Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comprehensive Analysis of the Immune Effect of TGFβ1 in Colorectal Adenocarcinoma:A TGFβ1-related Prognostic Model of Tumor Microenvironment

Yucheng Qian
Zhejiang University School of Medicine Second Affiliated Hospital
Lina Zhang
Zhejiang University School of Medicine Second Affiliated Hospital
Jihang Wen
Zhejiang University School of Medicine Second Affiliated Hospital
Yanxia Mei
Zhejiang University School of Medicine Second Affiliated Hospital
Jingsun Wei
Zhejiang University School of Medicine Second Affiliated Hospital
Wei Lu
Zhejiang University School of Medicine Second Affiliated Hospital
Kai Jiang
Zhejiang University School of Medicine Second Affiliated Hospital
Maxwell Hwang
Zhejiang University School of Medicine Second Affiliated Hospital
Tian Jin
Zhejiang University School of Medicine Second Affiliated Hospital
Dongliang Fu
Zhejiang University School of Medicine Second Affiliated Hospital
Yeting Hu
Zhejiang University School of Medicine Second Affiliated Hospital
Jun Li
Zhejiang University School of Medicine Second Affiliated Hospital
Kefeng Ding (dingkefeng@zju.edu.cn)
Zhejiang University School of Medicine Second Affiliated Hospital

---

Research Article

Keywords: colorectal cancer, TGFβ1, tumor microenvironment, prognostic model, TCGA

Posted Date: April 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-477295/v1

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

Comprehensive analysis of the immune effect of TGFβ1 in colorectal adenocarcinoma: A TGFβ1-related prognostic model of tumor microenvironment

Yucheng Qian[12], Lina zhang[12], Jihang Wen[12], Yanxia Mei[2], Jingsun Wei[12], Wei Lu[1], Maxwell Hwang[12], Kai Jiang[12], Tian Jin[12], Dongliang Fu[12] , Yeting Hu[12], Jun Li [12], Kefeng Ding[12]

1. Department of Colorectal Surgery and Oncology, Key Laboratory of Cancer Prevention and Intervention, Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
2. Zhejiang University Cancer Center

Correspondence: Jun Li email:2307016@zju.edu.cn, Kefeng Ding email: dingkefeng@zju.edu.cn

*Contributed equally
Yucheng Qian and Lina Zhang should be considered joint first author

Funding: the National Natural Science Foundation of China（81672916, 81802750, 81772545, 11932017）; National Key Research and Development Program of China (2017YFC0908200); Sharing the Sun-Clinical Research Project for Major Diseases of Bethune Charitable Foundation

Abstract:

Colorectal cancer is one of the most common cancer worldwide. Recently, tumor microenvironment (TME), especially its remoulding, is thought to control the colorectal cancer genesis and progression. In this study, we use ESTIMATEscore to make out the proportion of immune and stromal components in colorectal adenocarcinoma (CRA) samples from The Cancer Genome Atlas (TCGA) database. The differentially expressed genes (DEGs) were found by COX regression analysis and protein-protein interaction (PPI) network, among which TGFβ1 was supposed to be a prognosis factor and tumor environment indicator. Continuous analysis showed that TGFβ1 expression is positively correlated with lymph node metastasis (N stage) but negatively correlated with survival. Gene Set Enrichment Analysis (GSEA) revealed that the genes of the high-expression TGFβ1 group were mainly enriched in immune-related activities. Cluster analysis divided the samples into 2 subgroups. 24 HLA-related genes and 8 immune checkpoint genes were found upregulated in the high immunity group as well as TGFβ1, which suggests the possibility of novel therapies targeting immune checkpoints combined with TGFβ1. Tumor-infiltrating immune cell (TIC) profile of CRA patients was described by CIBERSORT analysis. Further analysis showed that the infiltration of Tregs and Neutrophils were positively correlated with TGFβ1 high expression. Then 3 TGFβ1-related genes were picked out to construct a prognostic model, which matches the survival data well.

Keywords: colorectal cancer; TGFβ1; tumor microenvironment; prognostic model; TCGA

Introduction:

Colorectal cancer (CRC) is the third common malignant tumor and the second leading cause of cancer death worldwide[1, 2]. Colorectal adenocarcinoma (CRA) is the major pathohistological type of CRC[3], accounting for about 90% of the number[4]. Recently, a tremendous advance has been made in CRC prevention and treatment including

surgery, chemotherapy, radiotherapy, and immunotherapy. The overall mortality of CRC has declined in developed countries because of early diagnosis and advanced therapies, however, it still causes huge losses in East Asia[5]. Besides, patients with CRC may suffer a poor clinical outcome, mainly due to unfavorable factors such as distant/regional metastasis, chemoresistance, and local recurrence. Thus, it is still important to explore carcinogenesis and new therapeutics in CRC.

Numerous pieces of evidence suggest that tumor microenvironment (TME) plays an important role in tumor biology, including local progression, distant metastasis, and drug resistance[6-8]. The tumor microenvironment consists of tumor cells, infiltrating inflammatory cells, tumor stroma, vessels, and various associated tissue[9]. Collaborative interactions between tumor cells and supporting cells contribute to the malignant behavior of cancer, such as overgrowth, immortalization, dysplasia, and immune privilege. It has been reported that supporting cells facilitate tumor matrix remodeling, cancer cell migration, drug resistance, neoangiogenesis, and evasion of immunosurveillance by producing various chemokines, growth factors, and cytokines [10]. TME exerts a vital effect on tumor progression and its remoulding may offer an unexpected therapeutic opportunity. Meanwhile, some studies persistently trace the impact of immune cells in TME on tumor progression and metastasis, finding that the tumor-infiltrating immune cell (TIC) in TME can be served as a promising indicator for the therapeutic effects[11-13]. Importantly, affluent TIL was now found to be in a positive correlation with improved prognosis and longer overall survival in patients with malignancy[14]. A recent multivariate analysis of cellular interaction in the tumor microenvironment based on the density, function, and localization of immune cells of colorectal cancer verified that immune response within the tumor influences clinical outcomes [15-17]. There was evidence that increasing the amount of TIL gradually cut down the risk of metastasis in colorectal cancer[18]. Tumor-induced cytotoxic activities of CD8+ TIL and migration of CD4+T helper cells were confirmed in colorectal cancer[19]. Furthermore, that cytotoxic T lymphocytes take effect in need of in-situ-activated T helper cells has been detected in colorectal liver metastases. Therefore, it is necessary to perform precise genetic analysis that could precisely describe the dynamic change of the immune and stromal components in TME[18]. In addition, there is a strong need for novel therapies in the battle against colorectal cancer progression and metastasis[20]. Immunotherapeutic strategies may be cheerful in this context as there is compelling evidence of cancer immunosurveillance[21].

Bioinformatics Analysis of TME-related genes reveal the component of cells in TME. In this article, transcriptome RNA-seq data were downloaded from TCGA database and divided with hierarchical cluster analysis into two subtypes, including immunity high group and immunity low group. Additionally, we match the tumor purity, ESTIMATE Score, Immune Score, and Stromal Score of the sample with its immunity status. 29 immune-related gene sets which represented various immune cell types, functions, and pathways was analyzed to describe the tumor immune environment of the samples. And we used the ssGSEA to quantify the activity or enrichment levels of immune cells, functions, or pathways in the cancer samples. Transcriptome RNA-seq data of CRA samples were calculated with ESTIMATE algorithms for the percentage of TICs and the proportion between the immune and stromal elements in CRA samples. DEGs found both in ImmuneScore and StromalScore were analyzed by protein-protein interaction (PPI) network and univariate COX regression analysis, and then intersection analysis was performed using the core nodes in the PPI network and the top significant factors in univariate COX regression. Only 1 gene TGFβ1 was found, and then we explore this gene by correlation analysis between survival and clinicopathological characteristics, expression analysis between tumor and normal tissue, univariate COX regression, Gene Ontology (GO) enrichment analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis, Gene Set Enrichment Analysis (GSEA), and connection analysis with TICs.

The transforming growth factor β (TGFβ) is a family of proteins including TGFβ, bone morphogenic proteins (BMPs), activins and inhibins, etc. The TGFβ family regulates many cellular functions, such as proliferation, apoptosis, differentiation, epithelial-mesenchymal transition (EMT), and migration. TGFβ suppresses tumor advance by inhibiting

cell cycle and promoting apoptosis in the early cancer stage. However, TGFβ exerts stimulative effects in tumor by prompting tumor invasiveness and metastasis in the late stage. TGFβ1 is a critical immunoregulatory cytokine that functions in the process of immune self-tolerance and T-cell homeostasis[22], which is secreted by various cells [23, 24]. Treg cells are the primary source and target of TGFβ1 in some autoimmune diseases [25]. Here we pick up differentially expressed genes (DEGs) between the immune element and stromal element of CRA samples and find that TGFβ1 might be a potential marker for the alteration of TME status in CRA.

Figure1|flow chart of this study.

RESULTS

ImmuneScore is related With the Clinic–Pathological Staging of CRA Patients

We matched the clinical information of CRA cases with ESTIMATEScore、StromalScore and ImmuneScore. The results are shown in Figure2, only ImmuneScore showed negative correlation with T classification (Figure 3A, p = 0.017) and M classification (Figure 3D, p=0.0035) of TMN stage. It suggests that the ratio of immune components was mainly associated with the progress of CRA, such as invasion and metastasis. Residual analysis in this part gain no significant results.

