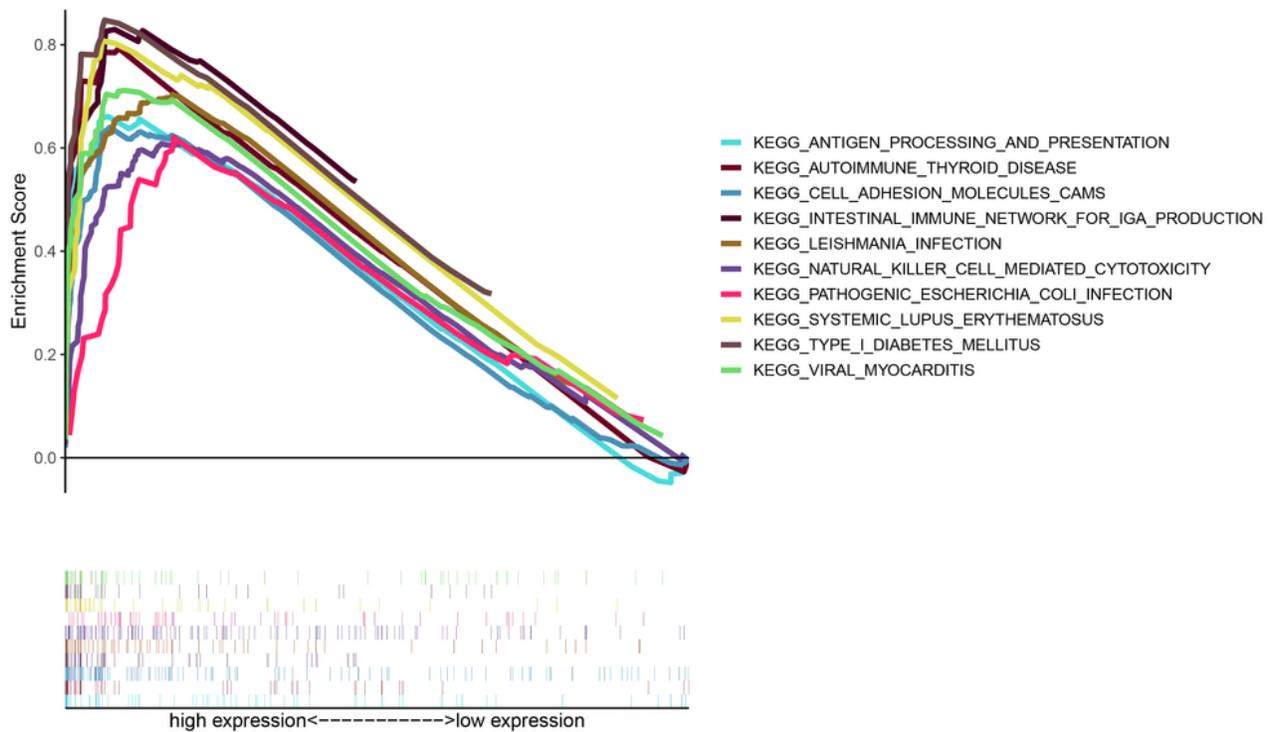


Figure 4

CFB expression levels and the survival and clinical correlation analysis of THCA patients. (A) The expression of CFB in the normal and tumor samples. (B) The expression level of CFB in paired normal and tumor samples. (C) Survival analysis with different CFB expression in high expression and low expression groups. The expression of CFB with stage (D), T classification (E), M classification (F), N classification (G).

