

Mining TCGA database for prognostic genes in head and neck squamous cell carcinoma microenvironment

Qiuchi Ran

Hainan Medical University

Shengrong Long

China medical university <https://orcid.org/0000-0002-5831-4235>

Yan Ye

Duoyi Network

Diwas Sunchuri

Hainan Medical University

Hong Liang

First Affiliated Hospital of Hainan Medical University

Zhuling Guo (✉ 604569033@qq.com)

Research article

Keywords: TCGA, tumor microenvironment, immune scores, HNSCC

Posted Date: September 24th, 2019

DOI: <https://doi.org/10.21203/rs.2.14729/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Journal of Dental Sciences on March 1st, 2021. See the published version at <https://doi.org/10.1016/j.jds.2020.09.017>.

Abstract

Background Head and neck squamous cell carcinoma (HNSCC) is one of the most common malignant tumors. Many previous reports have already shown that the extent of infiltrating immune and stromal cells in tumor tissues and the tumor microenvironment (TME) cells play a significant role in the overall prognosis. **Methods** The convenient access to The Cancer Genome Atlas (TCGA) database facilitates global gene expression profiling and database mining in a large-scale for potential correlation between genes and overall survival of a variety of malignancies including HNSCC. The quantification of the immune and stromal components in tumor tissues could be facilitated by calculating immune scores and stromal scores on the basis of Estimation of stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) algorithm could facilitate the quantification of the immune and stromal components in tumor tissues. The effects of genes involved in immune and stromal cells on prognosis were categorized. Prognosis associated genes of HNSCC patients were further identified. **Result** This study showed that GIMAP6, SELL, TIFAB, KCNA3, P2RY8 and CCR4 may mediate immune response, extracellular matrix, and immunoglobulin binding via neutrophil activation in HNSCC. **Conclusion** Depicting a comprehensive landscape of the TME characteristics of HNSCC may therefore help to interpret the responses of HNSCC to immunotherapies and provide new strategies for the treatment of cancers.

Background

Head and neck squamous cell carcinoma (HNSCC) is one of the most common malignant tumors(1). In recent years, surgical radiotherapy and chemotherapy for HNSCC have effectively improved the prognosis, but the global 5-year survival rate is still lower than 50%(2). Many studies have shown that the system dysfunction plays an important role in tumor occurrence and progression. Some studies have revealed the significance of tumor-related structures as well as upregulated signaling pathways in both cancer cells and the tumor microenvironment(TME)(3, 4). Tumor cell intrinsic genes especially master transcription factors dictate the initiation, progression, and evolution of HNSCC(5-7). Tumor microenvironment is the cellular milieu that facilitates cancer cells growth and maturation. Immune and stromal cells are two major types of non-tumor components. Studies have found that such non-tumor components are important for diagnostic and prognostic assessment of tumors.

To predict tumor purity, algorithms(8, 9) have been developed using gene expression data from The Cancer Genome Atlas (TCGA) database. Yoshihara et al.(8) designed an algorithm called Estimation of stromal and Immune cells in malignant Tumor tissues using Expression data (ESTIMATE). After analyzing specific gene expression signature of immune and stromal cells, immune and stromal scores were calculated to predict the infiltration of non-tumor cells. Subsequent reports have applied the ESTIMATE algorithm on prostate cancer(10), breast cancer(11), and colon cancer(12). Based on the

effectiveness of algorithms, this paper further investigated the development and progression of 546 HNSCC cases.

Methods

HNSCC datasets

We used level 3 mRNA expression data and clinical data on 546 HNSCC patient samples, including 44 normal samples and 502 cancer tissues samples. Survival data were available from TCGA data portal (<http://cancergenome.nih.gov/>), and data format is FPKM. Immune scores and stromal scores were calculated by applying the ESTIMATE algorithm to the downloaded database. Data were analyzed with the R (version 3.5.1) and R Bioconductor packages.

Identification of differentially expressed genes (DEGs)

Package limma(13) was used to analyze the data.. Fold change > 2 and False discovery rate (FDR) < 0.05 were set as the cutoffs to screen for differentially expressed genes (DEGs).