Figure 2|Relativity of ImmuneScore and StromalScore with clinicopathological characteristics. (A–C) correlation of ImmuneScore, StromalScore, and ESTIMATEScore with stage. (The p = 0.017, 0.8, and 0.31, separately, by Kruskal–Wallis rank sum test). (D–F) correlation of above three kinds of scores with T classification (p = 0.7, 0.36, 0.68 for ImmuneScore, StromalScore, and ESTIMATEScore, separately, by Kruskal–Wallis rank sum test). (G–I) correlation of scores with N classification. Similar to the above, p = 0.12, 0.14, 0.95, separately, with Kruskal–Wallis rank sum test. (J–L) correlation of scores with M classification (p = 0.0035, 0.56, 0.082 for ImmuneScore, StromalScore, and ESTIMATEScore separately by Wilcoxon rank sum test).

Common DEGs between ImmuneScore and StromalScore were mainly enriched to Immune-Related Genes

To ascertain the exact alterations of gene profile in TME both in immune and stromal components, samples were divided into high-score group and low-score group by comparing to the median. 1603 DEGs were found by ImmuneScore (samples with high score vs. low score). Among them, 1553 genes were up-regulated, 50 genes were down-regulated (Figure3C,D). Similarly, 1901 DEGs were found by StromalScore, consisting of 1889 up-regulated genes and 14 down-regulated genes (Figure3C,D). The Venn plot showed 1103 up-regulated genes shared by high ImmuneScore group and high StromalScore group. 5 down-regulated genes were shared by low score group both in ImmuneScore and StromalScore. These DEGs (total 1108 genes) were potential determinate factors of TME. Results of gene ontology (GO) enrichment analysis suggest that the DEGs mostly map to the immune-associated GO terms, such as immune response−regulating cell surface receptor signaling pathway and lymphocyte mediated immunity (Figure 3E,supplement1). The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis also showed the enrichment of immune response−regulating cell surface receptor signaling pathway, cytokine−cytokine receptor interaction and phagosome (Figure 3F,supplement2). Thus, the overall functions of DEGs seemed to mostly fall in immune-related activities, which indicate that the participance of immune factors was a critical feature of TME in CRA.

Figure 3| Heatmaps, Venn plots and enrichment analysis of GO and KEGG for DEGs. (A) Heatmap for DEGs from comparison between the high score group vs.the low score group in ImmuneScore. Row name of heatmap represent the specific gene, and column name means the ID of samples. DEGs were abtained using Wilcoxon rank sum test with q = 0.05 and fold-change >1 after log2 transformation as the significance threshold. (B) Heatmap for DEGs in StromalScore, similar with (A). (C,D) Venn plots show corporate up-regulated and down-regulated DEGs both in ImmuneScore and StromalScore, and q < 0.05 and fold-change >1 after log2 transformation as the DEGs significance filtering threshold. (E,F) GO and KEGG enrichment analysis for 1108 DEGs, with p and q < 0.05.

Corelation Analysis by PPI Network and Univariate COX Regression

PPI network was carried out based on the STRING database using Cytoscape software. The correlation among 1108 genes are presented in Figure 4A, and the bar graph showed the top 30 genes ranked by the quantity of nodes (Figure 4B). Univariate COX regression analysis was performed to pick up the significant survival factors among 1108 DEGs (Figure 4C). Next, the intersection analysis between the top genes in PPI network and the 21 significant factors of univariate COX regression was performed, and only 1 factors, TGFβ1, were selected from all above analyses (Figure 4D).

Figure 4| Protein–protein interaction network and univariate COX analysis. (A) Interaction network of nodes with interaction confidence value >0.95. (B)The top 30 genes ranked by the number of nodes. (C) Univariate COX regression analysis using 1108 DEGs, showing the significant factors with $p < 0.005$. (D) Venn plot circling the common factors shared by top 100 nodes in PPI and 21 significant factors in univariate COX.

The Expression Of TGFβ1 Influence the Survival and Clinical Characteristics in CRA Patients

TGFβ1 exert vital influence on the intracellular signaling of T lymphocytes. In this study, all CRA samples were distributed into TGFβ1 high-expression group or TGFβ1 low-expression group compared with the median. The survival curve indicated that CRA patients in low-expression group had better survival than that in high-expression group(Figure 5A).And then , the correlation of TGFβ1 expression with clinical characteristics was described with Wilcoxon rank sum test, showing that the expression of TGFβ1 in the tumor tissue was significantly higher than that in the normal tissue(Figure 5B) in pairing sample gained from the same patient (Figure 5B). All above results explicitly suggest that the expression of TGFβ1 is higher in tumor tissue than in normal tissue, and that this higher expression in TME bring worse outcomes. Besides, the expression of TGFβ1 increased along with the advance of N stage in TNM stages (Figures 5C–F), which probably indicates that the expression of TGFβ1 prompt metastasis in lymph node.

Figure 5 | Different expression of TGFβ1 in spicemen and its influence in survival and clinicopathological characteristics of CRA patients. (B) Different expression of TGF β 1 between pairing normal and tumor spicemen. Analyses were conducted with $p<0.05$ by Wilcoxon rank sum test. (A) Survival curve for CRA patients with various TGF β 1 expression. Patients were matched with high expression or low expression compared with the median. $p = 0.048$ by log-rank test. (C-F) The correlation of TGF β 1 expression with clinicopathological characteristics was revealed by Wilcoxon rank sum or Kruskal–Wallis rank sum test.

New classification divide the samples into 2 subgroups

Transcriptome data of colon cancer patients was downloaded from TCGA database and calculated with ssGSEA method. These samples were divided into 2 subgroups (low immunity: 186 samples, high immunity: 212 samples) (Figure 6A) by consensus cluster analysis, and ESTIMATEscore, Immunescore and Stromalscore were also involved in the heat map. Tumor purity describes the cancer cell proportion in the tumour, which is significantly higher in the immunity low group (Figure 6C). Meanwhile, 29 immune-related pathways and infiltrating immune cells were integrated to evaluate immunity level of TME (Figure 6B), which seems all gathered in the immunity high group. The RNA level comparison of HLA genes between the two immune subgroups was conducted (Figure 6D), showing notable preponderance in immunity high group. All above shows that the new classification divided the samples into two groups according to its immunity capability. 8 immune checkpoint genes (TIGIT, Tim-3, VISTA, BTLA, LAG-3, IDO1, PD-L1, PD-1) were found upregulated in the high immunity group ($P<0.001$) by univariate analysis (Figure 6A-L), which mostly inhibit the immune response of TIL. The strongly responsive microenvironment might improve the efficiency of immune checkpoint inhibitor in turn. Moreover, TGFβ1 is similarly up-regulated in the same group(Figure 6M), which suggests the possibility of novel therapies targeting immune checkpoint combined with TGFβ1.

Figure 6| Identification and validation of immune-related subgroups of colon cancer. A. cluster analysis divided samples into 2 subgroups, that are immunity low group and immunity high group. B.29 immune-related gene sets are enriched in ssGSEA with colon cancer,. These gene sets are composed of immune cells and immune processes. The heat map also shows tumor purity, ESTIMATE score, immune score and matrix score. Heatmap of sample clustering at consensus k = 2.C. Tumor purity of samples from the two immune subgroups (*** $P < 0.001$). D. RNA expression levels of HLA genes in samples from the two immune subgroups. E-L. 8 immune

checkpoint genes are significantly up-regulated in immunity high group (***P < 0.001). M. TGFβ1 is similarly upregulated in the immunity high group.

TGF β 1 Can Be an Prospective indictor For TME Modulation

GSEA was conducted in the TGFβ1 high-expression s divided by the median. As shown in Figure 7, the genes in high-expression group were mainly enriched in immune-related activities, such as allograft rejection, complement, adhesion molecules cams, T cell activation pathway and typical tumor pathways, which suggests that TGFβ1 might be a prospective indicator for the status of TME.

Figure 7 | GSEA for TGFβ1 high-expression group. (A) The enriched gene sets in HALLMARK collection of high TGFβ1 expression. Each line shows one particular gene set. Gene sets with NOM p < 0.05 and FDR q < 0.06 were regarded significant. Only some leading gene sets were shown in the plot. (B) The enriched gene sets in GSE by samples of TGFβ1 high expression. (C) Enriched gene sets in KEGG collection, the immunologic gene sets, by samples of high TGFβ1 expression. Only several leading gene sets are shown in plot. (D) Enriched gene sets in GO collection by the high BTK expression.

Correlation of TGFβ1 With the Proportion of TICs

To further demonstrate the correlation of TGFβ1 expression with the immune microenvironment, the ratio of tumor-infiltrating immune subsets was calculated by CIBERSORT algorithm. 21 types of immune cell profiles in CRA samples were constructed (Figure 8A-B). Then, correlation analysis showed that Neutrophils were significantly correlated with the expression of TGF1 (Figure 10C). Further analysis showed that Tregs and Neutrophils were positively correlated with TGFβ1 expression out of 21 types of immune cell (Figure10D-E). These results strongly suggest that the expression of TGFβ1 definitely influence the immune activity of TME.

Figure 8 | Connection between TGFβ1 expression and TIC profile. (A) Bar plot showing the ratio of 21 kinds of TICs in CRA samples. Column names were sample ID. (B) Heatmap showed the connection between 21 kinds of TICs and numbers in the tiny boxes represented the p value of correlation between two kinds of cells. The color in every tiny box showed correlation value between the two cells. And Pearson correlation coefficient was used for significant test.C. Violin graph shows the different proportion of 21 kinds of TIC between high TGFβ1 expression group and low TGFβ1 expression group compared to the median. D. The correlation between TGFβ1 expression and Tregs abundance was shown in the scatter plot, and P<0.05 was regarded significant. E. The correlation between TGFβ1 expression and Neutrophils abundance was shown in the scatter plot, and P<0.05 was regarded significant.

