

Identification of key genes related to macrophage infiltration in gastric cancer based on WGCNA

Haiyan Chen

xjtu <https://orcid.org/0000-0002-1059-475X>

Cangang Zhang

Xi'an Jiaotong University

Shuai Cao

Xi'an Jiaotong University

Meng Cao

Xi'an Jiaotong University

Nana Zhang

Xi'an Jiaotong University

Ayarick Vivian Adiila

Xi'an Jiaotong University

Jinjin Zhong

Xi'an Jiaotong University

Haibo Liu (✉ Haibo52099@sina.com)

Xi'an Jiaotong University

Yili Wang (✉ wangyili@mail.xjtu.edu.cn)

Xi'an Jiaotong University

Research article

Keywords: Macrophage, GC, WGCNA, Immune infiltration, Hub genes

Posted Date: August 19th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-56342/v1>

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

[Read Full License](#)

Abstract

Background: Gastric cancer (GC) is rampant around the world. Most of the GC cases are detected in advanced stages with poor prognosis. The identification of marker genes for early diagnosis is of great significance. Studying the tumor environment is helpful to acknowledge the process of tumorigenesis, development, and metastasis.

Methods: In GEO, 22 kinds of immune cell infiltration were calculated by CIBERSORT. Macrophages were discovered remarkably infiltrated higher in GC compared with normal tissues. WGCNA was utilized to construct the network and then identify key modules and genes related to macrophages in TCGA.

Results: Finally, 18 hub genes were verified. In the PPI bar chart, the top 3 genes were chosen as hub genes involved in most pathways. On the TIMER and THPA websites, it is verified that the expression levels of CYBB, CD86 and C3AR1 genes in tumor tissues were higher than those in normal tissues.

Conclusion: These genes may work as biomarkers or targets for accurate diagnosis and treatment of GC in the future. Our findings may be a new strategy for the treatment of GC.

Introduction

Gastric cancer (GC) is the third mortality rate among cancers worldwide[1]. Although the incidence of GC is declining in many countries[2], its dismal clinical outcome still threatens the health and lives of thousands of people. There are disparities between the five-year survival rates due to various factors, but the survival rates remain very low[3, 4]. Besides, it is worth noting that both of the morbidity and mortality of GC are still extraordinarily high in China[5]. The majority (95%) of stomach cancers are adenocarcinomas [6], and no obvious symptoms are observed in the early stage. Surgical resection is the most common treatment for GC, but with poor prognosis[7]. In recent years, immunotherapy has been given high expectations. Although immunotherapy has gradually been used as the first or second-line treatment in most kinds of cancers, it was not considered as the preferential treatment of GC. Lacking effective treatment is another problem of advanced GC. Thus, it is urgent to find effective methods for early diagnosis and treatment of GC.

The tumor microenvironment (TME) has an influence on the occurrence and progression of tumors. The complex interaction between tumor cells and tumor-related immune cells occurred in TME[8]. It is mainly composed of tumor-related fibroblasts, immune cells, extra-cellular matrix, various growth factors, inflammatory factors, special physical and chemical characteristics (such as low oxygen and low pH) and cancer cells. TME plays important role in the diagnosis, prognosis, and clinical treatment sensitivity of tumors. The cells in the microenvironment can be grouped into different categories, and also have complex and significant interactions with each other, and there are some robust cell infiltration patterns[9, 10]. It is closely related to clinical prognosis. Thus, the infiltrated immune cells could be used as drug targets to improve the survival rates of cancer patients which have been recognized in current

immunotherapy strategies. By understanding immune cells infiltration, we are capable of decrypting the proportion and functional potential of immune cells in tumor tissues.

With the development of bioinformatics, many algorithms and tools are utilized to explore TME[11]. Weighted gene co-expression network analysis (WGCNA) is used to find co-expressed gene modules, and to explore the association between the gene networks and the phenotypes of interest, as well as the core genes in the network[12]. Cell type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) deconvolutes the expression matrix of immune cell subtypes based on the principle of linear support vector regression. RNA-Seq data can be used to estimate immune cell infiltration in various cancers[13].

In this study, the WGCNA coexpression network in GC was constructed, and 18 key genes associated with macrophages were determined. Subsequently, the hub genes and signal pathways were found out in PPI. Moreover, we verified that these hub genes were highly expressed in GC at the gene and protein levels. Finally, immunohistochemistry (IHC) results of the collected clinical samples also showed that these genes were highly expressed in GC. In conclusion, CYBB, CD86, and C3AR1 were identified as the potential biomarkers in GC.

Materials And Methods

Data from the GEO cohort and preprocessing

From the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>), we downloaded publicly available raw microarray expression data of GSE13911 to acquire the transcriptional data of GC. Among the datasets of GC in GEO, GSE13911 has paired normal tissues, containing 31 normal samples and 38 tumor tissues. The chip standardization method was RMA (Robust Multichip Average algorithm), referring to the chip tutorial, the standardization process is mainly divided into 3 steps: Background correction (removing array auto-fluorescence), Quantile normalization (making all intensity distributions identical), and Probeset summarization (calculating one representative value per probeset).

Quantification of tissue immune cells

CIBERSORT, a deconvolution algorithm, can calculate the cell composition of complex tissues based on the standardized gene expression data, and change the method to energize the abundance of specific cell types. The composition of immune cells in breast and liver cancer tissues was verified and successfully evaluated by flow cytometry[14]. It provides 22 kinds of common infiltrating cell expression data LM22 as a reference. Then, we employed the CIBERSORT in the R packages to evaluate the infiltration of immune cells[12]. CIBERSORT was constructed based on microarray data, thus we applied CIBERSORT for quantitative analysis of immune cells in GEO data. Xcell (<https://xcell.ucsf.edu/>) is another powerful tool for quantifying the infiltration of immune cells in tissues[15]. It is more robust for RNA-seq data. For the TCGA data, Xcell was used for the quantitative analysis of immune cells.

Data preprocessing of GC in TCGA

RNA-seq and corresponding clinical data of GC from The Cancer Genome Atlas (TCGA) database (<https://genome-cancer.ucsc.edu/>) was acquired. The RNA-seq data was in the form of HTSeq-FPKM (fragments per kilobase of transcript per million). 32 normal tissues and 381 tumor tissues were screened for further study.

Co-expression network construction

A weight co-expression network was constructed by using the R package WGCNA[12]. WGCNA is suitable for complex data analysis in multiple samples. It can calculate the expression connection between genes, identify gene modules with similar expression patterns, analyze the relationship between the gene set and the sample phenotype, and plot the regulation between the genes in the gene set, and appraisal key regulatory genes. Through this analysis, we were able to identify the co-expressed gene set, which is called modules. We also associated modules with phenotype data for further analysis and discovered potential marker genes.

The first step was to filter the gene expression data. Missing values or genes with low expression were filtered. The samples with many missing values were also filtered. WGCNA has a built-in test gene and sample function, and a basic filter could be performed through this function. After basic filtering, we proceeded to check if there were samples of outliers and judged the clustering tree of the samples.

When constructing a co-expression network, the correlation coefficients between genes were multiplied to characterize their correlation. The power value was determined, which is the soft threshold or what we call the power value and the beta value.

Differentially expressed mRNAs between GC and non-tumorous tissues

Limma package in R was used to screen out the differentially expressed mRNA. FDR-Filter = 0.01 and logFC-filter = 1 were used as the critical values. Differentially expressed genes (DEGs) were screened out for follow-up research.

Correlation analysis of immune cell infiltration and clinical information

The corresponding clinical information organized in TCGA was analyzed. The clinical information matrix was combined with the immune cell expression matrix. Xcell (<https://xcell.ucsf.edu/>) calculated the content of 22 kinds of immune cells to visualize the changes of macrophages in different clinical stages[15]. Survival analysis of immune cells and the evaluation of their effects on tumors were attached.

Functional enrichment analysis

Gene ontology (GO) enrichment analysis was realized to explore the molecular functions (MF), biological processes (BP), and cellular components (CC) related to DEGs. From the Kyoto Encyclopedia of Genes

and Genomes (KEGG), numerous functions of hub genes were presented.

The validation in TIMER (Tumor IMMune Estimation Resource)

TIMER (<https://cistrome.shinyapps.io/timer/>) uses RNA-Seq expression profiling data to detect immune cell infiltration in tumor tissues[16]. It provides the infiltration of 6 kinds of immune cells (B cells, CD4⁺ T cells, CD8⁺ T cells, Neutrophils, Macrophages, and Dendritic cells). SCNA (somatic copy number alteration) module explores the relationship between somatic cell copy number variation and immune infiltration. TIMER uses GISTIC2.0 data to examine the effect of different gene copy states on immune infiltration compared with normal tissues. The SCNA module is grouped according to the CNA (copy number alteration) of a certain gene. This module explores the different levels of infiltration in 6 kinds of immune cells between the groups. The grouping of CNA is divided into five situations: deep deletion (-2), arm-level deletion (-1), diploid / normal (0), arm-level gain (1) and high amplification (2).

Gene expression level in GEPIA (Gene Expression Profiling Interactive Analysis)

GEPIA (<http://gepia.cancer-pku.cn/>) generates a gene expression boxplot and stages plots based on user-defined inputs. The specific gene expression of the normal tissues was compared with tumor tissues. Based on the section of the pathological stages, the violin diagram was generated according to the pathological stages of the cases[17].

