Microenvironment-related gene TNFSF13B predicts poor prognosis in kidney renal clear cell carcinoma

Mingzhe Jiang [Equal first author, 1] , **Jiaxing Lin** [Equal first author, 1] , **Haotian Xing** [1] , **Jun An** [1] , **Jieping Yang** [1] , **Biao Wang** [2] , **Meng Yu** [3] , **Yuyan Zhu** [Corresp. 1]

[1] Department of Urology, The First Hospital of China Medical University, Shenyang, Liaoning, China

[2] Department of Biochemistry and Molecular Biology, School of Life Sciences, China Medical University, Shenyang, Liaoning, China

[3] Key Laboratory of Transgenetic Animal Research, Liaoning Province, Department of Laboratory Animal Science, China Medical University, Shenyang, Liaoning, China

Corresponding Author: Yuyan Zhu
Email address: yyzhu@cmu.edu.cn

Background Kidney renal clear cell carcinoma (KIRC) affects the genitourinary system. Although treatment of KIRC in early stages can be highly successful, this type of cancer is difficult to detect until later stages of disease that are less easily treatable. Previous studies have focused on tumor cells, but have ignored the contributions of the tumor microenvironment. **Methods** We analyzed KIRC gene expression data from The Cancer Genome Atlas with the ESTIMATE algorithm to identify differentially expressed genes. Through protein-protein interaction network analysis, we identified clusters and picked genes from the clusters for further analysis. Differential expression, Kaplan-Meier, and univariate Cox analyses were used to select prognostic biomarkers. Gene Set Enrichment Analysis (GSEA) and Tumor Immune Estimation Resource (TIMER) analysis were used to explore the immune characteristics of these genes as biomarkers. **Results** Through ESTIMATE algorithm and other system biology tools, TNFSF13B was identified as a prognostic biomarker. TNFSF13B is highly expressed in tumors, and high expression of TNFSF13B leads to poor prognosis. Further GSEA and TIMER analysis revealed that the expression of TNFSF13B was related to immune signaling pathway and lymphocyte infiltration. Our findings strongly suggest that TNFSF13B may be a potential biomarker or target related to the tumor microenvironment for KIRC.

1 **Microenvironment-related gene TNFSF13B predicts poor prognosis in kidney renal clear**

2 **cell carcinoma**

3 Mingzhe Jiang [1]*, Jiaxing Lin [1]*, Haotian Xing [1], Jun An [1], Jieping Yang [1], Biao Wang [2], Meng

4 Yu [3], Yuyan Zhu [1]#

5 1 Department of Urology, The First Hospital of China Medical University, Shenyang, Liaoning,

6 China

7 2 Department of Biochemistry and Molecular Biology, School of Life Sciences, China Medical

8 University, Shenyang, Liaoning, China

9 3 Key Laboratory of Transgenetic Animal Research, Liaoning Province, Department of

10 Laboratory Animal Science, China Medical University, Liaoning, Shenyang, Liaoning, China

11 *Mingzhe Jiang and Jiaxing Lin contributed equally to this work

12 # Corresponding Author: Yuyan Zhu, No. 155 Nanjing North Street, Heping District, Shenyang,

13 Liaoning, 110001, China.

14 Email address: yyzhu@cmu.edu.cn

15 **Abstract**

16 **Background**

17 Kidney renal clear cell carcinoma (KIRC) affects the genitourinary system. Although treatment

18   of KIRC in early stages can be highly successful, this type of cancer is difficult to detect until

19   later stages of disease that are less easily treatable. Previous studies have focused on tumor cells,

20   but have ignored the contributions of the tumor microenvironment.

21   **Methods**

22   We analyzed KIRC gene expression data from The Cancer Genome Atlas with the ESTIMATE

23   algorithm to identify differentially expressed genes. Through protein-protein interaction network

24   analysis, we identified clusters and picked genes from the clusters for further analysis.

25   Differential expression, Kaplan-Meier, and univariate Cox analyses were used to select

26   prognostic biomarkers. Gene Set Enrichment Analysis (GSEA) and Tumor Immune Estimation

27   Resource (TIMER) analysis were used to explore the immune characteristics of these genes as

28   biomarkers.

29   **Results**

30   Through ESTIMATE algorithm and other system biology tools, TNFSF13B was identified as a

31   prognostic biomarker. TNFSF13B is highly expressed in tumors, and high expression of

32   TNFSF13B leads to poor prognosis. Further GSEA and TIMER analysis revealed that the

33   expression of TNFSF13B was related to immune signaling pathway and lymphocyte infiltration.

34   Our findings strongly suggest that TNFSF13B may be a potential biomarker or target related to

35   the tumor microenvironment for KIRC.

36   **Introduction**

37 Kidney renal clear cell carcinoma (KIRC) is a highly prevalent urinary system cancer,

38 accounting for 80% to 90% of lesions in the renal space and about 3% of all new cases of tumors

39 worldwide. About 100,000 patients die of renal cell carcinoma every year (Siegel et al., 2017),

40 and KIRC is the main pathological type of this carcinoma. Although biotherapy and surgical

41 treatment can be effective in early stages, renal cell carcinoma typically is not detected until later

42 stages. Additionally, KIRC typically progresses rapidly and exhibits early metastasis, with

43 overall poor patient prognosis. Therefore, identification of KIRC-related genes can provide

44 insight into the mechanism of renal cell carcinoma and facilitate early diagnosis and treatment of

45 renal cell carcinoma (Cairns, 2010).

46 To correlate tumor genetic composition with clinical prognosis, extensive databases of genome-

47 scale gene expression data have been established. The Cancer Genome Atlas (TCGA,

48 https://portal.gdc.cancer.gov/) is a public available repository of cancer genomic data, with data

49 relevant to several important cancers. These data can be analyzed to detect genomic changes that

50 can provide insight into cancer development. The expression patterns of different genes can be

51 compared to clinical diagnostic criteria and may help to identify markers that are relevant to

52 cancer diagnosis, may facilitate selection of effective treatment options, and can improve

53 prognosis (Krishnan et al., 2016).

54 Previous studies have shown that tumor cell intrinsic genes determine the occurrence,

55 development, and evolution of KIRC (Hsieh et al., 2017). However, the tumor microenvironment

56 generally has a serious impact on gene expression in cancer, as well as in KIRC and other solid

57   tumor tissues, which can affect clinical consequences (Chevrier et al., 2017). The tumor

58   microenvironment is complex, and includes inflammatory cytokines and mediators, extracellular

59   matrix (ECM) molecules, and immune, mesenchymal, and endothelial cells (Hanahan &

60   Coussens, 2012).

61   Of these components, immune and stromal cells are likely the key contributors to the tumor

62   microenvironment, and understanding how these cells affect tumor development is essential for

63   tumor diagnosis and assessment of prognosis. "Estimation of STromal and Immune cells in

64   MAlignant Tumours using Expression data" (ESTIMATE) is a method that uses gene expression

65   signatures to infer the level of non-tumor cells in tumor samples (Yoshihara et al., 2013). The

66   algorithm is based on single sample gene set enrichment analysis, and the levels of immune cells

67   and stromal cells are inferred by calculating immune score and matrix score. The ESTIMATE

68   scores reflect the purity of tumor cells, where the higher the ESTIMATE scores, the lower the

69   tumor purity. Previous studies have effectively used the ESTIMATE algorithm to analyze large

70   datasets to provide insight into melanoma (Yang et al., 2019), breast cancer (Chen et al., 2019),

71   and bladder cancer (Wang et al., 2019). However, despite numerous studies showing the efficacy

72   of this data-based algorithm and the ability to calculate immune and stromal scores for KIRC, no

73   detailed investigation has been conducted.