Heatmaps and clustering analysis

Using R package pheatmap, heatmaps and clustering were generated.

Enrichment analysis of DEGs

Functional enrichment analysis of DEGs was performed by R package clusterprofiler(14) to identify gene ontology(GO) categories by their biological processes (BP), molecular functions (MF), or cellular components (CC). The clusterprofiler was also used to perform pathway enrichment analysis with reference from KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways. FDR< 0.05 was used as the cut-off.

Overall survival curve

To illustrate the relationship between patients' overall survival and gene expression levels of DEGs, Kaplan-Meier plots were generated and relationship was tested by log-rank test.

Results

We downloaded gene expression profiles and clinical information of all 546 HNSCC patients from the TCGA database. Among them, 63(11.54%) patients were diagnosed with Grade 1, 310(56.78%) patients were Grade 2, 125(22.89%) patients were Grade 3, 7(1.28%) patients were Grade 4, and 41(7.51%) patients were of unknown grade. Based on ESTIMATE algorithm, stromal scores ranged from -1941.57 to 1947.28, and immune scores were distributed between -1057.57 to 2785.18, respectively (Figure 1a, b).

To find out the potential correlation of overall survival with immune scores and/or stromal scores, we divided the 500 HNSCC cases (only 500 of 502 cancer tissues samples have their grade clinical information) into top and bottom halves (high vs. low score groups) based on their scores. Kaplan-Meier survival curves (Figure 1c) showed that median overall survival of cases with the low score group of immune scores are shorter than the cases in the high score group ($p = 0.107$ in log-rank test). Consistently, cases with lower stromal scores also showed shorter median overall survival compared to patients with higher stromal scores (Figure 1d, $p = 0.814$ in log-rank test), although it was not statistically significant.

Comparison of gene expression profile with immune scores and stromal scores in HNSCC

To reveal the correlation of global gene expression profiles with immune scores and/or stromal scores, we compared Affymetrix microarray data of these 500 HNCSS cases obtained in TCGA database. Heatmaps in Figure 2 showed distinct gene expression profiles of cases belong to high vs. low immune scores/stromal scores groups. For comparison based on immune scores, 777 genes were upregulated and 161 genes downregulated in the high score than the low score group (fold change > 1.5 , $p < 0.05$). Similarly, for the high and low groups based on stromal scores, 986 genes were upregulated and 63 genes were downregulated in the high score group (fold change > 1.5 , $p < 0.05$). Moreover, Venn diagrams (Figure 2c, d) showed that 260 genes were commonly upregulated in the high scores groups, and 15 genes were commonly downregulated. It is worth mentioning that the DEGs extracted from the comparison of high vs. low immune scores groups covered the majority of genes extracted from the comparison based on stromal scores. Thus, we decided to focus on these DEGs for all subsequent analysis in this manuscript. To outline the potential function of the DEGs, we performed functional enrichment analysis of the 260 upregulated genes and 15 downregulated genes in high-immune scores group. Functional enrichment clustering of these genes showed strong association with immune response as well. Top GO terms identified neutrophil activation involved in and mediated immune response, extracellular matrix, and immunoglobulin binding (Figure 2e). In addition, all the pathways that were yielded from the KEGG identified Cytokine-cytokine receptor interaction, Complement and coagulation cascades and B cell receptor signaling pathway (Figure 2f)

Correlation of expression of individual DEGs in overall survival

To explore the potential roles of individual DEGs in overall survival, we generated Kaplan-Meier survival curves from TCGA database. Among the 275 DEGs in the high-immune scores group, a total of 59 DEGs were shown to significantly predict poor overall survival in log-rank test ($p < 0.001$, selected 6 genes are shown in Figure 3).

Discussion

In this work, we attempted to identify tumor microenvironment related genes contributing to HNSCC overall survival in the TCGA database. In particular, by comparing global gene expression in a large number of cases with high vs. low immune scores, we extracted 59 genes involved in extracellular matrix and immune response. By analyzing 275 differentially expressed genes yielded from comparison of high vs. low immune scores (or stromal scores) groups, we found that many of them were involved in tumor microenvironment, as shown by GO term analysis (Figure 2). This is in accordance with previous reports that the functions of immune cells and ECM molecules are closely interrelated in building tumor microenvironment in HNSCC(15-18). Finally, we analyzed overall survival analysis of these 275 genes and found that 59 genes were associated with poor outcomes in HNSCC patients.