THPA (The Human Protein Atlas database)

The Human Protein Atlas project (<https://www.proteinatlas.org/>) contains the protein expression of normal cells, tissues, and cancers. More than 50% of all human protein-encoding genes in line with 11,200 unique proteins were included[18]. In this website, IHC was used to compare the expression of different proteins between human normal tissues and cancer tissues.

IHC

Paraffin-embedded GC tissues were collected from the First Affiliated Hospital of Xi'an Jiaotong University. IHC was performed using the EnVision immunohistochemistry kit(Dako, Denmark, k5007). Firstly, the slides were toasted at 60°C overnight. Secondly, the slides were deparaffinized with xylene and rehydrated in gradient ethanol baths. 0.01M citrate buffer was applied to retrieve the antigen. A 3% H₂O₂ solution was dropped in methanol to block endogenous peroxidase. Then, slides were treated with 5% normal goat serum and incubated in the wet box at 37°C for 30 minutes to block non-specific antigen. Later, commercially available primary antibodies: CD86 (sc-28347), CYBB (sc-130543) (both from Santa Cruz Biotechnology, Inc.), and C3aR (ER1904-90) (Hangzhou HuaAn Biotechnology Co., Ltd.) were incubated at 4°C overnight. Sections were washed with TBST (Tris Borate Saline Tween-20) 5 minutes for 3 times. The tablets were incubated with biotinylated secondary antibody against mouse and rabbit at 37°C for 30 minutes. 3,3'-diaminobenzidine (DAB) was used as the chromogen. The slices were dyed with hematoxylin and mounted after dehydration. Images were digitally captured by ImageView software.

Results

As shown in Fig. 1, the flowchart presented the process. Firstly, we analyzed the GEO dataset GSE13911. Then, we found macrophage infiltration was different in GC and normal tissues. To verify this phenomenon, we downloaded related data from TCGA. In this section, we identified 3 hub genes relevant to macrophages by using WGCNA. Finally, the former results were verified by various websites and IHC.

GEO Data processing

Fig. 2A (barplot) showed the ratio of immune cells infiltration in 69 samples. 31 samples on the left are normal tissues, 38 samples on the right are tumor tissues. A heat map (Fig. 2B) of the cells infiltration about 69 samples presented that there was a significant difference in the distribution of some cells in normal and tumor tissues. For example, resting memory CD4⁺ T cells were high in the tumor group, while plasma cells and CD8⁺ T cells have higher proportion in the normal group. CorHeatmap (Fig. 2C) was used to observe the co-expression (red) or co-inhibition (blue) of various cells in tumor tissues. We could summarize from the picture that neutrophils and activated dendritic cells had the highest correlation and the correlation coefficient was 0.65. In contrast, resting and activated mast cells were most negatively correlated and the coefficient was -0.59. Then, 31 pairs of samples could be matched in the GSE13911 data set for pairwise comparison. The violin plot was a variation of the box plot, which combines the distribution kernel density estimation curve with the box plot. The outermost shape showed the density of the location so the distribution of the data could be found. The middle white dot represented the median, the thick black bar (black box) represented the quartile range (25% quantile and 75% quantile), and the thin black line extending from it represented the 95% confidence interval. Follicular helper T cells and M0 macrophages, activated dendritic cells, and activated mast cells were high in the tumor group (Fig. 2D). The difference was statistically significant. In the normal group, plasma cells and CD8⁺ T cells and resting mast cells were highly expressed. The most important finding was that M0 macrophages in the tumor were significantly higher than the normal group. Besides, the trends of M1 and M2 macrophages were also consistent with the general knowledge in immunology. Besides, the paired plot result was unanimous in the result of the violin plot but merely looked at a certain cell more directly (Additional file 1: Figure S1). The M0 macrophages were exclusively selected. In the 31 pairs of matched samples, it could be seen that except for individual cases, most of the M0 macrophage infiltration was higher in the tumor group than the normal group.

Immune infiltration of GC in TCGA

Combined with TCGA, Xcell website was used to analyze the infiltration of immune cells. The changes of macrophages in different clinical stages were observed.

In KS multi-group test, the p-value indicated the overall situation. The significant discrepancy stated the difference between each group. There was an upward trend in Fig. 3, which meant the macrophage fraction was increasing as the clinical-grade increased.

Differential analysis of genes in TCGA samples

In TCGA, 32 samples were in the normal group and 381 samples were in the tumor group. The following parameters (FDR-Filter=0.01, logFC-filter=1) were used to analyze the differences between the genes in TCGA samples. A total of 6227 DEGs were picked out.

WGCNA co-expression network analysis

WGCNA achieved the goal of quickly locking core genes by grouping genes (modules) and associating gene modules with phenotypes. Next, WGCNA was utilized to divide these genes into various modules. Macrophages and neutrophils were used as independent variables to calculate the modules related to macrophages. Finally, Module Membership (MM) > 0.8 and the Gene-Significance (GS) > 0.5 were used to screen out 18 genes that were positively associated with macrophages.

The first step of sample clustering in WGCNA was to screen out the samples with outlier expression and clusters those with a similar expression. It could be delineated from Fig. 4A that the TCGA-BR-6710-01A-11R-1884-13 sample was distinct from other samples. The filtering principle of the soft threshold is to make the network more scale-free. Filtered soft threshold, undirected networks with power less than 15 or directed networks with power less than 30 could not lead the scale-free network structure R^2 to reach to 0.8 or the average connectivity to drop to below 100. This may result from batch effects, sample heterogeneity, and complicated experimental conditions on expression, which needed to be removed.

The second step of the WGCNA gene clustering was to cluster genes with similar expression trends. By using a dynamic tree cut[19], a hierarchical clustering tree was formed. Its rationale was a module (pathway) based analysis. On the tree, each leaf represented a gene. The genes with similar expression data were tightly connected, formed a branch of the tree, and represented a gene module. After that, several modules were generated (Fig. 4B).

The third step was to choose a suitable cutting position. We calculated all integers from 1 to 20 as thresholds to test the optimal threshold. Among them, power Estimate was the best power value, fit Indices saved the characteristics of the network corresponding to each power and k is the connection degree value. The average value of the connection was visualized and then generated Fig. 4C, D. When the y-axis was equal to 0.9, the intersection with the curve was exactly equal to 4 so we chose 4 as the power value.

The fourth step was to merge the clustered gene modules and make the modules with similar expressions into a large module. We used 4 as a beta value to build a gene module. In Fig. 5A, red indicated a positive correlation and green indicated a negative correlation, the darker the color, the stronger the correlation.

The key step was to associate the genes of different modules with the content of macrophages and pick out the modules that were positively correlated with the content of macrophages. The MEpink module in Fig. 5B was selected as most positively correlated with macrophages and the MEbrown module was

negatively correlated with macrophages. This research focused on the modules that had a poor survival expectation and the MEpink module was positively correlated with the content of macrophages based on this module.

The sixth step could be realized in Fig. 5C that the MEpink module had a strong positive correlation with macrophages. Among them, there were 149 genes in the pink module. In case of $MM > 0.8$ and $GS > 0.5$, 18 genes most relevant to macrophage infiltration were selected for following research. Table 1 exhibited the function of 18 genes. A barplot was drawn (Additional file 2: Figure S2), and these genes were highly expressed in tumors. A total of 18 genes were all up-regulated in tumor tissues in Heat map (Additional file 3: Figure S3). From the correlation heat map (Fig. 5D), we could know the co-expression relation between these genes. All the values were greater than 0.5, indicating that these genes were correlated with each pair, which was also confirmed by WGCNA. After finding the modules associated with the phenotypes, the PPI bar chart (Fig. 5E) was obtained. Genes were at the core position, and the genes could be screened according to different conditions. Cytoscape software was used to form Fig. 5F. CD86, CYBB, and C3AR1 were identified as hub genes.

Verifying the tissue expression level in THPA and GEPIA database

We obtained the transcription level of real hub genes from THPA. As Additional file 4: Figure S4A exhibited, the expression level of three hub genes was remarkably high in cancer tissues, which also supported our suppose. CYBB, C3AR1, and CD86 were highly correlated, and had high expression in GC (Additional file 4: Figure S4C-E). There was a rising trend in Additional file 4: Figure S4B, but taking out any of them had little effect on survival ($p > 0.05$), indicating that, they might work together.

Verifying the protein expression level by IHC

After IHC staining, it was mostly intuitive that CD86 was highly up-regulated in GC tissues (Fig. 6A). Fig. 6B showed a rise of CYBB expression in tumor tissues compared to adjacent non-tumorous tissues. The positive cell density of C3AR1 was also higher in GC tissues than that in normal tissues (Fig. 6C). These results were consistent with THPA.

Verifying immune infiltration with TIMER

Using the TIMER website, we found the expression level of CD86, CYBB, and C3AR1 had a positive correlation with CD8⁺T Cells, CD4⁺T Cells, Macrophages, Neutrophils, and Dendritic cells (Fig. 7B). The degree of macrophage infiltration significantly affected the prognosis. Hence it is worthy of further research and exploration. These genes affected macrophages and had an impact on survival (5-year survival rate, 10-year survival rate and long term survival rate) (Fig. 7A). In order to explore the relationship between gene copy number variation and immune infiltration abundance, SCNA module was used to analyze the effect of different somatic copy number alterations of CD86, CYBB and C3AR1 on the immune cell infiltration in GC. It showed that these genes had a great influence on immune cell infiltration (Fig. 7C).