Construct of TGFβ1-associated immune prognostic model

TGFβ1-associated immune genes were retrieved from the online database TISIDB, including 17 immunoinhibitors (ADORA2A,CSF1R,CTLA4,HAVCR2,IL10,LAG3,PDCD1,PDCD1LG2,TIGIT,BTLA,CD244,CD274,CD96,IDO1,KDR,PVRL2,TGFBR1) and 34 Immunostimulators (C10of54,CD27,CD48,CD70,CD80,CD86,CXCR4,IL2RA,LTA,TNFRSF18,TNFRSF4,TNFRSF8, TNFRSF9,TNFSF13B,TNFSF14,TNFSF4,CD276,CD28,CD40,CD40LG,CXCL12,ENTPD1,ICOS,IL6,IL6R,KLRC1,KLRK1,MICB,PVR,TEME173,TNFRSF13B,TNFRSF13C,TNFRSF17,TNFSF9). Among these genes, 18 significant differential expression genes between tumor samples and normal counterpart were picked out by limma R package, shown in figure 9A-B. TGFβ1-related prognosis signature was built with risk scores based on the expression of genes multiplied by regression coefficients in the multivariate Cox regression analysis(P < 0.01 is significant). Samples were distributed into high risk group and low risk group according to the risk score , and the cut-off value derive from the "survminer" R package. This prognosis risk score model match with the survival data well (Figure 9C-E). The risk score was significantly connect with survival in CRC in the stepwise multivariate COX regression analysis (P < 0.001). And the area under the curve (AUC) values of the risk score were 0.622, 0.613 and 0.643 respectively of 1 year, 3years and 5 years(Figure 9D).

Table1|3 genes that constitute the TGFβ1-associated immune prognostic model and the corresponding risk factors

Figure 9| The prognostic value of TGFβ1-associated risk score. A. The heatmap of 18 significant differential expression TGFβ1- associated genes between tumor and normal tissue.(CSF1R,CTLA4,IL10,CD27,CD48,CD70,TNFSF4,BTLA,CD96,CD276,CD40LG,IL6R,PVR,TNFRSF13B,TNFRSF13C,TNFRSF17,TNFSF9). B.The volcano plot shows the distribution of 18 differential expression TGFβ1-associated genes mentioned above. C. The survival curve shows significantly diverse between high risk group and low risk group based on Kaplan–Meier curves.(P<0.001) D .Time-dependent receiver operating characteristic curves at 1 year,3 years and 5 years. E. The distribution and survival status of high risk group and low risk group and the expression of 3 predictive genes between the 2 groups. F. Stepwise multivariate COX regression analysis show risk score is significantly associated with survival.(P<0.001)

Discussion:

In this study, we try to find TME relevant gene that influence the survival and immunity therapy with data from TCGA database. TGFβ1 was identified to be involved in immune reaction, and continuous bioinformatics analysis suggest that TGFβ1 could be a signal for the status of TME in CRA patients. The tumor microenvironment is mainly consist of tumor stroma, infiltrating inflammatory cells and various associated supporting cells[9],and is fabricated and modified all the time by tumor itself. It has been fully demonstrated that the tumor microenvironment plays a critical role in tumor progression, which indicates remoulding the tumor microenvironment could be a promising therapic explore in turn. Besides, the condition of the tumor environment, which falls in immunosuppression or active can also suggest the patients' prognosis and therapy response.

Our results from the transcriptome analysis with CRA data from TCGA database implied that the immune components in TME is significantly correlated with the progression of CRA, such as metastasis. This highlighted the significance of exploring the association between tumor cells and immune cells, which provided a novel sight for new effective treatment. Recently, a great advance has been made in immunotherapy, as pleasant successes has been seen in immune checkpoint inhibitors. While numerous cancers have seen preliminary evidence of efficacy from immune checkpoint inhibitors, colorectal cancers remain the firm exception[26].It is previously acknowledged that colorectal cancer benefited rarely from immunotherapeutics[27].However, increasing data demonstrates that subsets of colorectal cancer patients, those with hypermutated tumors, which is called microsatellite instability (msi) tumor, may benefit from immune checkpoint inhibitors.

Transforming growth factor β 1 (TGFβ1) is a member of the large family of a series of structurally and functionally relevant proteins, which contains TGFβ1, TGFβ2 and TGFβ3 and also activins, nodal, inhibins, Mullerian-inhibiting substance (MIS), growth and differentiation factors (GDF) and bone morphogenetic proteins (BMPs)[28]. The TGFβ1 gene codes a raw peptide of 390 amino acids, which is subsequently disposed by furin proteases into the mature bioactive one[29]. TGFβ1 plays a suppressive role in the pre-malignant phase and accelerate tumor progress in the late stage by TGF-β/SMAD and Non-SMAD Signaling[25].In the progressive stage, TGFβ1 may affect T-cell activation, regulatory T (Treg) cell and effector-cell function and tumorigenesis to boost the malignant behavior[30-32].

Epithelial-mesenchymal transition（EMT）is stimulated by various cytokines, leading to decreased adhesion of cells, compelling motility, susceptibility to invasion and metastasis and resistance to chemotherapy[33] [34], among them TGF-β1 being a very powerful factor, which influences the expression and liveness of many transcription factors called EMT-TFs (Snail1/Snail, Snail2/Slug, Twist1/Twist etc.) [35]. Evaluation of gene expression profiles disclosed that TGF-β1 signaling is one of the most important gene pathway in liver metastases of colorectal cancer[36]. And EMT-TFs are the initiation of EMT in cell differentiation [37]. In fetal hepatocytes, TGF-β1 is associated with EMT and resistance to apoptosis[38, 39] [40].

TGFβ1 also inhibit tumour immunosurveillance to foster the neoplasm, particularly aiming at T cell and natural killer (NK) lymphocytes by block the generation and function of it[41-43] [44-46]. Moreover, TGF-β1 drive the TME against the cancer by induicing lymphocytes into suppressive subtypes, such as CD4+ regulatory T cells (Treg cell)[47]. Tregs participate in self-tolerance and immune suppression, discouraging the proliferation of CD4+ and CD8+ T cells as well as the production of IFN-γ[48, 49]. TGFβ1 in the tumor microenvironment turned pretumor (M2) type of tumor-associated macrophages (TAMs) into the antitumor (M1) phenotype, and similarly drove the tumor-associated neutrophil (TAN) phenotype from N1 to N2, increasing the propotion of protumor TANs [50]. Some TGFβ signaling inhibitors have been discovered to stop the aberrant behavior of TGF-β signaling in tumors, such as ligand traps, antisense oligonucleotides (AONs), neutralizing antibodies, receptor domain-immunoglobulin fusions and receptor kinase inhibitors, which are in pre-clinical development.

TGFβ1 has increasingly been arranged as an novel druggable target in several preclinical experiment [51]. As TGFβ1 signaling pathway is fundamental to the formation of tumor microenvironment and exert crucial influence on the immune inflammatory responses such as the Treg cells , combining modern immunotherapy trials with TGFβ1 signalling-related biomarkers is reasonable and necessary. The role PD-1/PD-L1 paly in immune activities is not fully understood, and TGFβ1 signaling pathway might enhance this process, which may a novel strategy deserving further investigation.

Materials and Methods

Raw Data

Transcriptome RNA-seq data of 437 CRA cases (39 normal samples; 398 tumor samples) and the corresponding clinical data were downloaded from TCGA database.

Generation of ImmuneScore, StromalScore, and ESTIMATEScore

ESTIMATE algorithm of R language version 3.5.1 loaded with estimate package was used to evaluate the proportion of immune and stromal component in TME for each sample, presenting three kinds of scores: ImmuneScore, StromalScore, and ESTIMATEScore, which respectively positive correlated with the ratio of immune, stromal, and the sum of both. The higher the respective score is, the larger the ratio of the corresponding component is in TME.

Survival Analysis

R package survival and survminer was used for the survival analysis. 398 tumor samples had a specific survival time record, were used for survival analysis. Kaplan–Meier method was used in the survival curve, with log rank as the statistical significance test; $p < 0.05$ was considered significant.

Generation of DEGs Between High-Score and Low-Score Groups According to ImmuneScore and StromalScore

tumor samples were divided into high-score group and low-score group comparison to the median score in according to ImmuneScore and StromalScore, respectively. Package limma was used to identify the different expression of the gene, and DEGs were found by the comparison between the high-score group vs. the low-score group. DEGs with fold change larger than 1 after transformation of log2 (high-score group/low-score group) and false discovery rate (FDR) <0.05 were considered significant.

GO and KEGG Enrichment Analysis

GO and KEGG enrichment analysis of 1108 DEGs were performed with clusterProfiler, enrichplot, and ggplot2 packages of R language. Only terms with both p- and q-value of <0.05 were considered significantly enriched.

Differential gene expression analysis

Heatmaps were produced by pheatmap R package. volcano plots were displayed by “ggplot2” R package, with an adjusted false discovery rate (FDR) < 0.05 and |log2(fold change)| > 1 as the thresholds. Volcano plots were displayed by “ggplot2” R package, with an adjusted false discovery rate (FDR) < 0.05 and |log2(fold change)| > 1 as the thresholds.

Difference Analysis of Scores With Clinical Stages

The clinicopathological data of the CRA samples were downloaded from TCGA database. The analysis was accomplished by R language, with Wilcoxon rank sum or Kruskal–Wallis rank sum test as the significance test.

PPI Network Construction

PPI network was produced by STRING database, and reconstruction with Cytoscape of version 3.6.1. Nodes with confidence of interactive relationship over 0.95 were used for network.

Univariate COX Regression Analysis

Survival package of R language was used for univariate COX regression. The top 21 genes ordered by p value from small to large in univariate COX were shown in the plot.

Multivariate COX Regression Analysis

multivariate cox regression analysis were conducted to estimate the prognostic value of the risk score, clinicopathological features including age, clinical stage, grade, and TNM stage with R software (version 3.5.1).