74   In this first report using the TCGA database to assess the KIRC cohort using the ESTIMATE

75   algorithm, we identified the gene related to the microenvironment that should be further

76   characterized for its ability to predict the prognosis of patients with KIRC.

77 **Materials and methods**

78 **Database**

79 The work flow of this study is shown in Fig. 1. The expression profiles FPKM (level 3) and

80 clinical data were obtained for KIRC patients using the TCGA website in March 1, 2019

81 (https://portal.gdc.cancer.gov/). The immune, stromal, and ESTIMATE scores of each sample

82 were determined according to the ESTIMATE algorithm. This tool uses expression data to

83 estimate the contributions of stromal and immune cells in malignant tumor tissue.

84 **Identification of differentially expressed gene (DEGs).**

85 The package "limma" (https://bioconductor.org/packages/limma/, version=3.42.2) was used for

86 expression data analysis (Ritchie et al., 2015). The input data are the expression data of KIRC.

87 According to the median value of the scores, the patients were divided into two groups for

88 difference analysis. We then used Wilcoxon test to test the difference in the expression of each

89 gene between the two groups. |Log FC| > 1 (FC: fold change) and false discovery rate (FDR) p <

90 0.05 were set as the cutoff values to screen DEG.

91 **Heatmap and cluster analysis**

92 The R package "pheatmap" (https://CRAN.R-project.org/package=pheatmap, version=1.0.12)

93 was used to generate a heatmap and perform cluster analysis. The gene expression profile data

94 were input, the gene expression heat map was drawn, and the genes were clustered according to

95 the sample groups.

96 **Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway**

97 **analysis**

98 The GO and KEGG analyses were performed using the R package "clusterProfiler"

99 (https://bioconductor.org/packages/clusterProfiler/, version=3.14.3) (Yu et al., 2012). FDR p

100 value < 0.05 was used to distinguish significantly enriched terms. Enter the genes, and used the

101 functions "enrichGO" and "enrichKEGG" for enrichment analysis. The results of enrichment

102 analysis are shown by bubble diagram.

103 **Construction of protein-protein interaction (PPI) network and identification of functional**

104 **clusters**

105 The PPI network was constructed using the Search Tool for the Retrieval of Interacting Genes

106 (STRING; https://string-db.org/) (Szklarczyk et al., 2019). Unconnected genes were removed

107 and the confidence was set to 0.4 to obtain the PPI network. We further analyzed the PPI

108 network with Cytoscape (Cytoscape_v3.6.1) (Shannon et al., 2003) and then Molecular Complex

109 Detection (MCODE) was used to identify functional clusters (Bader & Hogue, 2003). The

110 default parameters were set as follows: the degree cutoff value was 2, the node score was 0.2, the

111 k-score was 2, and the maximum depth was 100.

112 **Kaplan-Meier survival analysis**

113 Kaplan-Meier curve has time as the horizontal axis and survival rate as the vertical axis. The R

114 packages "survival" (https://CRAN.R-project.org/package=survival, version=3.1-8) and

115   "survminer" (https://CRAN.R-project.org/package=survminer, version=0.4.6) were applied for

116   survival analysis. A value of p < 0.05 means it is statistically significant. The input matrix

117   contained the sample's survival time, survival status, and gene expression data. Samples were

118   grouped using a median or optimal threshold for Kaplan-Meier regression analysis.

Univariate and multivariate Cox regression analysis

120   Cox analysis is a semi-parametric model of survival analysis, which is used to identify the risk

121   factors of disease and to analyze prognosis. Univariate and multivariate Cox analysis were

122   performed with the R package "survival". The input data was the same as the Kaplan-Meier

123   analysis, which was calculated using the function "cox" in package. The results of the analysis

124   can be displayed in a forest map.

Identification of practical prognostic biomarker

126   Three methods were used to select practical prognostic biomarker. The first was to find out that

127   the DEGs between tumor and normal tissue, the threshold was set to | Log FC | > 1, FDR-p <

128   0.05; the second was to use univariate Cox analysis to get the prognostic genes (p < 0.05); the

129   third was to use Kaplan-Meier analysis to find out the prognostic genes (p < 0.05). Finally, the

130   intersection of the three methods results was taken, and the Venn diagram is drawn with a Web

131   tool to show (http://bioinformatics.psb.ugent.be/webtools/Venn/). We used DEG, and univariate

132   COX here to screen prognostic markers for KIRC, and we did so because this is the most

133   common way to obtain prognostic markers. We added Kaplan-Meier analysis because it can

134   intuitively evaluate the prognosis of patients and has practical significance for clinical use. With

135   the intersection of the three meaningful results, practical prognostic markers can be obtained.

136   **Gene set enrichment analysis**

137   Gene Set Enrichment Analysis (GSEA) is a computational method that determines whether a

138   group of basically defined genes exhibit statistically significant differences between two

139   biological states (Aravind et al., 2005). In this study, according to the gene expression data, the

140   samples were divided into two groups for GSEA. One thousand genome replacement tests were

141   performed per analysis in GSEA_4.0.1 and the expression level of each individual gene was used

142   as a phenotypic tag. Values of p-value < 0.05 and FDR p-value < 0.05 were applied to define the

143   pathways with statistical significance. Finally, we used the R package "ggplot2"

144   (https://CRAN.R-project.org/package=ggplot2, version=3.2.1) to draw multiple pathways for

145   display, the input file is the result of GSEA's analysis..

146   **Exploration of immune characteristics and drug targets**

147   The R package "ggstatsplot" (https://CRAN.R-project.org/package=ggstatsplot, version=0.4.0)

148   was used to analyze the correlation between genes expression and the scores. We used the Tumor

149   Immune Estimation Resource (TIMER; https://cistrome.shinyapps.io/timer/) (Li et al., 2017) to

150   explore the relationship between genes and immune cells infiltration. Finally, we queried the

151   Tumor Immune System Interactions Database (TISIDB; http://cis.hku.hk/TISIDB) (Ru et al.,

152   2019) for drugs that can act on the genes, using the drug information in the database from

153   Drugbank database (https://www.drugbank.ca).

154   **Statistical Analysis**

155   All statistical analyses, including the operation of the ESTIMATE algorithm and gene

156   identification, were run using R software (R x64 3.5.1). All R packages were obtained from

157   CRAN (https://cran.r-project.org) and Bioconductor (http://www.bioconductor.org). The two

158   groups were compared by the Wilcoxon test, and comparison between multiple groups was

159   performed by Kruskal-Wallis test. Statistical significance was defined as $p < 0.05$.

160   **Results**

161   **Scores of the ESTIMATE algorithm are related to the survival and prognosis of kidney**

162   **renal clear cell carcinoma**

163   The gene expression profiles and clinical data were obtained for all KIRC patients available in

164   the TCGA database. Using the ESTIMATE algorithm, the stromal score range was -1433.77 to

165   1967.19, the immune score range was -693.96 to 3328.21, and the ESTIMATE score range was -

166   2127.72 to 5091.59. To make the Kaplan-Meier analysis more statistically significant, the

167   function "res.cut" in R package "survminer" was used to find the best segmentation point to

168   group the samples. The scores of each group were grouped according to the optimal

169   segmentation point calculated by the function. The best segmentation point based on immune

170   score was 1778.92. Relative to this point, the samples were divided into high and low immune

171   score groups for Kaplan-Meier overall survival analysis. The results showed that the prognosis of

172   the high immune score group was worse than the overall group (p=0.003, Fig 2A). In a similar

173   analysis, the best segmentation points for stromal and ESTIMATE scores were 830.37 and

174   2785.69, respectively. Kaplan-Meier analysis showed worse prognosis of the high score group

175   for both stromal and ESTIMATE score-based analysis (p <0.05, Fig 2B-3C). This shows that all

176   three scores predict the overall survival rate of patients. We generated a boxplot to show the

177   relationships between the three scores and pathological stages of KIRC, and then analyzed the

178   statistical significance by Kruskal-Wallis test. The results showed that the immune scores (Fig.

