

# Identification of key survival genes of the tumor microenvironment and immune infiltration in head and neck squamous cell carcinoma

**Min Wang**

Department of Otorhinolaryngology, The First Affiliated Hospital of Chongqing Medical University

**Tao Lu**

Department of Otorhinolaryngology, The First Affiliated Hospital of Chongqing Medical University

**Yanshi Li**

Department of Otorhinolaryngology, The First Affiliated Hospital of Chongqing Medical University

**Min Pan**

Department of Otorhinolaryngology, The First Affiliated Hospital of Chongqing Medical University

**Zhihai Wang**

Department of Otorhinolaryngology, The First Affiliated Hospital of Chongqing Medical University

**Guohua Hu** (✉ [hghcqmu@163.com](mailto:hghcqmu@163.com))

Department of Otorhinolaryngology, The First Affiliated Hospital of Chongqing Medical University

---

## Research Article

**Keywords:** Head and neck cancer, Tumor microenvironment, Immune scores, Bioinformatics analysis, Survival

**Posted Date:** March 13th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-283947/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** Head and neck cancer (HNC) are highly aggressive solid tumors with poor prognoses. The tumor microenvironment (TME) plays a critical role in angiogenesis, invasion, and metastasis of HNC. In the TME, immune and stromal cells influence tumor initiation, response, and therapy. Our study aimed to evaluate the progression and prognosis of HNC by analyzing the key genes involved in immunization and stromal cells.

**Methods:** Gene expression profiles, demographics, and survival data were downloaded from the TCGA database. Patients with HNC were divided into high immune/stromal score groups or low immune/stromal score groups based on the ESTIMATE algorithm. Differentially expressed genes (DEGs) were identified via functional enrichment analysis and protein-protein interaction networks, and survival analysis based on DEGs was also performed.

**Results:** A total of 522 patients with HNC were enrolled for analysis. The average age was 60.87, and one-third of the patients were HPV-positive. Kaplan-Meier survival analysis showed that patients' median survival time in the low-score group was shorter than that of the high-score group (625 vs. 680 days, log-rank,  $p = 0.1716$ ). According to immune scores, 925 genes were upregulated, and 72 genes were downregulated in the high-score group compared with the low-score group. Top Gene Ontology terms identified that T-cell costimulation, regulation of immune response, and the external side of the plasma membrane were the most involved pathways. Moreover, Kaplan-Meier analysis revealed that 480 DEGs were upregulated in the high-immune scores group, and a total of 126 DEGs were significantly associated with poor survival. Besides, we identified the hub genes of DEGs through protein-protein interactions and found that PTPRC, CD247, and CD4 are associated with immune infiltration and all-cause mortality.

**Conclusions:** We identified a series of TME-related genes significantly associated with overall mortality; this information is crucial for further understanding the role of TME and immune infiltration in the prognosis of HNC.

## Background

Cancer, a major public health burden, continues to increase in incidence and is predicted to be the leading cause of death worldwide in the 21st century <sup>1</sup>. Among all the cancer types, head and neck cancer (HNC) is a highly aggressive solid tumor with the 6th highest mortality rate. It is estimated that about 90% of HNC is head and neck squamous cell carcinomas (HNSCC). Smoking and excessive alcohol consumption are regarded as traditional risk factors for HNC. In recent years, HPV-induced HNC, particularly in oropharyngeal tumors, has drawn much attention <sup>2</sup>. Due to limited effective therapeutic options for HNC patients <sup>3</sup>, the 5-year mortality rate of HNC patients is around 50% <sup>4</sup>. Thus, developing novel therapeutic targets and strategies to guide first-line treatments is urgently required <sup>5</sup>.

The tumor microenvironment (TME) consists of a mass of heterogeneous cell types that infiltrate tumors and directly or indirectly interact with tumor cells<sup>6</sup>; these components play a critical role in angiogenesis, invasion, metastasis, and chronic inflammation<sup>7</sup>. Animal experiments have shown that HNC is an immunosuppressive disease; the main characteristics are the abnormal secretion of inflammatory cytokines as well as dysfunction of immune effector cells<sup>8</sup>.

Both innate and adaptive immune cells play essential roles in regulating immune surveillance and controlling tumor growth<sup>9</sup>, influencing tumor initiation, response, and therapy<sup>10</sup>. Elevated immune cell infiltration, including T- and B-lymphocytes, suggests they have prognostic relevance<sup>11</sup>. Thus, the characteristics of TME enable us to identify new prognostic and predictive biomarkers. The immune score is an example of how elements of the immune microenvironment can be quantified in a standardized way. It may serve as a prognostic index during cancer treatment<sup>5</sup>.

Gene profiling has entered the era of high-throughput methods with mega data. This provides opportunities to understand the cellular actions and functions precisely<sup>12</sup>. Implementation of these methods for HNC and other cancers in the context of precision oncology connects genetic information to cancer care<sup>13</sup>. Potential prognostic biomarkers in HNC can be identified based on analysis of microRNA expression profiles in the TCGA database<sup>14</sup>. For example, the T follicular helper signature was associated with favorable survival in HNC patients<sup>15,16</sup>.

In our study, using the Head and Neck Cancer (HNC) cohorts from The Cancer Genome Atlas (TCGA) database and taking advantage of ESTIMATE algorithm-derived immune scores, we identified a series of differentially expressed genes associated with TME, which could predict prognosis and survival of HNC patients.

## Methods

### Database

Level 3 data were obtained from The Cancer Genome Atlas (TCGA) database (<https://xenabrowser.net/datapages/>). The Illumina HiSeq 2000 RNA sequencing platform was used to analyze the gene expression profile of HNC, which was performed at the TCGA genome characterization center of the University of North Carolina. Demographic information, including age, gender, pathological type, and survival information, was also downloaded from the TCGA data portal. Immune scores and stromal scores of HNC were calculated by the Estimation of Stromal and Immune cells in malignant tumor tissues using Expression data (ESTIMATE) (<https://bioinformatics.mdanderson.org/public-software/estimate/>) tool<sup>17</sup>.

### Identification of differentially expressed genes (DEGs)

We used high versus low immune scores/stromal scores to distinguish gene expression profiles and used R software's limma package to analyze gene expression<sup>18</sup>. The screening criteria of the differentially

expressed genes (DEGs) defined as the gene expression was over 2-fold change and adjusted p-value < 0.05.

### **Heatmaps and clustering analysis**

We used Morpheus (<https://software.broadinstitute.org/morpheus>), an open-source, online versatile matrix visualization and analysis software, to generate the heatmaps and clustering. While we used the ggplot2 package of R 3.6.1 to draw the volcano plots.

### **Protein-protein interaction network**

The protein-protein interaction (PPI) network was generated by STRING (version 11.0) (<https://string-db.org>), and the combined score  $\geq 0.4$  will be considered as a significant interaction. Images were reconstructed using Cytoscape software (version 3.7.2), which was used for visualizing molecular interaction networks. Individual networks  $\geq$  eight nodes will be selected for further analysis. In addition, we also calculated the connectivity degree of each node in the network. The molecular complex detection (MCODE) app was then utilized for cluster finding based on the topology to locate densely connected regions. PPI networks were drawn using Cytoscape.

### **Kaplan-Meier survival analysis**

A Kaplan–Meier survival analysis was performed to investigate the relationship between the overall survival of patients with HNC and the DEGs. Gene expression level was divided into high or low expression. All statistical analyses were performed using GraphPad Prism (version 8.0.1).

### **GO and KEGG enrichment analysis of DEGs**

The above-screened DEGs were selected for functional enrichment analysis by The Database for Annotation, Visualization and Integrated Discovery (DAVID, version 6.8), Gene Ontology (GO) categories were identified through biological processes (BP), molecular functions (MF), and cellular components (CC). In addition, the DAVID database was also used for the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. The cut-off point was defined as a false discovery rate (FDR) < 0.05.

## **Results**

### **Basic information**

Finally, a total of 522 patients with HNC were included for analysis. The basic character of the included patients was summarized in Table 1. The mean age was 60.87 years (range 18–90), approximately two-thirds of the patients were male, and about 66.67% and 8.93% of the patients had histories of drinking and smoking, respectively. Approximately one-third of the patients were HPV-positive. Pathological tumor node metastasis staging results were as follows: 33 (6.32%) cases in stage I, 98 (18.77%) cases in stage II, 105 (20.11%) cases in stage III, and 286 (54.79%) cases in stage IV. According to the ESTIMATE algorithm, stromal and immune scores were  $-436.34$  (interquartile range (IRQ),  $-904.04-136.41$ ), and  $361.16$  (IQR,  $-160.20-1038.52$ ), respectively. Average immune scores were higher than stromal scores in

all pathology stages ( $p < 0.05$ ) (Fig. 1). A previous study has shown that TP53 mutation was associated with decreased overall survival in patients with HNC<sup>19</sup>. We then calculated immune scores and stromal scores in HNC patients with or without *p53* mutation. Compared with *p53* wildtype patients, the immune score in *p53* mutation patients was significantly higher ( $p < 0.001$ ), while no significant difference was noticed in stromal scores between these two groups of patients (Fig. 2A and 2B). Moreover, to determine the relationship between immune scores/stromal scores and all-cause mortality, patients with HNC were classified into high immune/stromal scores and low immune/stromal score groups based on the median value of immune scores or stromal scores. There was no significant difference in age, sex, HPV infection, or tumor pathology stage between the two groups (Table 1). Kaplan-Meier survival analysis revealed that the median overall survival time of patients in the immune low-score group was shorter than in the high-score group (Fig. 2C, 625 vs. 680 days, log-rank test,  $p = 0.1716$ ). While compared to patients with higher stromal scores, patients with lower stromal scores also had shorter median overall survival times (Fig. 2D, 623 vs. 701 days, log-rank test,  $p = 0.8528$ ).

