Identification of S1PR4 as an immune modulator for favorable prognosis in HNSCC through unbiased machine learning

Chenshen Huang[1, 2, 3†], Zeda Zhao[1†], Fengshuo Zhu[4†], Ying Qian[1], Dajia Lin[2], Junyin Zheng[2], Hong Chen[2], Ning Wang[5], Wujin Li[6], Xiangyi Kong[7], Qi Huang[3*], Juliang Qin[1*], and Bing Du[1*]

[1] Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai, China.

[2] Department of Gastrointestinal Surgery, Fujian Provincial Hospital, Fuzhou, China.

[3] Department of General Surgery, Tongji Hospital, Tongji University School of Medicine, Shanghai, China.

[4] Department of Oral Maxillofacial-Head and Neck Oncology, Shanghai Ninth People's Hospital, College of Stomatology, Shanghai, China. Jiao Tong University School of Medicine, National Clinical Research Center for Oral Disease, Shanghai Key Laboratory of Stomatology and Shanghai Research Institute of Stomatology, Shanghai, China.

[5] Huzhou Central Hospital, Affiliated Hospital of Zhejiang University, Huzhou, China

[6] Department of Thoracic Surgery, Fujian Provincial Hospital, Fuzhou, China.

[7] University College London, London, UK

† Contributed equally.

*** Correspondence:**

Bing Du, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, 500 Dongchuan Road, Shanghai, China (bdu@bio.ecnu.edu.cn)

Juliang Qin, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, 500 Dongchuan Road, Shanghai, China (jlqin@bio.ecnu.edu.cn)

Qi Huang, Department of General Surgery, Tongji Hospital, Tongji University School of Medicine, 389 Xincun Road, Shanghai, China (hqhq0078@163.com)

Keywords: S1PR4; CD8 T cell; Immune modulator; HNSCC; Unbiased machine learning

Abstract

As the largest family of membrane proteins, G protein-coupled receptor (GPCR) has become the most prominent family of pharmacological targets. Whereas, only a little GPCRs have been well defined in physiological and pathological functions. Thus, an efficient way to identify key GPCRs in tumor formation is urgent needed. Here, the head and neck squamous cell carcinoma (HNSCC) patients were classified into two identical subtypes by the characteristics of GPCR with unbiased machine learning method. Interestingly, these two subtypes showed significant differences in prognosis, gene expression, and immune microenvironment, especially the infiltration of $CD8^+$ T cell. Based on these differences, we screened three potential key regulators (S1PR4, S1PR5, and GPR87) from all GPCRs, and constructed a prognosis-predicting nomogram for HNSCC patients. Further, we identified S1PR4 as the key GPCR in determining the two subtypes for positively correlated with both proportion and cytotoxicity of $CD8^+$ T cells in HNSCC, and was mainly expressed in a subset of $CX3CR1^+CD8^+$ T cells. Moreover, we also demonstrated S1PR4 as an immune modulator for favorable prognosis of HNSCC patients. We found that S1PR4 was highly expressed in $CD8^+$ T cells from the tumor of HNSCC patients which was greatly associated with better prognosis, and S1PR4 expressing accompanied with higher T cell cytotoxic markers expression (IFNG and GZMB). Most importantly, S1PR4 was co-localization with CX3CR1, which has been identified as the most cytotoxic marker of $CD8^+$ T cell. Taken together, these results identified S1PR4 as a key indicator for cytotoxicity and proportion of tumor infiltrating $CD8^+$ T cells, and confirmed the prognostic value of S1PR4 in HNSCC. Our findings contributed to the knowledge of S1PR4 in anti-tumor immunity, providing a potential GPCR-targeted therapeutic option for future HNSCC treatment.

Introduction

G protein-coupled receptors (GPCRs) belong to a superfamily of cell surface signaling proteins, which were found to be overexpressed in primary and metastatic tumor cells of head and neck squamous cell carcinoma (HNSCC), non-small cell lung cancer, breast, prostate and gastric tumors, and GPCRs variants can lead to increased cancer risk[1, 2]. Meanwhile, GPCRs can be activated by various important active mediators in the immune or inflammatory response such as chemokines, S1P and lysophosphatidic acid (LPA), play a key role in cell proliferation, transformation, angiogenesis and metastasis as well as in inflammation-associated cancer[1]. However, some GPCRs, such as OX1R, were shown to promote robust apoptosis in various colon cancer cells[3]. So far, in human genome, more than 800 membrane proteins have been identified as GPCRs, and the role of numerous GPCRs in tumor formation and progression is still unclear[4]. Accordingly, an improved understanding of GPCRs' involvement in tumor formation and progression may contribute to the development of a new generation of anti-tumor drugs.

Due to the heterogeneity of tumors, different types of tumors often have their own characteristics. For this reason, the therapeutic targets developed for specific tumor

types often have better clinical effects. In our study, we focused on the head and neck squamous cell carcinoma (HNSCC) that is the sixth most common malignancy worldwide, which accounts for about 380,000 deaths and 600,000 new cases annually[5, 6]. Head and neck squamous cell carcinoma (HNSCC), has highly heterogeneous and immunosuppressive features, and is characterized by high morbidity and low survival rates with limited therapeutic options outside of regional surgery, conventional cytotoxic chemotherapy, and irradiation[7-9]. Meanwhile, tumor microenvironment (TME) play a key role in HNSCC progression. The tumor microenvironment of HNSCC has typical features, such as an amplification in circulating Tregs, high level of TAMs, and express high levels of TGFβ and VEGF, which result in Treg cell activation, NK cell suppression, and suppression of DC maturation, and T-lymphocyte inactivation and dysregulation[10, 11]. High levels of cytotoxic $CD8^+$ T lymphocytes is a favorable prognostic factor in oropharyngeal cancer[12]. Most importantly, some GPCRs, such as CX3CR1, also known as G-protein coupled receptor 13 (GPR13), could improve cytotoxic of $CD8^+$ T cells[13, 14]. Therefore, our study would like to explore the potential key GPCRs involved in HNSCC development, providing a potential GPCR-targeted therapeutic option for HNSCC treatment.

Recently, multiple machine learning methods, including Non-negative Matrix Factorization (NMF), least absolute shrinkage and selection operator (LASSO), and Random Forest (RF), have been widely used for exploration of key biomarkers in tumor patients[15]. With the methods above, previous studies could classify the tumor patients into different molecular subtypes, and could further identify the potential key biomarkers by comparing the classified subtypes. However, in previous studies, the gene list of specific biological function or pathway (e.g. ferroptosis, N6-methyladenosine, or hypoxia) was often determined in advance. After that, the molecular subtypes of patients were further classified based on these pre-selected genes. In other words, at the beginning of these studies, the biological differences had been limited to a specific biological function or pathway, to which the potential key biomarkers were anticipated to be related. In our study, we decided not to limit the biological function or pathway of genes, and not to anticipate biological differences between molecular subtypes. Thus, we classified the patients into different molecular subtypes based on GPCRs, which were determined by their structures instead of specific biological functions, and were regarded as the most prominent family of pharmacological targets.

Here, we first classified HNSCC patients into two different GPCR-based molecular subtypes only from a perspective of statistics. Then, we compared the differences between two subtypes from clinical and biological perspective, and further identified potential key GPCRs in determining these two subtypes. After that, we also explored the prognostic value of the key GPCR, and its association with immune microenvironment. Eventually, we validated our findings using an independent cohort of 59 HNSCC patients from our institution via immunohistochemistry (IHC), immunofluorescence (IF), and Kaplan-Meier survival analysis (K-M analysis). We find that S1PR4 could as a key regulator to improve the anti-tumor response of tumor infiltrating $CD8^+$ T cell.

Materials and Methods

Ethics

In this study, an independent cohort was used for validation, including 59 HNSCC patients with a median follow-up time of 4.17 years. All the patients provided the written informed consent before surgery. This study was reviewed and approved by Shanghai Ninth People's Hospital, Shanghai, China. The reference number of ethics approval was SH9H-2022-T189-1.

Data Collection

We retrieved the gene expression profiles of TCGA database through UCSC Xena (http://xena.ucsc.edu/). Htseq-count profiles and Fragments per Kilobase per Million mapped reads (FPKM) profiles of TCGA-HNSCC cohort were downloaded, including 500 primary HNSCC samples, 2 metastatic HNSCC sample, and 44 solid normal tissue samples. After that, the Htseq-count and FPKM profiles of the primary HNSCC samples were retrieved. The survival data and clinical demographic information of the primary HNSCC patients was also acquired through UCSC Xena.