179   2D, p < 0.001) and ESTIMATE scores (Fig. 2F, p = 0.006) were both significant in different

180   pathological stages, and increased with the increase of pathological stages. However, there were

181   no differences in the stromal scores among different stages (Fig. 2E, p=0.140).

182   **Mining differential genes in gene expression profiles according to immune score and**

183   **stromal score**

184   539 KIRC samples were first grouped according to the median of immune score and then a

185   heatmap of gene expression was constructed (Fig. 3A). In the differential gene analysis of

186   immune score, 512 genes were up-regulated and 147 genes were down-regulated (|log FC| > 1, p

187   < 0.05). Similarly, the samples were grouped according to the median of stromal score and

188   constructing heat maps of gene expression (Fig. 3B). 259 genes were up-regulated and 152 genes

189   were down-regulated in the groups with high and low stromal scores (|log FC| > 1, p < 0.05).

190   Since both immune and stromal factors are part of the immune microenvironment, we took the

191   intersection between immune up-regulated genes and stromal up-regulated genes. Then we also

192   took the intersection between immune down-regulated genes and stromal down-regulated genes.

193   The Venn map shows that 48 genes are generally up-regulated (Fig. 3C) and 47 genes generally

194   are down-regulated (Fig. 3D). We decided to focus on these 95 DEGs, which we propose are

195   genes that are important to the microenvironment. We investigated the function of these genes by

196   GO and KEGG enrichment analysis. The most significant term of GO enrichment analysis is

197   Humoral immune response, while that of KEGG enrichment analysis is Cytokine-cytokine

198   receptor interaction. These results suggest that these genes are related to immunity (Fig. 3E-3F).

199   **Construction of PPI network and identification of functional clusters**

200   To better understand the interaction of DEGs in the microenvironment, we used the STRING

201   tool to construct a PPI network for the 95 DEGs and visualize the network with Cytoscape (Fig.

202   4A). The network consists of 53 nodes and 110 edges, and 42 genes are not displayed in the

203   network because they are not linked to the other genes. We then used the tool MCODE of

204   Cytoscape to identify the functional clusters; this tool aggregates highly related genes into a

205   cluster. A total of four clusters which contain 24 genes were identified (Fig. 4B-4F). We took the

206   24 genes for the next step of analysis.

207   **Identification of KIRC practical prognostic biomarker**

208   We used three methods to screen practical prognostic biomarkers from the 24 genes. The

209   expression profiles of 24 genes were analyzed to find the DEGs of tumor and normal tissues. A

210   total of seven DEGs (CD163, CCL18, TNFSF13B, IL7R, IL2RA, PROZ, and PAH) were found

211 (|logFC| > 1, p < 0.05). According to the median expression of each gene, the genes were divided

212 into high expression group and low expression group. Kaplan-Meier analysis of the 24 genes

213 showed that seven genes (TNFSF13B, XCR1, IGLL5, PAH, IL10, F7, and AMBP) showed

214 significant differences in expression (p<0.05). Univariate Cox analysis of the 24 genes showed

215 that expression levels of 11 genes (CCL13, IL2RA, TNFSF13B, IGLL5, CCL19, SPIB, CCL21,

216 CR1, TNFRSF17, CD19, POU2AF1) were significantly different (p<0.05). The results of the

217 three analyses were combined in a Venn diagram (Fig. 5). TNFSF13B is the only gene found in

218 all three analysis.

High expression of TNFSF13B leads to poor prognosis of KIRC

220 The expression of TNFSF13B in tumor tissue is higher than that in normal tissue (p < 0.001, Fig.

221 6A). High expression of TNFSF13B leads to poor prognosis of KIRC (p=00019, Fig. 6B). The

222 clinical factors and TNFSF13B were analyzed by univariate and multivariate regression analysis.

223 Univariate Cox analysis showed that Age, Grade, Stage, T (T stage), N (Lymph node), M

224 (Tumor metastasis), and TNFSF13B (p<0.001) were prognostic factors (Fig. 6C), it has very

225 significant statistical significance. Multivariate Cox analysis showed that Grade and Age were

226 independent prognostic factors (Fig. 6D). The result indicated that TNFSF13B was a prognostic

227 factor, but it could not analyze the prognosis independently of other clinical factors. The

228 relationship between gene expression and stage/grade was analyzed by box diagram, and

229 Kruskal-Wallis test was used for comparison among groups. The results showed that the

230 TNFSF13B expression was different in different stages and grades (p < 0.001, Fig. 6E-6F), and

231 the TNFSF13B expression increased with the increase of stage and grade.

232 **There is a positive correlation between TNFSF13B and tumor microenvironment.**

233 By analyzing the correlation between immune/stromal/ESTIMATE score and TNFSF13B

234 expression, it was found that TNFSF13B was positively correlated with immune (R = 0.67, p <

235 0.001, Fig 7A), stromal (R = 0.43, p < 0.001, Fig 7B), and ESTIMATE (R = 0.64, p < 0.001, Fig

236 7C). We next queried the relationship between TNFSF13B and immune cell infiltration in the

237 TIMER database (Fig. 7D-7J), and found that its expression was positively correlated with tumor

238 purity and positively correlated with immune cells. TNFSF13B expression value showed

239 maximum partial spearmanm's correlation with dendritic cells (partial cor = 0.634, p =1.35E-52).

240 **Gene set enrichment analysis of TNFSF13B**

241 GSEA is a powerful analytical method used to interpret gene expression data. We sequentially

242 analyzed gene enrichment and performed GSEA for TNFSF13B. We used the Molecular

243 Signatures Database (MSigDB, http://www.broadinstitute.org/msigdb) of Collection

244 (c2.cp.kegg.v7.0.symbols.gmt) (A Liberzon et al. 2011; A Liberzon et al. 2015). We set the

245 threshold for FDR p-value< 0.05 and p-value< 0.05. Significantly enriched signal transduction

246 pathways were identified based on normalized enrichment score (NES). We identified pathways

247 with high NES that are related to immune function. Because few meaningful signal pathways

248 were enriched by low gene expression, we only present the signal pathways enriched for high

249 gene expression (Fig. 8A). There was clear enrichment in immune-related signaling pathways.

250 Finally, we found three drugs that have effects on TNFSF13B in the TISIDB database (Fig. 8B),

251 1, 4-dimethyl Dioxide (DB03316), Citric Acid (DB04272), and Belimumab (DB08879). Among

252 them, Belimumab is an intravenous immunosuppressant and the only target for the adjuvant

253 treatment of systemic lupus erythematosus (SLE) (LJ Scott et al. 2012).

254 **Discussion**

255 Previous experiments of KIRC focused on animal tumor models, in vitro tumor cell lines, or

256 patient tumor samples (Capitanio & Montorsi, 2016). However, the KIRC microenvironment

257 must also be considered. With increased exploitation of whole genome sequencing, TCGA has

258 been established as an extensive tumor database for free use by the research community. These

259 resources facilitate data analysis of many cancers, including KIRC (Krishnan et al., 2016).