Table 1  
The basic information of the included head and neck cancer

	Total patients (n = 522)	Low immune score group (n = 261)	High immune score group (n = 261)	P value
Age	60.87(19–90)	60.3(19–88)	61.2(24–90)	0.776
Sex (male, %)	385(73.7%)	202(77.39%)	183(70.1%)	0.059
Alcohol	348(66.67%)	173(66.28%)	174(66.67%)	0.926
Smoking	20/224(8.93%)	11/116(9.48)	9/108(7.63%)	0.763
HPV	39/112(34.82%)	12/46(26.09%)	27/66(40.90%)	0.105
Stage				0.294
Stage I	35(6.71%)	21(4.02%)	14(2.68%)	
Stage II	98(18.77%)	48(9.20%)	50(9.58%)	
Stage III	105(20.11%)	58(11.11%)	47(9.00%)	
Stage IV	286(54.79%)	134(25.67%)	152(29.12%)	

### Association between gene expression and ESTIMATE scores

To further understand the relationship between the gene expression and immune scores and/or stromal scores, we generated distinct gene expression profiles of HNC based on high vs. low immune/stromal score groups. The heat maps in Fig. 3A and 3B show the expression patterns of genes differentially

changed in the immune and stromal score groups, respectively. Differentially expressed genes (DEGs) were then identified from the comparison of high vs. low immune score and stromal score groups. According to the immune scores, a total of 925 genes were upregulated, while a total of 72 genes were downregulated ( $FC > 2$ ,  $p < 0.05$ ). For the stromal scores, there were about 1414 genes that were upregulated, and 26 genes were downregulated ( $FC > 2$ ,  $p < 0.05$ ). The Venn diagrams were then generated and indicated that 425 genes were commonly upregulated, while only 2 genes were commonly downregulated in the high-scoring groups (Fig. 3C and 3D). A volcano plot of the acquired data highlights that more genes are upregulated (right side), and fewer genes are downregulated (left side) (Fig. 4A and 4B). To describe the potential function of the identified DEGs, the commonly upregulated genes (The genes are listed in Supplementary Table S1) in the high-immune scores group were selected for Gene Ontology (GO) enrichment analysis. The results indicated that the top GO terms associated with these genes included T cell costimulation, regulation of immune response, external side of the plasma membrane, integral component of the plasma membrane, transmembrane signaling receptor activity, receptor activity for biological processes (Fig. 3E), cellular component (Fig. 3F), and molecular function (Fig. 3G).

### **Association between DEGs and overall mortality**

Kaplan-Meier survival curves were then obtained to further analyze the association between the individual DEGs and all-cause mortality. During an average of 1.77 years (IQR 1.04–3.23) follow-up, a total of 222 patients died. Kaplan-Meier survival analysis showed that from the 480 upregulated DEGs in the high-immune scores group, about 176 DEGs (Supplementary Table S2) were significantly associated with all-cause mortality ( $p < 0.05$ ; representative genes are shown in Fig. 5).

### **Protein-protein interactions**

To further search the hub genes of DEGs, we used the STRING tool to construct protein-protein interaction (PPI) networks and the MCODE app to identify gene clusters, respectively. The constructed PPI networks and the top five clusters are shown in Figs. 6 and 7, respectively. Then, we named these modules PTPRC, CD247, CD4, FCGR3A, and FCGR1A (Table 2) for ease of description. In the PTPRC module, the network forms 461 edges, involving 36 nodes; CTLA4, ITGAL, CD40LG, IL10RA, CD28, and SLAMF1 were the prominent nodes, which have the most connections with other members of the module. In the CD247 module, 26 nodes and 124 edges were observed. Among the clusters, CD247, CD3D, IKZF1, CD3G, PLEK, and HLA-DRB1 had higher degree values. For the CD4 module, CD4, IFNG, CXCR3, CD48, CD5, and CXCR5 had higher degree values. In addition, we performed a functional enrichment analysis for the five hub genes, and the results showed that 19 GO terms were significantly enriched in the biological process (BP), 7 GO terms were significantly enriched in the cellular compartment (CC), and 4 GO terms were significantly enriched in molecular function (MF) ( $FDR < 0.05$ ,  $-\log FDR > 1.30$ ). Consistent with previous results, most of the genes are mainly involved in immune function. The top GO terms included T cell costimulation, regulation of immune response (Fig. 8A), external side of the plasma membrane, integral component of the plasma membrane (Fig. 8B), transmembrane signaling receptor activity, and receptor activity (Fig. 8C). Besides, the KEGG pathways analysis was also performed (Fig. 8D). In addition, A

circos graph was constructed to visualize the prognosis-related genes involving in GO analysis according to the logFC value (Fig. 9).

Table 2  
Functional roles of five hub genes with degree  $\geq 8$

Cluster	Nodes	Edges	Gene
1	36	461	SLAMF1, CD28, CD40LG, IL10RA, ITGAL, CTLA4, IL2RG, TLR7, TLR8, VCAM1, PDCD1, SPN, PRF1, GZMB, CD3E, CCL5, CCR5, KLRK1, CXCL10, CD8A, GZMA, CXCL9, LAG3, SELL, TBX21, CXCL13, KLRD1, CD52, CCL19, IDO1, CD2, KLRB1, CCR2, FASLG, PTPRC, CD27
2	26	124	TRAT1, PLEK, MPEP1, CYBB, HLA-DRB1, C1QA, C1QB, HLA-DMB, HLA-DMA, SIGLEC1, CD3D, HLA-DRB5, HLA-DQB1, HLA-DQA1, CD247, HLA-DPB1, HLA-DPA1, CD3G, IKZF1, GRAP2, GZMH, NKG7, HLA-DOA, HLA-DQA2,
3	24	82	SH2D1A, GBP5, ICOS, HCST, CD5, IFNG, GBP4, CIITA, CXCR3, FCGR1B, TNFSF13B, CD8B, GZMK, CXCL11, BTLA, CXCR5, CXCR6, IL21R, CD226, TIGIT, CD4, CD7, ITK, CD48
4	15	40	FCGR3A, EVI2B, MS4A6A, SASH3, DOCK2, NCKAP1L, CD84, FGL2, WDFY4, MND4, GPR174, P2RY10, CD53, AOA4, CD74
5	9	16	CD163, CLEC10A, CD37, FCGR1A, IGSF6, IRF8, HLA-DRA, MS4A1, C1QC

### The relationship between the hub genes and immune infiltration and survival

To further analyze the association of the five hub genes with immune infiltration. We found that all the 5 hub genes are significantly associated with tumor purity, B cell, CD8<sup>+</sup> T cell, CD4<sup>+</sup> T cell, macrophage, neutrophil, and dendritic cell (Fig. 10). Kaplan-Meier survival analysis showed that low PTPRC, CD247, and CD4 groups had poor overall survival except for FCGR3A and FCGR1A (Fig. 11). Also, multivariate Cox regression analysis further confirmed that PTPRC, CD247, and CD4 are associated with all-cause mortality (Table 3).

Table 3  
Univariate and multivariate cox regression analysis of five hub genes

Gene	Overall survival			
	Univariate analysis		Multivariate analysis	
	HR (95% CI)	P Value	HR (95% CI)	P Value
PTPRC	0.884(0.809–0.996)	0.006	0.854(0.761–0.958)	0.007
CD247	0.809(0.72–0.909)	0.001	0.791(0.681–0.919)	0.002
CD4	0.903(0.818–0.996)	0.041	0.861(0.760–0.976)	0.019
FCGR3A	1.017(0.932–1.11)	0.703	0.979(0.882–1.087)	0.694
FCGR1A	0.997(0.879–1.13)	0.959	0.947(0.815-1.100)	0.474

## Discussion

In our study, we try to investigate the relationship between TME-related genes and overall survival in patients with HNC using the TCGA database. Notably, we found that several genes were significantly associated with overall mortality in patients with HNC. Thus, we identified immune-related overexpressed genes related to a poor outcome, which is relevant to the personalized treatment of HNC patients.