Gene function analysis

From the GO enrichment pathway map, we acquainted these 18 genes that were mainly enriched in several pathways (Fig. 8A). These genes mainly regulated the production of tumor necrosis superfamily cytokines. Table 2 exhibited the details. In GC, another GO-Biological Process enrichment analysis of DEGs was displayed in Additional file 5: Table S1. It was delineated that these 18 genes were mainly enriched in the phagosome pathway in the KEGG enrichment pathway map (Fig. 8B). GO-Cellular Components, GO-Molecular Function, and KEGG pathway enrichment analysis of DEGs in GC samples were showed in Table 3.

ClueGO as a Cytoscape plugin that integrated gene ontology (GO) terminology and KEGG/BioCarta pathway were used to create a functionally ordered GO/pathway network[20]. After screening ($p < 0.05$), C3AR1 participated in the regulation of mononuclear cell migration and monocyte chemotaxis. CD86 and CYBB involved in the regulation of cytokine biosynthetic process and interleukin-2 production (Fig. 8C).

Discussion

Although the incidence of GC continues to decline in Western countries, it is still a common tumor in developing countries. At present, surgery is still the main treatment for localized GC. However, most GC is diagnosed at advanced stages, the recurrence rate remains very high. Also, combined chemotherapy for patients with GC usually causes drug resistance. In recent years, immunotherapy has become one of the most promising cancer treatment strategies and had significant effects on several types of tumors[21-23]. Immunotherapy for GC also has promising prospects[24].

Tumor-associated macrophages (TAMs) are one of the most abundant immune cell populations in the TME[25-27]. TAMs might be a potential marker for poor prognosis. While the mechanisms of TAMs in GC is still unclear, further research is required .

In our study, we first calculated the infiltration of various immune cells in GC tissues and matched normal tissues(GSE13911) by using CIBERSORT. We found that the M0 macrophage infiltrated in the tumor tissues was significantly higher than the normal tissues ($p=0.001$). Subsequently, we found that the degree of macrophage infiltration had a positive correlation with disease progression by analyzing the data of TCGA. We suspected that tumor progression was related to the degree of macrophage infiltration. To verify our hypothesis, the WGCNA algorithm was performed to find 18 genes that were positively related to the degree of macrophage infiltration in GC. Furthermore, we identified three hub genes: CD86, CYBB, and C3AR1 based on the protein-protein interaction network for further studies. We used TIMER to explore the effect of various immune cell infiltration levels on the survival of GC patients. Results showed that the more TAM infiltration, the worse the short-term and long-term survival of patients. Moreover, we found that these three genes affected various tumor-infiltrating immune cells, especially tumor-infiltrating macrophages. Again, these hub genes were found to be highly expressed in tumor tissues than normal tissues by using the THPA database, which supported our results. As the tumor progressed, these hub genes expressions were confirmed to rise in GEPIA. Then, we collected some tumor tissues and adjacent

tissue samples in the clinic. Through IHC, we also confirmed that these genes were highly expressed in tumor tissues. Finally, we conducted GO and KEGG enrichment analysis on these 18 genes and found these genes were related to the production of TNF and IL-2 and the formation of phagosomes.

Cytochrome B-245 Beta Chain (CYBB) is expressed in eosinophils, neutrophils, and B lymphocytes, et al[28], responding to many inflammatory cytokines and stimuli such as IFN- γ , LPS, and TNF- α [29, 30]. It has been proposed as a primary component of the microbicidal oxidase system of phagocytes[31]. CYBB can lead to Immunodeficiency 34 (IMD34) and Granulomatous Disease. The terminal of a respiratory chain needs it. Cellular pH can be regulated by CYBB.

C3AR1 is activated by its natural ligand, C3a, which is a 77 amino acid split product converted by C3 protein and is traditionally considered to be mainly pro-inflammatory[32]. When the complement system is activated, an anaphylatoxin called Complement C3a Receptor 1(C3AR1) is released. After stimulation, chemotaxis, granule enzyme, and superoxide anion were produced[33]. It is widely expressed in a variety of differentiated hematopoietic cell lines, in the lung, spleen, ovary, placenta, small intestine, whole brain, heart, and endothelial cells. Also, it was mainly expressed in lymphoid tissues[34].

CD86 is a member of the immunoglobulin superfamily. It works as a ligand for two proteins-CD28 antigen and cytotoxic T-lymphocyte-associated protein 4(CTLA-4). It is widely known that the CD86 molecule (B7-2) belongs to the B-7 family. Together with CD80 molecule (B7-1), it is expressed on antigen-presenting cells(APCs, like dendritic cells, macrophages, and B-cells)[35]. Diseases related to CD86 include acute myocarditis and myocarditis. CD86, as a costimulatory molecule, participates in the formation and regulation of immune synapses[36]. Dai et al showed that CD86 expression in CLL was lower than that in normal B cells, but its role in CLL cell survival was not clear[37, 38]. In general, C3AR1 is related to complement function. CYBB and TNF are pro-inflammatory genes. CYBB is related to interaction between neutrophils and macrophages. CD86 is a macrophage activation marker. Its high expression indicates that macrophages are activated.

The articles on identification of GC related prognostic genes by WGCNA have sprung up. For instance, one research explored the prognosis of GC, but not related to macrophages[39]. Another study aimed to identify potential vital miRNAs and modules associated with the progression of GC[40]. Our study revealed the important genes associated with macrophages in GC for the first time.

However, our study had limitations that should be acknowledged. To begin with, the defect of this paper was that several other modules could be used to identify the hub genes, but we only chose one on consideration. Again, we did not differentiate the types of macrophages although M1 and M2 macrophages have opposite biological effects, and they were derived from M0 macrophages in different tumor microenvironments. But using the Xcell result, we did not differentiate the types of macrophages. Last but not least, we analyzed the results based on bioinformatics methods and only predicted the relevant results. Therefore, further molecular biological experiments were required to confirm the function of the hub genes associated with GC.

In summary, the macrophage immune infiltration was higher in the GC tissue than its paired normal tissues. First of all, we combined immune infiltration with WGCNA. Subsequently, we applied the TCGA to seek the genes which could explain the source of this disparity. Combined with WGCNA, the key modules and genes associated with macrophages were identified. Lastly, three hub genes were discovered and validated, which can well predict the degree of macrophage infiltration in GC. These genes can be used as drug targets to reduce macrophage infiltration and improve survival rate in the future.

Additional Files

Additional file 1: Figure S1. The paired plot of the samples in GEO. CD8+T cells and plasma cells, resting Mast cells were lower in tumor group. On the contrary, activated Dendritic cells and follicular helper T cells, macrophages M0 were higher in tumor group. * $P < 0.05$ is considered statistically significant.

Additional file 2: Figure S2. Barplot of the 18 genes. 18 genes were up-regulated in tumor group. All of them were statistically different. Type N represents normal group, type T represents tumor group.

Additional file 3: Figure S3. Heat map of 18 genes. The 18 genes were up-regulated in tumor tissues.

Additional file 4: Figure S4. The verification of hub genes on protein level by THPA and GEPIA. (A) In GEPIA the expression level of CD86, CYBB and C3AR1 was showed on the left. Translational level of GC and normal tissues were demonstrated on the upper side and the lower side in THPA. (B) Violin plot of CD86, CYBB and C3AR1 in different stages. (C-E) Co-expression of every two genes.

Additional file 5: Table S1. Other GO-Biological Process enrichment analysis of DEGs in STAD samples.

Abbreviations

GC, gastric cancer; TME, tumor microenvironment; WGCNA, weighted gene co-expression network analysis; CIBERSORT, Cell type Identification by Estimating Relative Subsets of RNA Transcripts; IHC, immunohistochemistry; GEO, Gene Expression Omnibus; RMA, Robust Multichip Average; TCGA, The Cancer Genome Atlas; DEGs, differentially expressed genes; GO, Gene Ontology; MF, molecular functions; BP, biological processes; CC, cellular components; KEGG, the Kyoto Encyclopedia of Genes and Genomes; TIMER, Tumor IMMune Estimation Resource; SCNA, somatic copy number alteration; CNA, copy number alteration; GEPIA, Gene Expression Profiling Interactive Analysis; THPA, The Human Protein Atlas database; FPKM, fragments per kilobase of transcript per million; MM, Module Membership; GS, Gene-Significance; TBST, Tris Borate Saline Tween-20; DAB, diaminobenzidine

Declarations

Acknowledgements

Not applicable.

Authors' contributions

HYC and CGZ conceived the idea. HYC wrote the scripts. CGZ carried out the gene expression bioinformatics analyses. All coauthors are involved in interpreting the results, developing the discussion section, and writing the manuscript. All coauthors have read and agreed with the final manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets analyzed during the current study are available in the GEO repository under accession numbers GSE13911 and TCGA (<https://genome-cancer.ucsc.edu/>). The immunohistochemistry images and antibody staining levels are available in TPHA (<https://www.proteinatlas.org>).

Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the Ethical committee of the First Affiliated Hospital of Xi'an Jiaotong University (No: XJTU1AF2018LSK-292) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute for Cancer Research, School of Basic Medical Science of Xi'an Jiaotong University, 710061 Xi'an, Shaanxi, China. ²Department of Pathogenic Microbiology and Immunology, School of Basic Medical Sciences, Xi'an Jiaotong University, 710061 Xi'an, Shaanxi, China.

³Department of Orthopedics, Second Affiliated Hospital of Xi'an Jiaotong University, 710061 Xi'an, Shaanxi, China. ⁴Department of Hematology, The First Affiliated Hospital of Xi'an Jiaotong University, 710061 Xi'an, Shaanxi, China.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019;69(1):7-34.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(1):7-30.

3. Youn HG, An JY, Choi MG, Noh JH, Sohn TS, Kim S. Recurrence after curative resection of early gastric cancer. *Annals of surgical oncology*. 2010;17(2):448-54.
4. Fields R, Strong V, Gönen M, Goodman K, Rizk N, Kelsen D, et al. Recurrence and survival after pathologic complete response to preoperative therapy followed by surgery for gastric or gastroesophageal adenocarcinoma. *British journal of cancer*. 2011;104(12):1840-7.
5. Zhao L-L, Huang H, Wang Y, Wang T-B, Zhou H, Ma F-H, et al. Lifestyle factors and long-term survival of gastric cancer patients: A large bidirectional cohort study from China. *World Journal of Gastroenterology*. 2020;26(14):1613.
6. Ma J, Shen H, Kapesa L, Zeng S. Lauren classification and individualized chemotherapy in gastric cancer. *Oncol Lett*. 2016;11(5):2959-64.
7. Liu C, Pan C, Shen J, Wang H, Yong L. Identification of serum amyloid A in the serum of gastric cancer patients by protein expression profiling. *Oncology letters*. 2012;3(6):1259-62.
8. Sperlich J, Kerr R, Teusch N. The marine natural product pseudopterosin blocks cytokine release of triple-negative breast cancer and monocytic leukemia cells by inhibiting NF- κ B signaling. *Marine drugs*. 2017;15(9):262.
9. Tamborero D, Rubio-Perez C, Muiños F, Sabarinathan R, Piulats JM, Muntasell A, et al. A pan-cancer landscape of interactions between solid tumors and infiltrating immune cell populations. *Clinical Cancer Research*. 2018;24(15):3717-28.
10. Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The cancer genome atlas pan-cancer analysis project. *Nature genetics*. 2013;45(10):1113.
11. Lin J, Yu M, Xu X, Wang Y, Xing H, An J, et al. Identification of biomarkers related to CD8+ T cell infiltration with gene co-expression network in clear cell renal cell carcinoma. *Aging (Albany NY)*. 2020;12(4):3694.
12. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*. 2008;9(1):559.
13. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nature methods*. 2015;12(5):453-7.
14. Hackl H, Charoentong P, Finotello F, Trajanoski Z. Computational genomics tools for dissecting tumour-immune cell interactions. *Nature Reviews Genetics*. 2016;17(8):441.
15. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. 2017;18(1):220.
16. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer research*. 2017;77(21):e108-e10.
17. Zhou L, Liu S, Li X, Yin M, Li S, Long H. Diagnostic and prognostic value of CEP55 in clear cell renal cell carcinoma as determined by bioinformatics analysis. *Molecular medicine reports*. 2019;19(5):3485-96.

18. Zhou Z, Cheng Y, Jiang Y, Liu S, Zhang M, Liu J, et al. Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis. *International journal of biological sciences*. 2018;14(2):124.
19. Wang Y, Chen L, Wang G, Cheng S, Qian K, Liu X, et al. Fifteen hub genes associated with progression and prognosis of clear cell renal cell carcinoma identified by coexpression analysis. *Journal of cellular physiology*. 2019;234(7):10225-37.
20. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009;25(8):1091-3.
21. Kerr WG, Chisholm JD. The Next Generation of Immunotherapy for Cancer: Small Molecules Could Make Big Waves. *J Immunol*. 2019;202(1):11-9.
22. Yang Y. Cancer immunotherapy: harnessing the immune system to battle cancer. *J Clin Invest*. 2015;125(9):3335-7.
23. Granier C, Karaki S, Roussel H, Badoual C, Tran T, Anson M, et al. [Cancer immunotherapy: Rational and recent breakthroughs]. *Rev Med Interne*. 2016;37(10):694-700.
24. Wu X, Gu Z, Chen Y, Chen B, Chen W, Weng L, et al. Application of PD-1 Blockade in Cancer Immunotherapy. *Comput Struct Biotechnol J*. 2019;17:661-74.
25. Solinas G, Germano G, Mantovani A, Allavena P. Tumor-associated macrophages (TAM) as major players of the cancer-related inflammation. *Journal of leukocyte biology*. 2009;86(5):1065-73.
26. Pang L, Han S, Jiao Y, Jiang S, He X, Li P. Bu Fei Decoction attenuates the tumor associated macrophage stimulated proliferation, migration, invasion and immunosuppression of non-small cell lung cancer, partially via IL-10 and PD-L1 regulation. *Int J Oncol*. 2017;51(1):25-38.
27. Franklin RA, Li MO. Ontogeny of Tumor-associated Macrophages and Its Implication in Cancer Regulation. *Trends Cancer*. 2016;2(1):20-34.
28. Nauseef WM, Borregaard N. Neutrophils at work. *Nature Immunology*. 2014;15(7):602.
29. Newburger JW, Burns JC, Beiser AS, Loscalzo J. Altered lipid profile after Kawasaki syndrome. *Circulation*. 1991;84(2):625-31.
30. Frazão JB, Thain A, Zhu Z, Luengo M, Condino-Neto A, Newburger PE. Regulation of CYBB Gene Expression in Human Phagocytes by a Distant Upstream NF- κ B Binding Site. *J Cell Biochem*. 2015;116(9):2008-17.
31. Belambri SA, Rolas L, Raad H, Hurtado-Nedelec M, Dang PMC, El-Benna J. NADPH oxidase activation in neutrophils: role of the phosphorylation of its subunits. *European journal of clinical investigation*. 2018;48:e12951.
32. Coulthard LG, Woodruff TM. Is the complement activation product C3a a proinflammatory molecule? Re-evaluating the evidence and the myth. *J Immunol*. 2015;194(8):3542-8.
33. Gaudet P, Livstone MS, Lewis SE, Thomas PD. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief Bioinform*. 2011;12(5):449-62.

34. Brennan FH, Jogia T, Gillespie ER, Blomster LV, Ruitenber MJ. Complement receptor C3aR1 controls neutrophil mobilization following spinal cord injury through physiological antagonism of CXCR2. *JCI Insight*. 2019;4(9):e98254.
35. Takács F, Tolnai-Kriston C, Hernádfői M, Szabó O, Szalóki G, Szepesi Á, et al. The Effect of CD86 Expression on the Proliferation and the Survival of CLL Cells. *Pathology & Oncology Research*. 2019;25(2):647-52.
36. Huemer M, Rebhandl S, Zaborsky N, Gassner FJ, Hainzl S, Weiss L, et al. AID induces intraclonal diversity and genomic damage in CD86(+) chronic lymphocytic leukemia cells. *Eur J Immunol*. 2014;44(12):3747-57.
37. Brzostek J, Gascoigne NR, Rybakin V. Cell Type-Specific Regulation of Immunological Synapse Dynamics by B7 Ligand Recognition. *Front Immunol*. 2016;7:24.
38. Dai ZS, Chen QF, Lu HZ, Xie Y. Defective expression and modulation of B7-2/CD86 on B cells in B cell chronic lymphocytic leukemia. *Int J Hematol*. 2009;89(5):656-63.
39. Zhao X, Cai H, Wang X, Ma L. Discovery of signature genes in gastric cancer associated with prognosis. *Neoplasma*. 2016;63(2):239-45.
40. Gong C, Hu Y, Zhou M, Yao M, Ning Z, Wang Z, et al. Identification of specific modules and hub genes associated with the progression of gastric cancer. *Carcinogenesis*. 2019;40(10):1269-77.

Tables

Table 1. Functional roles of 18 hub genes.