260 The goal of this study was to identify genes that are related to the microenvironment and overall

261 survival of KIRC. Gene expression patterns were compared for cases with high and low immune

262 scores, and 95 genes were identified as potentially involved in functions related to the

263 extracellular matrix or to immune response. To assess prognostic ability, we analyzed 95

264 differentially expressed genes from groups with high immune score (or high stromal score), and

265 used GO and KEGG enrichment analysis to determine that many of these genes are components

266 of the tumor microenvironment. Next, we constructed a protein-protein interaction network of

267 these 95 genes, and identified four clusters which contained 24 genes. We performed differential

268 expression, Kaplan-Meier, and univariate Cox analyses for the 24 genes. The TNFRSF13B gene

269 was identified in all three analyses. TNFRSF13B was overexpressed in tumors, Kaplan-Meier

270   analysis showed that high expression of TNFRSF13B led to poor prognosis, and univariate Cox

271   showed that TNFRSF13B was a risk factor. The expression of TNFRSF13B is positively

272   correlated with tumor stage and grade. These results suggest that TNFRSF13B is a potential

273   prognostic marker.

274   TNFRSF13B is an essential factor in the tumor microenvironment of KIRC. There was a high

275   positive correlation between TNFRSF13B and ESTIMATE score, indicating that the more

276   TNFRSF13B expression, the less the content of tumor cells, the more the content of immune

277   cells and stromal cells. Renal cell carcinoma has a good effect on immunotherapy (29120911), so

278   we inquired about the relationship between TNFRSF13B and six kinds of immune cells.

279   TNFRSF13B has a high correlation with immune cells. GSEA results showed that TNFRSF13B

280   was involved in many immune-related pathways, including the T cell receptor signaling pathway

281   and so on. TNFSF13B belongs to the tumor necrosis factor (TNF) ligand family, and is also

282   named BAFF.   TNFSF13B variants are associated with multiple sclerosis and SLE (Steri et al.,

283   2017), and TNFSF13B is a key factor in the survival of B lymphocytes (Cremasco et al., 2014).

284   Previous study suggested a potential role of these TNFSF members including TNFRSF13B in

285   renal tumor biology (Pelekanou et al., 2011). Recently, it has been reported that TNFSF13B is

286   related to the prognosis and immune infiltration of KIRC (Li et al., 2019). This article starts

287   directly from TNFSF13B to explore its prognosis and immune infiltration. However, we started

288   from the tumor microenvironment, went through many aspects of screening, and finally selected

289   the representative gene TNFSF13B. Their research and our exploration have jointly confirmed

290 the important role of TNFSF13B in KIRC. Through our research, TNFSF13B can not only

291 predict the prognosis of KIRC, but also be an important medium for KIRC to transmit

292 information between the tumor cell and tumor microenvironment, and is a potential therapeutic

293 target.

294 The interaction between KIRC and the tumor microenvironment seriously affects the progress

295 and evolution of this carcinoma, and subsequently affects the recurrence and overall prognosis of

296 patients. Previous reports have provided a detailed analysis of how the activation of tumor

297 intrinsic genes shapes the tumor microenvironment. In our current work, we analyzed the genetic

298 characteristics of the microenvironment that affect the development of KIRC. Our results may

299 help explain the complex interactions between the tumor and the tumor environment in kidney

300 renal clear cell carcinoma. Of course, our research also has some shortcomings. In the future, we

301 need more data to prove the prognostic ability of TNFSF13B and also need experiments to prove

302 the relationship between TNFSF13B and microenvironment.

303 **Conclusions**

304 We analyzed KIRC gene expression data from The Cancer Genome Atlas with the ESTIMATE

305 algorithm to identify the differentially expressed genes. Through multiple analyses, we identified

306 TNFSF13B as a potential biomarker and target for kidney renal clear cell carcinoma. Further

307 study of TNFSF13B may improve understanding of the potential relationships between the tumor

308 microenvironment and the prognosis of kidney renal clear cell carcinoma.

309 **Acknowledgements**

310 We appreciate the data obtained from the TCGA database in this study.

311 **Abbreviations:** KIRC - Kidney renal clear cell carcinoma, TCGA - The Cancer Genome Atlas,

312 ESTIMATE -using expression data to estimate stromal and immune cells in malignant tumor

313 tissue, DEGs - differentially expressed genes, FC - fold change, GSEA - Gene Set Enrichment

314 Analysis. FDR - false discovery rate, PPI - protein-protein interaction, TIMER - Tumor Immune

315 Estimation Resource.

316 **References**

317 Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, and Mesirov Jp. 2011. Molecular
318   signatures database (MSigDB) 3.0. *Bioinformatics (Oxford, England)* 27:1739-1740. DOI
319   10.1093/bioinformatics/btr260

320 Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov Jp, and Tamayo P. 2015. The Molecular Signatures
321   Database (MSigDB) hallmark gene set collection. *Cell systems* 1:417-425. DOI 10.1016/j.cels.2015.12.004

322 Krishnan B, Rose Tl, Kardos J, Milowsky Mi, and Kim Wy. 2016. Intrinsic Genomic Differences Between African
323   American and White Patients With Clear Cell Renal Cell Carcinoma. *JAMA oncology* 2:664-667. DOI
324   10.1001/jamaoncol.2016.0005

325 Ru B, Wong Cn, Tong Y, Zhong Jy, Zhong Ssw, Wu Wc, Chu Kc, Wong Cy, Lau Cy, Chen I, Chan Nw, and Zhang
326   J. 2019. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics
327   (Oxford, England)* 35:4200-4202. DOI 10.1093/bioinformatics/btz210

328 Wang B, Pan W, Yang M, Yang W, He W, Chen X, Bi J, Jiang N, Huang J, and Lin T. 2019. Programmed death
329   ligand-1 is associated with tumor infiltrating lymphocytes and poorer survival in urothelial cell carcinoma
330   of the bladder. *Cancer Science* 110:489-498. DOI 10.1111/cas.13887

331 Chen Ch, Lu Ys, Cheng Al, Huang Cs, Kuo Wh, Wang My, Chao M, Chen Ic, Kuo Cw, Lu Tp, and Lin Ch. 2019.
332   Disparity in Tumor Immune Microenvironment of Breast Cancer and Prognostic Impact: Asian versus
333   Western Populations. *The oncologist* undefined:undefined. DOI 10.1634/theoncologist.2019-0123

334   Hanahan D, and Coussens Lm. 2012. Accessories to the crime: functions of cells recruited to the tumor
335   microenvironment. *Cancer Cell* 21:309-322. DOI 10.1016/j.ccr.2012.02.022

336   Szklarczyk D, Gable Al, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva Nt, Morris Jh, Bork
337   P, Jensen Lj, and Mering Cv. 2019. STRING v11: protein-protein association networks with increased
338   coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*
339   47:D607-D613. DOI 10.1093/nar/gky1131

340   Yu G, Wang Lg, Han Y, and He Qy. 2012. clusterProfiler: an R package for comparing biological themes among
341   gene clusters. *Omics : a journal of integrative biology* 16:284-287. DOI 10.1089/omi.2011.0118

342   Bader Gd, and Hogue Cw. 2003. An automated method for finding molecular complexes in large protein interaction
343   networks. *BMC Bioinformatics* 4:2. DOI 10.1186/1471-2105-4-2

344   Hsieh Jj, Purdue Mp, Signoretti S, Swanton C, Albiges L, Schmidinger M, Heng Dy, Larkin J, and Ficarra V. 2017.
345   Renal cell carcinoma. *Nature reviews Disease primers* 3:17009. DOI 10.1038/nrdp.2017.9

346   Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird Pw,
347   Levine Da, Carter Sl, Getz G, Stemke-Hale K, Mills Gb, and Verhaak Rg. 2013. Inferring tumour purity
348   and stromal and immune cell admixture from expression data. *Nature Communications* 4:2612. DOI
349   10.1038/ncomms3612