A



B

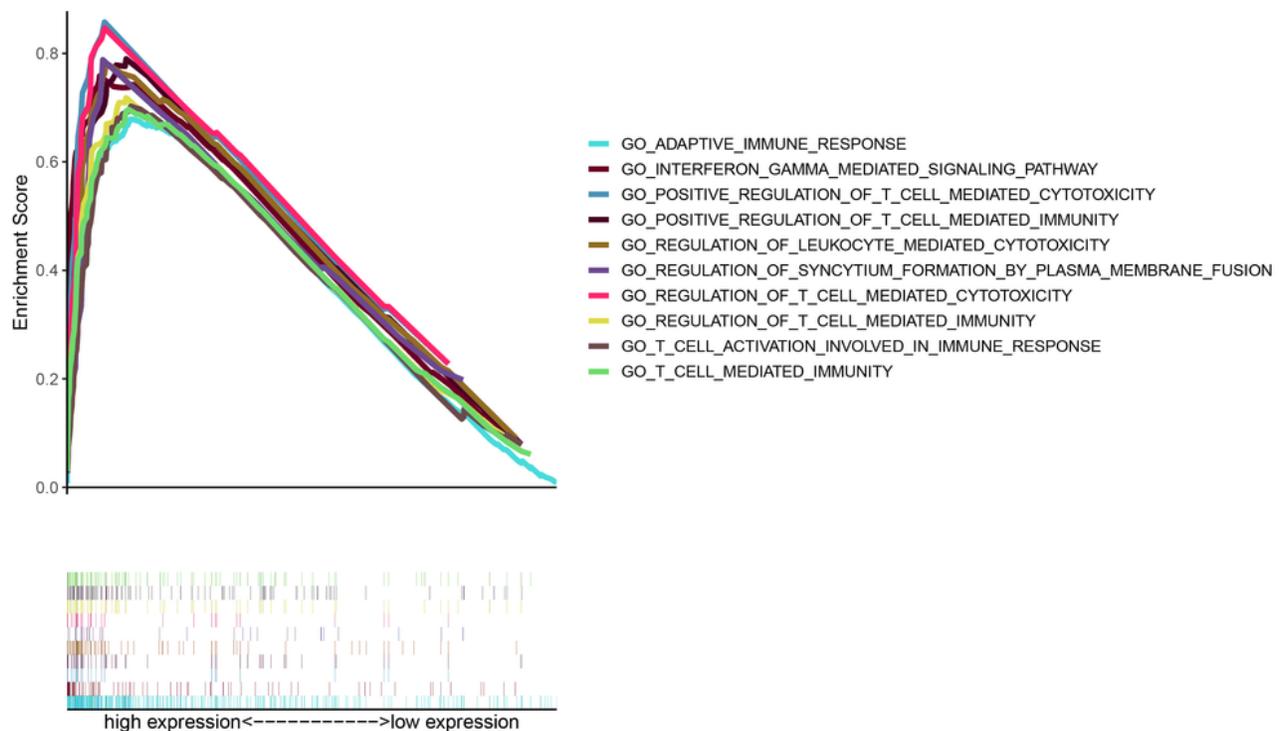


Figure 5

GSEA for CFB with high expression and low expression. (A) The enriched gene sets in KEGG collection by the high expression of CFB. Each line represents one particular gene set with different color, and up-regulated genes and the down-regulated genes located in the left and right of x-axis, respectively. Only gene sets with NOM $p < 0.05$ and FDR $q < 0.05$ were displayed. And only the top 10 leading gene sets were displayed in the plot. (B) The enriched gene sets in GO collection by the high expression of CFB.