No	Gene symbol	Full name	Function
1	HAVCR2	Hepatitis A virus cellular receptor 2	HAVCR2 is an immune checkpoint and together with other inhibitory receptors including programmed cell death protein 1 and lymphocyte activation gene 3 protein mediates the CD8+ T-cell exhaustion.
2	SLAMF8	SLAM family member 8	This gene encodes a member of the CD2 family of cell surface proteins involved in lymphocyte activation. These proteins are characterized by Ig domains.
3	CD300LF	CD300 molecule like family member f	Members of this family are cell surface glycoproteins with a single IgV-like extracellular domain and are involved in the regulation of immune response.
4	CD86	CD86 molecule	Receptor involved in the costimulatory signal essential for T-lymphocyte proliferation and interleukin-2 production, by binding CD28 or CTLA-4.
5	SIGLEC7	Sialic acid-binding Ig like lectin 7	Putative adhesion molecule that mediates sialic-acid dependent binding to cells. Preferentially binds to alpha-2,3- and alpha-2,6-linked sialic acid.
6	PILRA	Paired immunoglobulin like type 2 receptor alpha	PILRA is thought to act as a cellular signaling inhibitory receptor by recruiting cytoplasmic phosphatases like PTPN6/SHP-1 and PTPN11/SHP-2.
7	CYBB	Cytochrome b-245 beta chain	It is the terminal component of a respiratory chain that transfers single electrons from cytoplasmic NADPH across the plasma membrane to molecular oxygen on the exterior.
8	PLA2G7	Phospholipase A2 group VII	The protein encoded by this gene is a secreted enzyme that catalyzes the degradation of platelet-activating factor to biologically inactive products.
9	SIGLEC9	Sialic acid-binding Ig like lectin 9	Putative adhesion molecule that mediates sialic-acid dependent binding to cells.
10	C3AR1	Complement C3a receptor 1	Receptor for the chemotactic and inflammatory peptide anaphylatoxin C3a. This receptor stimulates chemotaxis, granule enzyme release, and superoxide anion production.
11	FCGR1A	Fc fragment of IgG receptor Ia	This gene encodes a protein that plays an important role in the immune response. This protein is a high-affinity Fc-gamma receptor.
12	HK3	Hexokinase 3	This gene encodes hexokinase 3. Similar to hexokinases 1 and 2, this allosteric enzyme is inhibited by its product glucose-6-phosphate.
13	LILRB4	Leukocyte immunoglobulin-like receptor B4	This gene is a member of the leukocyte immunoglobulin-like receptor family, which is found in a gene cluster at chromosomal region 19q13.4.
14	CLEC7A	C-type lectin domain containing 7A	This gene is closely linked to other CTL/CTLD superfamily members on chromosome 12p13 in the natural killer gene complex region.
15	FPR3	Formyl Peptide Receptor 3	This gene includes G protein-coupled receptor activity and N-formyl peptide receptor activity. An important paralog of this gene is FPR2.
16	NCF2	Neutrophil cytosolic factor 2	This gene encodes neutrophil cytosolic factor 2, the 67-kilodalton cytosolic subunit of the multi-protein NADPH oxidase complex found in neutrophils.
17	IGSF6	Immunoglobulin Superfamily Member 6	This gene related to transmembrane signaling receptor activity.
18	GPR141	G protein-coupled receptor 141	GPR141 is a member of the rhodopsin family of G protein-coupled receptors (GPRs)

Table2. Top 20 GO-Biological Process enrichment analysis of DEGs in GC samples.

Ontology	Term	Description	Count	p.adjust
BP	GO:0045730	respiratory burst	4	2.42E-05
BP	GO:1903557	positive regulation of tumor necrosis factor superfamily cytokine production	4	4.15E-04
BP	GO:0042554	superoxide anion generation	3	1.09E-03
BP	GO:0071675	regulation of mononuclear cell migration	3	1.60E-03
BP	GO:1903555	regulation of tumor necrosis factor superfamily cytokine production	4	1.60E-03
BP	GO:0071706	tumor necrosis factor superfamily cytokine production	4	1.60E-03
BP	GO:0032663	regulation of interleukin-2 production	3	1.60E-03
BP	GO:0032623	interleukin-2 production	3	2.12E-03
BP	GO:0042108	positive regulation of cytokine biosynthetic process	3	2.39E-03
BP	GO:0006801	superoxide metabolic process	3	2.58E-03
BP	GO:0002479	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	3	2.58E-03
BP	GO:0042590	antigen processing and presentation of exogenous peptide antigen via MHC class I	3	2.58E-03
BP	GO:0002683	negative regulation of immune system process	5	2.58E-03
BP	GO:0001819	positive regulation of cytokine production	5	2.58E-03
BP	GO:0002430	complement receptor-mediated signaling pathway	2	2.58E-03
BP	GO:0045086	positive regulation of interleukin-2 biosynthetic process	2	2.58E-03
BP	GO:0032760	positive regulation of tumor necrosis factor production	3	2.58E-03
BP	GO:0051249	regulation of lymphocyte activation	5	2.58E-03
BP	GO:0071674	mononuclear cell migration	3	2.73E-03
BP	GO:0002474	antigen processing and presentation of peptide antigen via MHC class I	3	3.15E-03

Table 3. GO-Cellular Components, GO-Molecular Function, and KEGG pathway enrichment analysis of DEGs in GC samples.

Ontology	Term	Description	Count	p.adjust
CC	GO:0043020	NADPH oxidase complex	2	3.69E-03
CC	GO:0030667	secretory granule membrane	3	3.46E-02
CC	GO:0030139	endocytic vesicle	3	3.46E-02
CC	GO:0035579	specific granule membrane	2	3.46E-02
CC	GO:1990204	oxidoreductase complex	2	4.15E-02
CC	GO:0045335	phagocytic vesicle	2	4.76E-02
MF	GO:0016175	superoxide-generating NADPH oxidase activity	2	2.32E-03
MF	GO:0050664	oxidoreductase activity, acting on NAD(P)H, oxygen as acceptor	2	2.32E-03
MF	GO:0030246	Carbohydrate-binding	4	2.32E-03
MF	GO:0042287	MHC protein binding	2	9.31E-03
MF	GO:0016651	oxidoreductase activity, acting on NAD(P)H	2	4.90E-02
MF	GO:0009055	electron transfer activity	2	4.90E-02
KEGG	hsa04145	Phagosome	4	1.65E-03
KEGG	hsa05140	Leishmaniasis	3	2.98E-03
KEGG	hsa05150	Staphylococcus aureus infection	3	3.82E-03
KEGG	hsa04380	Osteoclast differentiation	3	6.69E-03

Figures

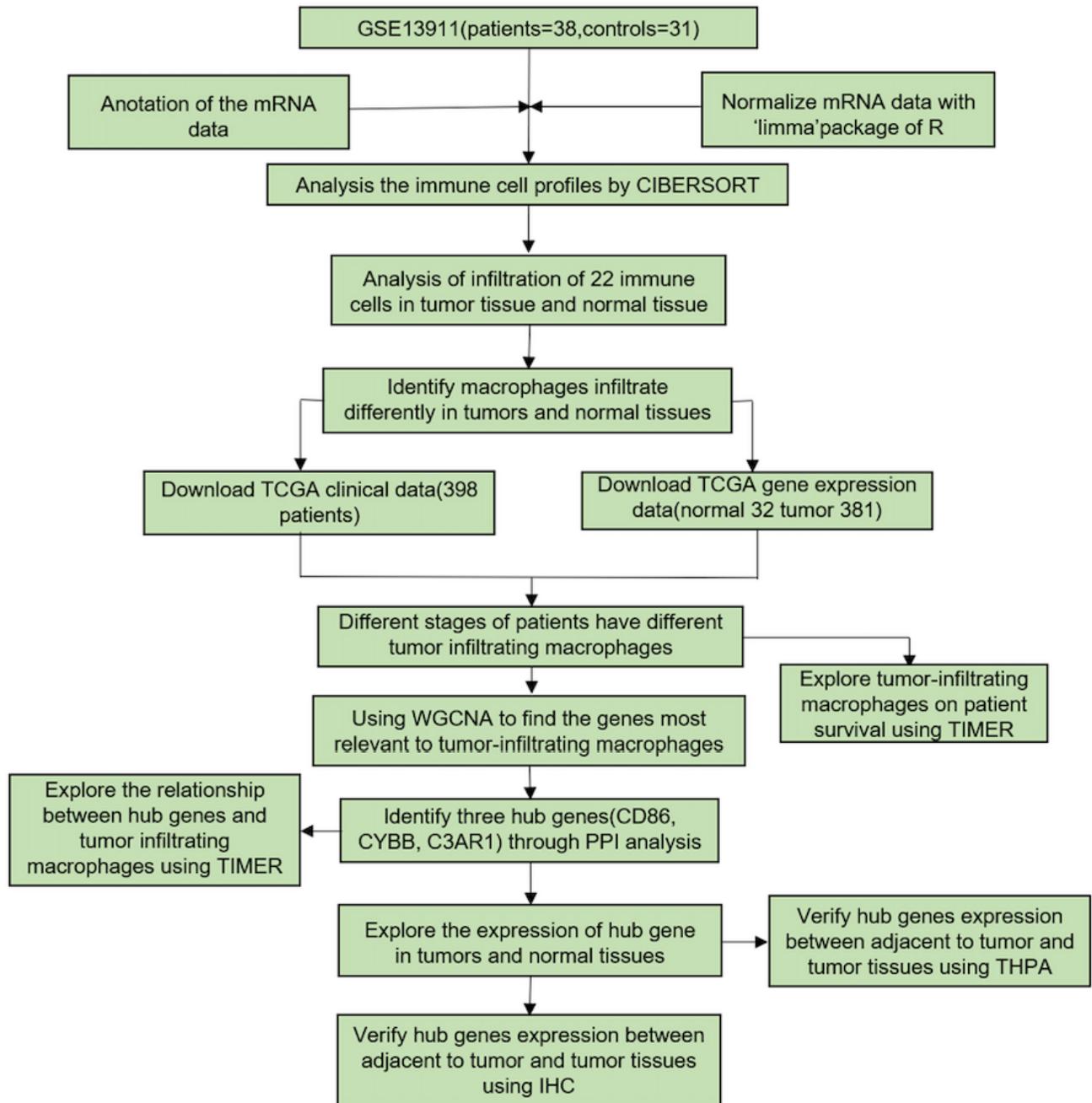


Figure 1

The flowchart of the data processing. GSE13911 was used to analyze the immune cell profiles by CIBERSORT. Corresponding TCGA gene expression and clinical data were used to find the genes related to tumor-infiltrating macrophages by WGCNA. Through PPI analysis, 3 hub genes were identified. The relationship between hub genes and tumor-infiltrating macrophages was explored by TIMER. The expression of hub genes in tumor and normal tissues was identified in GEPIA. THPA was utilized to verify hub genes expression between adjacent non-tumorous tumor and tumor tissues. This result was validated by IHC.