The comprehensive molecular catalog on HNSCC was firstly described by the TCGA consortium in 2015<sup>20</sup>, providing oncologists and clinicians with a comprehensive understanding of the pathogenesis of HNC. Data mining the TCGA database can reveal the effects of gene expression on tumor resistance, treatment, and prognosis, guiding us into the era of precision cancer medicine<sup>21</sup>. Using this information, optimal treatment strategies can be developed for these patients<sup>22</sup>. A previous study indicated that tumors could be further subclassified based on genomic profiling<sup>2</sup>. Researchers using a data mining strategy have identified genes that can predict the prognosis of HNC<sup>23</sup>. A subgroup of genetically distinct HPV-negative HNC with favorable prognoses was confirmed<sup>2</sup>; however, the role and prognostic value of TME-related genes had not been fully elucidated. In oral cancer nodal disease (OCND), immune-related gene expression has prognostic value with regard to node-free survival (NFS)<sup>24</sup>. Lecerf *et al.* retrospectively evaluated the expression of 46 immune-related genes and immune-cell subpopulation genes among 96 patients with HNSCC and concluded that overexpression of OX40L and PDGFRB was associated with poor outcomes<sup>25</sup>. Meanwhile, in patients with oral squamous cell carcinoma (OSCC), FAP, IFI27, LAMC2, MMP1, SPINK5, and ZNF662 genes can be regarded as valuable therapeutic biomarkers associated with the development and progression of OSCC<sup>22</sup>.

A recent study has demonstrated that based on cellular and molecular characteristics of the TME, there are two broad categories of tumors that escape from the immune system,<sup>26</sup>. Among these, the tumor

stroma is an active player in a variety of cancers<sup>27</sup>. A previous study confirmed that immune cells in the TME fail to exert anti-tumor effects and promote tumor growth<sup>28</sup>. The adaptive immune response is involved in tumor development through immune interactions in the TME, including with T lymphocytes, macrophages, and dendritic cells<sup>29</sup>. Increasing evidence shows that the TME of HNC has a highly immunosuppressive effect, mediated by soluble and cell-related inhibitory mediators and recruitment of host immunosuppressive cells<sup>30</sup>. However, not all gene-related immune cells have a strong prognostic value. For example, the PD-L1/CD8A/CYT subsets were unable to distinguish between the number of mutations and neoantigens in the glioblastoma multiforme, prostate adenocarcinoma, or head, neck, and lung squamous cell carcinomas<sup>31</sup>. However, in our study comparing high vs. low immune/stromal scores, we identified several DEGs associated with overall mortality. This method is quite different from previous literature reports, and to our knowledge, is the first to implement this strategy in HNC.

PPI mapping provides a valuable framework to facilitate understanding of the functional organization of the proteome<sup>32</sup>. In our study, using the PPI network, we found five clusters of genes that were highly upregulated in the TME and all significantly associated with immune infiltration. The most evident is PTPRC, also known as CD45. It has been reported that CD45 plays a critical role in the initiation of T cell receptor signaling via activation of the Src family protein-tyrosine kinases Lck and Fyn<sup>33</sup>, and is highly expressed in lymphocytes. A population study has revealed that patients with high CD45 expression had better local recurrence-free survival and disease-specific survival<sup>34</sup>. Another important module is CD247, which is mostly expressed in normal lymphoid tissue and peripheral blood mononuclear cells (PBMCs)<sup>35</sup>. The protein encoded by this gene is the T-cell receptor zeta, which combines with other T-cell receptor subunits and CD3 subunits to form the T-cell receptor-CD3 complex. This complex that is present on the T-lymphocyte cell surface plays an essential role in the adaptive immune response. Abnormal expression of CD247 is related to differentiation and classification in ovarian cancer<sup>36</sup>. Previous research has shown that CD247 is related to papillary thyroid carcinoma (PTC) clinical phenotypes in the TCGA database<sup>37</sup>. CD247 is also part of the TCR signal-triggering module<sup>38</sup>. However, the role of CD247 in HNC has not been widely studied. In our analysis, we found that CD247 level is associated with overall mortality in patients with HNC. This may provide an additional prognostic marker for HNC. Further investigation of these genes will enable us to gain a new and more comprehensive understanding of the potential relationship between TME and the prognosis of HNC.

Some limitations of this study should be mentioned. First, we did not verify whether the genes identified from the TCGA database could predict poor survival in HNC patients from other databases. Second, we only performed bioinformatics analysis from the open online database, while the biological functions of these genes were not further studied.

## Conclusions

In summary, we performed a bioinformatics analysis for HNC, based on immune scores, using the TCGA database. We extracted a list of TME-related genes significantly associated with survival in patients with

HNC; although some genes have already been reported, the remaining genes may be regarded as additional prognostic biomarkers for HNC.

## Abbreviations

HNC: Head and neck cancer;

TME: The tumor microenvironment;

DEGs: Differentially expressed genes;

HNSCC: Head and Neck squamous cell carcinoma;

TCGA: The Cancer Genome Atlas;

GDC: Genomic Data Commons;

PPI: Protein-Protein Interaction;

MCODE: Molecular complex detection;

DAVID: Database for Annotation, Visualization and Integrated Discovery;

BP: Biological processes;

MF: Molecular functions;

CC: Cellular components;

KEGG: Kyoto Encyclopedia of Genes and Genomes;

FDR: False discovery rate;

GO: Gene Ontology;

OCND: Oral cancer nodal disease;

NFS: Node-free survival;

OSCC: Oral squamous cell carcinoma;

PBMCs: Peripheral blood mononuclear cells;

PTC: Papillary thyroid carcinoma;

## Declarations

## Ethics approval and consent to participate

The research didn't involve human specimens and animal experiments, with no ethics-related issues.

## Consent for publication

Not applicable.

## Availability of data and materials

All the data used in the study are obtained from the TCGA database (<https://xenabrowser.net/datapages/>), which is open and available to all.

## Competing interests

The authors declare that they have no competing interests.

## Funding

No specific funding was disclosed.

## Authors' contributions

Conceived and designed the experiments: G.H.H., Writing-original draft: M.W., T.L. Writing-review, and editing: Y.S.L., M.P., Z.H.W., and All authors read and approved the final manuscript.

## Acknowledgments

Not applicable.

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68: 394–424.
2. Leemans CR, Snijders PJ, Brakenhoff RH. The molecular landscape of head and neck cancer. *Nat Rev Cancer.* 2018;18(5):269–282.
3. Canning M, Guo G, Yu M, et al. Heterogeneity of the head and neck squamous cell carcinoma immune landscape and its impact on immunotherapy. *Front Cell Dev Biol.* 2019;7: 52.
4. Braakhuis BJ, Leemans CR, Visser O. Incidence and survival trends of head and neck squamous cell carcinoma in the Netherlands between 1989 and 2011. *Oral Oncol.* 2014;50: 670–675.
5. Taube JM, Galon J, Sholl LM, et al. Implications of the tumor immune microenvironment for staging and therapeutics. *Mod Pathol.* 2018;31: 214–234.

6. Balkwill FR, Capasso M, Hagemann T. The tumor microenvironment at a glance. *J Cell Sci.* 2012;125: 5591–5596.
7. Denton AE, Roberts EW, Fearon DT. Stromal cells in the tumor microenvironment. *Adv Exp Med Biol.* 2018;1060: 99–114.
8. Gavrielatou N, Doumas S, Economopoulou P, Foukas PG, Psyrii A. Biomarkers for immunotherapy response in head and neck cancer. *Cancer Treat Rev.* 2020;84: 101977.
9. Chen SMY, Krinsky AL, Woolaver RA, Wang X, Chen Z, Wang JH. Tumor immune microenvironment in head and neck cancers. *Mol Carcinog.* 2020.
10. Binnewies M, Roberts EW, Kersten K, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med.* 2018;24: 541–550.
11. Schneider K, Marbaix E, Bouzin C, et al. Immune cell infiltration in head and neck squamous cell carcinoma and patient outcome: a retrospective study. *Acta Oncol.* 2018;57: 1165–1172.
12. Macgregor PF, Squire JA. Application of microarrays to the analysis of gene expression in cancer. *Clin Chem.* 2002;48: 1170–1177.
13. Morris LG. Abstract IA11: Interactions between the genome and immune microenvironment in head and neck cancer: *AACR.* 2017. 10.1158/1557–3265.AACRAHNS17-IA11
14. Wang X, Yin Z, Zhao Y, He M, Dong C, Zhong M. Identifying potential prognostic biomarkers in head and neck cancer based on the analysis of microRNA expression profiles in TCGA database. *Mol Med Rep.* 2020;21: 1647–1657.
15. Cillo AR, Kurten CHL, Tabib T, et al. Immune Landscape of Viral- and Carcinogen-Driven Head and Neck Cancer. *Immunity.* 2020;52: 183–199.e189.
16. Wong N, Khwaja SS, Baker CM, et al. Prognostic microRNA signatures derived from The Cancer Genome Atlas for head and neck squamous cell carcinomas. *Cancer Med.* 2016;5: 1619–1628.
17. Yoshihara K, Shahmoradgoli M, Martinez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4: 2612.
18. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43: e47.
19. Poeta ML, Manola J, Goldwasser MA, et al. TP53 mutations and survival in squamous-cell carcinoma of the head and neck. *N Engl J Med.* 2007;357:2552–2561.
20. Network CGA. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature.* 2015;517: 576–582.
21. Cheng PF, Dummer R, Levesque MP. Data mining The Cancer Genome Atlas in the era of precision cancer medicine. *Swiss Med Wkly.* 2015;145: w14183.
22. De Paz D, Kao H-K, Huang Y, Chang K-P. Prognostic stratification of patients with advanced oral cavity squamous cell carcinoma. *Current oncology reports.* 2017;19: 65.
23. Zhao C, Zou H, Zhang J, Wang J, Liu H. An integrated methylation and gene expression microarray analysis reveals significant prognostic biomarkers in oral squamous cell carcinoma. *Oncol Rep.*