NMF Clustering

A total of 500 HNSCC patients were clustered into different subtypes using NMF method. The R package “NMF” was used to perform NMF clustering with the “brunet” criterion and 100 interactions[16, 17]. The number of clusters, rank k, was set as 2 to 10. The minimum number of patients in each subtype was set as 10. The NMF rank survey was performed for the optimal cluster numbers, and the stability of clusters was reflected by cophenetic correlation coefficient. According to literature investigation, the cophenetic correlation coefficient was currently the most common approach for determining the optimal values of rank k[16]. The values of rank k could be selected where the cophenetic correlation coefficient started decreasing, or was found to decreased most sharply. In this study, both conditions occurred at rank $k = 2$. Thus, we selected rank $k = 2$ as the optimal value, clustering the samples into two subtypes.

Identification of Differentially Expressed Genes

We used the R package “edgeR”[18] to identify the differentially expressed genes (DEGs) between the molecular subtypes. For the screening of DEGs, statistical significance was set at FDR $p < 0.05$. And the log2 (Fold Change) > 1 or < -1 were defined either upregulated or downregulated, respectively. Heatmap and volcano plot were used for the visualization of DEGs based on the R package “pheatmap” (https://CRAN.R-project.org/package=pheatmap) and the R package “EnhanceVolcano” (https://github.com/kevinblighe/EnhancedVolcano).

Gene Set Enrichment Analysis

The algorithm of Gene Set Enrichment Analysis (GSEA) was used to explore the difference in pathway expression between the molecular subtypes[19]. The gene sets of

HALLMARK pathways, KEGG pathways, and GO pathways were identified by Molecular Signatures Database (http://www.gsea-msigdb.org/). And the above gene sets were obtained by using the R package "msigdbr" (https://CRAN.R-project.org/package=msigdbr). The R package "clusterProfiler"[20, 21] and "enrichplot" (https://github.com/GuangchuangYu/enrichplot) were used to perform GSEA algorithm between the molecular subtypes. Statistical significance was set at p.adjust < 0.05.

Estimation of Immune Infiltrating Microenvironment

The ESTIMATE scores were employed for evaluating the immune infiltrating microenvironment for each sample. After that, the MCPcounter algorithm[22] was used to further estimate the proportions of infiltrating immune cells in HNSCC samples. And to minimize the bias of using a single algorithm, another four algorithms (CIBERSORT[23], CIBERSORT-ABS[24], QUANTISEQ[25], and TIMER algorithms[26]) were further performed for validation. The R package "limma"[27] was used to identify the immune cells with significant changes in proportion between the molecular subtypes. And the R package "factoextra" (https://CRAN.R-project.org/package=factoextra) was used for Principal Component Analysis (PCA).

Construction of Prediction Nomogram

The R package "glmnet"[28] was used to perform the LASSO regression, ensuring that there was no overfitting in the model. After that, the reduced multi-cox model was constructed by the R packages "Survival" (https://CRAN.R-project.org/package=survival). Then, based on the model, the prediction nomogram was visualized by the R package "rms" (https://CRAN.R-project.org/package=rms). With the R package "nomogramEx" (https://CRAN.R-project.org/package=nomogramEx), the acceptable accuracy of prediction nomogram was validated by calibration curves of 3-year-survival and 5-year-survival.

Integrative Analysis for Key GPCR Exploration

To explore the key GPCRs, we proposed a strategy of "classification before analysis". We first clustered the HNSCC patients into different subtypes based on the genes with a similar structure, instead of the genes related to specific biological functions/pathways. Then, from multiple perspectives, we confirmed the differences of characteristics between the subtypes. After that, we analyzed the principle of clustering, and identified the GPCRs that might be essential in the procedure of clustering. Eventually, among these GPCRs, we searched for the key GPCR which was closely correlated with the featured characteristics.

The detailed steps included: **1)** The list of currently known GPCRs was downloaded from the GPCR NaVa database (http://nava.liacs.nl)[29], and gene expressions of the recorded GPCRs were retrieved from the Htseq-count profiles of primary HNSCC patients. **2)** These GPCRs were filtered by their average expressions, and the GPCRs with an average expression > 1 were further filtered by univariate cox analysis. **3)** Based on the retained GPCRs above, we performed NMF method to cluster the HNSCC patients into different subtypes. The K-M analysis was used to confirm the

difference in prognosis between the subtypes. The differences of characteristics (gene expressions, pathway expressions, tumor microenvironment, and immune cells) between the subtypes were also investigated by using the methods which had been described in former sections. **4)** After confirming the significant differences between the subtypes, we then analyzing the principle of clustering. Because the subtypes were clustered based on the featured GPCRs in step 3, these GPCRs were further investigated. We used the LASSO regression and the reduced muli-cox regression to screen the potential key GPCRs. **5)** With the R package "randomForest" (https://CRAN.R-project.org/package=randomForest), we employed the RF method to compare the importance of the potential key GPCRs. The prognostic value of these GPCRs were also compared by using K-M analysis, and the cutoffs were determined by either maximally selected rank statistics or the median gene expressions. The Pearson correlation analysis of these GPCRs and the $CD8^+$ T-cell-related markers (CD8A, CD8B, IFNG, and GZMB) was also performed.

Multi-database Analysis

To identify the cell subset in which the key GPCR was mainly expressed, we acquired and analyzed the single cell RNA-seq profiles of T cells in pan-cancer through the scDVA data portal (http://cancer-pku.cn:3838/PanC_T/). The meta-cluster of $CD8^+$ T cells, proposed by Zheng et al.[13], was visualized by the default settings of scDVA data portal. The distribution of the key GPCR was analyzed and visualized by the "Embedding" function of the scDVA data portal.

In this study, we also used multiple databases to validate our results, preventing the bias of only using a single database. Prognostic value of the key GPCR was validated by several databases, including the Timer database (http://timer.cistrome.org/)[26, 30, 31], the LinkedOmics database (https://linkedomics.org/)[32], and the HumanProteinAtlas database (https://www.proteinatlas.org/). Correlations between the key GPCR and several marker genes were validated by the Gepia database (http://gepia.cancer-pku.cn/)[33].

IHC Staining

IHC staining was performed to detect the protein expressions of IFNG and GZMB in our independent HNSCC cohort. The 59 paraffin-embedded HNSCC samples were incubated overnight at 4°C with the primary antibodies of IFNG (Abcam, ab231036) and GZMB(Abcam, ab134933), respectively. Then, slides were washed for three times, and further incubated with the secondary antibody for 1 hour at room temperature. After that, DAB (Service Bio, G1211) was used to stain the slides, and hematoxylin was used for nuclear staining. The IHC staining results were detected with the microscope from the ServiceBio platform (ServiceBio, Wuhan, China).

IF Staining

Parraffin-embedded tissues of 59 HNSCC samples from our independent cohort were used for the IF staining. The primary antibodies, including S1PR4 (Affinity, DF4872), CD8 (ServiceBio, GB13068), CX3CR1 (ServiceBio, GB11861), IFNG (Abcam,

ab231036), and GZMB (Abcam, ab134933), were used to detect the target proteins. DAPI (ServiceBio, G1012) was used for counterstaining of the nuclei. The IF statining results were detected with the microscope from the ServieBio platform (ServiceBio, Wuhan, China).

Statistical Analysis

Statistical significance was set at adjusted $p < 0.05$, or $p < 0.05$. The correlation threshold was set at $r < -0.2$ or $r > 0.2$ in Pearson or Spearman correlation analysis. The Shapiro-wilk normality test was applied for checking the variable normality. For the variables which distributed non-normally, we applied wilcoxon rank-test for comparisons of two independent groups. For the normally distributed variables, we applied Student's *t*-test for comparisons of two independent groups. In the K-M analysis, the optimal cutoffs were determined by the median gene expressions, or by the maximally selected rank statistics. When the K-M analysis was performed in TCGA-HNSCC cohort, which had a relatively larger number of patients, both methods of cutoff determination were used. When the K-M analysis was performed in our independent cohort, which only included 59 patients, we chose to use the median gene expressions as cutoffs. The method of maximally selected rank statistics was performed by the R package "survminer" (https://cran.r-project.org/package=survminer). All the statistical analyses in this study was implemented in the software R (Version 3.5.1, http://www.r-project.org/).

Results

Identification of GPCR-Based Molecular Subtypes *via* NMF Method

NMF is an effective machine learning method, which has been widely used in computational biology for molecular pattern discovery, class comparison, and class prediction[34, 35]. Here, we applied NMF method to explore the GPCR-based molecular subtypes in TCGA-HNSCC cohort.