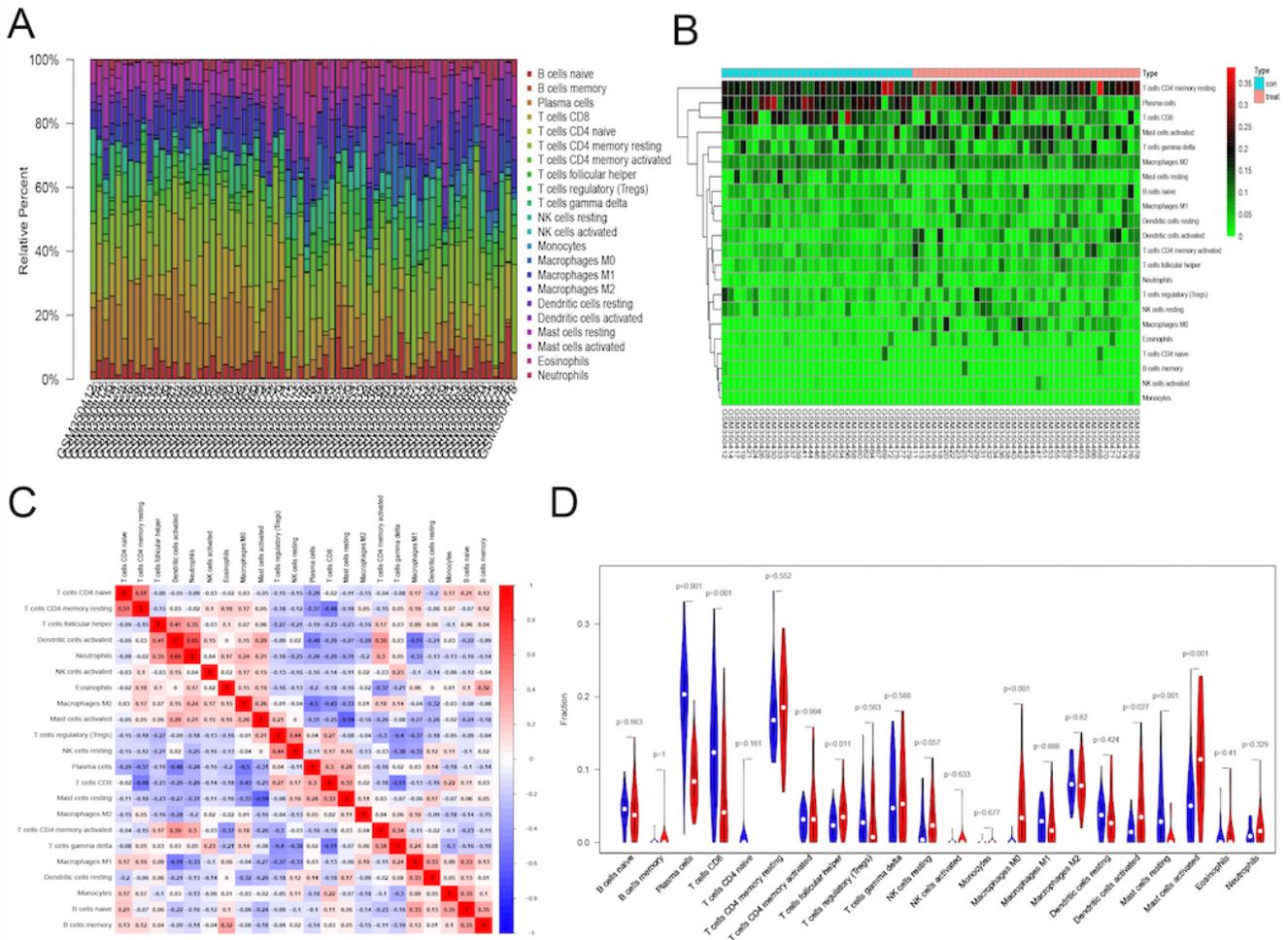


Figure 2

CIBERSORT algorithm for immune infiltration analysis of GEO data. (A) relative percent of 22 kinds of immune cells in GEO. (B) A heatmap of the cell infiltration data in 69 samples. (C) CorHeatmap of 22 immune cells in normal and tumor tissues. (D) A comparison of normal and tumor tissues.

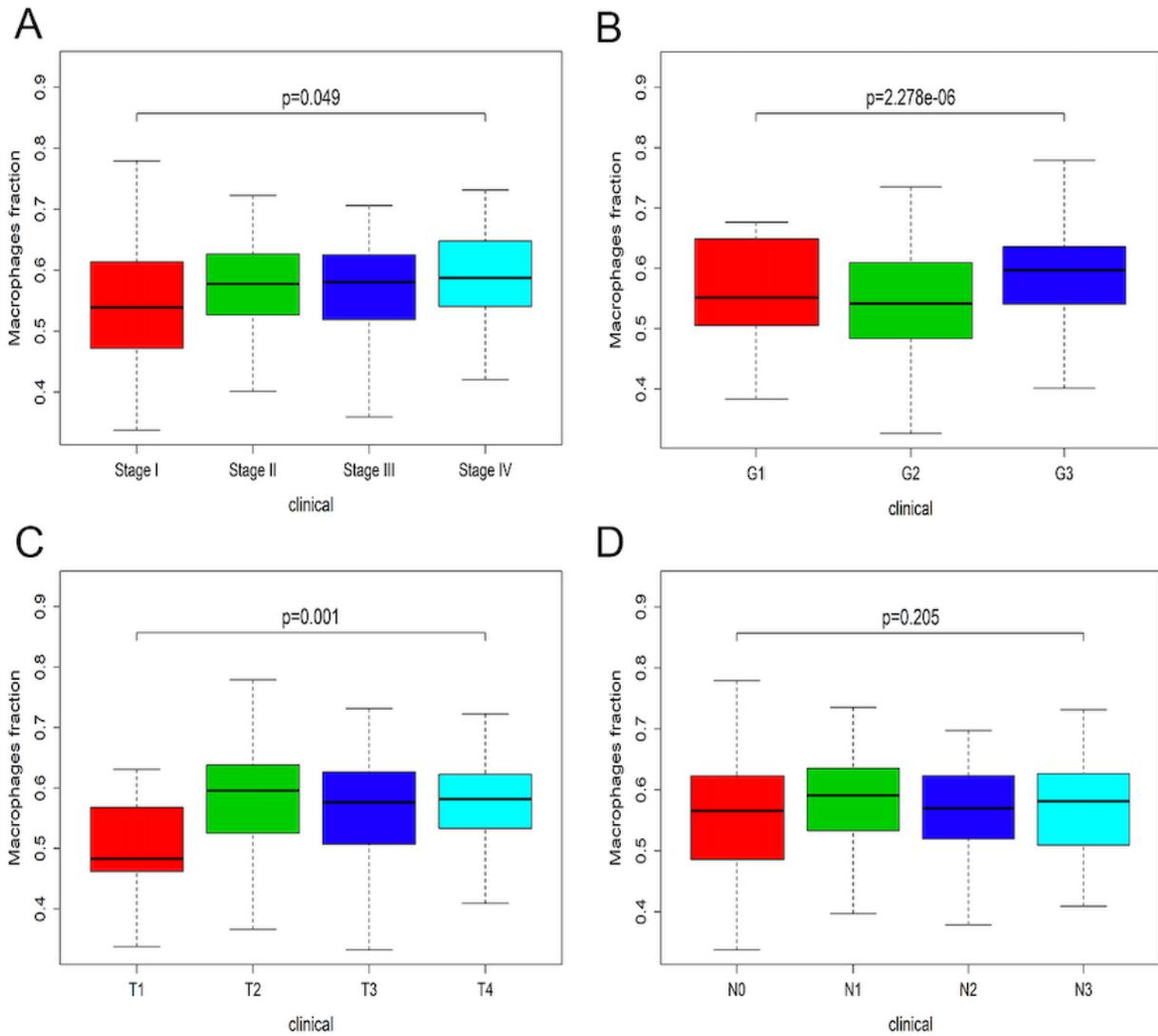


Figure 3

Clinical stage of macrophage in TCGA. (A) macrophage fraction in the I-IV stage. (B) macrophage fraction in G1-G3 grade. (C) macrophage fraction in the T1-T4 stage. (D) macrophage fraction in N0-N3 stage.