- 2018;40: 2637–2647.
24. Liu KYP, Lu XJD, Zhu Y, Yip S, Poh CF. Altered Immune-Related Gene Expressions Indicate Oral Cancer Nodal Disease. *J Dent Res*. 2018;97: 709–716.
  25. Lecerf C, Kamal M, Vacher S, et al. Immune gene expression in head and neck squamous cell carcinoma patients. *Eur J Cancer*. 2019;121: 210–223.
  26. Gajewski TF, Schreiber H, Fu Y-X. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol*. 2013;14: 1014–1022.
  27. Joyce JA. Therapeutic targeting of the tumor microenvironment. *Cancer cell*. 2005;7: 513–520.
  28. Whiteside TL. The tumor microenvironment and its role in promoting tumor growth. *Oncogene*. 2008;27: 5904–5912.
  29. Curry JM, Sprandio J, Cognetti D, et al. Tumor microenvironment in head and neck squamous cell carcinoma. *Semin Oncol*. 2014:217–234.
  30. Tong CC, Kao J, Sikora AG. Recognizing and reversing the immunosuppressive tumor microenvironment of head and neck cancer. *Immunol Res*. 2012;54: 266–274.
  31. Chen YP, Zhang Y, Lv JW, et al. Genomic analysis of tumor microenvironment immune types across 14 solid cancer types: immunotherapeutic implications. *Theranostics*. 2017;7: 3585–3594.
  32. Stelzl U, Worm U, Lalowski M, et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell*. 2005;122: 957–968.
  33. Rheinländer A, Schraven B, Bommhardt U. CD45 in human physiology and clinical medicine. *Immunology letters*. 2018;196: 22–32.
  34. Camacho M, Aguero A, Sumarroca A, et al. Prognostic value of CD45 transcriptional expression in head and neck cancer. *Eur Arch Otorhinolaryngol*. 2018;275: 225–232.
  35. Seitzer U, Kayser K, Höhn H, et al. Reduced T-cell receptor CD3zeta-chain protein and sustained CD3epsilon expression at the site of mycobacterial infection. *Immunology*. 2001;104: 269–277.
  36. Ye W, Zhou Y, Xu B, et al. CD247 expression is associated with differentiation and classification in ovarian cancer. *Medicine*. 2019;98: e18407.
  37. Han J, Chen M, Wang Y, et al. Identification of Biomarkers Based on Differentially Expressed Genes in Papillary Thyroid Carcinoma. *Sci Rep*. 2018;8: 9912.
  38. Wang Q, Li P, Wu W. A systematic analysis of immune genes and overall survival in cancer patients. *BMC Cancer*. 2019;19: 1225.

## Figures

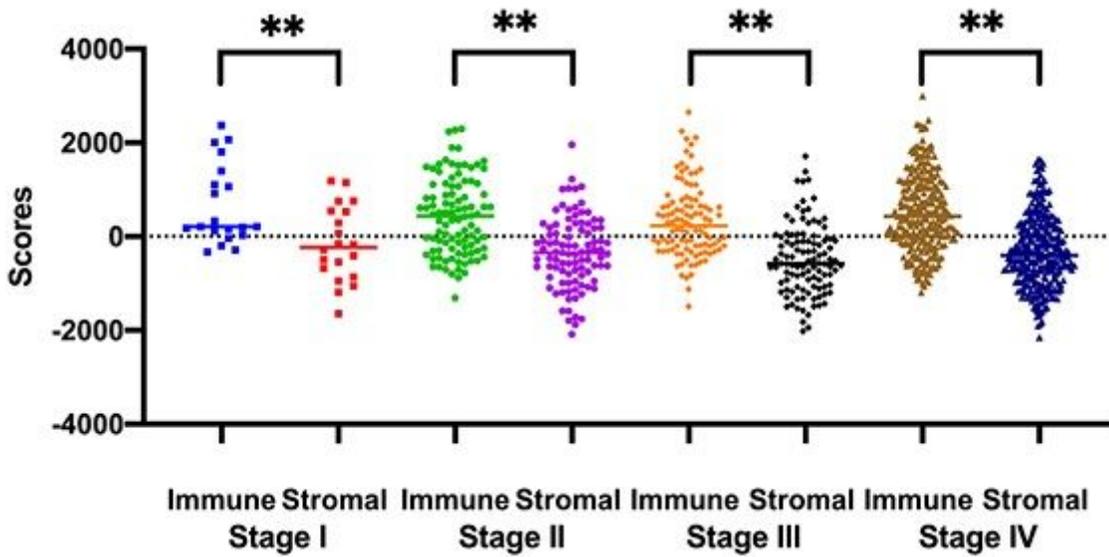


Figure 1

Comparison between immune scores and stromal scores in HNC patients at different pathological stages.

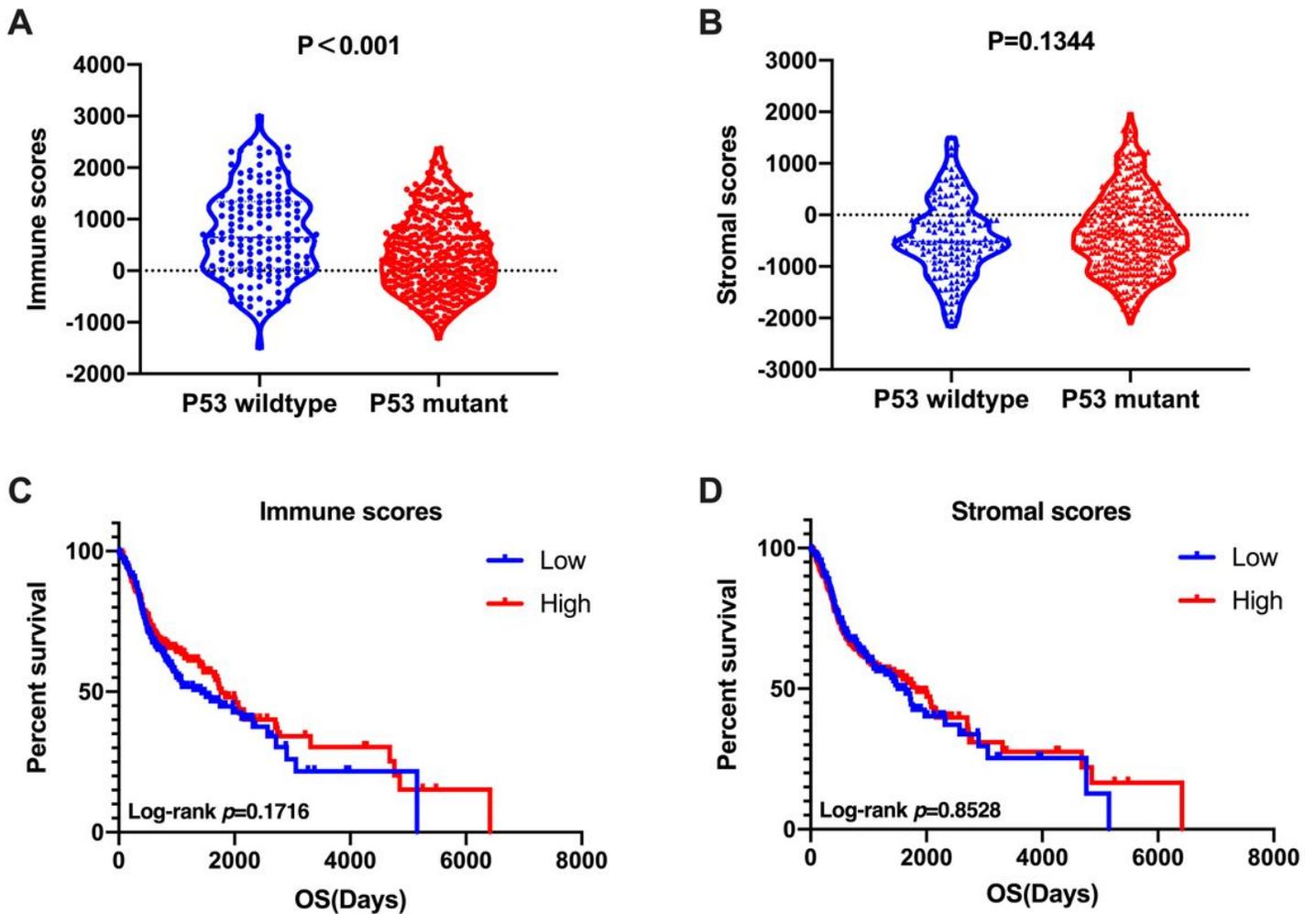


Figure 2

Immune scores and stromal scores were associated with HNC and overall scores. (A) Distribution of immune scores for p53 mutant and p53 wildtype HNC patients. Violin plot showing a significant association between p53 mutation status and immune scores ( $p < 0.001$ ). (B) Distribution of stromal scores for p53 mutant and p53 wildtype HNC patients. Violin plot showing no significant association between p53 mutation status and stromal scores ( $p=0.1344$ ). (C) HNC patients were divided into two groups based on their immune scores. As shown in the Kaplan-Meier survival curve, the median survival of the low-score group was longer than that of the high-score group (680 vs. 625 days,  $p = 0.1716$  in log-rank test). (D) HNC patients were divided into two groups based on their stromal scores; the median survival of the low-score group was longer than that of the high-score group (701 vs. 623 days,  $p= 0.8528$  in log-rank test); however, this difference was not statistically significant.