Gene Set Enrichment Analysis

Hallmark, KEGG ,GSE and GO gene sets collections were downloaded from Molecular Signatures Database as the target sets with which GSEA performed using the software gsea-3.0 downloaded from Broad Institute. The whole transcriptome of all tumor samples was used for GSEA, and only gene sets with NOM $p < 0.05$ and FDR $q < 0.06$ were considered as significant.

TICs Profile

CIBERSORT package was used to evaluate the TIC profile in all tumor samples, followed by quality check that 154 tumor samples with $p < 0.05$ were chosen for the following analysis.

ssGESA

A set of gene signatures of immune cell populations from previously reported articles were applied to individual cancer samples. The calculation method used in this study includes immune cell types and immune pathways related to innate immunity and adaptive immunity, a total of 29 immune gene sets, including immune cell types and functions, regulatory T (Treg) cells, immune checkpoint, cytokine and cytokine receptor (CCR), human leukocyte antigen (HLA) , proinflammatory, para-inflammation (PI) , tumor-infiltrating lymphocytes (TILs). R package “gsva” is used to quantified the infiltration levels of immune cell types.

hierarchical agglomerative clustering

In order to study the correlation between the immunity of colon cancer and the clinical phenotype, we divided the samples from TCGA into 2 different groups (high and low immunity) based on "TCS genetic clustering" (50 iterations, 80% resampling rate) with ssGSEA scores. R package "estimate" is included to calculate the Immunescore, Stromalscore and ESTIMATEscore of each sample. Tumor purity is sescribed by Mann-Whitney U test.

ROC curve

The time-dependent receiver operating characteristic (ROC) curve was produced to evaluate sensitivity and specificity of the prognostic model using the "timeROC" R software package. The area under the curve (AUC) was calculated.

Acknowledgments

The authors would like to thank the TCGA and TISIDB databases for the availability of the data.

Author Contributions

YQ conceived the research, and YQ and LZ designed the research process and conducted bioinformatics analysis. YQ and LZ downloaded and collated the data in the article and wrote the first draft of the article. YQ, LZ and JW performed statistical analysis on the data. JW revised the article strictly to obtain the necessary knowledge and administrative support. YM, JW, KJ, DF, MH, WL, TJ, QX, YH, JL and KD reviewed and edited the manuscript. All authors read and approved the final manuscript.

Competing of Interest

The author claims that this study was conducted without any commercial or financial relationship, that is, without potential competing interests.

Ethics approval and consent to participate

This article does not contain any studies with human participants or animals performed by any of the authors.

Funding:

This study was funded by grants from the National Natural Science Foundation of China（81672916, 81802750, 81772545, 11932017）; National Key Research and Development Program of China (2017YFC0908200) and Sharing the Sun-Clinical Research Project for Major Diseases of Bethune Charitable Foundation.

Availability of data and material

All data used in this study are publicly available. TCGA data are available through the Genome Data Commons (https://gdc.cancer.gov/).

.

1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2020.* CA Cancer J Clin, 2020. **70**(1): p. 7-30.
2. *Erratum: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer J Clin, 2020. **70**(4): p. 313.
3. Engstrand, J., et al., *Colorectal cancer liver metastases - a population-based study on incidence, management and survival.* BMC Cancer, 2018. **18**(1): p. 78.
4. Evdokimova, A., S. Laurent, and L. Dorthu, *Recurrent Acute Pancreatitis Caused by Mucin-Producing Liver Metastases.* J Belg Soc Radiol, 2018. **102**(1): p. 19.
5. Center, M.M., et al., *Worldwide variations in colorectal cancer.* CA Cancer J Clin, 2009. **59**(6): p. 366-78.
6. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
7. Bissell, M.J. and W.C. Hines, *Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression.* Nat Med, 2011. **17**(3): p. 320-9.
8. Quail, D.F. and J.A. Joyce, *Microenvironmental regulation of tumor progression and metastasis.* Nat Med, 2013. **19**(11): p. 1423-37.
9. Whiteside, T.L., *The tumor microenvironment and its role in promoting tumor growth.* Oncogene, 2008. **27**(45): p. 5904-12.
10. Bussard, K.M., et al., *Tumor-associated stromal cells as key contributors to the tumor microenvironment.* Breast Cancer Res, 2016. **18**(1): p. 84.
11. Whiteside, T.L., *Tumor-infiltrating lymphocytes in human malignancies*. 1993: RG Landes Company.
12. Mihm Jr, M.C., C.G. Clemente, and N. Cascinelli, *Tumor infiltrating lymphocytes in lymph node melanoma metastases: a histopathologic prognostic indicator and an expression of local immune response.* Laboratory investigation; a journal of technical methods and pathology, 1996. **74**(1): p. 43-47.
13. Gajewski, T.F., H. Schreiber, and Y.X. Fu, *Innate and adaptive immune cells in the tumor microenvironment.* Nat Immunol, 2013. **14**(10): p. 1014-22.
14. Reichert, T.E., et al., *Absent or low expression of the ζ chain in T cells at the tumor site correlates with poor survival in patients with oral carcinoma.* Cancer research, 1998. **58**(23): p. 5344-5347.
15. Pages, F., et al., *Effector memory T cells, early metastasis, and survival in colorectal cancer.* N Engl J Med, 2005. **353**(25): p. 2654-66.
16. Galon, J., et al., *Type, density, and location of immune cells within human colorectal tumors predict clinical outcome.* Science, 2006. **313**(5795): p. 1960-4.
17. Galon, J., W.H. Fridman, and F. Pages, *The adaptive immunologic microenvironment in colorectal cancer: a novel perspective.* Cancer Res, 2007. **67**(5): p. 1883-6.

18. Laghi, L., et al., *CD3+ cells at the invasive margin of deeply invading (pT3-T4) colorectal cancer and risk of post-surgical metastasis: a longitudinal study.* Lancet Oncol, 2009. **10**(9): p. 877-84.
19. Koch, M., et al., *Tumor infiltrating T lymphocytes in colorectal cancer: Tumor-selective activation and cytotoxic activity in situ.* Ann Surg, 2006. **244**(6): p. 986-92; discussion 992-3.
20. Mocellin, S., et al., *Part I: Vaccines for solid tumours.* Lancet Oncol, 2004. **5**(11): p. 681-9.
21. Yang, L. and D.P. Carbone, *Tumor-host immune interactions and dendritic cell dysfunction.* Adv Cancer Res, 2004. **92**: p. 13-27.
22. Bommireddy, R. and T. Doetschman, *TGF-beta, T-cell tolerance and anti-CD3 therapy.* Trends Mol Med, 2004. **10**(1): p. 3-9.
23. Li, M.O., et al., *Transforming growth factor-beta regulation of immune responses.* Annu Rev Immunol, 2006. **24**: p. 99-146.
24. Schmidt-Weber, C.B. and K. Blaser, *Regulation and role of transforming growth factor-β in immune tolerance induction and inflammation.* Current opinion in immunology, 2004. **16**(6): p. 709-716.
25. Colak, S. and P. Ten Dijke, *Targeting TGF-beta Signaling in Cancer.* Trends Cancer, 2017. **3**(1): p. 56-71.
26. Boland, P.M. and W.W. Ma, *Immunotherapy for Colorectal Cancer.* Cancers (Basel), 2017. **9**(5).
27. Le, D.T., et al., *PD-1 Blockade in Tumors with Mismatch-Repair Deficiency.* N Engl J Med, 2015. **372**(26): p. 2509-20.
28. Derynck, R. and E.H. Budi, *Specificity, versatility, and control of TGF-beta family signaling.* Sci Signal, 2019. **12**(570).
29. Derynck, R., et al., *Human transforming growth factor-beta complementary DNA sequence and expression in normal and transformed cells.* Nature, 1985. **316**(6030): p. 701-5.
30. Roberts, A.B., et al., *Transforming growth factor-beta: possible roles in carcinogenesis.* Br J Cancer, 1988. **57**(6): p. 594-600.
31. Goumans, M.J., et al., *Balancing the activation state of the endothelium via two distinct TGF-beta type I receptors.* EMBO J, 2002. **21**(7): p. 1743-53.
32. Breier, G., et al., *Transforming growth factor-beta and Ras regulate the VEGF/VEGF-receptor system during tumor angiogenesis.* Int J Cancer, 2002. **97**(2): p. 142-8.
33. Tsai, J.H. and J. Yang, *Epithelial-mesenchymal plasticity in carcinoma metastasis.* Genes Dev, 2013. **27**(20): p. 2192-206.
34. van Staalduinen, J., et al., *Epithelial-mesenchymal-transition-inducing transcription factors: new targets for tackling chemoresistance in cancer?* Oncogene, 2018. **37**(48): p. 6195-6211.
35. Peinado, H., M. Quintanilla, and A. Cano, *Transforming growth factor beta-1 induces snail transcription factor in epithelial cell lines: mechanisms for epithelial mesenchymal transitions.* J Biol Chem, 2003. **278**(23): p. 21113-23.
36. Steinbichler, T.B., et al., *The role of exosomes in cancer metastasis.* Semin Cancer Biol, 2017. **44**: p. 170-181.