We are particularly interested in CCR4 and GIMAP6. CCR4 has been proven that it may have an important role in HNSCC progression, regional lymph node metastasis and recurrence. Significant progress has been made on the correlation of overall survival with gene expression in HNSCCs(19-21). Many of these experiments were done in tumor cell lines, animal models, or patients' tumor samples. GIMAP6 has also been reported to have a potential role in tumor evolution mechanisms related to inflammation and microenvironment(22). However, the complexity of HNSCC and HNSCC microenvironment demands more comprehensive analysis consisting of larger cohorts. The interaction between HNSCC and its tumor microenvironment critically affects tumor evolution, which subsequently impacts tumor subtype classification, recurrence, drug resistance, and the overall prognosis of patients.

In the current work, we focused on genes characteristics of microenvironment, which in turn affect the development of HNSCC and hence contribute to patients' overall survival. Thus, our results may provide additional data in decoding the complex interaction of tumor and tumor environment in HNSCC.

In summary, we extracted a list of tumor microenvironment related genes, through the detailed analysis of TCGA applied by ESTIMATE algorithm based on immune scores. These genes were validated in an independent HNSCC cohort could be useful for determining the prognosis of HNSCC patients. Some of the previously ignored genes can be potential additional biomarkers for HNSCC. In addition to this, testing whether these new combined set of genes, provide a strong predictor of survival than individual genes would be extremely interesting. Lastly, further investigation of these genes could lead to new insights into the potential association of tumor microenvironment with HNSCC prognosis in a comprehensive manner.

Abbreviations

HNSCC: Head and neck squamous cell carcinoma; TCGA: The Cancer Genome Atlas; ESTIMATE: Estimation of stromal and Immune cells in Malignant Tumor tissues using Expression; TME: tumor microenvironment; GO: Gene ontology; KEGG: Kyoto encyclopedia of genes and genomes; FDR: False discovery rate; DEG: Differentially expressed gene; BP: biological processes; CC: cellular components; MF: molecular functions; FPKM: Fragments per kilobase million.

Declarations

Acknowledgements

All the authors thank Dr. Guangyu Li for editing this manuscript.

Authors' contributions

QCR, SRL, HL and ZLG participated in the research design of the manuscript. QCR, SRL and YY were responsible for study selection, quality assessment, data extraction and data synthesis. QCR, DS and ZLG drafted the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by Hainan Provincial Natural Science Foundation of China (817330), Young Talents' Science and Technology Innovation Project of Hainan Association for Science and Technology (QCXM201817), Education Department of Hainan Province (Hnky2017-31), Medical Education Research Project of Medical Education Branch of Chinese Medical Association (2018B-N12003), Course Construction Project of Hainan Medical University (HYJW201834)