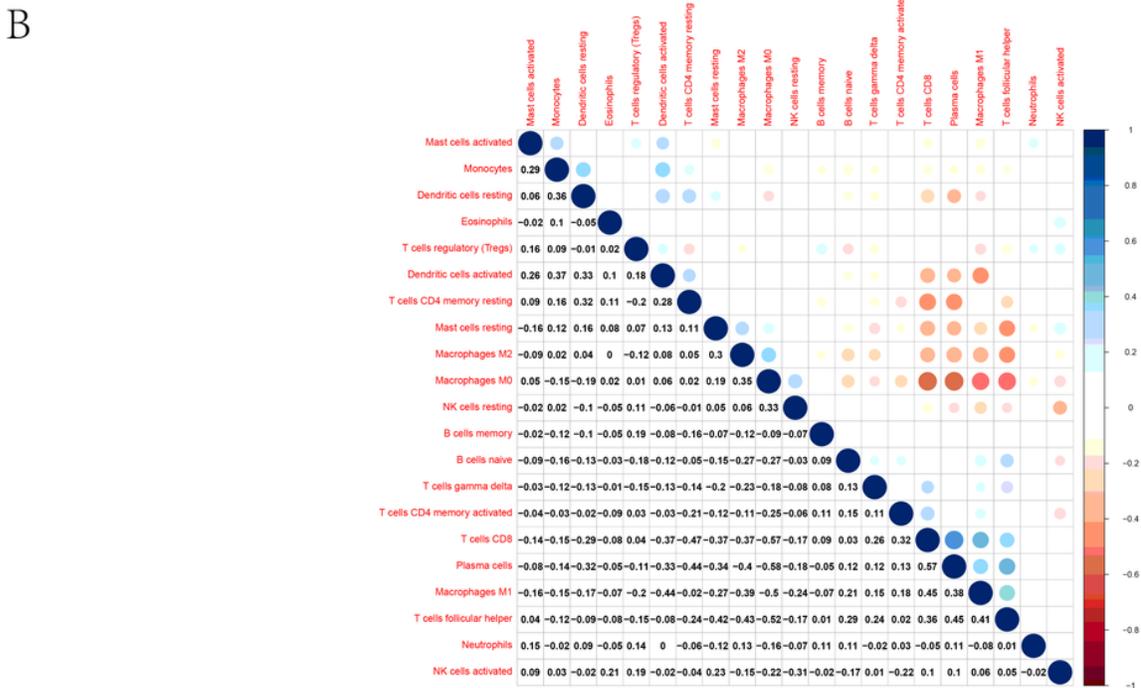
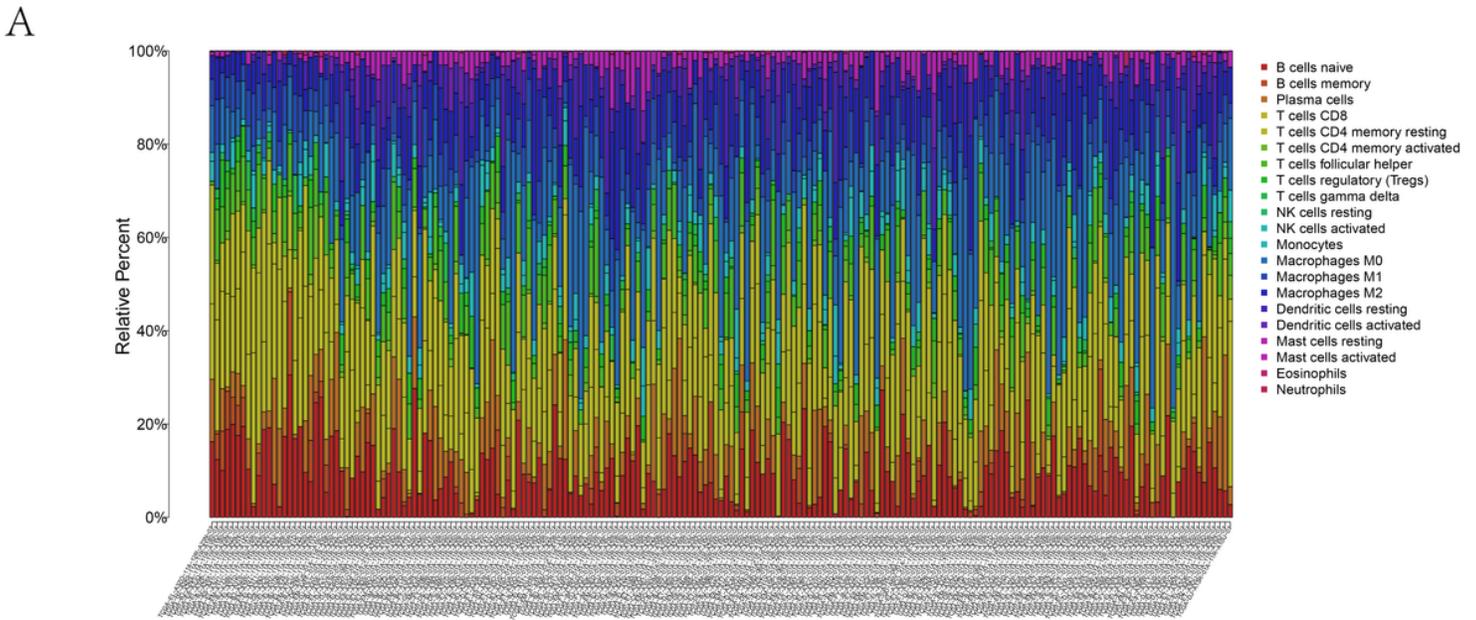


Figure 6

TIC profile and correlation analysis in TME. (A) Bar plot showing the components of TICs in THCA samples. (B) The proportions of TICs in THCA samples were displayed in correlation matrix.

Expression of CFB in different GEO datasets. The expression levels of CFB in normal and PTC tissues from the GSE33630, GSE29265, GSE35570 and GSE27155.

Supplementary Files

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

- [SupplementaryFigure.tif](#)