37. Nieto, M.A., et al., *Emt: 2016.* Cell, 2016. **166**(1): p. 21-45.
38. Del Castillo, G., et al., *Autocrine production of TGF-beta confers resistance to apoptosis after an epithelial-mesenchymal transition process in hepatocytes: Role of EGF receptor ligands.* Exp Cell Res, 2006. **312**(15): p. 2860-71.
39. Caja, L., et al., *The transforming growth factor-beta (TGF-beta) mediates acquisition of a mesenchymal stem cell-like phenotype in human liver cells.* J Cell Physiol, 2011. **226**(5): p. 1214-23.
40. Franco, D.L., et al., *Snail1 suppresses TGF-beta-induced apoptosis and is sufficient to trigger EMT in hepatocytes.* J Cell Sci, 2010. **123**(Pt 20): p. 3467-77.
41. Schmidt, A., N. Oberle, and P.H. Krammer, *Molecular mechanisms of treg-mediated T cell suppression.* Front Immunol, 2012. **3**: p. 51.
42. Thomas, D.A. and J. Massague, *TGF-beta directly targets cytotoxic T cell functions during tumor evasion of immune surveillance.* Cancer Cell, 2005. **8**(5): p. 369-80.
43. Gorelik, L. and R.A. Flavell, *Immune-mediated eradication of tumors through the blockade of transforming growth factor-beta signaling in T cells.* Nat Med, 2001. **7**(10): p. 1118-22.
44. Imai, K., et al., *Inhibition of dendritic cell migration by transforming growth factor-beta1 increases tumor-draining lymph node metastasis.* J Exp Clin Cancer Res, 2012. **31**: p. 3.
45. Ito, M., et al., *Tumor-derived TGFbeta-1 induces dendritic cell apoptosis in the sentinel lymph node.* J Immunol, 2006. **176**(9): p. 5637-43.
46. Kobie, J.J., et al., *Transforming growth factor beta inhibits the antigen-presenting functions and antitumor activity of dendritic cell vaccines.* Cancer Res, 2003. **63**(8): p. 1860-4.
47. Ohkura, N., Y. Kitagawa, and S. Sakaguchi, *Development and maintenance of regulatory T cells.* Immunity, 2013. **38**(3): p. 414-23.
48. Somasundaram, R., et al., *Inhibition of cytolytic T lymphocyte proliferation by autologous CD4+/CD25+ regulatory T cells in a colorectal carcinoma patient is mediated by transforming growth factor-beta.* Cancer Res, 2002. **62**(18): p. 5267-72.
49. Murakami, M., et al., *CD25+CD4+ T cells contribute to the control of memory CD8+ T cells.* Proc Natl Acad Sci U S A, 2002. **99**(13): p. 8832-7.
50. Fridlender, Z.G., et al., *Polarization of tumor-associated neutrophil phenotype by TGF-beta: "N1" versus "N2" TAN.* Cancer Cell, 2009. **16**(3): p. 183-94.
51. Giannelli, G., et al., *Moving towards personalised therapy in patients with hepatocellular carcinoma: the role of the microenvironment.* Gut, 2014. **63**(10): p. 1668-76.

Figures

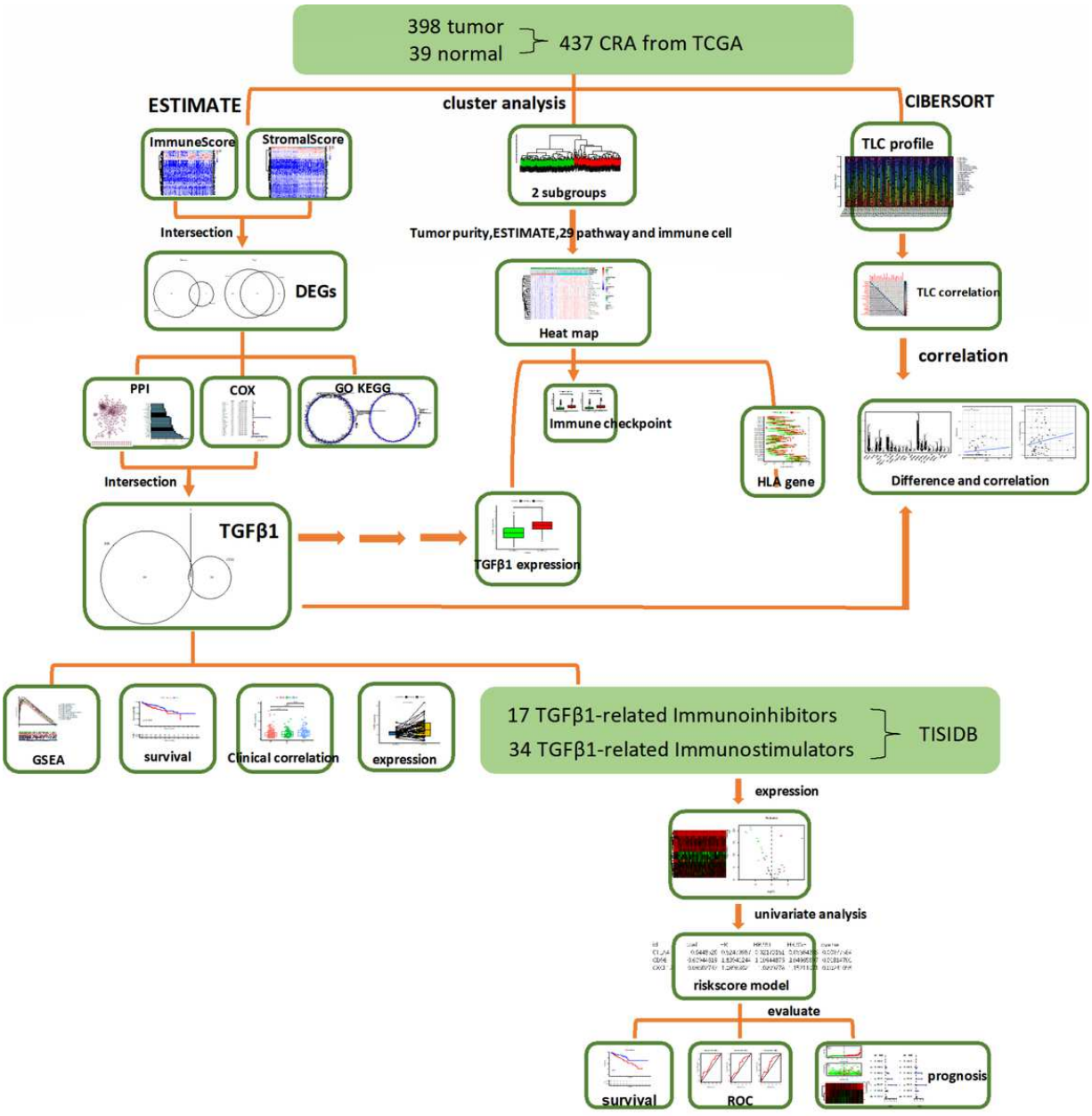

Figure 1

flow chart of this study

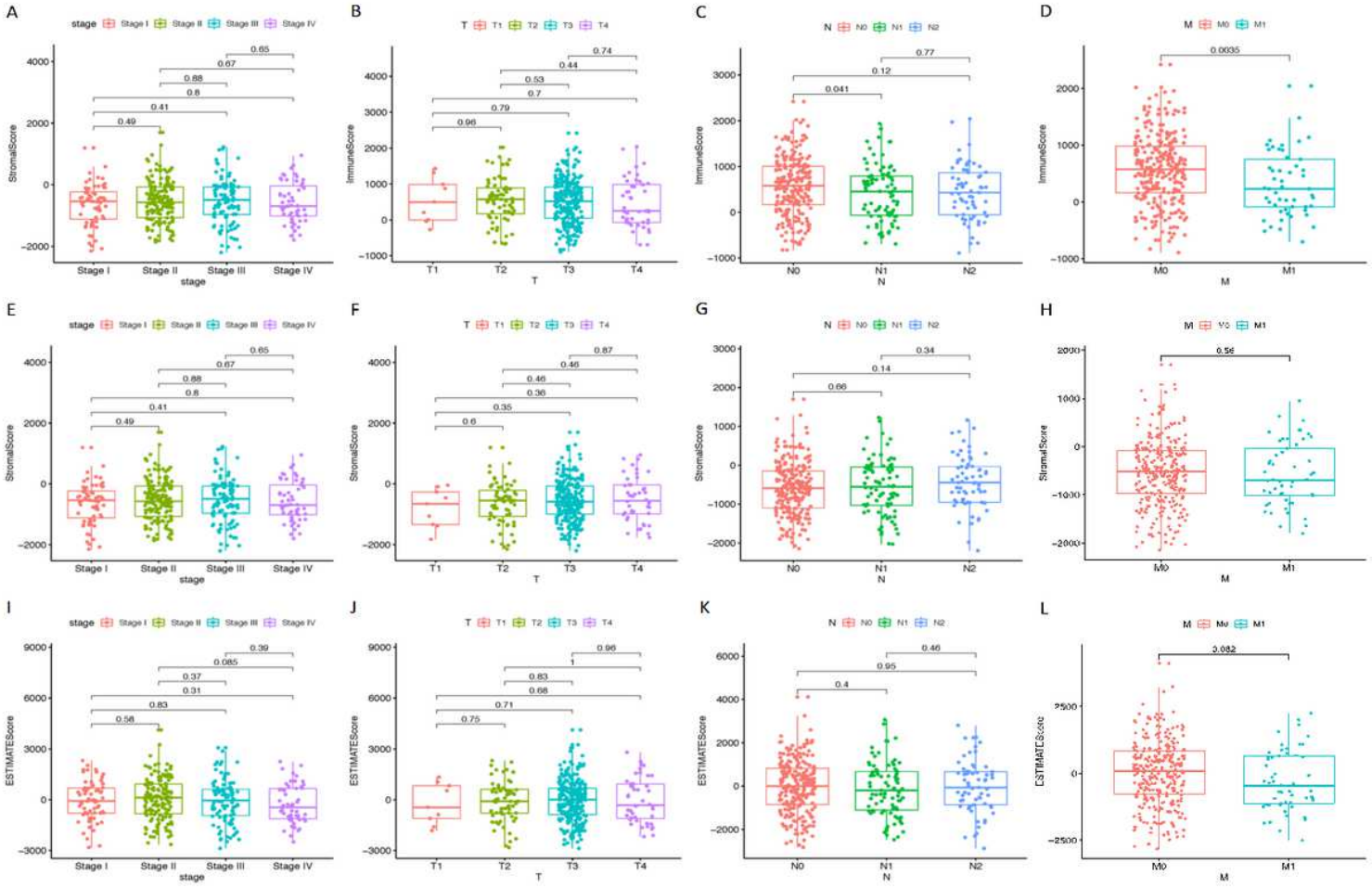

Figure 2

Relativity of ImmuneScore and StromalScore with clinicopathological characteristics. (A–C) correlation of ImmuneScore, StromalScore, and ESTIMATEScore with stage. (The p = 0.017, 0.8, and 0.31, separately, by Kruskal–Wallis rank sum test). (D–F) correlation of above three kinds of scores with T classification (p = 0.7, 0.36, 0.68 for ImmuneScore, StromalScore, and ESTIMATEScore, separately, by Kruskal–Wallis rank sum test). (G–I) correlation of scores with N classification. Similar to the above, p = 0.12, 0.14, 0.95, separately, with Kruskal–Wallis rank sum test. (J–L) correlation of scores with M classification (p = 0.0035, 0.56, 0.082 for ImmuneScore, StromalScore, and ESTIMATEScore separately by Wilcoxon