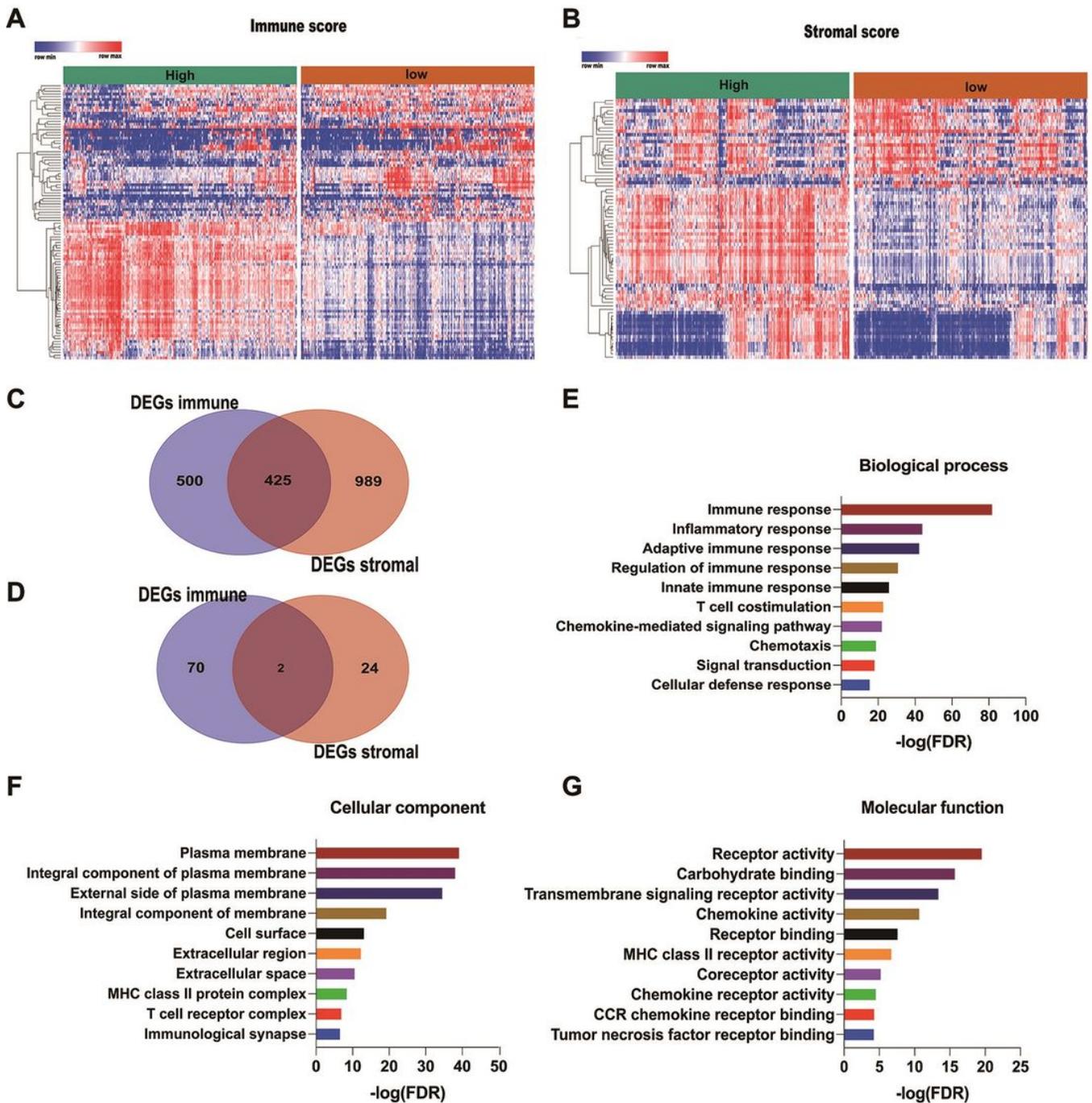
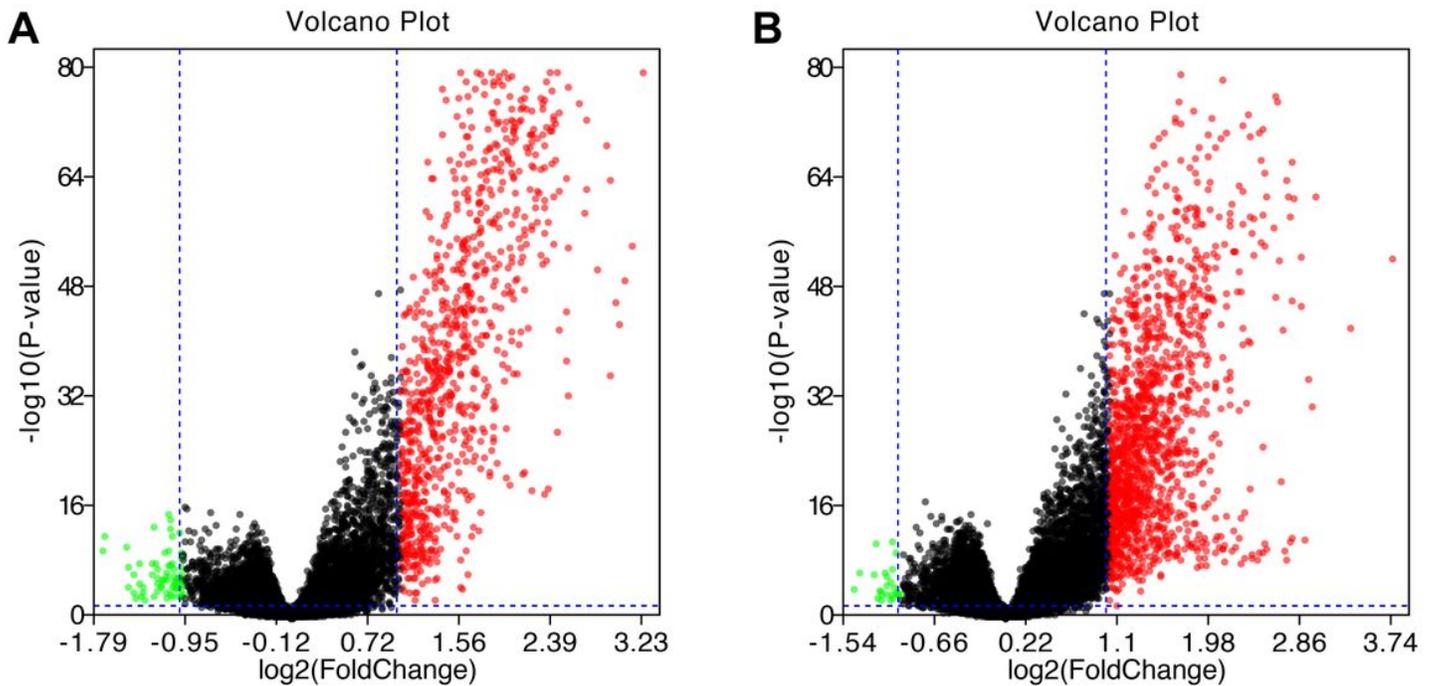