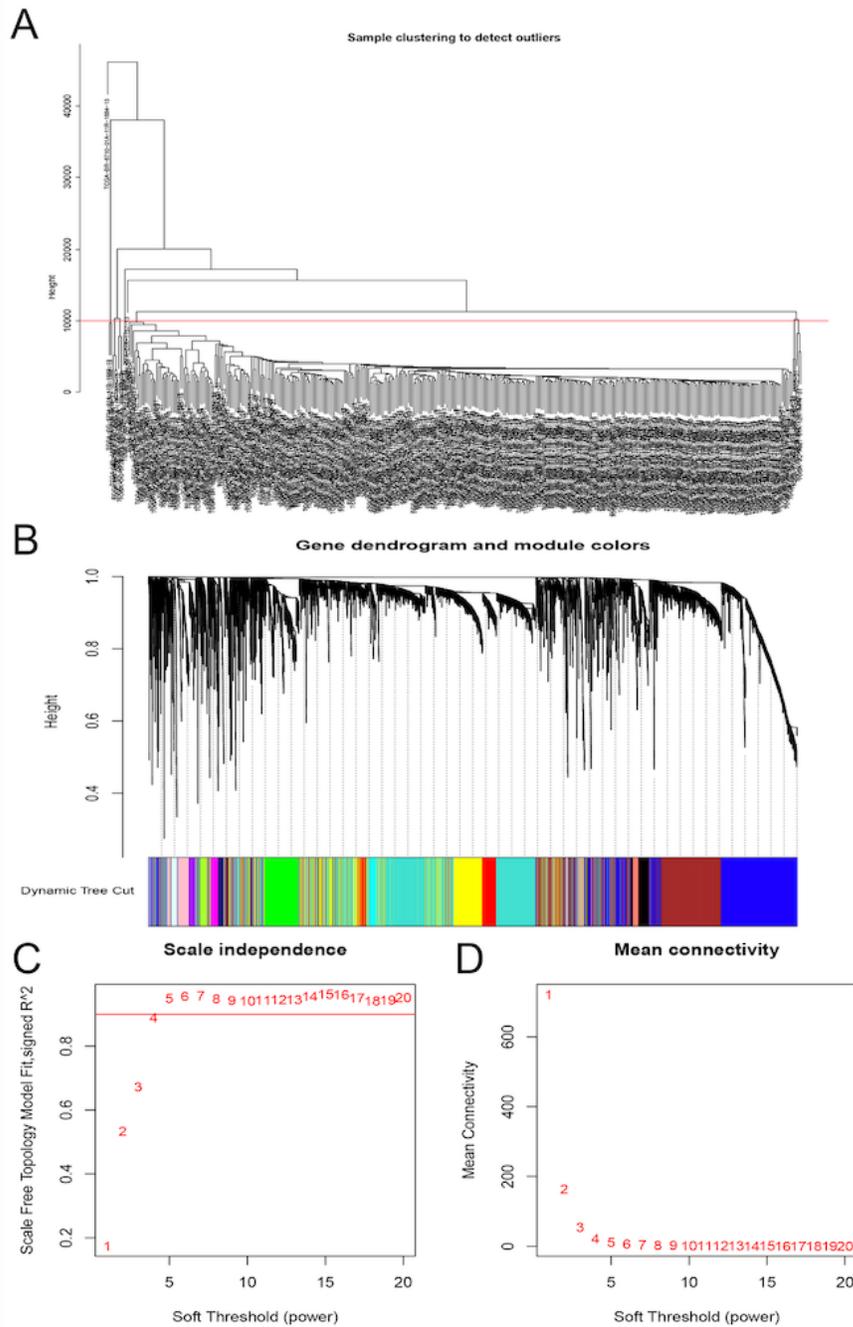


Figure 4

The determination of soft threshold power. (A) Screening out the samples with outlier expression. (B) By hierarchical clustering, genes were divided into different modules. Different colors equal to different modules. (C) Scale-free fit index analysis of 1-20 soft threshold power (β). The horizontal axis was the soft threshold (power), and the vertical axis was the evaluation parameter of the scale-free network. The

higher the value, the more the network conforms to the non-scale feature (non-scale). (D) The average connectivity of soft threshold power was analyzed.

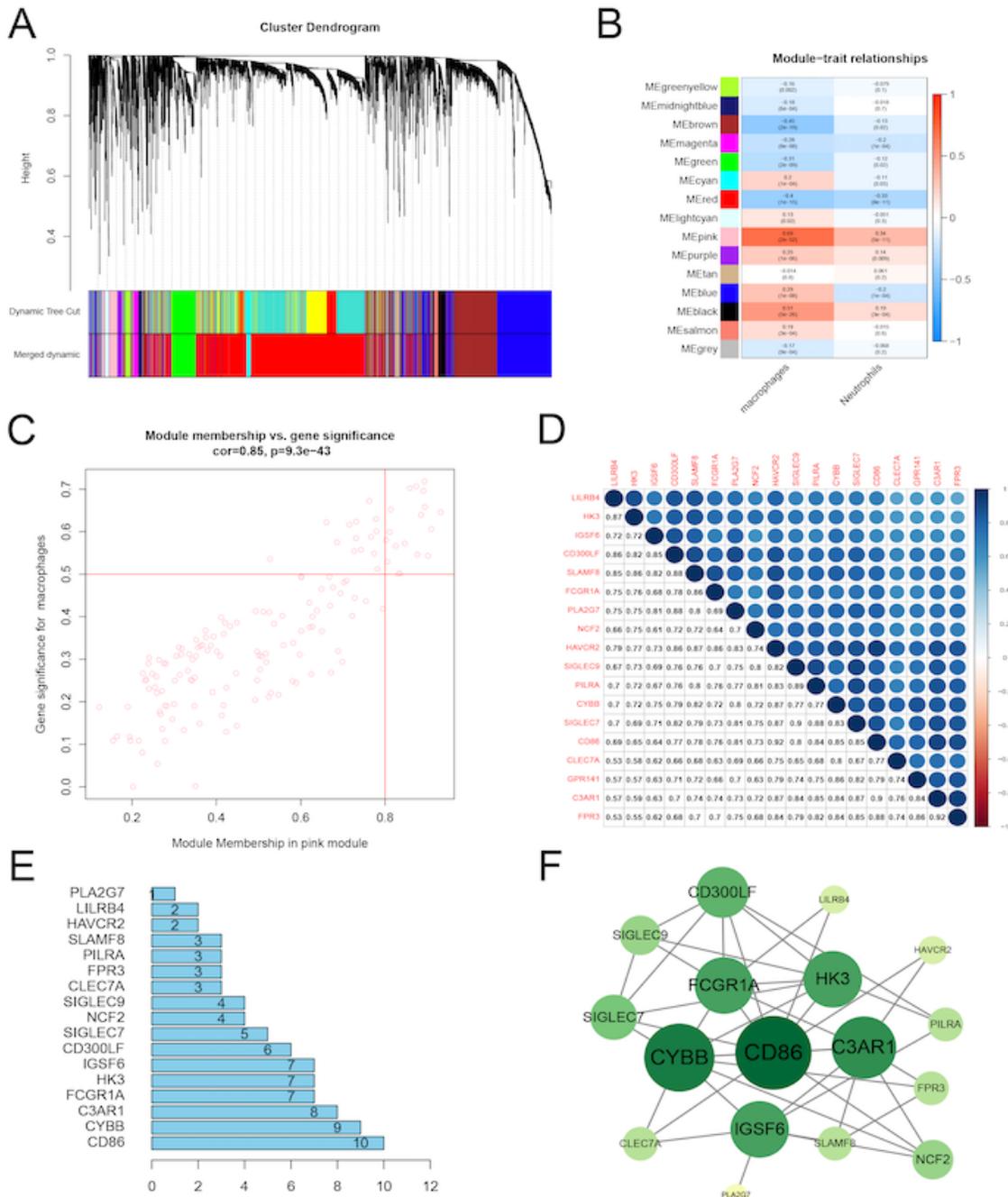


Figure 5

The identification of key modules and genes. (A) Co-expression gene modules were acquired from a hierarchical cluster dendrogram. (B) Heatmap displays correlations of module eigengenes with macrophages. (C) In the MEpink module, each dot represented a gene. The upper right corner of the

scatter plot showed genes under the condition of $MM > 0.8$ and $GS > 0.5$. (D) the correlation heat map of 18 genes. (E) Core genes in PPI. (F) Co-expression of key genes and upstream genes.