350   Scott Lj, Burness Cb, and McCormack Pl. 2012. Belimumab: a guide to its use in systemic lupus erythematosus.
351   *BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy* 26:195-199. DOI
352   10.2165/11209060-000000000-00000

353   Steri M, Orrù V, Idda Ml, Pitzalis M, Pala M, Zara I, Sidore C, Faà V, Floris M, Deiana M, Asunis I, Porcu E,
354   Mulas A, Piras Mg, Lobina M, Lai S, Marongiu M, Serra V, Marongiu M, Sole G, Busonero F, Maschio A,
355   Cusano R, Cuccuru G, Deidda F, Poddie F, Farina G, Dei M, Virdis F, Olla S, Satta Ma, Pani M, Delitala A,
356   Cocco E, Frau J, Coghe G, Lorefice L, Fenu G, Ferrigno P, Ban M, Barizzone N, Leone M, Guerini Fr,
357   Piga M, Firinu D, Kockum I, Lima Bomfim I, Olsson T, Alfredsson L, Suarez A, Carreira Pe, Castillo-
358   Palma Mj, Marcus Jh, Congia M, Angius A, Melis M, Gonzalez A, Alarcón Riquelme Me, da Silva Bm,
359   Marchini M, Danieli Mg, Del Giacco S, Mathieu A, Pani A, Montgomery Sb, Rosati G, Hillert J, Sawcer S,
360   D'Alfonso S, Todd Ja, Novembre J, Abecasis Gr, Whalen Mb, Marrosu Mg, Meloni A, Sanna S, Gorospe
361   M, Schlessinger D, Fiorillo E, Zoledziewska M, and Cucca F. 2017. Overexpression of the Cytokine BAFF
362   and Autoimmunity Risk. *The New England journal of medicine* 376:1615-1626. DOI
363   10.1056/NEJMoa1610528

364   Ritchie Me, Phipson B, Wu D, Hu Y, Law Cw, Shi W, and Smyth Gk. 2015. limma powers differential expression

365   analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43:e47. DOI
366   10.1093/nar/gkv007

367   Cairns P. 2010. Renal cell carcinoma. *Cancer biomarkers : section A of Disease markers* 9:461-473. DOI
368   10.3233/CBM-2011-0176

369   Shannon P, Markiel A, Ozier O, Baliga Ns, Wang Jt, Ramage D, Amin N, Schwikowski B, and Ideker T. 2003.
370   Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome*
371   *Research* 13:2498-2504. DOI 10.1101/gr.1239303

372   Siegel Rl, Miller Kd, and Jemal A. 2017. Cancer Statistics, 2017. *CA: A Cancer Journal for Clinicians* 67:7-30. DOI
373   10.3322/caac.21387

374   Chevrier S, Levine Jh, Zanotelli Vrt, Silina K, Schulz D, Bacac M, Ries Ch, Ailles L, Jewett Mas, Moch H, van den
375   Broek M, Beisel C, Stadler Mb, Gedye C, Reis B, Pe'er D, and Bodenmiller B. 2017. An Immune Atlas of
376   Clear Cell Renal Cell Carcinoma. *Cell* 169:736-749.e718. DOI 10.1016/j.cell.2017.04.016

377   Yang S, Liu T, Nan H, Wang Y, Chen H, Zhang X, Zhang Y, Shen B, Qian P, Xu S, Sui J, and Liang G. 2019.
378   Comprehensive analysis of prognostic immune-related genes in the tumor microenvironment of cutaneous
379   melanoma. *Journal of Cellular Physiology*. DOI 10.1002/jcp.29018

380   Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette,
381   Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov. 2005. Gene set
382   enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.
383   *Proceedings of the National Academy of Sciences* 102:15545-15550. DOI 10.1073/pnas.0506580102

384   Li T, Fan J, Wang B, Traugh N, Chen Q, Liu Js, Li B, and Liu Xs. 2017. TIMER: A Web Server for Comprehensive
385   Analysis of Tumor-Infiltrating Immune Cells. *Cancer Research* 77:e108-e110. DOI 10.1158/0008-
386   5472.can-17-0307

387   Capitanio U, and Montorsi F. 2016. Renal cancer. *Lancet (London, England)* 387:894-906. DOI 10.1016/S0140-
388   6736(15)00046-X

389   Cremasco V, Woodruff Mc, Onder L, Cupovic J, Nieves-Bonilla Jm, Schildberg Fa, Chang J, Cremasco F, Harvey
390   Cj, Wucherpfennig K, Ludewig B, Carroll Mc, and Turley Sj. 2014. B cell homeostasis and follicle
391   confines are governed by fibroblastic reticular cells. *Nature Immunology* 15:973-981. DOI 10.1038/ni.2965

392   Pelekanou V, Notas G, Theodoropoulou K, Kampa M, Takos D, Alexaki Vi, Radojicic J, Sofras F, Tsapis A,
393   Stathopoulos En, and Castanas E. 2011. Detection of the TNFSF members BAFF, APRIL, TWEAK and
394   their receptors in normal kidney and renal cell carcinomas. *Analytical cellular pathology (Amsterdam)*

395   34:49-60. DOI 10.3233/ACP-2011-0001

396   Li S, Hou Jq, Wang Z, Chang Jk, Xu Wb. 2019. TNFSF13B correlated with poor prognosis and immune

397   infiltrates in KIRC: a study based on TCGA data. International Journal of Clinical and Experimental
398   Medicine. DOI 12(10):12057-12067.

399

Figure 1

Work flow used in this analysis

The flow chart reflects the research idea of this study.
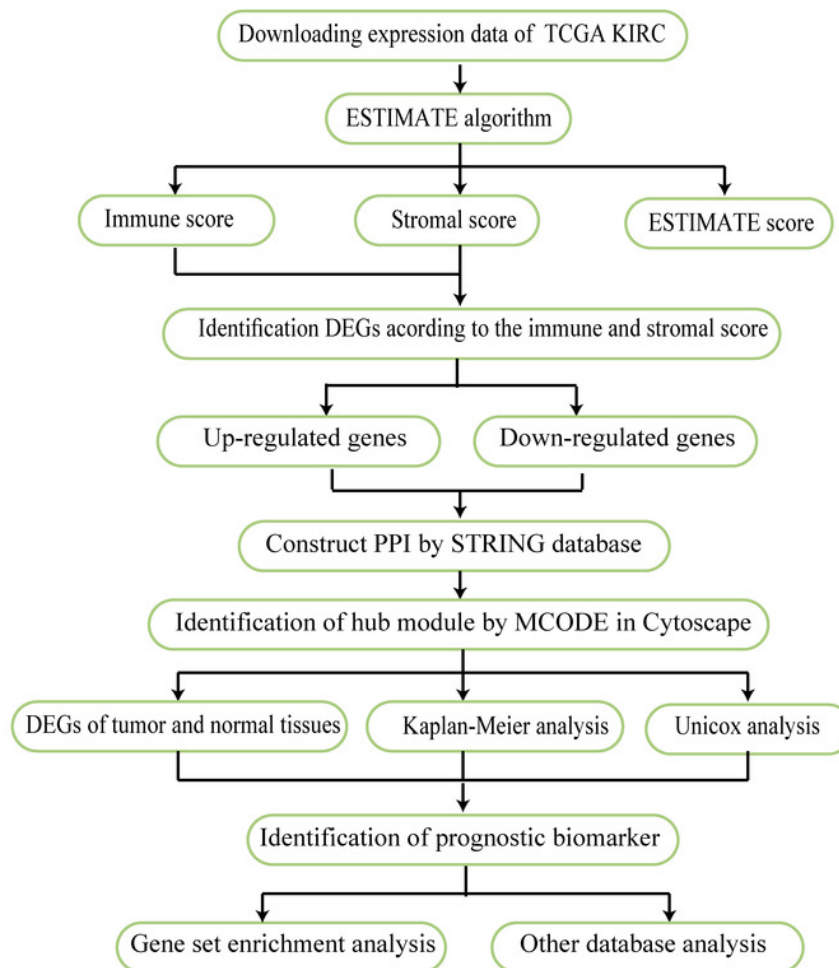
Figure 2

The scores obtained from the ESTIMATE algorithm were related to the KIRC overall survival rate and stage.