We first obtained a list of currently known GPCRs from GPCR NaVa database (http://nava.liacs.nl). And 90 GPCRs, with an average expression > 1 in TCGA-HNSCC cohort, were filtered for further analysis. Further filtering procedures were performed using univariate cox analysis, and 17 GPCRs with significant prognostic values of 5-year survival were retained for the NMF clustering ($p < 0.05$, **Figure 1A**). We performed NMF rank survey (**Figure 1B** and **Figure S1**) for clustering prediction, and noticed that the magnitude of cophenetic correlation coefficient began to fall sharply at rank $k = 2$ (**Figure 1B**). According to the sharp decrease, the substantially less stability was indicated to be achieved with more than two clusters. On the other hand, the consensus maps of NMF clustering (rank k was set as 2 to 10) also indicated that the clustering results were better when rank $k = 2$ (**Figure S1**). Therefore, we decided to select rank $k = 2$ as the optimal value, and clustered the HNSCC patients into two GPCR-based molecular subtypes (C1 and C2, **Figure 1C**). The gene expressions of 17 featured GPCRs used in NMF clustering were visualized through

heatmap (**Figure 1D**) and volcano plot (**Figure 1E**). After NMF clustering, there were 363 patients in subtype C1, and 137 patients in subtype C2.

Differences of the Characteristics between GPCR-Based Molecular Subtypes

Because the NMF is a machine learning method without considering biological significance, we then further explored the difference between the identified subtypes (C1 and C2) from the perspective of biomedicine.

Firstly, we compared the prognostic difference between the two subtypes. According to K-M analysis, the overall survival (OS) time of HNSCC patients in subtype C1 was significantly shorter ($p < 0.05$, **Figure 2A**). Similar, we also found that the progression free survival (PFS) time of subtype C1 tended to be shorter ($p < 0.05$, **Figure 2B**). Considering both OS and PFS, we found that the prognosis of patients in subtype C2 was significantly better.

Then, we explored the differences in gene and pathway expression between the two subtypes. With the R package edgeR, a total of 914 genes were identified as DEGs, including 681 and 233 upregulated and downregulated DEGs, respectively (**Figure 2C** and **Figure 2D**). To explore the difference of pathway expression, we performed the GSEA algorithm on KEGG pathways and HALLMARK pathways. As shown in **Figure 2E** and **Figure 2F**, many immune-related pathways were found to be upregulated in subtype C2 ($p < 0.05$). Then, we also compared several classical series of genes related to antigen presentation, cell receptor, ligand, cell adhesion, co-inhibition, and co-stimulation between the two subtypes. Most of these genes were significantly upregulated in subtype C2 (**Figure 2G**). Therefore, we hypothesized that the difference of tumor immune microenvironment (TIME) might be one of the main differences between the two subtypes.

Cytotoxicity and Proportion of $CD8^+$ T Cells were enhanced in Subtype C2

We then proceeded to explore the difference of TIME between the GPCR-based molecular subtypes. The ESTIMATE scores were important indicators for the overall changes in TIME. We found that three immune-related ESTIMATE scores (ESTIMATEScores, ImmuneScores, and StromalScores) were significantly increased in subtype C2, while the tumor-related ESTIMATE score (TumorPurity) was relatively lower in subtype C2 ($p < 0.001$, **Figure 3A**). We also used the method proposed by Petitprez et al.[36] to estimate the gene signatures related to the functions of TIME, and the gene expressions related to immune checkpoints. The genes associated with T cell activation and survival, Class I MHC, regulatory gene signatures, tertiary lymphoid structures (TLSs), and several immune checkpoint-related genes were all found to be higher in subtype C2 (**Figure 3B**). The above findings indicated that the subtype C2 had an “immune and TLS high” characteristic, which was once described and proved to be correlated with better prognosis by Petitprez et al.[36] in sarcoma.

After discovering the differences at overall level of TIME, we further investigated the differences of tumor-infiltrating immune cell (TIIC) proportions between the two subtypes. To render the TIICs of each HNSCC patients comparable, we estimated the scaled proportion of TIICs using the MCPcounter algorithm (**Figure 3C**). The

proportion of TIICs differed significantly between subtypes (**Figure 3C** and **Figure 3D**). And the volcano plot indicated that, among all the TIICs, the largest difference in proportion was $CD8^+$ T cells (**Figure 3E**). The significant difference of $CD8^+$ T cells was validated using four algorithms, including CIBERSORT (**Figure 3F**), CIBERSORT-ABS (**Figure 3G**), QUANTISEQ (**Figure 3H**), and TIMER (**Figure 3I**). And all the algorithms above confirmed that $CD8^+$ T cells had the largest range of changes in proportion between the two subtypes. In short, the subtype C2 had a significant higher proportion of $CD8^+$ T cells. Moreover, the GSEA analysis was performed to demonstrate the difference of T cell functions. The functional pathways of $CD8^+$ T cells were all significantly upregulated in subtype C2 ($p < 0.05$, **Figure 3J**). In summary, both the proportion and the cytotoxicity of $CD8^+$ T cells were significantly elevated in the subtype C2, which was consistent with our previous result that subtype C2 had a better prognosis.

Identification of the Key GPCR S1PR4 in HNSCC

Next, we aimed to identify potential key GPCRs by analyzing the principle of grouping two GPCR-based molecular subtypes. Considering that the subtypes were clustered based on GPCRs, we further investigated the 17 featured GPCRs used in the initial NMF clustering. First, the survival analysis indicated that all the 17 GPCRs showed prognostic values of HNSCC patients' OS (**Figure S2**). Then, these GPCRs were further processed using LASSO method (**Figure 4A** and **Figure 4B**). Based on the results of LASSO regression, five GPCRs (CXCR6, GPR87, MRGPRF, S1PR4, and S1PR5) were regarded as eligible, and were further included into reduced-cox regression model (C-Index = 0.61, $p < 0.05$, **Figure 4C**). The reduced-cox model indicated that GPR87 ($p = 0.048$), S1PR4 ($p = 0.030$), and S1PR5 ($p = 0.029$) could effectively evaluate the HNSCC patients' prognosis of OS. To be more exact, HNSCC patients with higher S1PR4 or S1PR5 expression tended to have better prognosis of OS (Hazard ratio < 1), while patients with higher GPR87 expression tended to have poorer prognosis of OS (Hazard ratio > 1). Moreover, we also constructed a nomogram based on GPR87, S1PR4, and S1PR5 (**Figure 4D**). And the acceptable accuracy of this nomogram was validated by the calibration curves of 3-year-survival (**Figure 4E**) and 5-year-survival (**Figure 4F**) of HNSCC patients.

Considering that GPR87, S1PR4, and S1PR5 were all potential key GPCRs in HNSCC, we further compared these three genes from several perspectives to identify the most important one. Firstly, all of the three GPCRs were found to effectively evaluate the prognosis of OS ($p < 0.05$, **Figure 5A**) and PFS ($p < 0.05$, **Figure 5B**) in HNSCC patients. Secondly, we found that only GPR87 ($p < 0.05$) and S1PR4 ($p < 0.05$) were identified as DEGs between two subtypes (**Figure 5B**). Although S1PR5 was one of the featured GPCRs in NMF clustering, it was not differentially expressed between subtypes (**Figure 5B**). Thirdly, we also performed another machine learning approach, the RF method, to compare the importance of these three genes in classifying the two molecular subtypes. The results of RF method identified S1PR4 as the most important GPCR (**Figure 5D**), suggesting that S1PR4 was a key GPCR determining the aforementioned subtypes. Most patients of subtype C1 belonged to the S1PR4-Low-

Expression groups (**Figure S3**). More importantly, the correlation analysis indicated that, among the three GPCRs, only the S1PR4 was significantly correlated with the proportion and cytotoxicity markers of $CD8^+$ T cells ($r > 0.4$, $p < 0.05$, **Figure 5E**). Collectively, we identified S1PR4 as the potential key GPCR in HNSCC, and chose S1PR4 for further analysis.

S1PR4 as a Key Indicator for Cytotoxicity and Proportion of $CD8^+$ T cells

In aforementioned results, we had identified the correlations between S1PR4 and $CD8^+$ T-cell-related markers (CD8A, CD8B, IFNG, and GZMB). To confirm these correlations, we used several algorithms to estimate the scores of the cytotoxicity and proportion of $CD8^+$ T cells. Consistent with our previous results, the S1PR4 was found to be significantly correlated with these scores ($p < 0.05$, **Figure 6A**), which indicated that S1PR4 might be a key indicator for cytotoxicity and proportion of $CD8^+$ T cells. Therefore, we decided to explore the relationship between S1PR4 and $CD8^+$ T cells more comprehensively.