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* 2018;68(1):7-30.
2. Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. *Nat Rev Cancer.* 2011;11(1):9-22.
3. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med.* 2013;19(11):1423-37.
4. Su S, Chen J, Yao H, Liu J, Yu S, Lao L, et al. CD10(+)/GPR77(+) Cancer-Associated Fibroblasts Promote Cancer Formation and Chemoresistance by Sustaining Cancer Stemness. *Cell.* 2018;172(4):841-56 e16.
5. Curry JM, Sprandio J, Cognetti D, Luginbuhl A, Bar-ad V, Pribitkin E, et al. Tumor microenvironment in head and neck squamous cell carcinoma. *Semin Oncol.* 2014;41(2):217-34.
6. Baruah P, Lee M, Odutoye T, Williamson P, Hyde N, Kaski JC, et al. Decreased levels of alternative co-stimulatory receptors OX40 and 4-1BB characterise T cells from head and neck cancer patients. *Immunobiology.* 2012;217(7):669-75.
7. Whiteside TL. Immunobiology of head and neck cancer. *Cancer Metastasis Rev.* 2005;24(1):95-105.
8. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612.
9. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol.* 2012;30(5):413-21.
10. Shah N, Wang P, Wongvipat J, Karthaus WR, Abida W, Armenia J, et al. Regulation of the glucocorticoid receptor via a BET-dependent enhancer drives antiandrogen resistance in prostate cancer. *Elife.* 2017;6.
11. Priedigkeit N, Watters RJ, Lucas PC, Basudan A, Bhargava R, Horne W, et al. Exome-capture RNA sequencing of decade-old breast cancers and matched decalcified bone metastases. *JCI Insight.* 2017;2(17).
12. Alonso MH, Ausso S, Lopez-Doriga A, Cordero D, Guino E, Sole X, et al. Comprehensive analysis of copy number aberrations in microsatellite stable colon cancer in view of stromal component. *Br J Cancer.* 2017;117(3):421-31.
13. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
14. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284-7.
15. Lu P, Takai K, Weaver VM, Werb Z. Extracellular matrix degradation and remodeling in development and disease. *Cold Spring Harb Perspect Biol.* 2011;3(12).
16. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol.* 2014;15(12):786-801.
17. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep.* 2014;15(12):1243-53.

18. Peranzoni E, Rivas-Caicedo A, Bougherara H, Salmon H, Donnadieu E. Positive and negative influence of the matrix architecture on antitumor immune surveillance. *Cell Mol Life Sci.* 2013;70(23):4431-48.
19. Gonzalez-Arriagada WA, Lozano-Burgos C, Zuniga-Moreta R, Gonzalez-Diaz P, Coletta RD. Clinicopathological significance of chemokine receptor (CCR1, CCR3, CCR4, CCR5, CCR7 and CXCR4) expression in head and neck squamous cell carcinomas. *J Oral Pathol Med.* 2018;47(8):755-63.
20. Sun W, Li WJ, Wei FQ, Wong TS, Lei WB, Zhu XL, et al. Blockade of MCP-1/CCR4 signaling-induced recruitment of activated regulatory cells evokes an antitumor immune response in head and neck squamous cell carcinoma. *Oncotarget.* 2016;7(25):37714-27.
21. Tsujikawa T, Yaguchi T, Ohmura G, Ohta S, Kobayashi A, Kawamura N, et al. Autocrine and paracrine loops between cancer cells and macrophages promote lymph node metastasis via CCR4/CCL22 in head and neck squamous cell carcinoma. *Int J Cancer.* 2013;132(12):2755-66.
22. Sanati N, Iancu OD, Wu G, Jacobs JE, McWeeney SK. Network-Based Predictors of Progression in Head and Neck Squamous Cell Carcinoma. *Front Genet.* 2018;9:183.