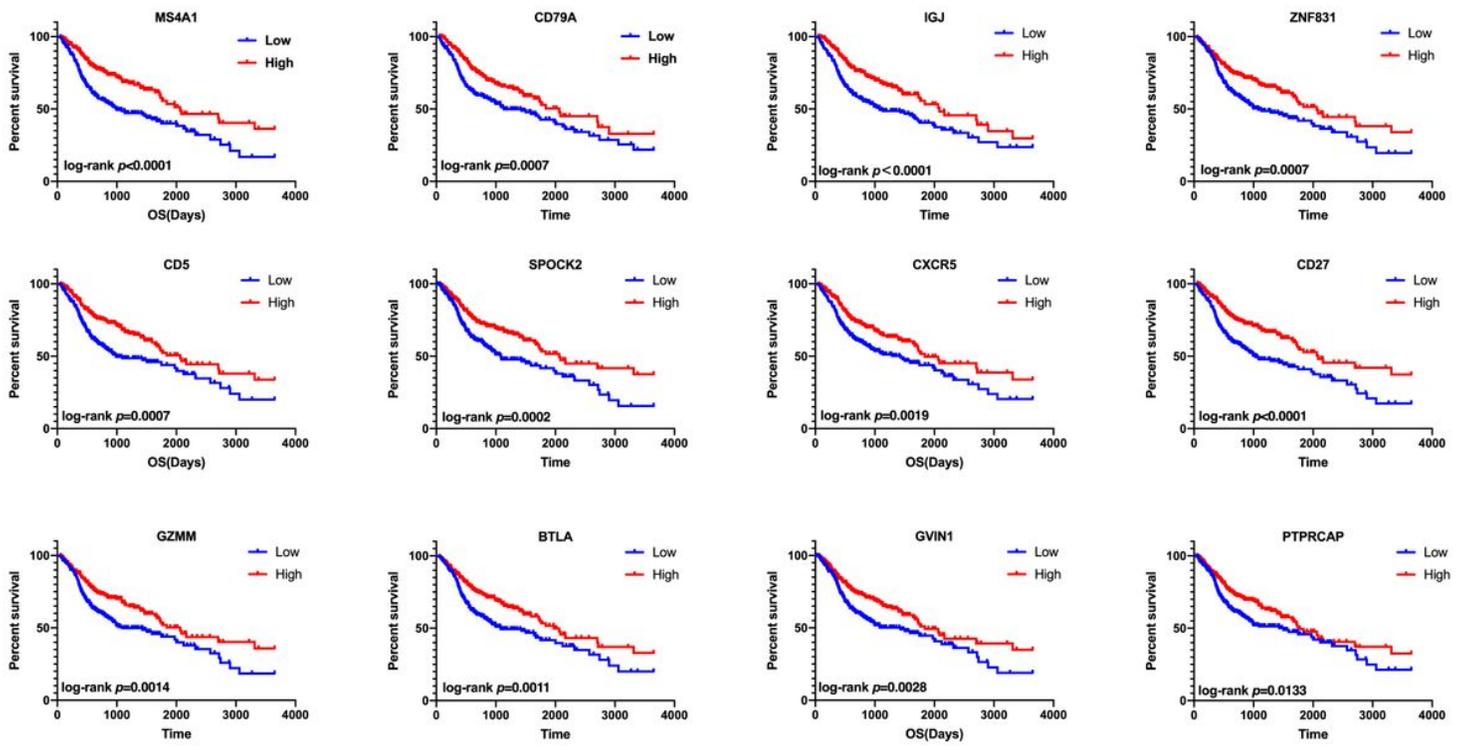
Figure 3

Comparison of gene expression profiles with immune scores and stromal scores in HNC. (A) Heatmap of DEGs in high vs. low immune scores ( $p < 0.05$ , fold change  $> 1.5$ ); genes with higher expression are shown in red, lower expression in green color, and genes with the same expression level are shown in black color. (B) Heatmap of the DEGs in high vs. low stromal scores ( $p < 0.05$ , fold change  $> 1.5$ ); relative expression levels are shown as in (A). (C) Venn diagrams of upregulated DEGs immune score groups. (D) Venn diagrams of downregulated DEGs immune score groups. (E) The biological process of top GO terms, (F) Cellular component, (G) Molecular function. False discovery rate (FDR) of GO analysis was acquired using the DAVID functional annotation tool.  $p < 0.05$ .



**Figure 4**

Volcano plot of gene expression profile data in (A) high vs. low immune score groups and (B) high vs. low stromal score groups.



**Figure 5**

Association between representative DEGs and overall survival in patients with HNC.

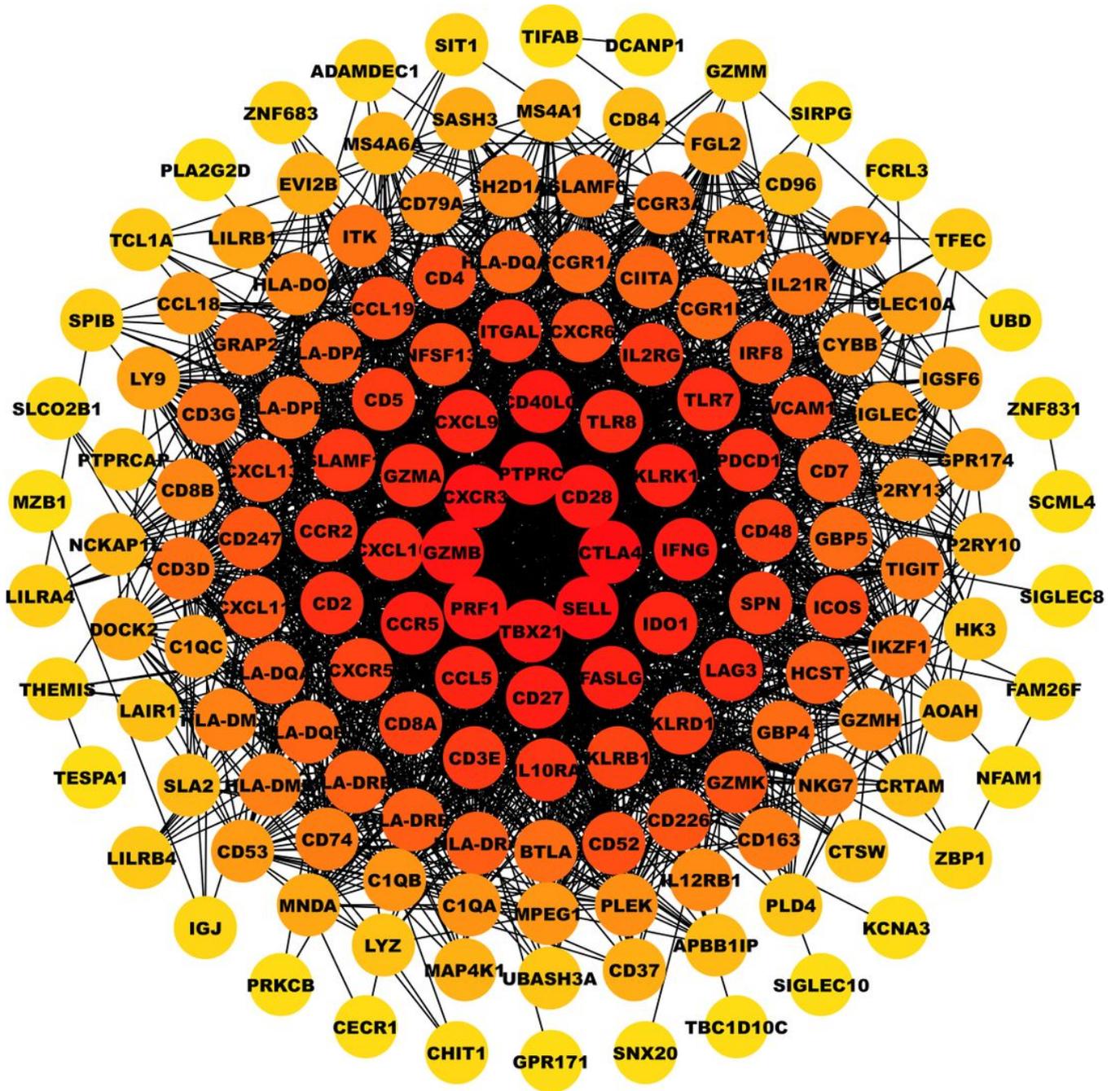


Figure 6

The total PPI network of DEGs.