(A) Kaplan-Meier analysis of immune score. A value of p< 0.05 means it is statistically significant. CI: confidence interval (B) Stromal score. (C) ESTIMATE score. (D) Box diagram of immune score in different stages. Kruskal-Wallis test was used for comparison among groups. A value of p<0.05 indicates that it is statistically significant. (E) Stromal score. (F) ESTIMATE score.
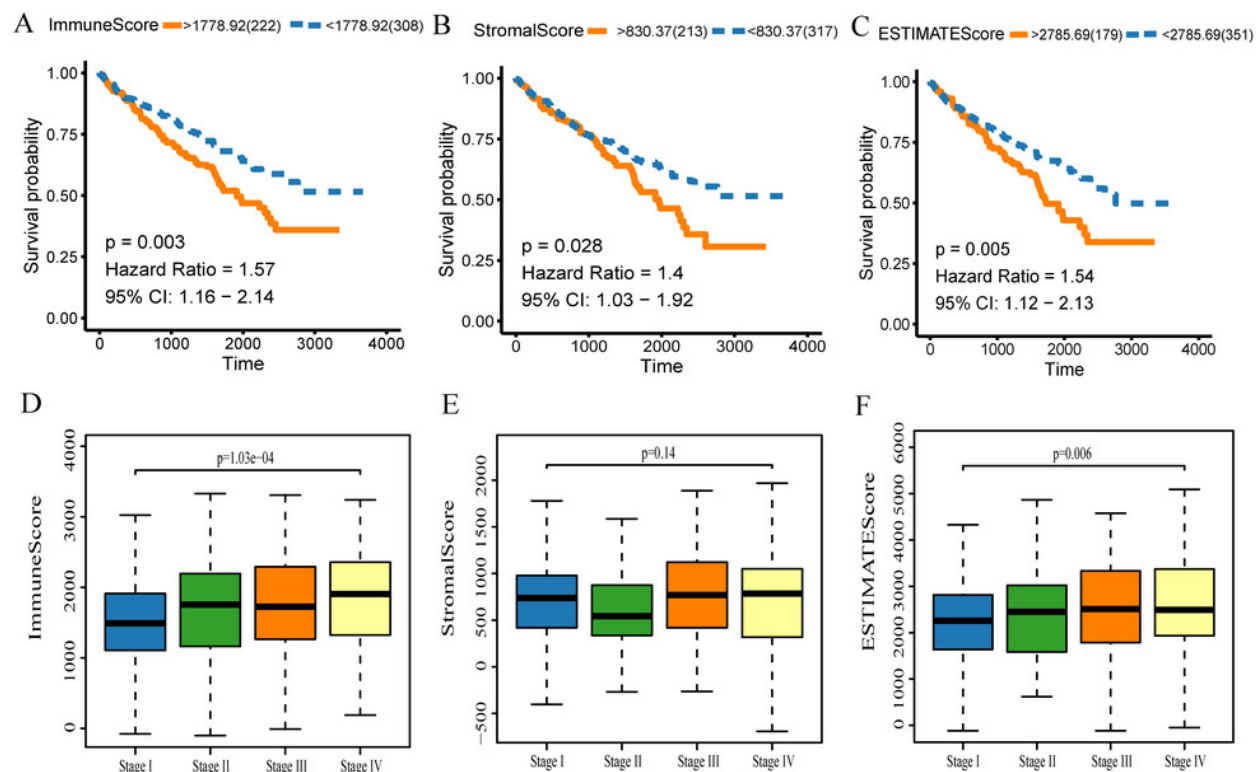
Figure 3

Comparison of expression profiles for genes with low or high immune and stromal scores.

(A) The heatmap of DEGs between high and low immune scores. (B)The heatmap of DEGs between high and low stromal scores. (C) Venn diagram shows the number of up-regulated DEGs. (D) Venn diagram shows the number of down-regulated DEGs. (E) A bubble chart shows the top 20 enriched GO terms. The GO analysis had an FDR of less than 0. 05. (F) The KEGG enrichment analysis exhibited an FDR of the six pathways that were less than 0.05.
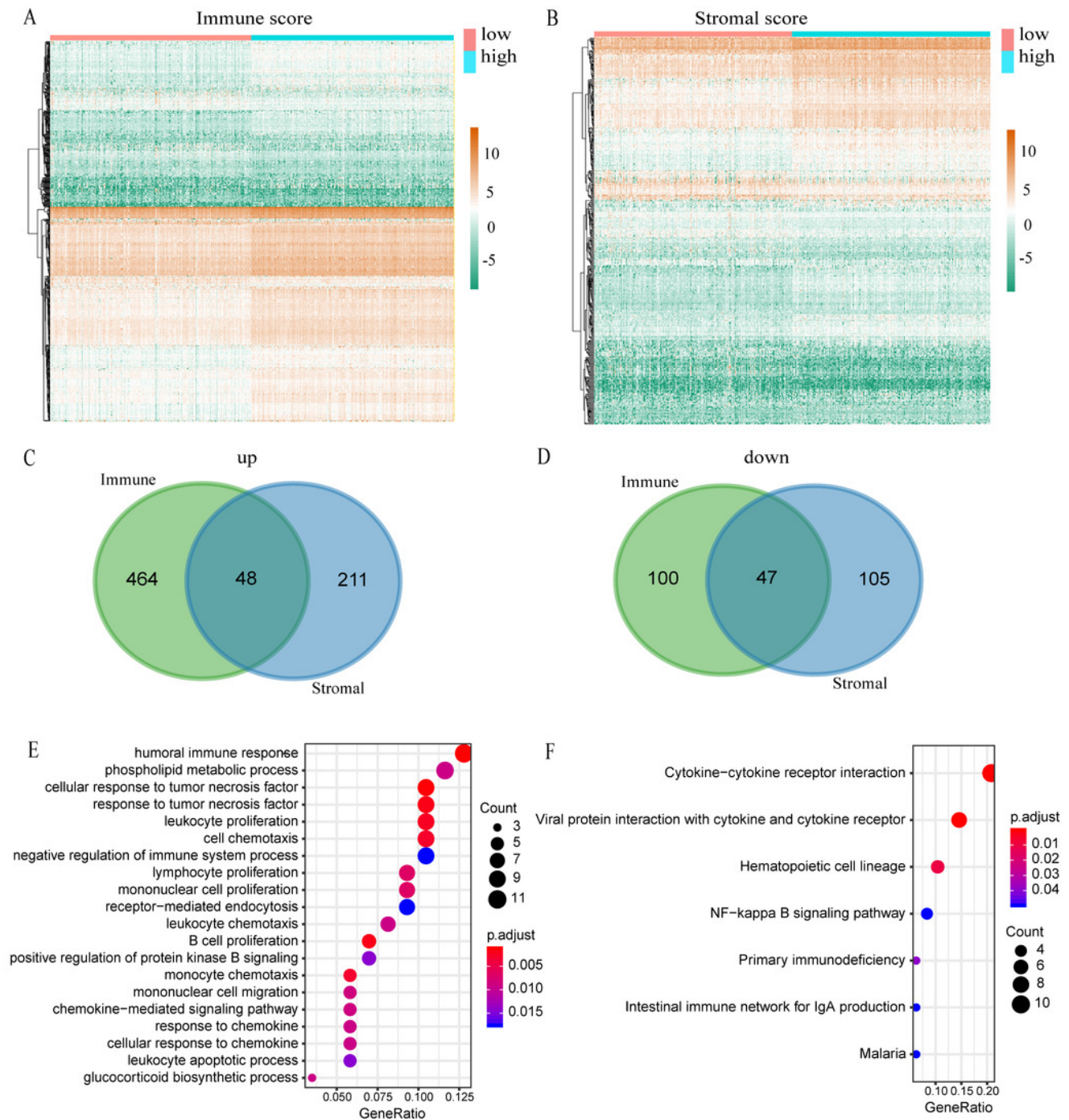
Figure 4

Construction of PPI network and identification of clusters.