On the one hand, we used the public single cell RNA-seq profiles to explore the S1PR4 expression in $CD8^+$ T cells. In previous studies, the $CX3CR1^+CD8^+$ T cells had been proved to have the highest cytotoxicity score among $CD8^+$ T cells[13, 14]. Interestingly, in our study, S1PR4 was found to be mainly expressed in a subset of $CX3CR1^+CD8^+$ T cells among $CD8^+$ T cells (**Figure 6B**). And the correlation between S1PR4 and CX3CR1 was also validated through the bulk RNA-seq profiles of TCGA ($p < 0.05$, **Figure 6C**). On the other hand, recent studies had proved that the TLS structure was an important indicator for the cytotoxicity and proportion of $CD8^+$ T cells in TIME. And the gene signatures of TLS had been already defined by either single gene (CXCL13) or multiple genes (CCL2, CCL3, CCL4, CCL5, CCL8, CCL18, CCL19, CCL21, CXCL9, CXCL10, CXCL11, and CXCL13)[36, 37]. Therefore, we also investigated the correlation between the TLS and S1PR4. We found that the expression of S1PR4 was significantly correlated with all the TLS signatures ($p < 0.05$, **Figure 6D**), including CXCL13 ($r = 0.5$, $p < 0.05$, **Figure 6E**).

Multiple Database Validation

To prevent the bias from only using a single database, multiple databases were used for validation. The prognostic value of S1PR4 was validated by using TIMER database ($p < 0.05$, **Figure 6F**), LinkedOmics database ($p < 0.05$, **Figure 6G**), and HumanProteinAtlas database ($p < 0.05$, **Figure 6H**), respectively. Moreover, we validated the correlations between S1PR4 and $CD8^+$T-cell-related markers (CD8A, CD8B, IFNG, GZMB, and CX3CR1) using the Gepia database (**Figure 6I-M**). Similarly, the correlations between S1PR4 and TLS signatures were also validated (**Figure 6N** and **Figure S4**).

Independent HNSCC Cohort Validation

To validate our bioinformatic findings, an independent cohort from our institution was used. A total of 59 HNSCC patients, with a median follow-up time of 4.17 years, were included in this independent cohort.

First, we explored the correlations among S1PR4, $CD8^+$ T cells, and several cytotoxic markers (IFNG and GZMB). Based on IHC staining, the IFNG and GZMB expression were quantitated in protein level (**Figure 7A**). Similarly, the protein expression of CD8 and S1PR4 were also quantitated in the 59 HNSCC patients by IF staining (**Figure 7B**). The heatmap showed the relative expression of the above markers (**Figure 7C**). And the correlation heatmap showed that S1PR4 was significantly correlated with CD8 ($r = 0.82$, $p < 0.05$), IFNG ($r = 0.25$, $p < 0.05$), and GZMB ($r = 0.38$, $p < 0.05$), indicating that S1PR4 was essential for both proportion and cytotoxicity of $CD8^+$ T cells (**Figure 7D**).

Then, we performed K-M analysis, confirming that HNSCC patients with higher S1PR4 protein expression tended to have better prognosis of OS ($p < 0.05$, **Figure 7E**), which was consistence with our bioinformatic findings.

After statistically confirming the correlations among S1PR4, $CD8^+$ T cells, and the cytotoxic markers, we further explored the spatial distributions among S1PR4 protein, $CD8^+$ T cells, and the cytotoxic markers. Based on IF staining experiments, the co-localization of S1PR4 with $CD8^+$ T cells was confirmed (**Figure 7B**), suggesting that S1PR4 might be essential for $CD8^+$ T cells. In addition, considering that we had found that S1PR4 was mainly expressed in a subset of $CX3CR1^+CD8^+$ T cells, we further validated the co-localization between S1PR4 and CX3CR1 (**Figure 7F**). According to the aforementioned literature research results, the $CX3CR1^+CD8^+$ T cells had already been identified as the most cytotoxicitic subset of $CD8^+$ T cells. Therefore, we also confirmed the co-localization between S1PR4 and the important cytotoxic markers of $CD8^+$ T cells (IFNG and GZMB, **Figure 7G**). Collectively, the S1PR4 was found to be expressed in the most cytotoxic subset of $CD8^+$ T cells ($CX3CR1^+CD8^+$ T cells), and was also found to have co-localizations with the important cytotoxic markers (IFNG and GZMB) of $CD8^+$ T cells. In short, S1PR4 was found to play an important role in the $CX3CR1^+CD8^+$ T cell-mediated tumor killing, which was essential for anti-tumor immune regulation in tumor microenvironment.

Discussion

The GPCR surperfamily is one of the most important families of human proteins. Regulating a wide range of physiological processes, it had been known to be essential in the progression of multiple tumor types, including HNSCC. In addition, as cell surface proteins, GPCRs had broad prospects in the field of drug development. However, currently, the biological functions of many GPCRs still remain unclear, and further study of its involvement in tumor progression is still warranted. Therefore, our study aimed to explore the key GPCRs involved in HNSCC progression, providing a potential GPCR-targeted therapeutic option for tumor treatment.

In this study, we adopted an interesting strategy of “classification before analysis”. First, from a statistical perspective, we classified the HNSCC patients into two molecular subtypes only based on their gene expression patterns. Then, from clinical and biological perspectives, we comprehensively compared the two subtypes in terms of prognosis, gene expression, pathway expression, TIME, and immune cell infiltration.

Eventually, by exploring the featured genes between the two subtypes, we identified S1PR4 as a potential key GPCR in HNSCC.

It is widely known that the GPCR superfamily is defined based on the characteristic of seven-transmembrane structure, and the biological function of each GPCR varies. When using the strategy of "classification before analysis", we did not limit the biological differences in the initial classification of molecular subtypes. Therefore, in the process of classification, we could only determine that there were statistical differences in HNSCC patients' gene expression patterns, but could not determine the differences in specific biological functions between the two molecular subtypes. For comparison, previous similar studies often obtain the gene list of specific biological function or pathway in advance, and then classify the molecular subtypes based on these selected genes. Compared with this strategy of "determining the biological functions before classification", our strategy did not anticipate biological differences between subtypes, which made our strategy relatively more objective.

With the strategy above, our study identified S1PR4 as a key GPCR in HNSCC. The GPCR S1PR4 belongs to the endothelial differentiation G protein coupled (EDG) receptors, which ligands includes lysophosphingolipids and sphingolipids[38]. In the past decades, EDG receptors have been proved to be participated in cell signal transductions of multiple cell types, indicating its importance in physiological and pathological regulations. As a member of EDG receptors, S1PR4 binds sphingosine-1-phosphate (S1P) as ligand, and has been found to play a role in immune regulation.

In previous studies, S1PR4 had been reported to regulate the migration and proliferation of immune cells, especially T cells. For instance, Graeler et al. and Xiong et al. had proved that S1PR4 plays an important role in promoting T cell migration[39, 40]. And Schulze et al. demonstrated that the deletion of S1PR4 would affect S1P induced $CD8^+$ T cell migration[41]. Meanwhile, Olesch et al. showed that the ablation of S1PR4 promoted the expansion of $CD8^+$ T cells[42]. In short, S1PR4 could not only promote $CD8^+$ T cell migration, but also inhibit $CD8^+$ T cell expansion. Therefore, the relationship between S1PR4 and the proportion of tumor-infiltrating $CD8^+$ T cells depended on the joint action of the above two regulatory mechanisms, which had opposite effects on the regulation of $CD8^+$ T cell proportion. Of note, a previous study presumed that S1PR4 might play a minor role in lymphocyte migration rather than expansion[43]. And a recent study further endorsed the possibility that S1PR4 could drive CXCR4, affecting lymphocyte differentiation (eg., expansion) rather than lymphocyte trafficking[44]. However, the relationship between S1PR4 and the proportion of $CD8^+$ T cells in HNSCC remains largely unknown. Our study identified that, at the tissue level of HNSCC samples, S1PR4 is positively correlated with the proportion of tumor-infiltrating $CD8^+$ T cells, which contributed to anti-tumor immunity. Considering that HNSCC was a special tumor type with relatively more lymphocytes infiltration, we hypothesized that S1PR4 might play a major role in $CD8^+$ T cell migration rather than expansion in HNSCC patients.

Moreover, we also identified that S1PR4 was positively correlated with the cytotoxicity of $CD8^+$ T cells in HNSCC, and further found that S1PR4 was mainly expressed in a subset of $CX3CR1^+CD8^+$ T cells. Interestingly, in the literature, the

$CX3CR^+CD8^+$ T cells were identified as a high migration subset, and were further reported to have the highest cytotoxicity score among $CD8^+$ T cells[13, 14]. On the other hand, in recent years, many studies had confirmed that the TLS was an important effecter of anti-tumor immunity, and could promote the cytotoxicity of $CD8^+$ T cells[36, 45, 46]. In our study, S1PR4 was also found to be positively correlated with gene signatures of TLSs in HNSCC. Collectively, our findings were in accordance with the above literatures, suggesting that S1PR4, mainly expressed in a $CX3CR1^+CD8^+$ T cells, might be essential for the cytotoxic functions of tumor-infiltrating $CD8^+$ T cells.