Figures

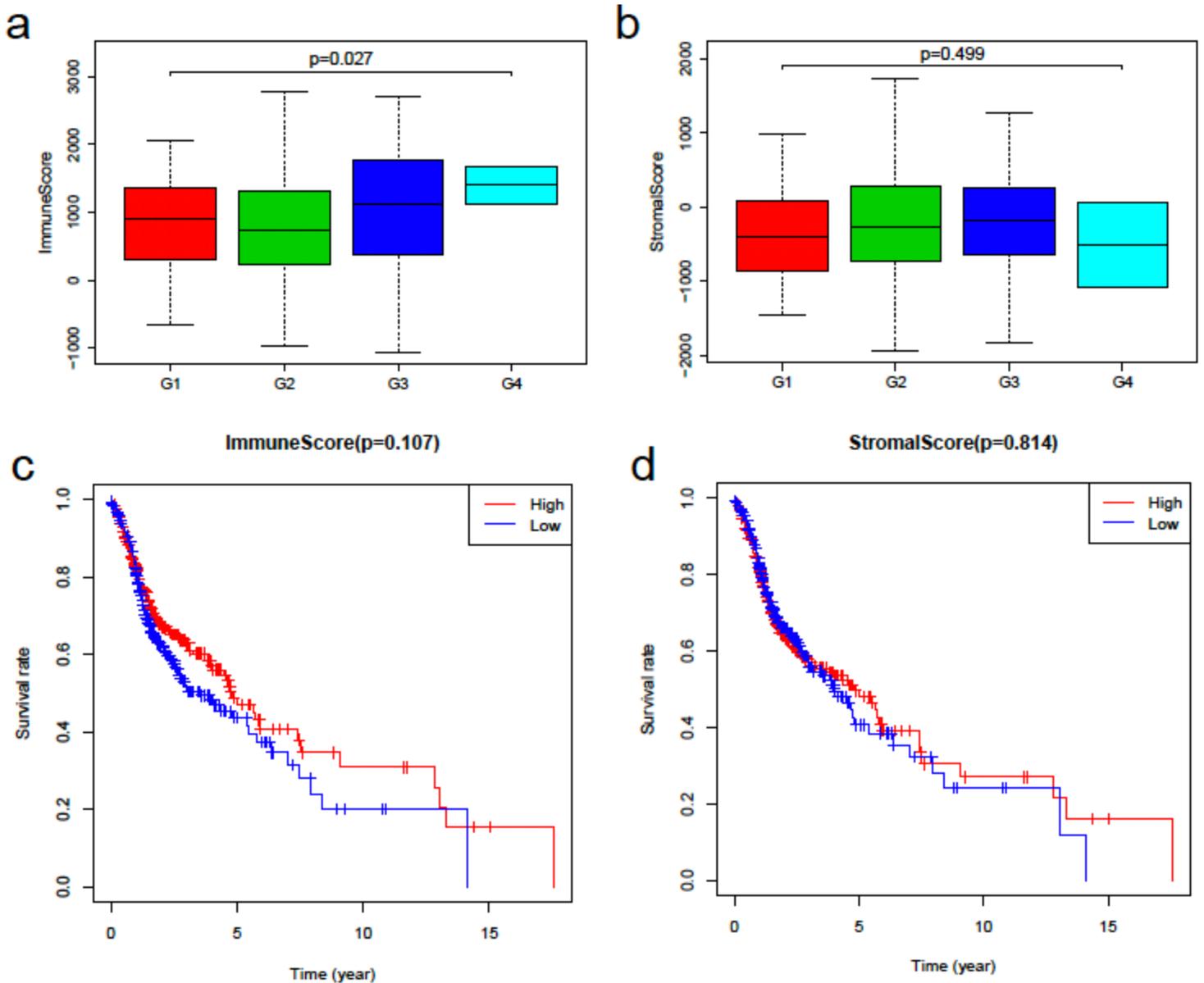


Figure 1

Immune scores and stromal scores are associated with HNSCC grades and their overall survival. (a) Distribution of immune scores of HNSCC grades. Boxplot shows that there is significant association between HNSCC grades and the level of immune scores ($p < 0.05$). (b) Distribution of stromal scores of HNSCC grades. Box-plot shows that there is not significant association between HNSCC grades and the level of stromal scores ($p < 0.05$). (c) HNSCC cases were divided into two groups based on their immune scores: the top half of cases with higher immune scores and the bottom half of cases with lower immune scores. As shown in the Kaplan-Meier survival curve, median survival of the low score group is shorter than high score group, as indicated by the log-rank test, p value is 0.107. (d) HNSCC cases were divided into two groups based on their stromal scores: the top half of 209 cases and the bottom half of 208 cases. The median survival of the low score group is shorter than the high score group, however, it is not statistically different as indicated by the log-rank test $p = 0.814$