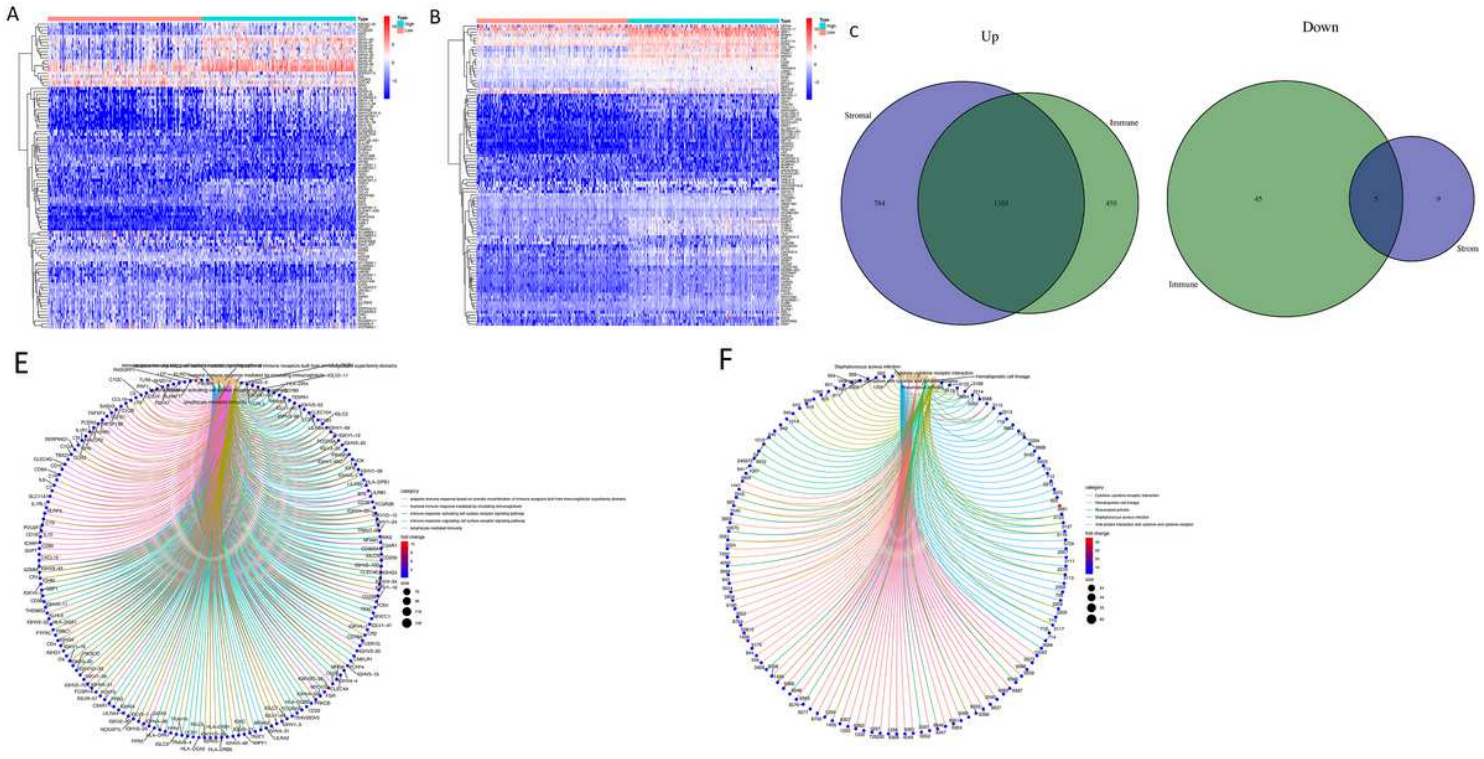

Figure 3

Heatmaps, Venn plots and enrichment analysis of GO and KEGG for DEGs. (A) Heatmap for DEGs from comparison between the high score group vs.the low score group in ImmuneScore. Row name of heatmap represent the specific gene, and column name means the ID of samples. DEGs were abtained using Wilcoxon rank sum test with q = 0.05 and fold-change >1 after log2 transformation as the significance threshold. (B) Heatmap for DEGs in StromalScore, similar with (A). (C,D) Venn plots show corporate up-regulated and down-regulated DEGs both in ImmuneScore and StromalScore, and q < 0.05 and fold-change >1 after log2 transformation as the DEGs significance filtering threshold. (E,F) GO and KEGG enrichment analysis for 1108 DEGs, with p and q < 0.05

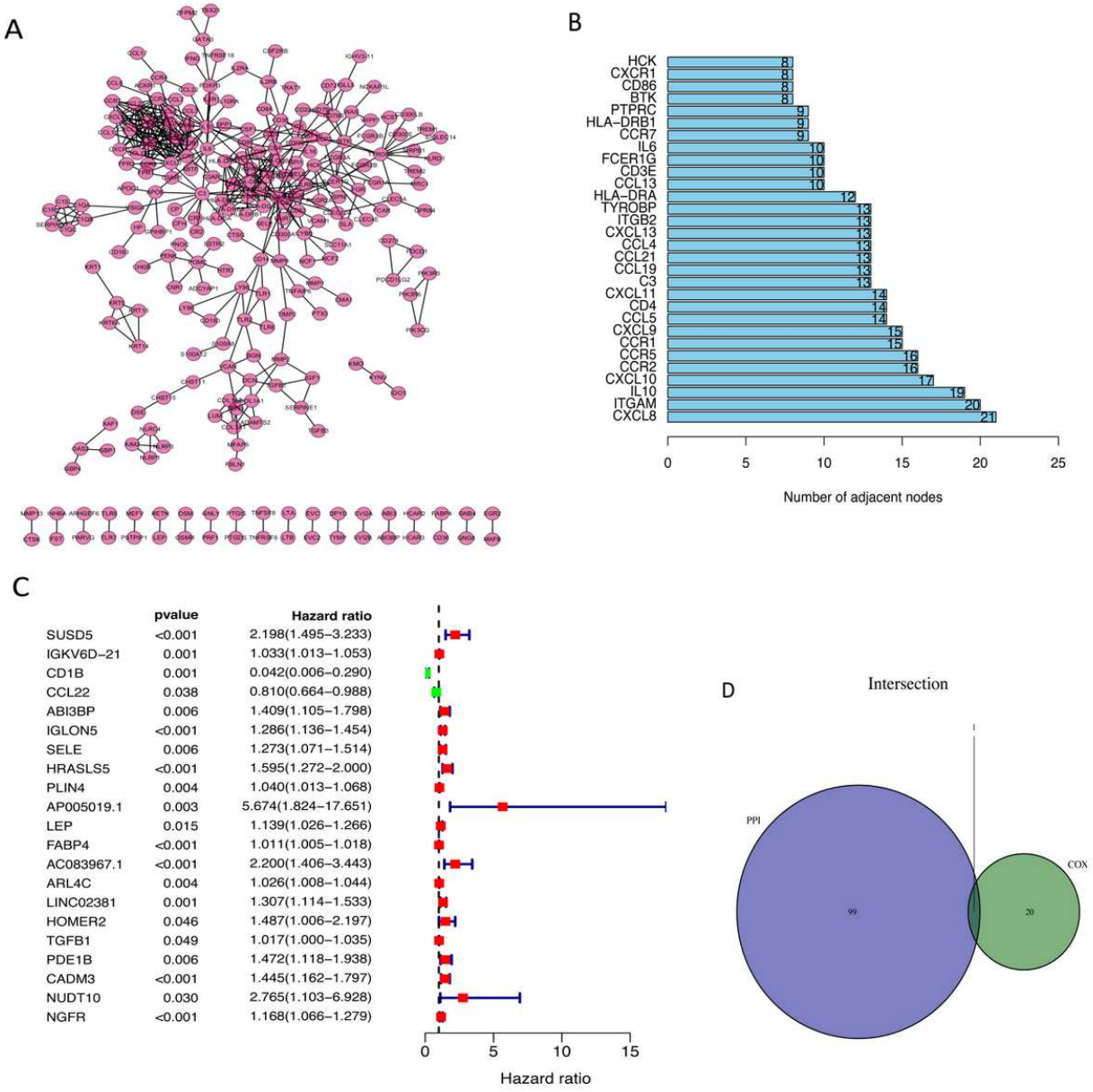

Figure 4

Protein–protein interaction network and univariate COX analysis. (A) Interaction network of nodes with interaction confidence value >0.95. (B)The top 30 genes ranked by the number of nodes. (C) Univariate COX regression analysis using 1108 DEGs, showing the significant factors with $p < 0.005$. (D) Venn plot circling the common factors shared by top 100 nodes in PPI and 21 significant factors in univariate COX