Several limitations in this study need to be addressed. First, this study only focused on the relationship between S1PR4 and $CD8^+$ T cells, and did not further analyzed other types of immune cells. Second, most of the findings in this study were based on bioinformatics, and did not further investigate the underlying mechanisms. In future studies, the biological mechanisms of S1PR4 regulating $CD8^+$ T cells' cytotoxicity should be further investigated experimentally. Despite the limitations mentioned above, this study was the first to infer that S1PR4 might be a key GPCR in HNSCC. We found that S1PR4 could effectively evaluated the prognosis of both OS and PFS for HNSCC patients. And we also found that S1PR4 was mainly expressed in $CX3CR1^+CD8^+$ T cells, and was positively correlated with both proportion and cytotoxicity of $CD8^+$ T cells in HNSCC. The above findings were eventually validated using an independent cohort from our institution, further suggesting that S1PR4 might be a potential therapeutic target for HNSCC treatment.

In summary, the present study proposed a strategy of "classification before analysis", and identified S1PR4 as a key GPCR for HNSCC patients. We found that S1PR4 was a key GPCR affecting HNSCC patients' prognosis, and further constructed a prognosis-predicting nomogram based on three variables including S1PR4. We also found that S1PR4 was mainly expressed in $CX3CR1^+CD8^+$ T cells, and was positively correlated with both proportion and cytotoxicity of $CD8^+$ T cells in HNSCC. Our findings contributed to the knowledge of S1PR4 in anti-tumor immunity, and might provide a potential GPCR-targeted therapeutic option for future HNSCC treatment.

Competing interests.

The authors declare no competing interests.

Author Contributions

Conception/design: CH, HC. Literature research: CH, YQ, DL, JZ. Data acquisition: CH, ZZ, FZ, YQ, WN. Data analysis/interpretation: CH, ZZ, FZ, HC, YQ, XK. Experimental studies: CH, ZZ, FZ, YQ. Writing, original draft: CH. Writing, review and editing: CH, ZZ, QH, JQ, BD. All authors read and approved the submitted version of manuscript.

Acknowledgements

The authors thank Yue Tian (East China Normal University, Shanghai, China) and Runzhi Huang (Tongji University, Shanghai, China) for generously sharing their experience.

Funding

This study was supported by National Natural Science Foundation of China (Grant No. 31770969 to BD, Grant No. 81902892 to JQ, and Grant No. 82173099 to JQ).

Data Accessibility

Publicly available datasets were analyzed in this study. Gene expression profiles of HNSCC patients could be acquired from UCSC Xena (https://xenabrowser.net/). Further inquiries could be directed to the corresponding authors.

Ethics Statement

This study was reviewed and approved by Shanghai Ninth People's Hospital, Shanghai, China. The reference number of ethics approval was SH9H-2022-T189-1.

References

1. *Lappano, R. and M. Maggiolini, G protein-coupled receptors: novel targets for drug discovery in cancer. Nat Rev Drug Discov, 2011.* ***10****(1): p. 47-60.*
2. *Lappano, R. and M. Maggiolini, GPCRs and cancer. Acta Pharmacol Sin, 2012.* ***33****(3): p. 351-62.*
3. *Voisin, T., et al., Aberrant expression of OX1 receptors for orexins in colon cancers and liver metastases: an openable gate to apoptosis. Cancer Res, 2011.* ***71****(9): p. 3341-51.*
4. *Jassal, B., et al., The systematic annotation of the three main GPCR families in Reactome. Database (Oxford), 2010.* ***2010****: p. baq018.*
5. *Miyauchi, S., et al., Immune Modulation of Head and Neck Squamous Cell Carcinoma and the Tumor Microenvironment by Conventional Therapeutics. Clin Cancer Res, 2019.* ***25****(14): p. 4211-4223.*
6. *Solomon, B., R.J. Young, and D. Rischin, Head and neck squamous cell carcinoma: Genomics and emerging biomarkers for immunomodulatory cancer treatments. Semin Cancer Biol, 2018.* ***52****(Pt 2): p. 228-240.*
7. *Bhat, A.A., et al., Tumor microenvironment: an evil nexus promoting aggressive head and neck squamous cell carcinoma and avenue for targeted therapy. Signal Transduct Target Ther, 2021.* ***6****(1): p. 12.*
8. *Elmusrati, A., J. Wang, and C.Y. Wang, Tumor microenvironment and immune evasion in head and neck squamous cell carcinoma. Int J Oral Sci, 2021.* ***13****(1):*

p. 24.

9. *Johnson, D.E., et al., Head and neck squamous cell carcinoma. Nat Rev Dis Primers, 2020.* ***6****(1): p. 92.*
10. *Hanakawa, H., et al., Regulatory T-cell infiltration in tongue squamous cell carcinoma. Acta Otolaryngol, 2014.* ***134****(8): p. 859-64.*
11. *Seiwert, T.Y. and E.E. Cohen, Targeting angiogenesis in head and neck cancer. Semin Oncol, 2008.* ***35****(3): p. 274-85.*
12. *Sasidharan Nair, V. and E. Elkord, Immune checkpoint inhibitors in cancer therapy: a focus on T-regulatory cells. Immunol Cell Biol, 2018.* ***96****(1): p. 21-33.*
13. *Zheng, L., et al., Pan-cancer single-cell landscape of tumor-infiltrating T cells. Science, 2021.* ***374****(6574): p. abe6474.*
14. *Liu, Y., et al., Tumour heterogeneity and intercellular networks of nasopharyngeal carcinoma at single cell resolution. Nat Commun, 2021.* ***12****(1): p. 741.*
15. *Pucher, B.M., O.A. Zeleznik, and G.G. Thallinger, Comparison and evaluation of integrative methods for the analysis of multilevel omics data: a study based on simulated and experimental cancer data. Brief Bioinform, 2019.* ***20****(2): p. 671-681.*
16. *Brunet, J.P., et al., Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A, 2004.* ***101****(12): p. 4164-9.*
17. *Gaujoux, R. and C. Seoighe, A flexible R package for nonnegative matrix factorization. BMC Bioinformatics, 2010.* ***11****: p. 367.*
18. *Robinson, M.D., D.J. McCarthy, and G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 2010.* ***26****(1): p. 139-40.*
19. *Subramanian, A., et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 2005.* ***102****(43): p. 15545-50.*
20. *Yu, G., et al., clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS, 2012.* ***16****(5): p. 284-7.*
21. *Wu, T., et al., clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb), 2021.* ***2****(3): p. 100141.*
22. *Becht, E., et al., Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol, 2016.* ***17****(1): p. 218.*
23. *Newman, A.M., et al., Robust enumeration of cell subsets from tissue expression profiles. Nat Methods, 2015.* ***12****(5): p. 453-7.*
24. *Newman, A.M., et al., Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol, 2019.* ***37****(7): p. 773-782.*
25. *Plattner, C., F. Finotello, and D. Rieder, Deconvoluting tumor-infiltrating immune cells from RNA-seq data using quanTIseq. Methods Enzymol, 2020.* ***636****: p. 261-285.*
26. *Li, B., et al., Comprehensive analyses of tumor immunity: implications for*