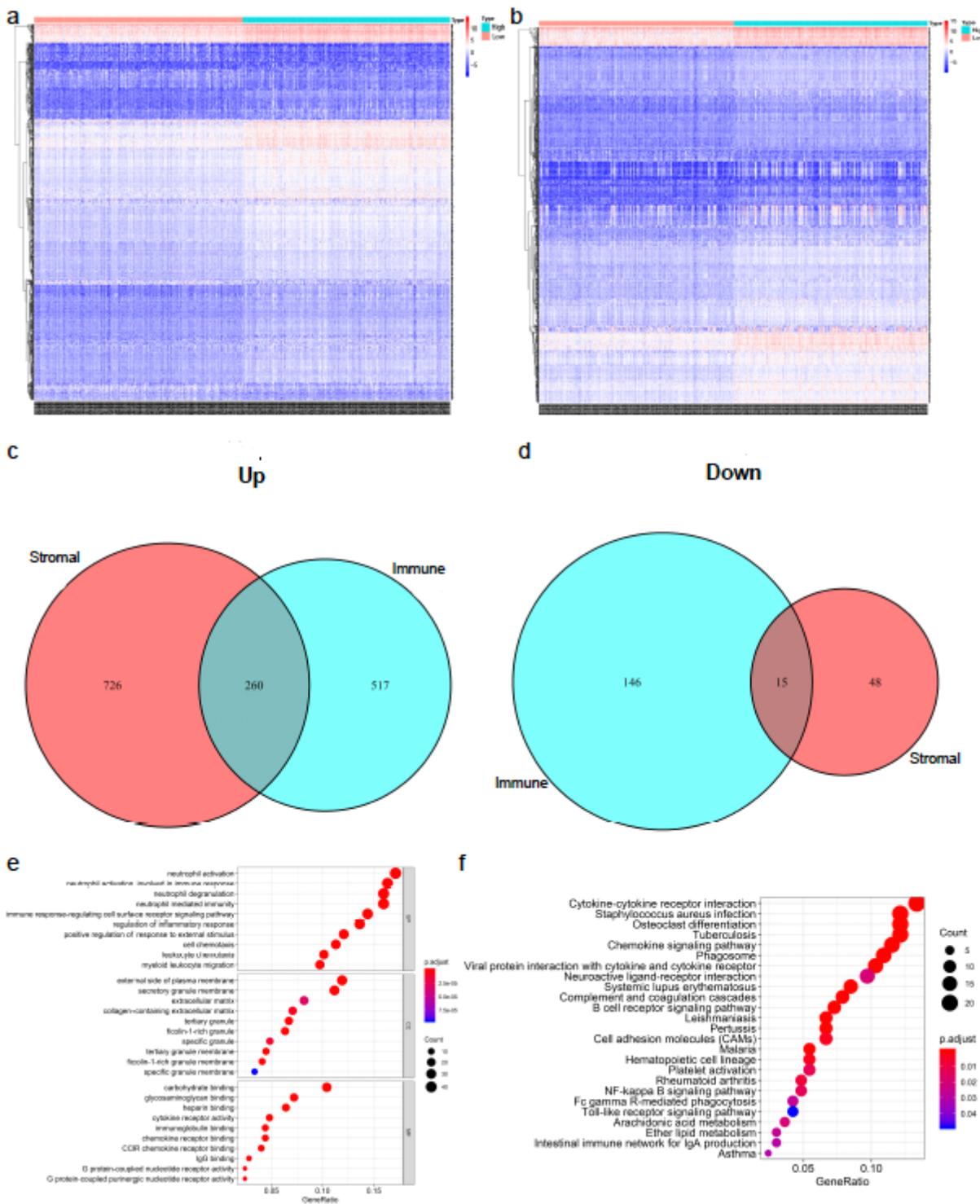


Figure 2

Comparison of gene expression profile with immune scores and stromal scores in HNSCC. Heatmaps were drawn based on the average linkage method and Pearson distance measurement method. Genes with higher expression are shown in red, lower expression are shown in blue, genes with same expression level are in white. (a) Heatmap of the DEGs of immune scores of top half (high score) vs. bottom half (low score). $p < 0.05$, fold change > 1.5). (b) Heatmap of the DEGs of stromal scores of top half (high score) vs. bottom half (low score). $p < 0.05$, fold change > 1.5). (c, d) Venn diagrams showing the number

of commonly upregulated (c) or downregulated (d) DEGs in stromal and immune score groups. (e, f) Top 10 BP, CC, MF terms and top 25 KEGG pathways. enrichment analysis was performed by R package clusterprofiler (p. adjusted < 0.05).