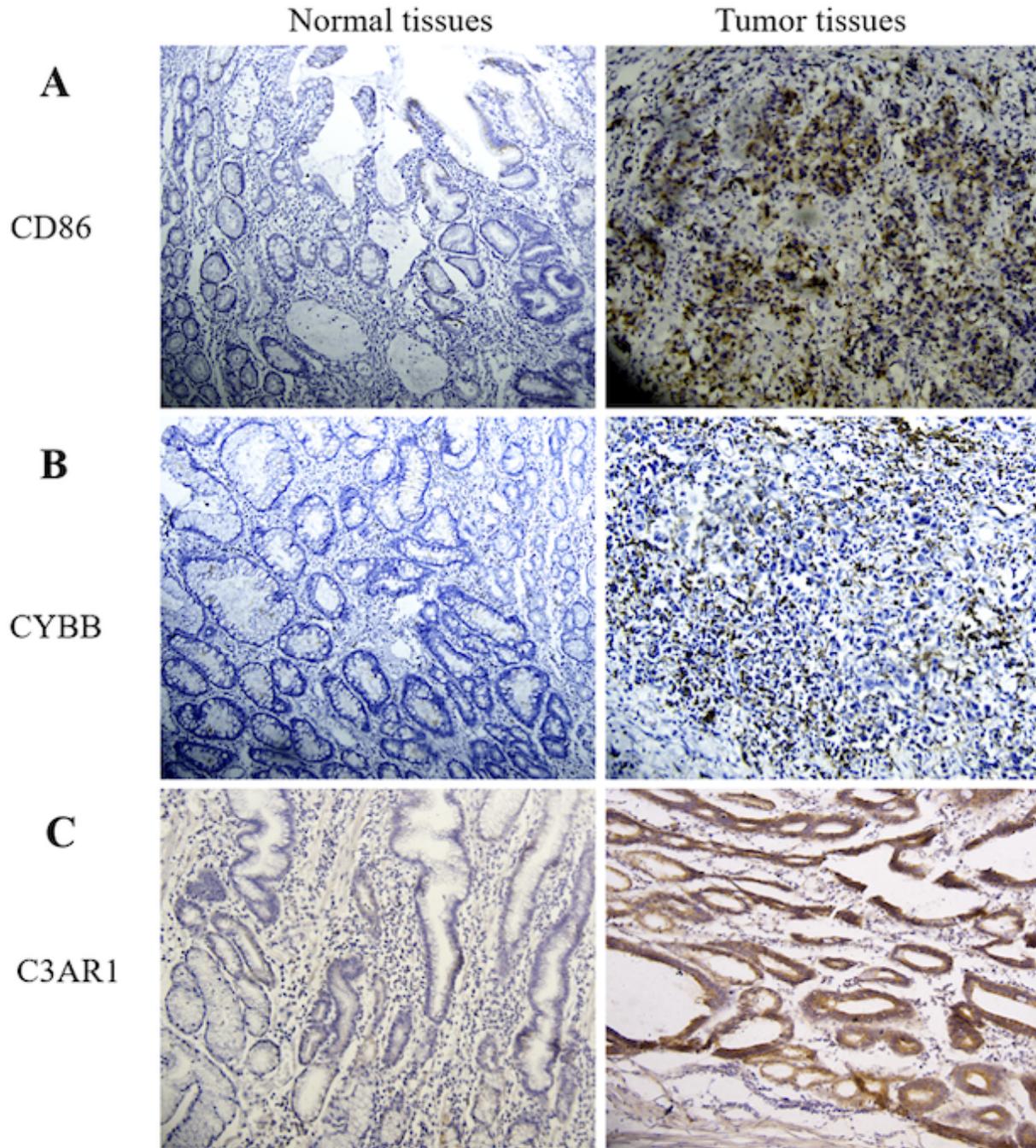


Figure 6

IHC of CD86, CYBB and C3AR1 in GC and normal tissues. (A) The expression level of CD86 in normal and GC tissues. (B) Images of CYBB IHC (Original magnification, $\times 20$). (C) IHC for C3AR1.

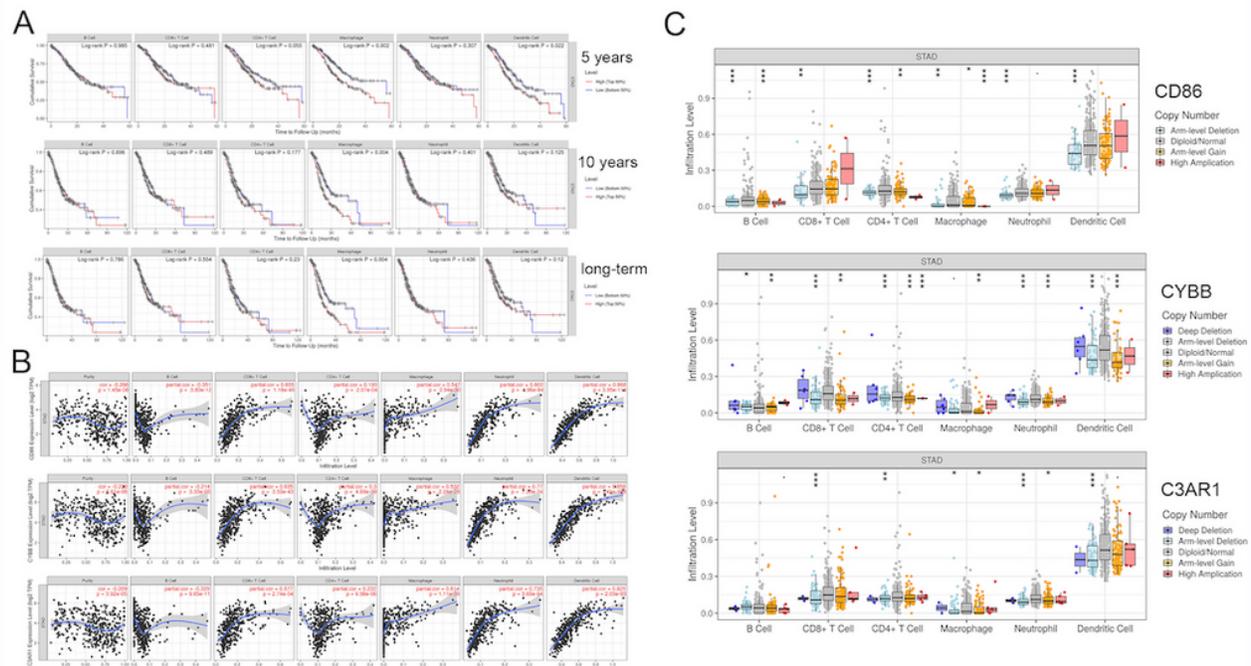


Figure 7

6 kinds of immune cells infiltration situation in TIMER. (A) The effect of CD86, CYBB, C3AR1 expression on six kinds of immune cells. (B) The survival of immune cells. (C) Mutations influence the immune infiltration(* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

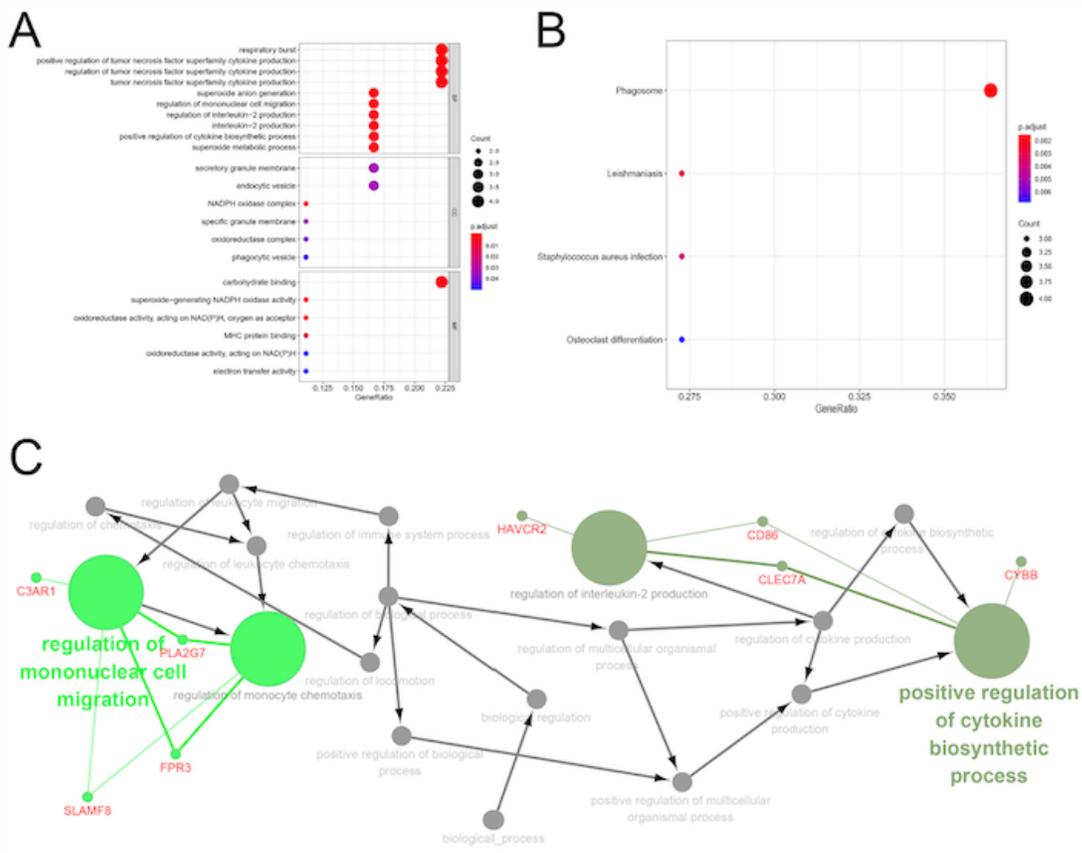


Figure 8

Functional enrichment of key genes. (A) The blue module was significantly enriched in GO annotations. (B) The blue module was significantly enriched in KEGG pathways. (C) Filtering the path with ClueGO.

Supplementary Files

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

- [SupplementaryMaterial.docx](#)