*cancer immunotherapy. Genome Biol, 2016. **17**(1): p. 174.*

27. *Ritchie, M.E., et al., limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res, 2015. **43**(7): p. e47.*
28. *Friedman, J., T. Hastie, and R. Tibshirani, Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw, 2010. **33**(1): p. 1-22.*
29. *Kazius, J., et al., GPCR NaVa database: natural variants in human G protein-coupled receptors. Hum Mutat, 2008. **29**(1): p. 39-44.*
30. *Li, T., et al., TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res, 2017. **77**(21): p. e108-e110.*
31. *Li, T., et al., TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res, 2020. **48**(W1): p. W509-W514.*
32. *Vasaikar, S.V., et al., LinkedOmics: analyzing multi-omics data within and across 32 cancer types. Nucleic Acids Res, 2018. **46**(D1): p. D956-D963.*
33. *Tang, Z., et al., GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res, 2017. **45**(W1): p. W98-W102.*
34. *Devarajan, K., Nonnegative matrix factorization: an analytical and interpretive tool in computational biology. PLoS Comput Biol, 2008. **4**(7): p. e1000029.*
35. *Huang, J., et al., Evaluation of gene-drug common module identification methods using pharmacogenomics data. Brief Bioinform, 2021. **22**(3).*
36. *Petitprez, F., et al., B cells are associated with survival and immunotherapy response in sarcoma. Nature, 2020. **577**(7791): p. 556-560.*
37. *Coppola, D., et al., Unique ectopic lymph node-like structures present in human primary colorectal carcinoma are identified by immune gene array profiling. Am J Pathol, 2011. **179**(1): p. 37-45.*
38. *Graler, M.H., et al., The sphingosine 1-phosphate receptor S1P4 regulates cell shape and motility via coupling to Gi and G12/13. J Cell Biochem, 2003. **89**(3): p. 507-19.*
39. *Graeler, M., G. Shankar, and E.J. Goetzl, Cutting edge: suppression of T cell chemotaxis by sphingosine 1-phosphate. J Immunol, 2002. **169**(8): p. 4084-7.*
40. *Xiong, Y., et al., CD4 T cell sphingosine 1-phosphate receptor (S1PR)1 and S1PR4 and endothelial S1PR2 regulate afferent lymphatic migration. Sci Immunol, 2019. **4**(33).*
41. *Schulze, T., et al., Sphingosine-1-phospate receptor 4 (S1P(4)) deficiency profoundly affects dendritic cell function and TH17-cell differentiation in a murine model. FASEB J, 2011. **25**(11): p. 4024-36.*
42. *Olesch, C., et al., S1PR4 ablation reduces tumor growth and improves chemotherapy via CD8+ T cell expansion. J Clin Invest, 2020. **130**(10): p. 5461-5476.*
43. *Wang, W., M.H. Graeler, and E.J. Goetzl, Type 4 sphingosine 1-phosphate G protein-coupled receptor (S1P4) transduces S1P effects on T cell proliferation and cytokine secretion without signaling migration. FASEB J, 2005. **19**(12): p. 1731-3.*
44. *Burkard, T., et al., Enhanced CXCR4 Expression of Human CD8(Low) T*

Lymphocytes Is Driven by S1P(4). Front Immunol, 2021. ***12****: p. 668884.*
45. *Helmink, B.A., et al., B cells and tertiary lymphoid structures promote immunotherapy response. Nature, 2020.* ***577****(7791): p. 549-555.*
46. *Cabrita, R., et al., Tertiary lymphoid structures improve immunotherapy and survival in melanoma. Nature, 2020.* ***577****(7791): p. 561-565.*

Figure Legends

Figure 1 | Identification of GPCR-Based Molecular Subtypes via NMF Method.
(A) Filtered by univariate cox analysis, 17 GPCRs were found to have significant prognostic values of 5-year survival ($p < 0.05$), and were retained for further analysis.
(B) The NMF rank survey was performed for clustering prediction (rank k was set as 2 to 10). The cophenetic correlation coefficient started decresing at rank k = 2, and was found to decreased most sharply at rank k = 2. Similarly, the sharp decrease of dispersion value was also found at rank k = 2.
(C) Consensus map of NMF clustering (rank k = 2). The HNSCC patients were clustered into two GPCR-based molecular subtypes, including 363 patients in subtype C1, and 137 patients in subtype C2.
(D) Gene expression of the 17 featured GPCRs was visualized through heatmap.
(E) Volcano plot of the 17 featured GPCRs. The GPCRs which were differentially expressed between the subtypes were highlighted in red.
GPCR, G protein-coupled receptor; NMF, Non-negative Matrix Factorization; HNSCC, head and neck squamous cell carcinoma.

Figure 2 | Differences of the Characteristics between GPCR-Based Molecular Subtypes.
(A) Kaplan-Meier plot showed that the OS time of HNSCC patients in subtype C1 was significantly shorter ($p < 0.05$). Patients in Subtype C2 tended to have better prognosis of OS.
(B) Kaplan-Meier plot showed that the PFS time of HNSCC patients in subtype C1 was significantly shorter ($p < 0.05$). Patients in Subtype C2 tended to have better prognosis of PFS.
(C, D) A total of 914 genes were identified as DEGs between the subtypes. The gene expression of DEGs was visualized through heatmap **(C)**. The DEGs between subtype C1 and subtype C2 were highlighted in red **(D)**, including 681 and 233 upregulated and downregulated, respectively.
(E, F) GSEA analysis was performed on KEGG pathways **(E)** and HALLMARK pathways **(F)**. The $p < 0.05$ was considered significant. **(G)** Seven classical series of genes were compared between the two subtypes. Scattered dots represented the relative gene expression values.
*GPCR, G protein-coupled receptor; OS, overall survival; HNSCC, head and neck squamous cell carcinoma; PFS, progression free survival; DEG, differentially expressed gene; GSEA, Gene Set Enrichment Analysis. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$; ns, not significant.*

713
714 **Figure 3 | Cytotoxicity and Proportion of CD8$^+$ T Cells were Upregulated in**
715 **Subtype C2**
716 **(A)** Violin plot showed the ESTIMATE scores of the two subtypes. Three immune-
717 related ESTIMATE scores (ESTIMATEScores, ImmuneScores, and StromalScores)
718 were significantly upregulated in subtype C2, while the tumor-related ESTIMATE
719 score (TumorPurity) was relatively lower in subtype C2 ($p < 0.05$).
720 **(B)** Heatmap for the expression of gene signatures, which were related to the functions
721 of TIME or to the immune checkpoints. The gene signatures associated with T cell
722 activation and survival, Class I MHC, regulatory gene signatures, TLSs, and several
723 immune checkpoint-related genes were all found to be higher in subtype C2.
724 **(C)** Heatmap for the proportion of TIICs, which were estimated by the MCPcounter
725 algorithm.
726 **(D)** PCA plot showed that the proportion of TIICs differed significantly between
727 subtypes.
728 **(E)** Volcano plot of the TIICs, which were estimated by the MCPcounter algorithm. The
729 CD8$^+$ T cells showed the largest value of fold-change, and the smallest value of p-value,
730 indicating that among all the estimated TIICs, the largest difference in proportion was
731 CD8$^+$ T cells. The proportion of CD8$^+$ T cells was significantly higher in subtype C2 (p
732 < 0.0001).
733 **(F-I)** Vocano plots of the TIICs, which were estimated by four algorithms, including
734 CIBERSORT **(F)**, CIBERSORT-ABS **(G)**, QUANTISEQ **(H)**, and TIMER **(I)**. The
735 CD8$^+$ T cells were all found to have the largest value of fold-change, as well as the
736 smallest value of p-value.
737 **(J)** GSEA analysis of the GO pathways indicated that functional pathways, especially
738 the cytotoxic function, of CD8$^+$ T cells were significantly upregulated in subtype C2 (p
739 < 0.05).
740 *TIME, tumor immune microenvironment; TLS, tertiary lymphoid structure; TIIC,*
741 *tumor-infiltrating immune cell. *** $p < 0.001$; **** $p < 0.0001$.*
742
743 **Figure 4 | GPR87, S1PR4, and S1PR5 were all potential key GPCRs in HNSCC.**
744 **(A, B)** LASSO regression was applied to prevent overfitting. Trajectories of the
745 independent variables were visualized in **(A)**, showing the coefficient of each
746 independent variable at different lambda. The confidence interval at each lambda was
747 visualized in **(B)**.
748 **(C)** Reduced multi-cox regression model was constructed based on the results of
749 LASSO regression (C-Index = 0.61, $p < 0.05$). Three variables, including GPR87
750 (Hazard ratio > 1, $p = 0.048$), S1PR4 (Hazard ratio < 1, $p = 0.030$), and S1PR5 (Hazard
751 ratio < 1, $p = 0.029$), were identified as the potential key GPCRs predicting the
752 prognosis of HNSCC.
753 **(D)** According to the reduced multi-cox model in **(C)**, a nomogram was constructed
754 based on GPR87, S1PR4, and S1PR5.
755 **(E, F)** Calibration curves of 3-year survival **(E)** and 5-year survival **(F)** indicated the
756 acceptable accuracy.

*GPCR, G protein-coupled receptor; HNSCC, head and neck squamous cell carcinoma; LASSO, Least absolute shrinkage and selection operator. * $p < 0.05$.*

Figure 5 | Identification of the Key GPCR S1PR4.

(A) Kaplan-Meier plots revealed that all of the three GPCRs (GPR87, S1PR4, and S1PR5) could effectively evaluate the prognosis of OS for the HNSCC patients. HNSCC patients with higher S1PR4 ($p < 0.0001$) or S1PR5 ($p < 0.05$) expression tended to have better prognosis of OS, while patients with higher GPR87 ($p < 0.05$) expression tended to have poorer prognosis of OS.

(B) Kaplan-Meier plots revealed that all of the three GPCRs (GPR87, S1PR4, and S1PR5) could effectively evaluate the prognosis of PFS for the HNSCC patients. HNSCC patients with higher S1PR4 ($p < 0.001$) or S1PR5 ($p < 0.01$) expression tended to have better prognosis of PFS, while patients with higher GPR87 ($p < 0.01$) expression tended to have poorer prognosis of PFS.