(A) The PPI network. (B) Cluster 1. (C) Cluster 2. (D) Cluster 3. (E) Cluster 4.
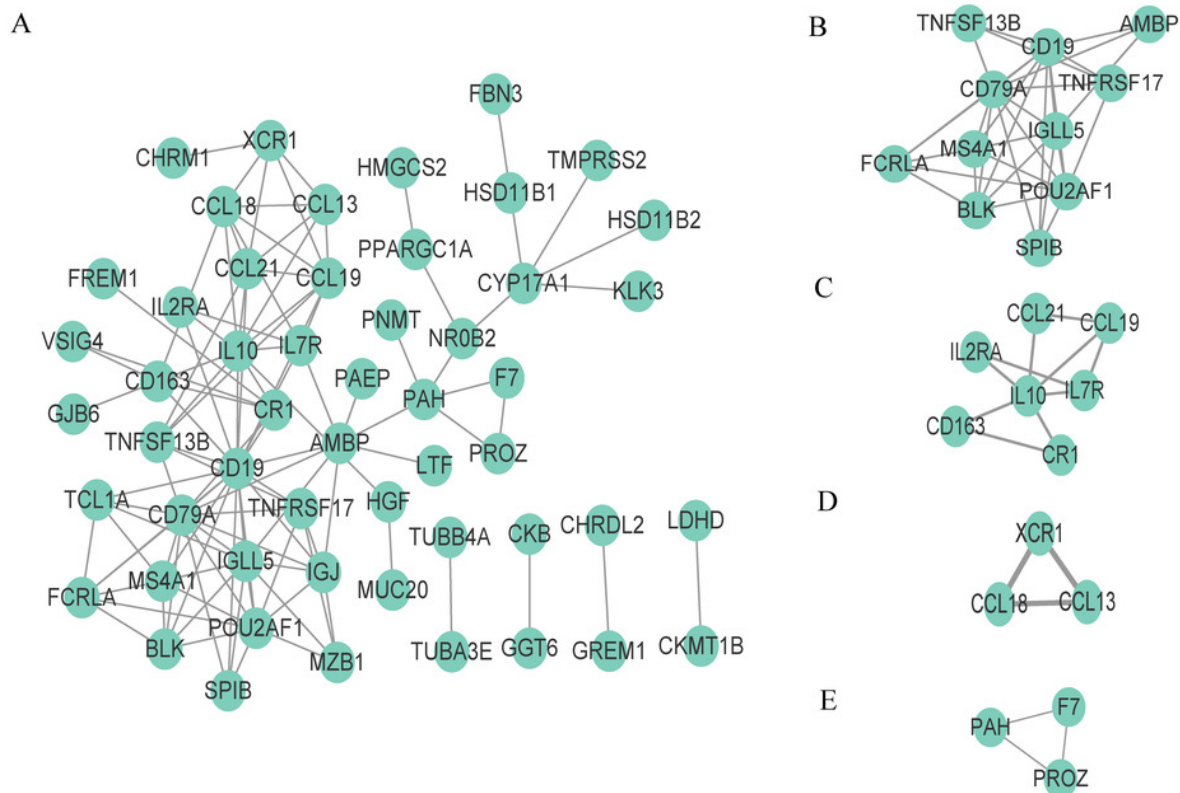
Figure 5

Venn analysis of results from three analyses.

Blue circles indicate differentially expressed genes of the tumor and normal tissue (|logFC>1|, P<0.05). Red circles indicate genes with significant results from univariate analysis (p<0.05). Green circles mean genes with significant results from Kaplan-Meier analysis (p<0.05).
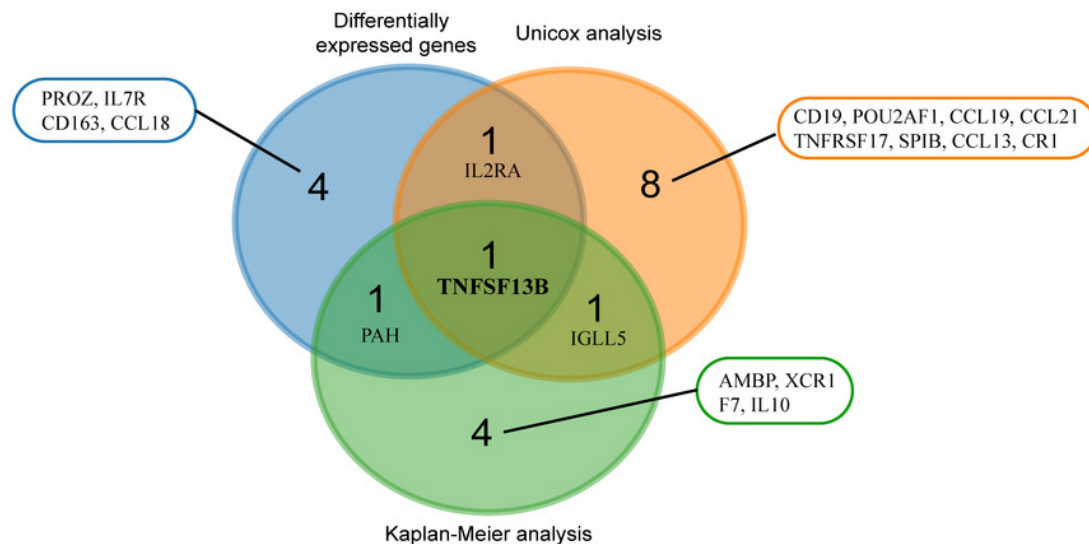
Figure 6

Analysis of the relationship between TNFS13B and clinical factors.