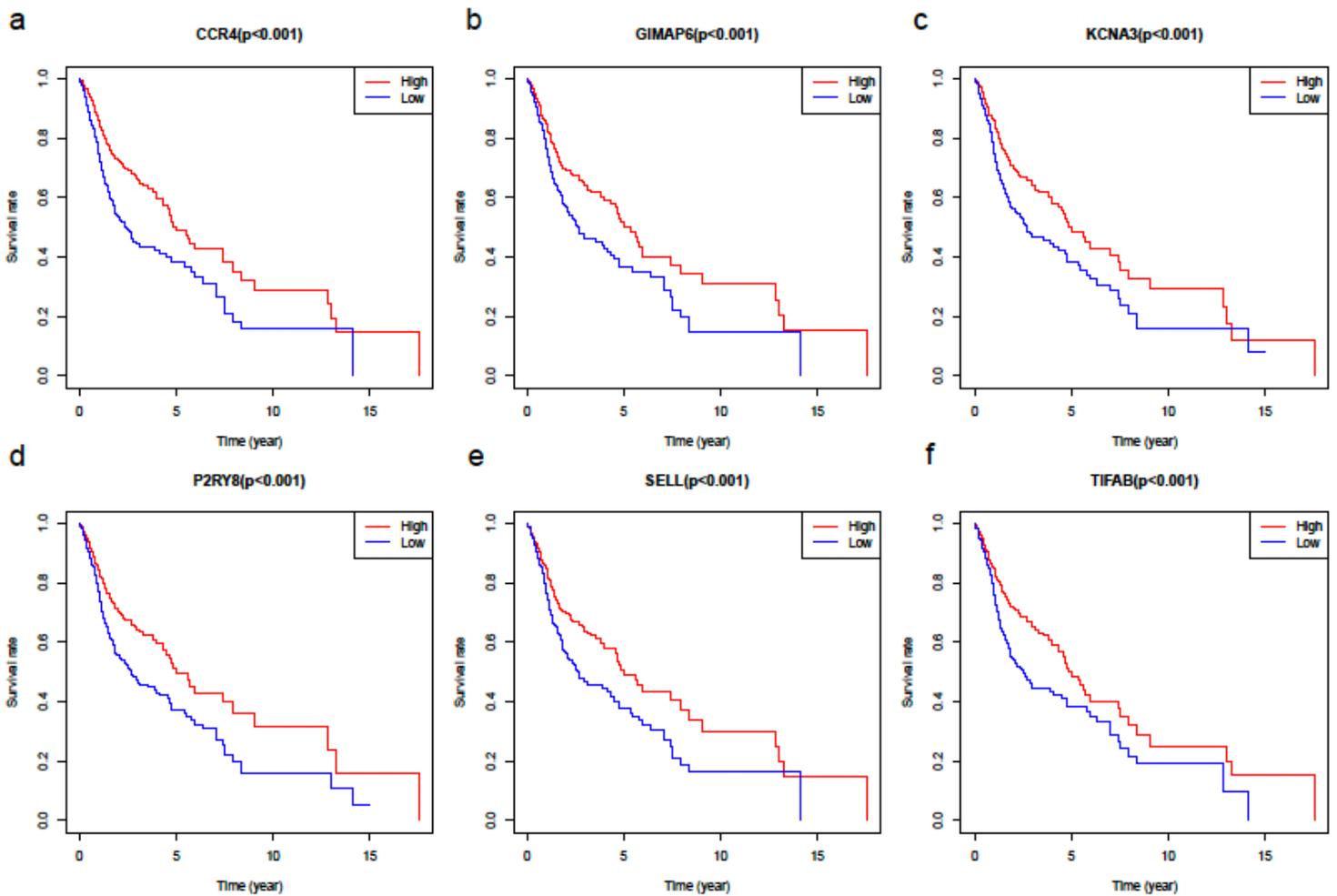


Figure 3

DEGs extracted from TCGA database with overall survival. Kaplan-Meier survival curves were generated for selected DEGs extracted from the comparison of groups of high (red line) and low (blue line) gene expression. p < 0.001 in Log-rank test.