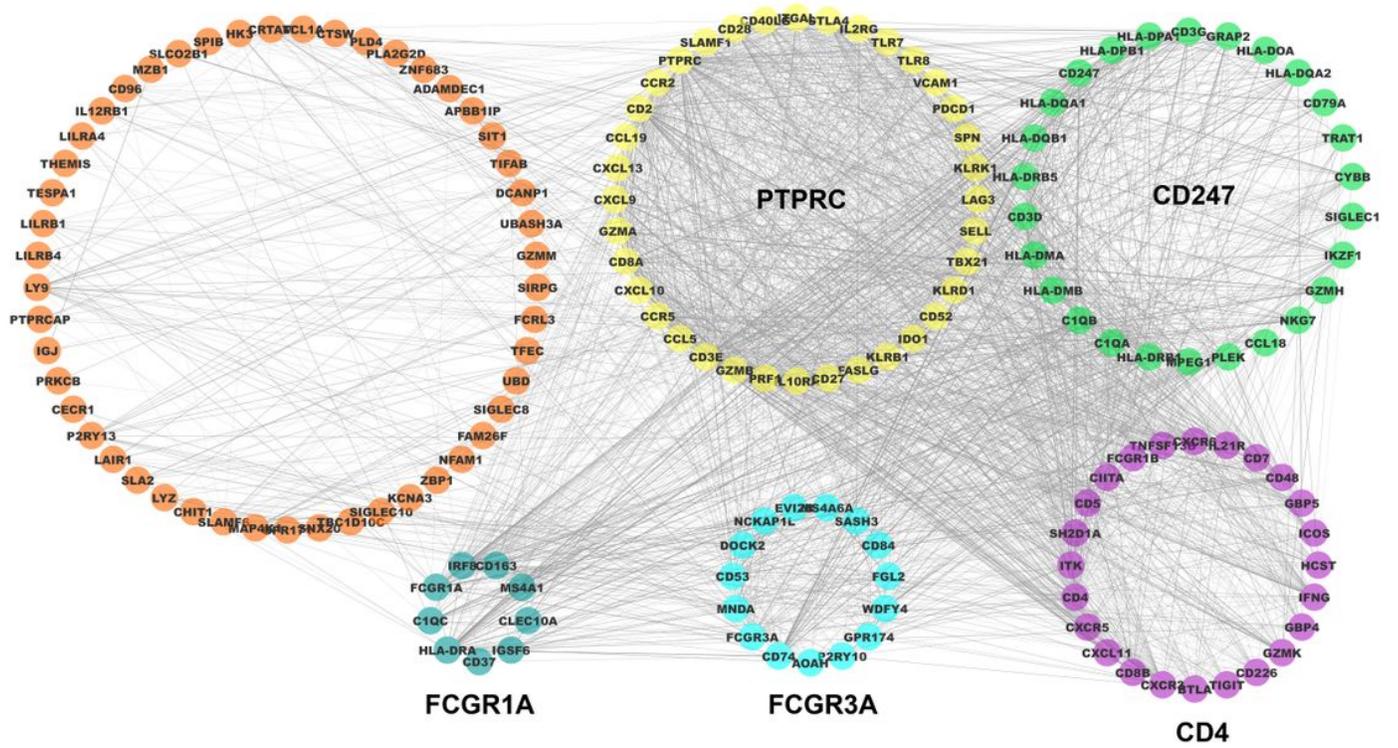
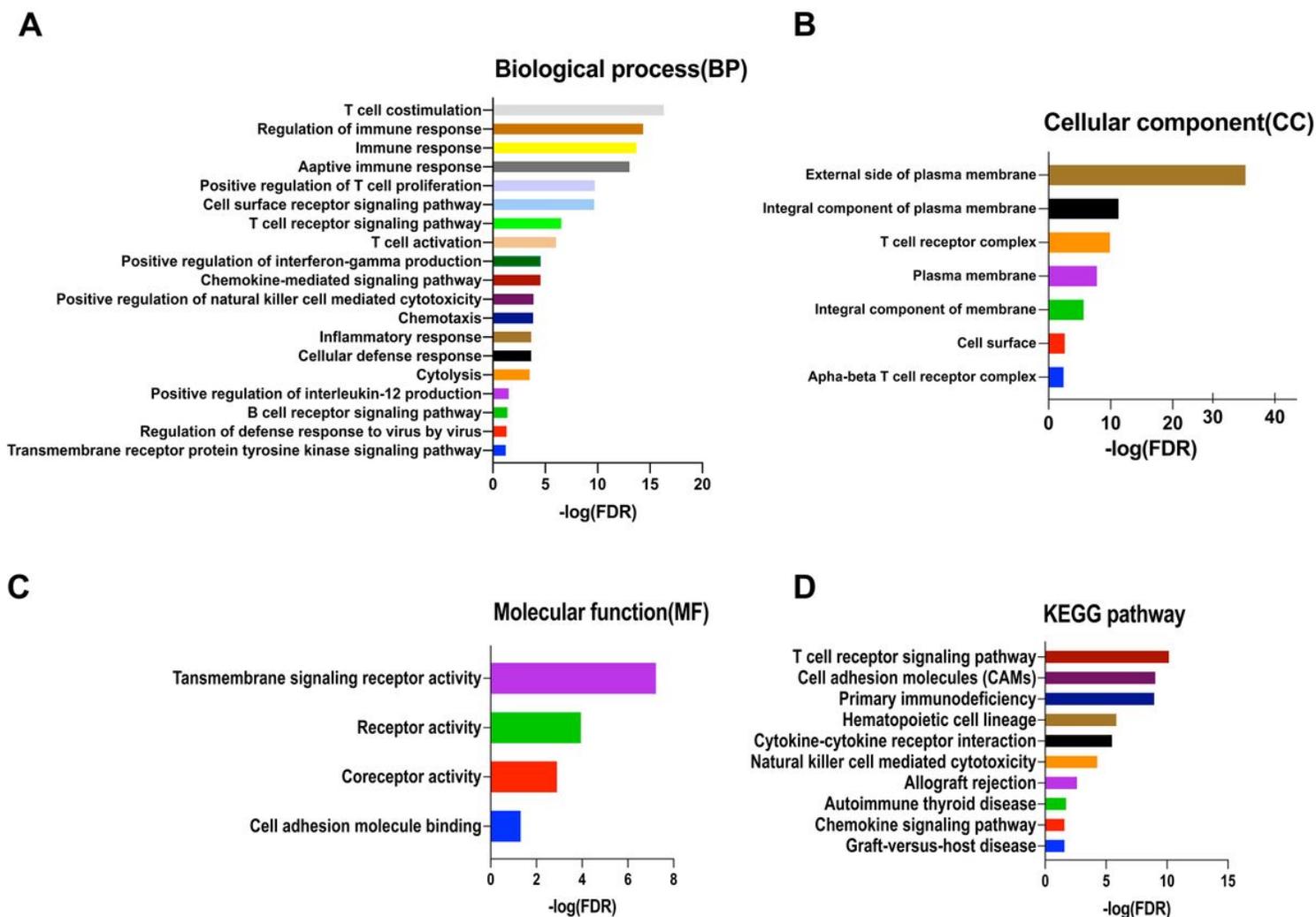


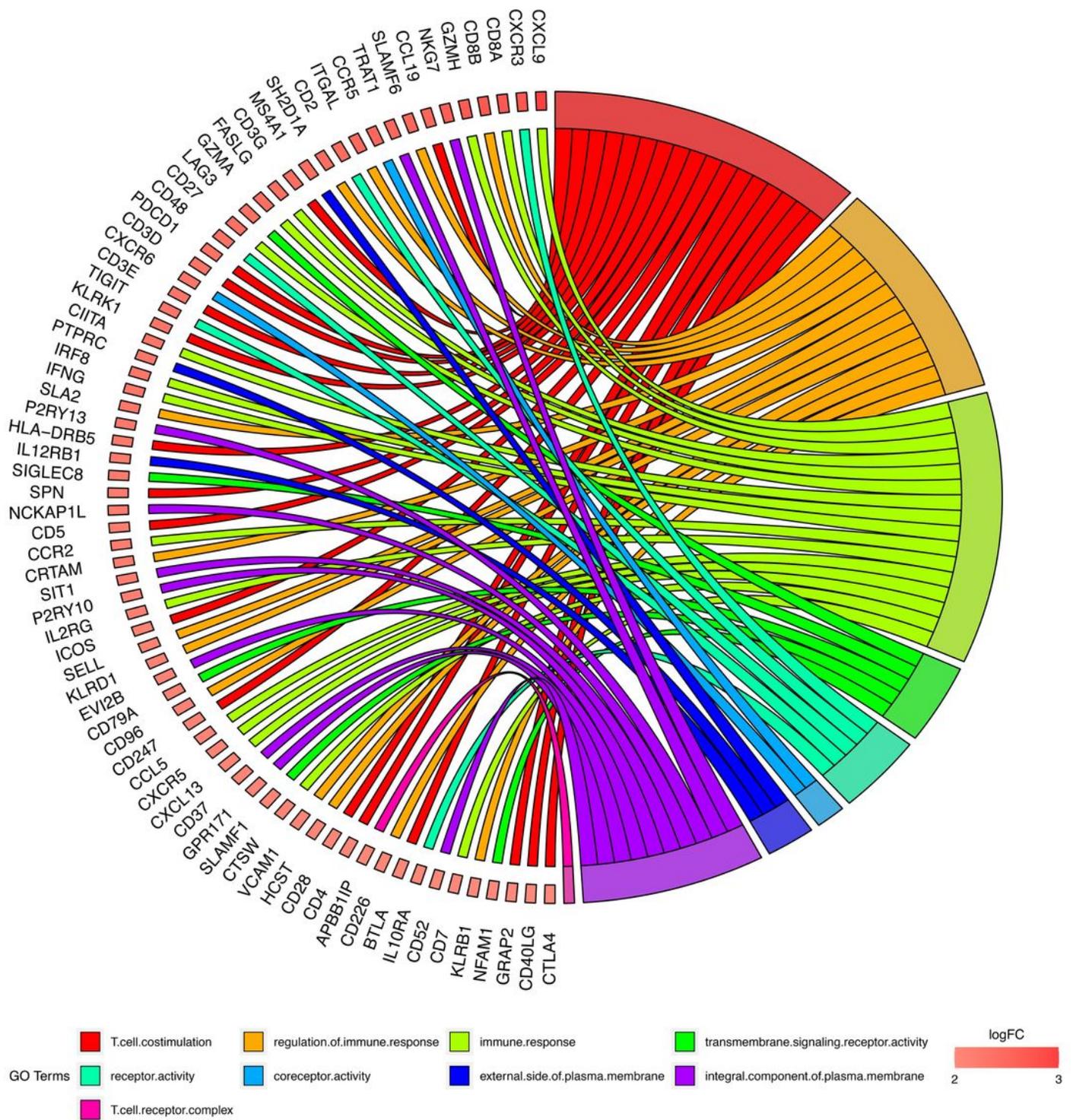
Figure 7

Top five PPI networks of PTPRC, CD247, CD4, FCGR3A, and FCGR1A modules.