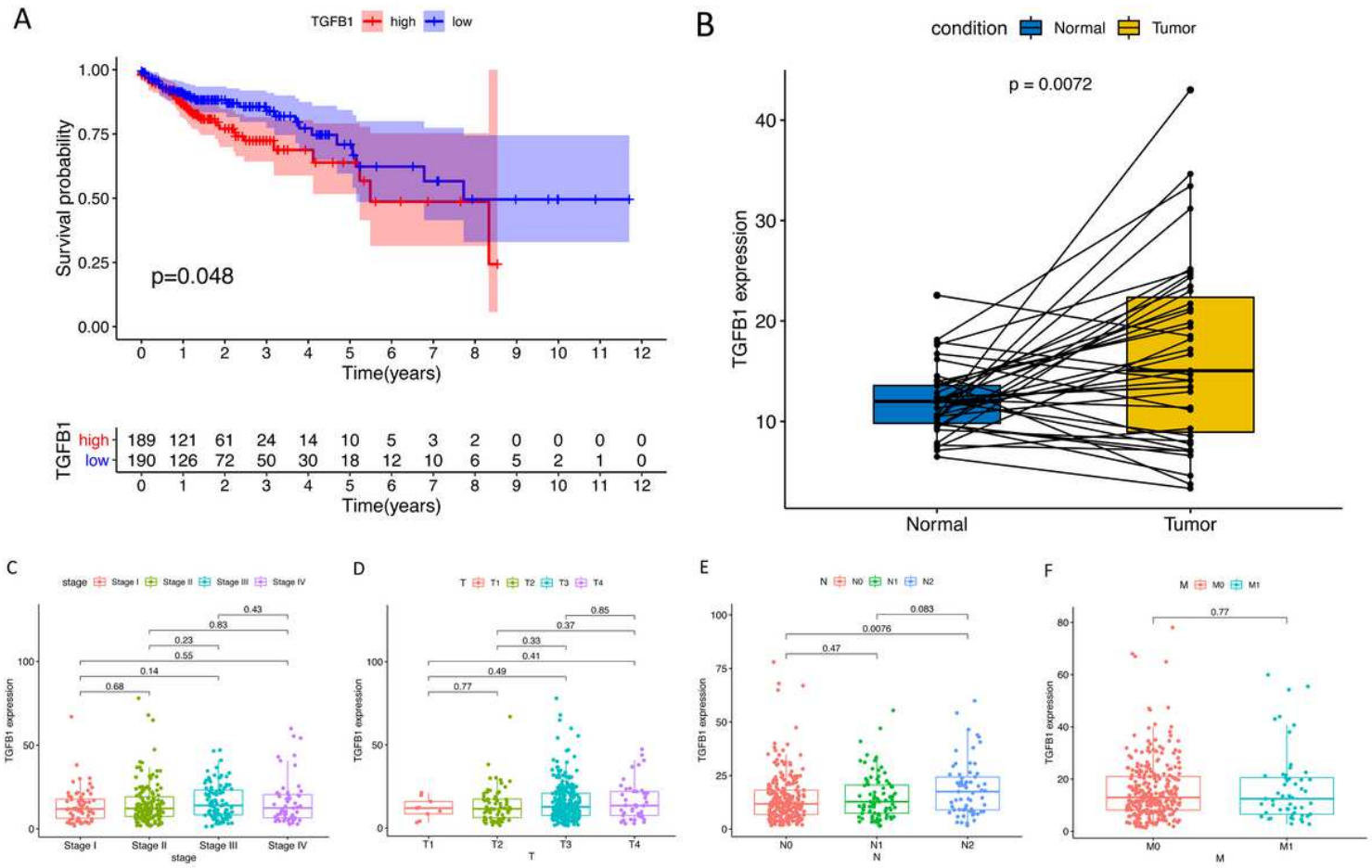

Figure 5

Different expression of TGFβ1 in spicemen and its influence in survival and clinicopathological characteristics of CRA patients. (B) Different expression of TGF β1 between pairing normal and tumor spicemen. Analyses were conducted with $p<0.05$ by Wilcoxon rank sum test. (A) Survival curve for CRA patients with various TGFβ1 expression. Patients were matched with high expression or low expression compared with the median. $p = 0.048$ by log-rank test. (C-F) The correlation of TGFβ 1 expression with clinicopathological characteristics was revealed by Wilcoxon rank sum or Kruskal–Wallis rank sum test

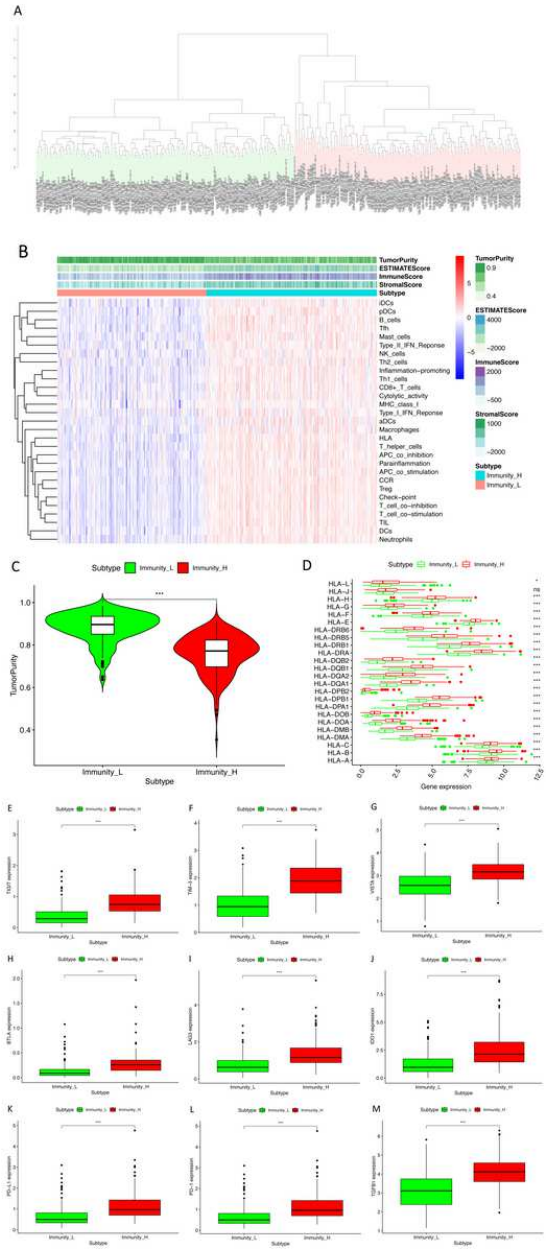

Figure 6

Identification and validation of immune-related subgroups of colon cancer. A. cluster analysis divided samples into 2 subgroups, that are immunity low group and immunity high group. B.29 immune-related gene sets are enriched in ssGSEA with colon cancer,. These gene sets are composed of immune cells and immune processes. The heat map also shows tumor purity, ESTIMATE score, immune score and matrix score. Heatmap of sample clustering at consensus k = 2.C. Tumor purity of samples from the two immune subgroups (*** P < 0.001). D. RNA expression levels of HLA genes in samples from the two immune subgroups. E-L. 8 immunecheckpoint genes are significantly up-regulated in immunity high group (***P < 0.001). M. TGFβ1 is similarly upregulated in the immunity high group