(C) Relative expression of all the three GPCRs (GPR87, S1PR4, and S1PR5) in subtype C1 and C2. GPR87 was significantly downregulated in subtype C2 ($p < 0.0001$), while the expression of S1PR4 was significantly higher in subtype C2 ($p < 0.0001$). S1PR5 was not differentially expressed between two subtypes ($p > 0.05$).

(D) The RF method was performed to compare the importance of the three GPCRs (GPR87, S1PR4, and S1PR5), revealing that S1PR4 was the most important GPCR in determining the two subtypes.

(E) Correlation heatmap of the three GPCRs (GPR87, S1PR4, and S1PR5) and the $CD8^+$ T cell-related markers. S1PR4 was strongly correlated with the proportion (CD8A and CD8B) and cytotoxicity markers (IFNG and GZMB) of $CD8^+$ T cells ($r > 0.4$). However, both of GPR87 and S1PR4 were weakly correlated with the above $CD8^+$ T cell-related markers ($r < 0.2$).

*GPCR, G protein-coupled receptor; OS, overall survival; HNSCC, head and neck squamous cell carcinoma; PFS, progression free survival; RF, Random Forest. **** $p < 0.0001$; ns, not significant.*

Figure 6 | Multiple Database analysis of the relationship between S1PR4 and $CD8^+$ T cells.

(A) Correlation heatmap indicated that S1PR4 was significantly correlated with the algorithm-estimated scores of $CD8^+$ T cell cytotoxicity and proportion ($p < 0.05$). The score of $CD8^+$ T cell cytotoxicity was estimated by the MCPcounter algorithm. And the scores of $CD8^+$ T cell proportion were estimated by five algorithms, including MCPcounter, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, and TIMER.

(B) S1PR4 was mainly expressed in a subset of $CX3CR1^+CD8^+$ T cells (marked in dark green).

(C) The dot plot indicated that S1PR4 was significantly correlated with CX3CR1 ($r = 0.2$, $p < 0.0001$).

(D) Correlation heatmap of S1PR4 and the gene signatures of TLS (CCL2, CCL3, CCL4, CCL5, CCL8, CCL18, CCL19, CCL21, CXCL9, CXCL10, CXCL11, and CXCL13). S1PR4 was significantly correlated with all the TLS signatures ($p < 0.05$).

(E) The dot plot indicated that S1PR4 was significantly correlated with CXCL13 ($r = 0.5$, $p < 0.0001$).
(F-H) The TIMER database **(F)** indicated that S1PR4 could effectively evaluated the prognosis of HNSCC patients ($p < 0.05$). Similarly, the prognostic value of S1PR4 was also validated by the LinkedOmics database **(G)** and the HumanProteinAtlas database **(H)**. All the three database indicated that patients with higher S1PR4 expression tended to have better prognosis ($p < 0.05$).
(I-N) The Gepia database was used to validate the correlations between S1PR4 and $CD8^+$T-cell-related markers, including CD8A **(I)**, CD8B **(J)**, IFNG **(K)**, GZMB **(L)**, and CX3CR1 **(M)**. Similarly, the Gepia database was also used to validate the correlation between S1PR4 and the TLS signature CXCL13 **(N)**.
TLS, tertiary lymphoid structure; head and neck squamous cell carcinoma.

Figure 7 | Independent HNSCC Cohort Validation.
(A) Representative images of GZMB and IFNG IHC staining in the independent HNSCC cohort.
(B) Representative images of S1PR4 and CD8 IF staining in the independent HNSCC cohort, showing the co-localization between S1PR4 and CD8.
(C) Heatmap showed the relative protein expression of S1PR4, CD8, GZMB, and IFNG in the independent HNSCC cohort.
(D) Correlation heatmap of S1PR4, CD8, GZMB, and IFNG. In protein level, S1PR4 was found to have the strongest correlation with CD8 ($r = 0.82$, $p < 0.05$). The S1PR4 was also found to be significantly correlated with both IFNG ($r = 0.25$, $p < 0.05$) and GZMB ($r = 0.38$, $p < 0.05$).
(E) Kaplan-Meier plot revealed that, in protein level, the HNSCC patients with higher S1PR4 expression tended to have better prognosis ($p = 0.00041$).
(F) Representative images of the co-localization between S1PR4 and the cytotoxic markers of $CD8^+$ T cells (IFNG and GZMB).
(G) Representative images of the co-localization between S1PR4 and CX3CR1.
HNSC, Head and neck squamous cell carcinoma.

Supplementary Material

Figure S1 | Consensus map of NMF clustering (rank k was set as 2 to 10).
Figure S2 | Kaplan-Meier plots indicated that all the 17 GPCRs showed prognostic values of HNSCC patitents' OS.
Figure S3 | Most patients of subtype C1 belonged to the S1PR4-Low-Expression groups.
Figure S4 | The Gepia database was used to validate the correlations between S1PR4 and the gene signatures of TLS (CCL2, CCL3, CCL4, CCL5, CCL8, CCL18, CCL19, CCL21, CXCL9, CXCL10, CXCL11, and CXCL13).

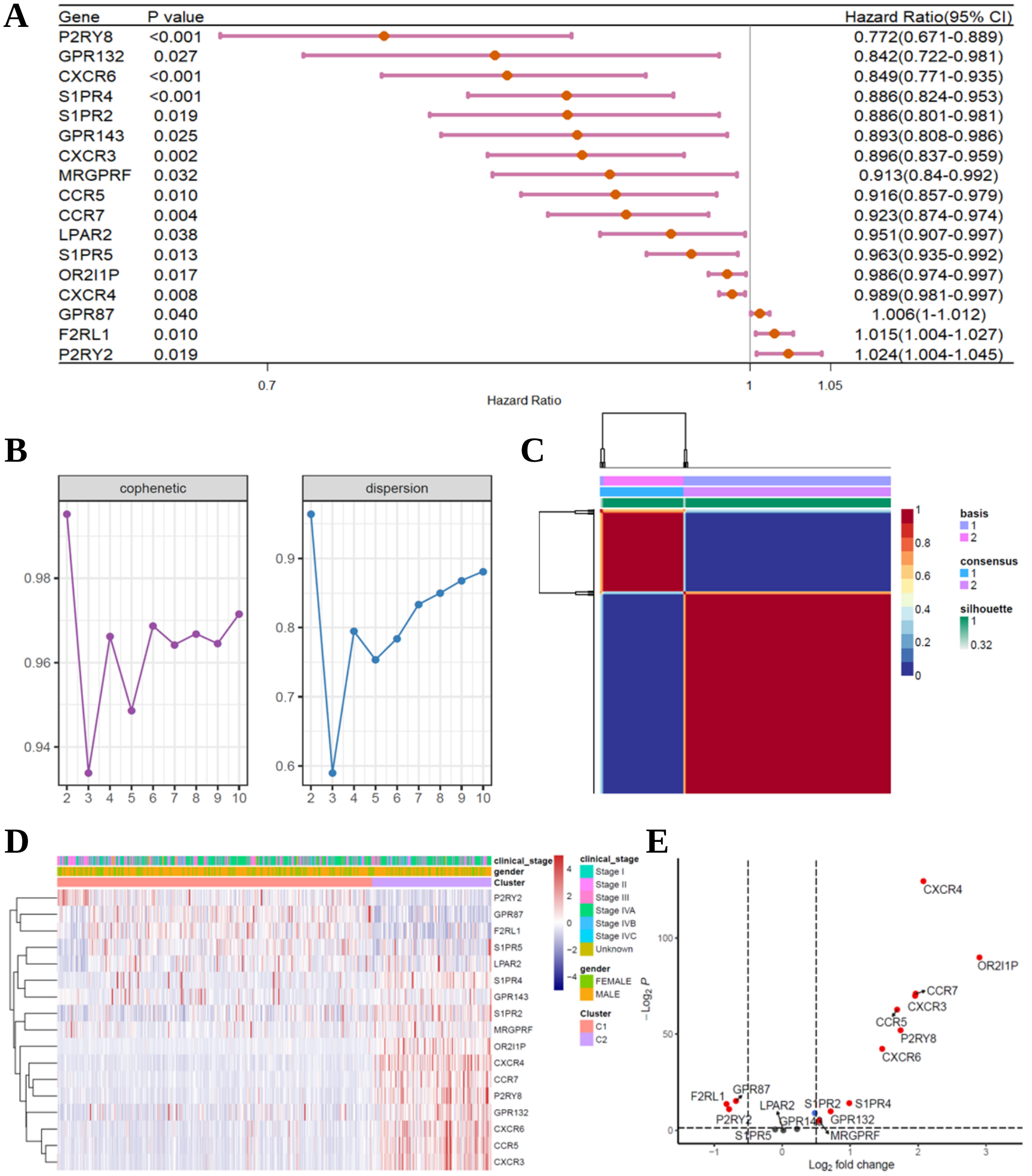
A
Gene
P value
Hazard Ratio(95% CI)
P2RY8 <0.001 0.772(0.671-0.889)
GPR132 0.027 0.842(0.722-0.981)
CXCR6 <0.001 0.849(0.771-0.935)
S1PR4 <0.001 0.886(0.824-0.953)
S1PR2 0.019 0.886(0.801-0.981)
GPR143 0.025 0.893(0.808-0.986)
CXCR3 0.002 0.896(0.837-0.959)
MRGPRF 0.032 0.913(0.84-0.992)
CCR5 0.010 0.916(0.857-0.979)
CCR7 0.004 0.923(0.874-0.974)
LPAR2 0.038 0.951(0.907-0.997)
S1PR5 0.013 0.963(0.935-0.992)
OR2I1P 0.017 0.986(0.974-0.997)
CXCR4 0.008 0.989(0.981-0.997)
GPR87 0.040 1.006(1-1.012)
F2RL1 0.010 1.015(1.004-1.027)
P2RY2 0.019 1.024(1.004-1.045)
0.7
1
1.05
Hazard Ratio
B
cophenetic
dispersion
C
basis
consensus
silhouette
D
clinical_stage
gender
Cluster
Stage I
Stage II
Stage III
Stage IVA
Stage IVB
Stage IVC
Unknown
FEMALE
MALE
C1
C2
E
CXCR4
OR2I1P
CCR7
CXCR3
CCR5
P2RY8
CXCR6
F2RL1
GPR87
LPAR2
S1PR2
S1PR4
P2RY2
GPR143
GPR132
S1PR5
MRGPRF
Log2 fold change
−Log2 P