**Figure 8**

GO term and KEGG pathway analysis for DEGs were significantly associated with overall survival. (A) Biological processes, (B) Cellular component, (C) Molecular function, (D) KEGG pathway.



**Figure 9**

The circos graph of the prognosis-related genes involving in GO analysis.

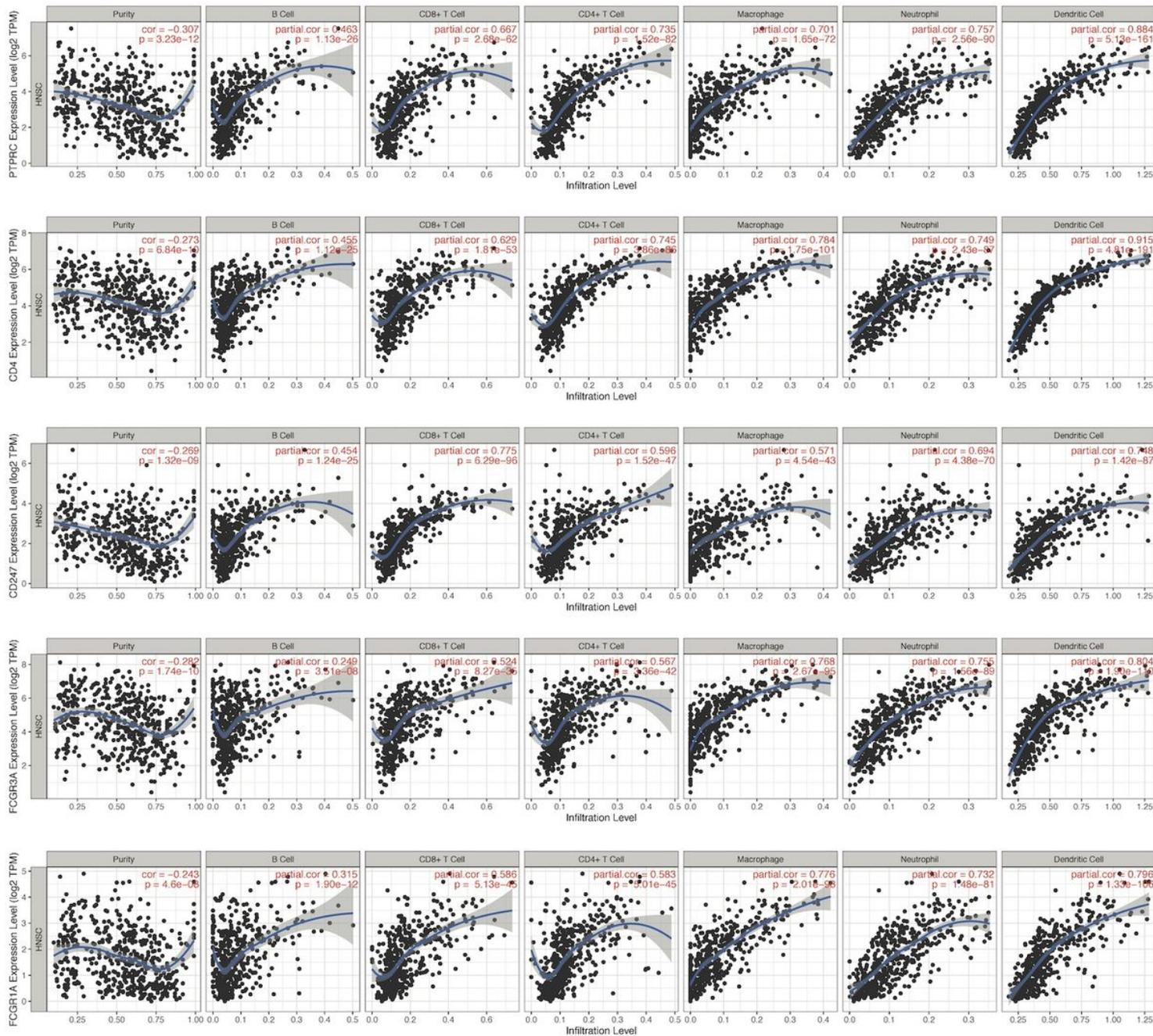
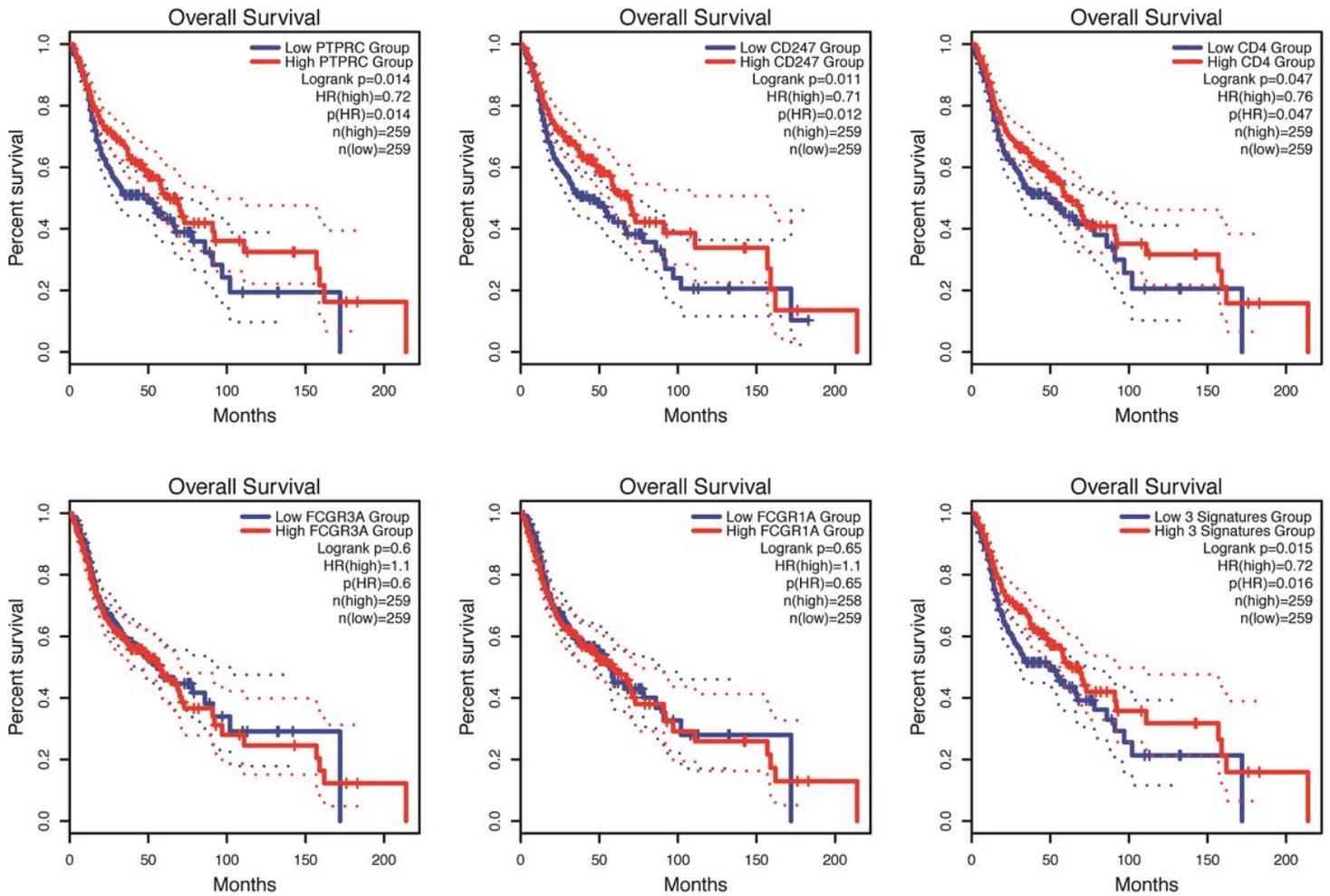


Figure 10

The association of the five hub genes with immune infiltration.



**Figure 11**

The five hub genes and overall mortality.

## Supplementary Files

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

- [SupplementaryTableS1.docx](#)
- [SupplementaryTableS2.docx](#)