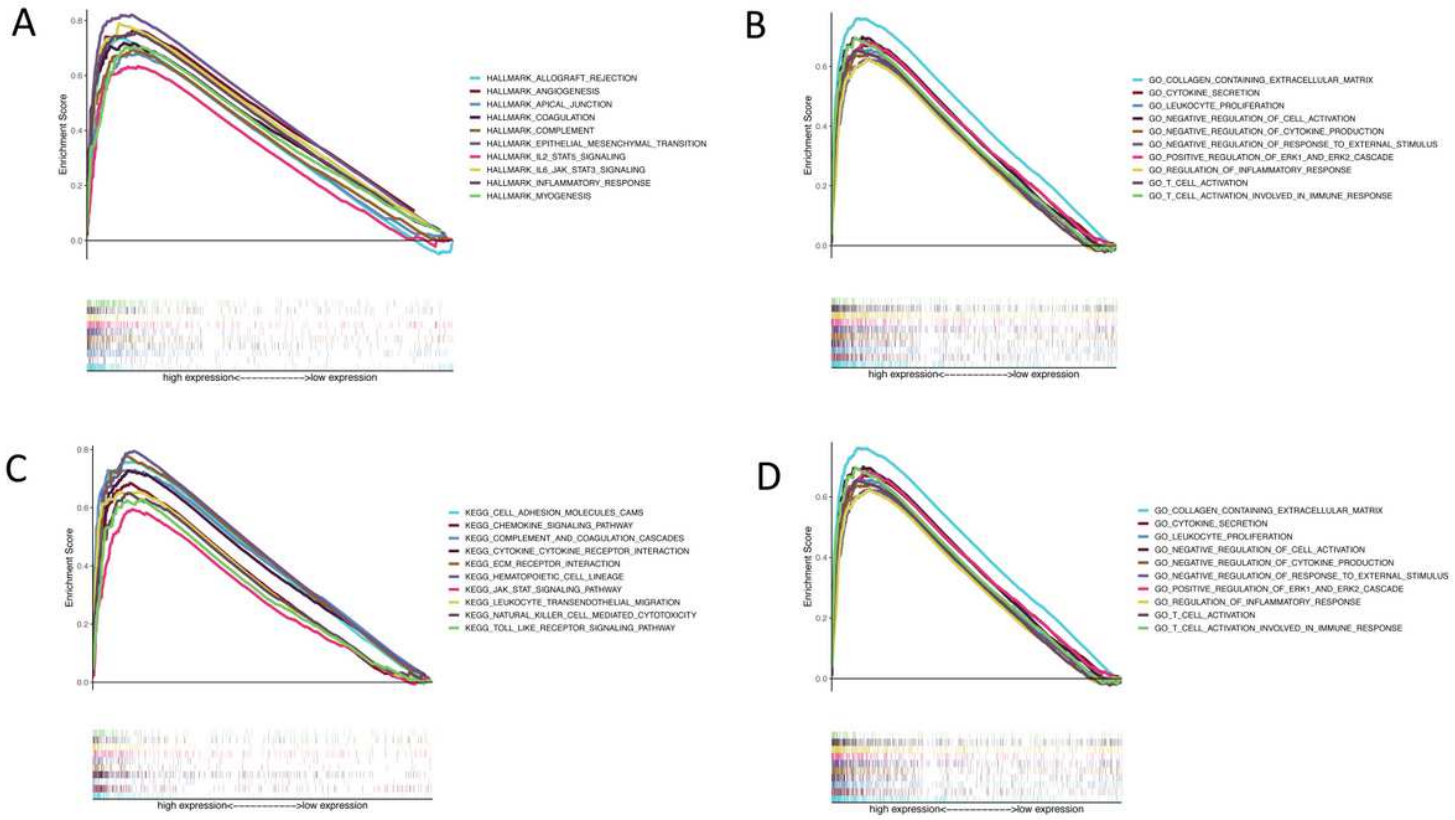

Figure 7

GSEA for TGFβ1 high-expression group. (A) The enriched gene sets in HALLMARK collection of high TGFβ1 expression. Each line shows one particular gene set. Gene sets with NOM $p < 0.05$ and FDR $q < 0.06$ were regarded significant. Only some leading gene sets were shown in the plot. (B) The enriched gene sets in GSE by samples of TGFβ1 high expression. (C) Enriched gene sets in KEGG collection, the immunologic gene sets, by samples of high TGFβ1 expression. Only several leading gene sets are shown in plot. (D) Enriched gene sets in GO collection by the high

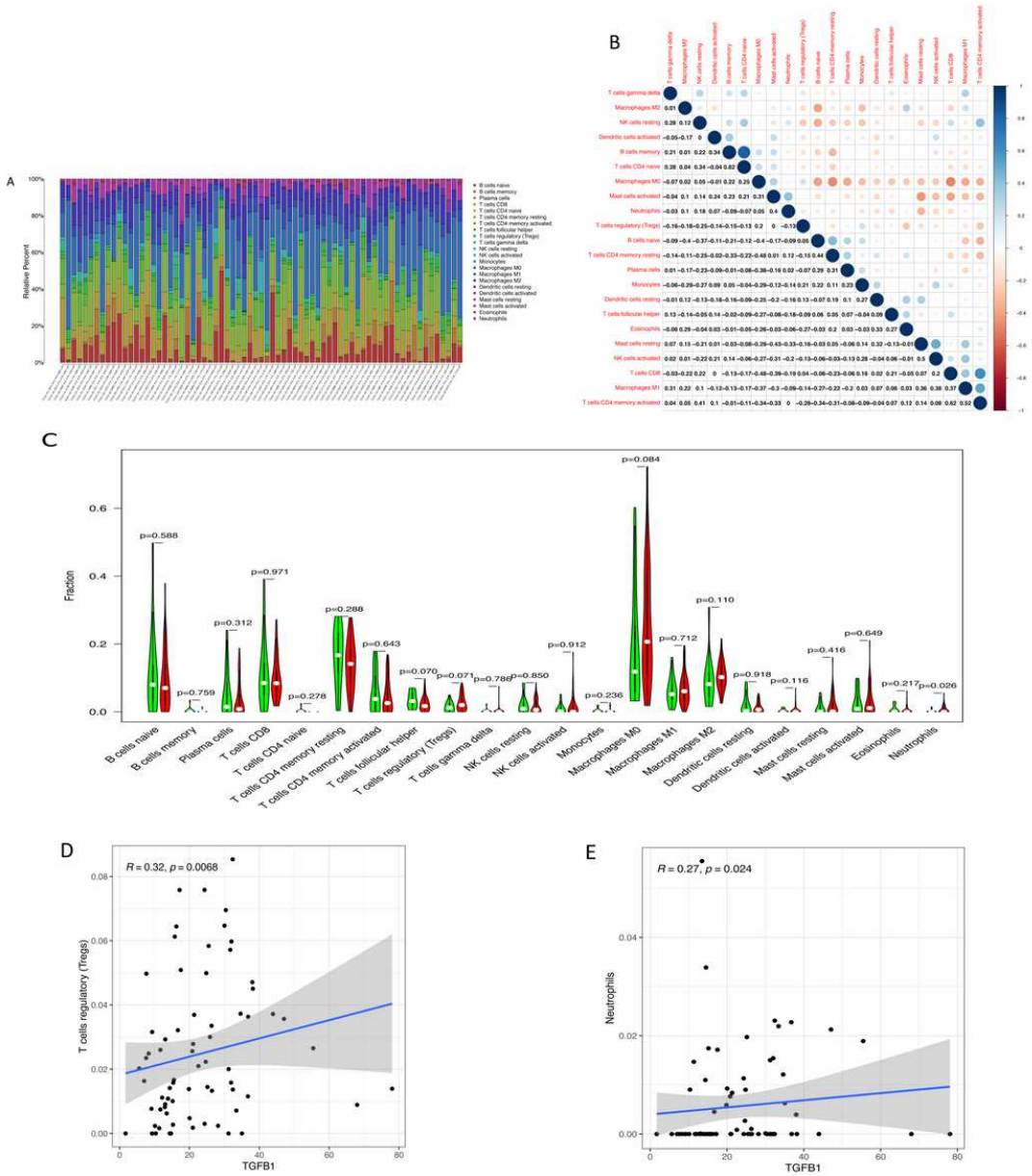

Figure 8

Connection between TGFβ1 expression and TIC profile. (A) Bar plot showing the ratio of 21 kinds of TICs in CRA samples. Column names were sample ID. (B) Heatmap showed the connection between 21 kinds of TICs and numbers in the tiny boxes represented the p value of correlation between two kinds of cells. The color in every tiny box showed correlation value between the two cells. And Pearson correlation coefficient was used for significant test.C. Violin graph shows the different proportion of 21 kinds of TIC between high TGFβ1 expression group and low TGFβ1 expression group compared to the median. D. The correlation between TGFβ1 expression and Tregs abundance was shown in the scatter plot, and P<0.05 was regarded significant. E. The correlation between TGFβ1 expression and Neutrophils abundance was shown in the scatter plot, and P<0.05 was regarded significant

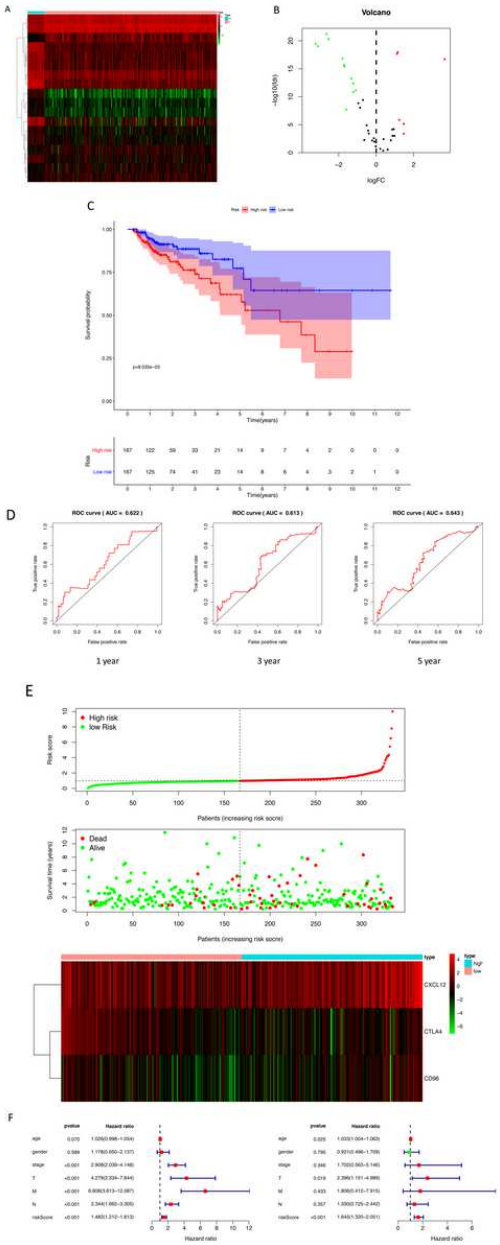

Figure 9

The prognostic value of TGFβ1-associated risk score. A. The heatmap of 18 significant differential expression TGFβ1- associated genes between tumor and normal tissue. (CSF1R,CTLA4,IL10,CD27,CD48,CD70,TNFSF4,BTLA,CD96,CD276,CD40LG,IL6R,PVR,TNFRSF13B,TNFRSF13C,TNFRSF17,TNFSF9). B.The volcano plot shows the distribution of 18 differential expression TGFβ1-associated genes mentioned above. C. The survival curve shows significantly diverse between high risk group and low risk group based on Kaplan–Meier curves.(P<0.001) D .Time-dependent receiver operating characteristic curves at 1 year,3 years and 5 years. E. The distribution and survival status of high risk group and low risk group and the expression of 3 predictive genes between the 2 groups. F. Stepwise multivariate COX regression analysis show risk score is significantly associated with survival.(P<0.001)

Supplementary Files

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

- Supplementary1.tif
- Supplementary2.tif

- Table1.xls