(A) Scatter plot of TNFS13B expression in tumor and normal tissues. Wilcoxon test was used to compare the two groups. (B) Kaplan-Meier analysis of TNFS13B. (C) Forest plot of univariate COX regression analysis for clinical factors. (D) Forest plot of multivariate COX regression for clinical factors. (E-D) Box diagram of TNFSF13B expression under different grades and stages. Kruskal-Wallis test was used for comparison among groups. A value of $p<0.05$ indicates that it is statistically significant.
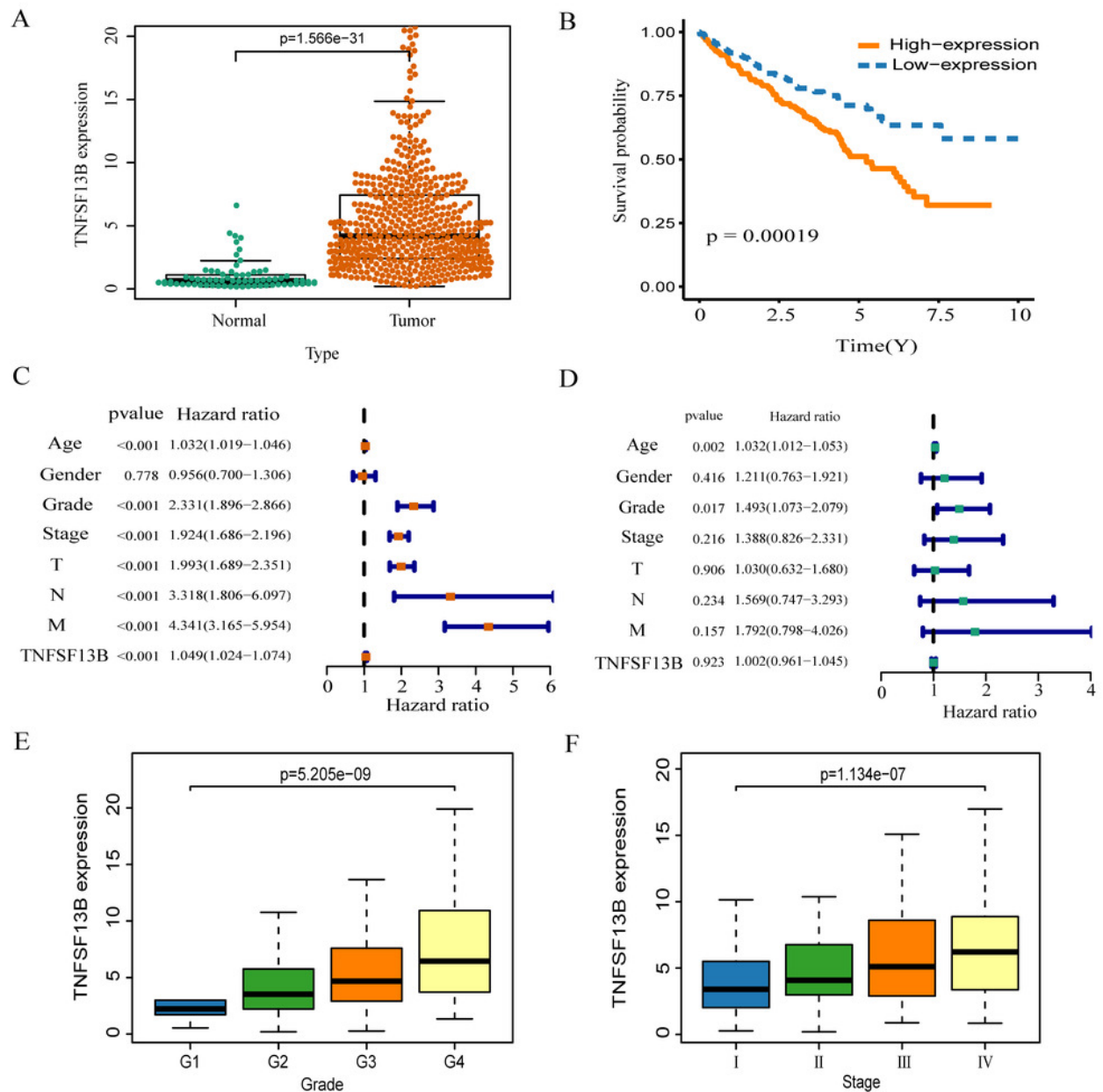
Figure 7

Tumor microenvironment exploration of TNFSF13B.

(A) Correlation scatters diagram of the immune score and TNFSF13B expression. (B) Stromal score. (C) ESTIMATE score. (D) Correlation scatters Diagram of Tumor purity and TNFSF13B expression. (E) B cell. (F) CD8+ T cell. (G) CD4+ T cell. (H) Macrophage. (I) Neutrophil. (J) Dendritic cell. A value of p<0.05 indicates that it is statistically.
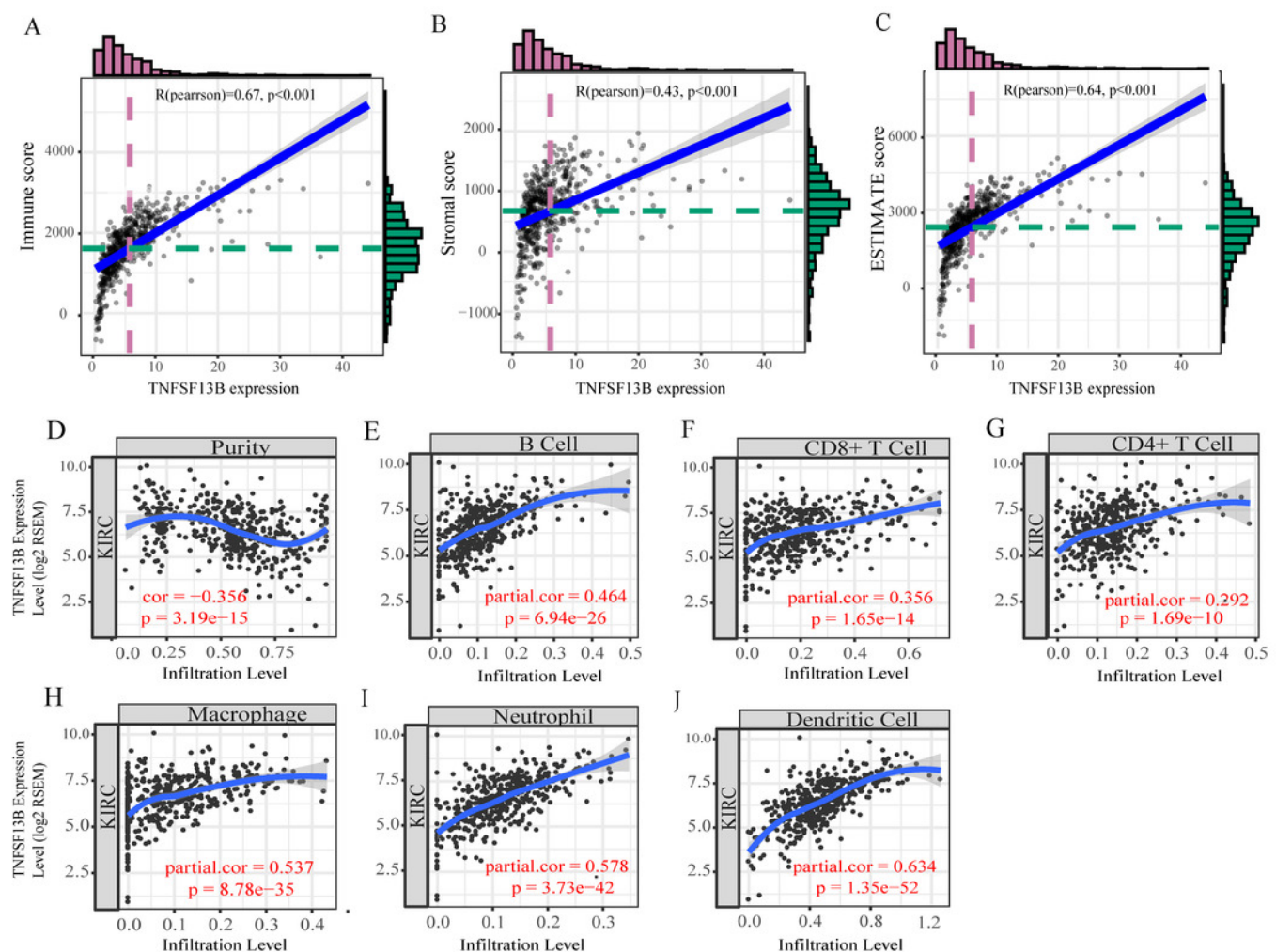
Figure 8

Gene set enrichment analysis and drug target network.

(A) A broken line above represents an enrichment pathway, and the straight line below represents every gene in the pathway. (B) Network diagram of drugs and acting genes.