Figure 1

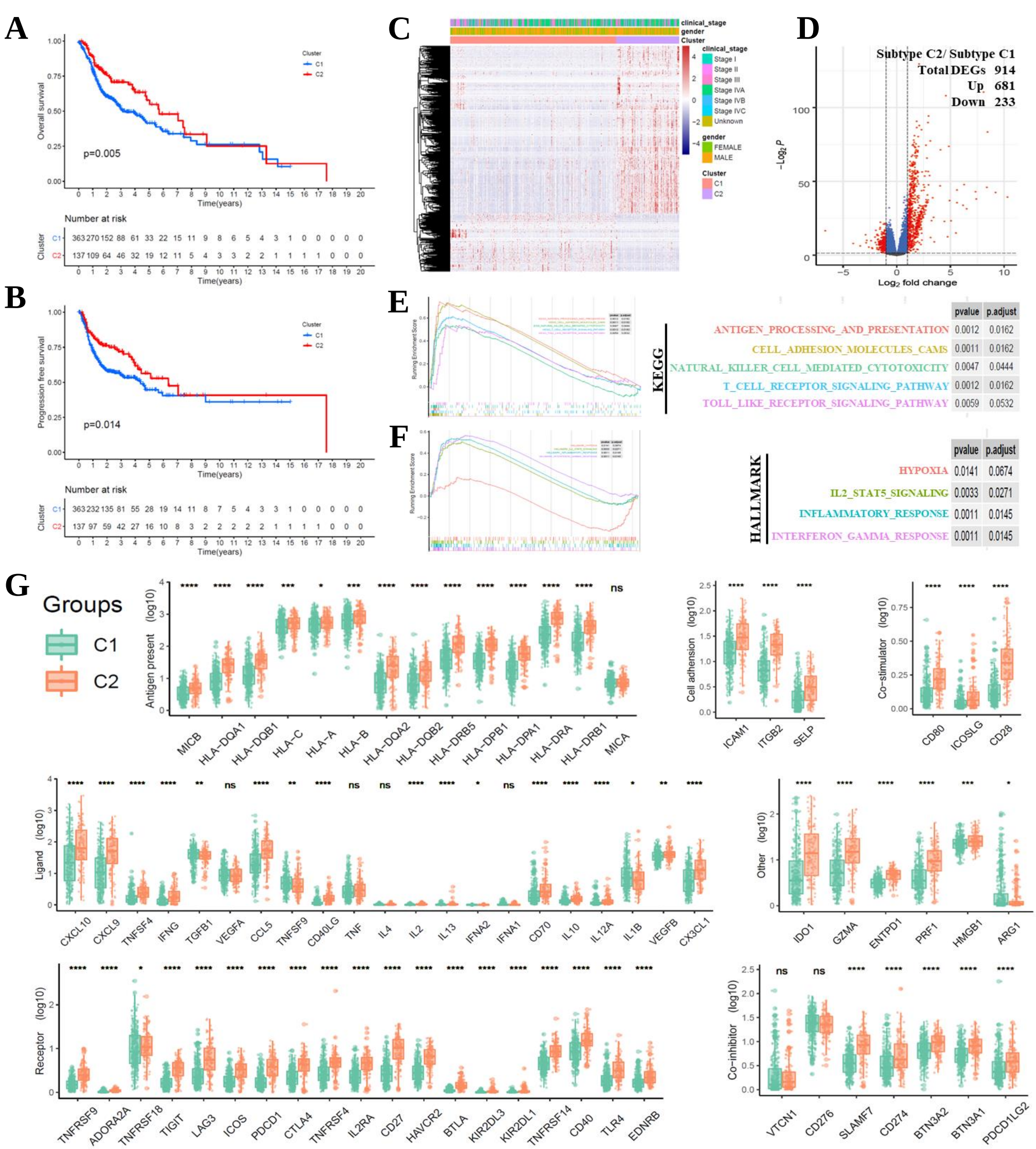
A
Overall survival
Cluster
C1
C2
p=0.005
Time(years)
Number at risk
Cluster C1 363 270 152 88 61 33 22 15 11 9 8 6 5 4 3 1 0 0 0 0 0
C2 137 109 64 46 32 19 12 11 5 4 3 3 2 2 1 1 1 1 0 0 0
B
Progression free survival
p=0.014
Number at risk
Cluster C1 363 232 135 81 55 28 19 14 11 8 7 5 4 3 3 1 0 0 0 0 0
C2 137 97 59 42 27 16 10 8 3 2 2 2 2 2 1 1 1 1 0 0 0
C
clinical_stage
gender
Cluster
Stage I
Stage II
Stage III
Stage IVA
Stage IVB
Stage IVC
Unknown
FEMALE
MALE
D
Subtype C2/ Subtype C1
Total DEGs 914
Up 681
Down 233
Log2 fold change
E
KEGG
pvalue p.adjust
ANTIGEN_PROCESSING_AND_PRESENTATION 0.0012 0.0162
CELL_ADHESION_MOLECULES_CAMS 0.0011 0.0162
NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY 0.0047 0.0444
T_CELL_RECEPTOR_SIGNALING_PATHWAY 0.0012 0.0162
TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY 0.0059 0.0532
F
HALLMARK
pvalue p.adjust
HYPOXIA 0.0141 0.0674
IL2_STAT5_SIGNALING 0.0033 0.0271
INFLAMMATORY_RESPONSE 0.0011 0.0145
INTERFERON_GAMMA_RESPONSE 0.0011 0.0145
G
Groups
C1
C2
Antigen present (log10)
MICB HLA-DQA1 HLA-DQB1 HLA-C HLA-A HLA-B HLA-DQA2 HLA-DQB2 HLA-DRB5 HLA-DPB1 HLA-DPA1 HLA-DRA HLA-DRB1 MICA
Cell adhension (log10)
ICAM1 ITGB2 SELP
Co-stimulator (log10)
CD80 ICOSLG CD28
Ligand (log10)
CXCL10 CXCL9 TNFSF4 IFNG TGFB1 VEGFA CCL5 TNFSF9 CD40LG TNF IL4 IL2 IL13 IFNA2 IFNA1 CD70 IL10 IL12A IL1B VEGFB CX3CL1
Other (log10)
IDO1 GZMA ENTPD1 PRF1 HMGB1 ARG1
Receptor (log10)
TNFRSF9 ADORA2A TNFRSF18 TIGIT LAG3 ICOS PDCD1 CTLA4 TNFRSF4 IL2RA CD27 HAVCR2 BTLA KIR2DL3 KIR2DL1 TNFRSF14 CD40 TLR4 EDNRB
Co-inhibitor (log10)
VTCN1 CD276 SLAMF7 CD274 BTN3A2 BTN3A1 PDCD1LG2

Figure 2

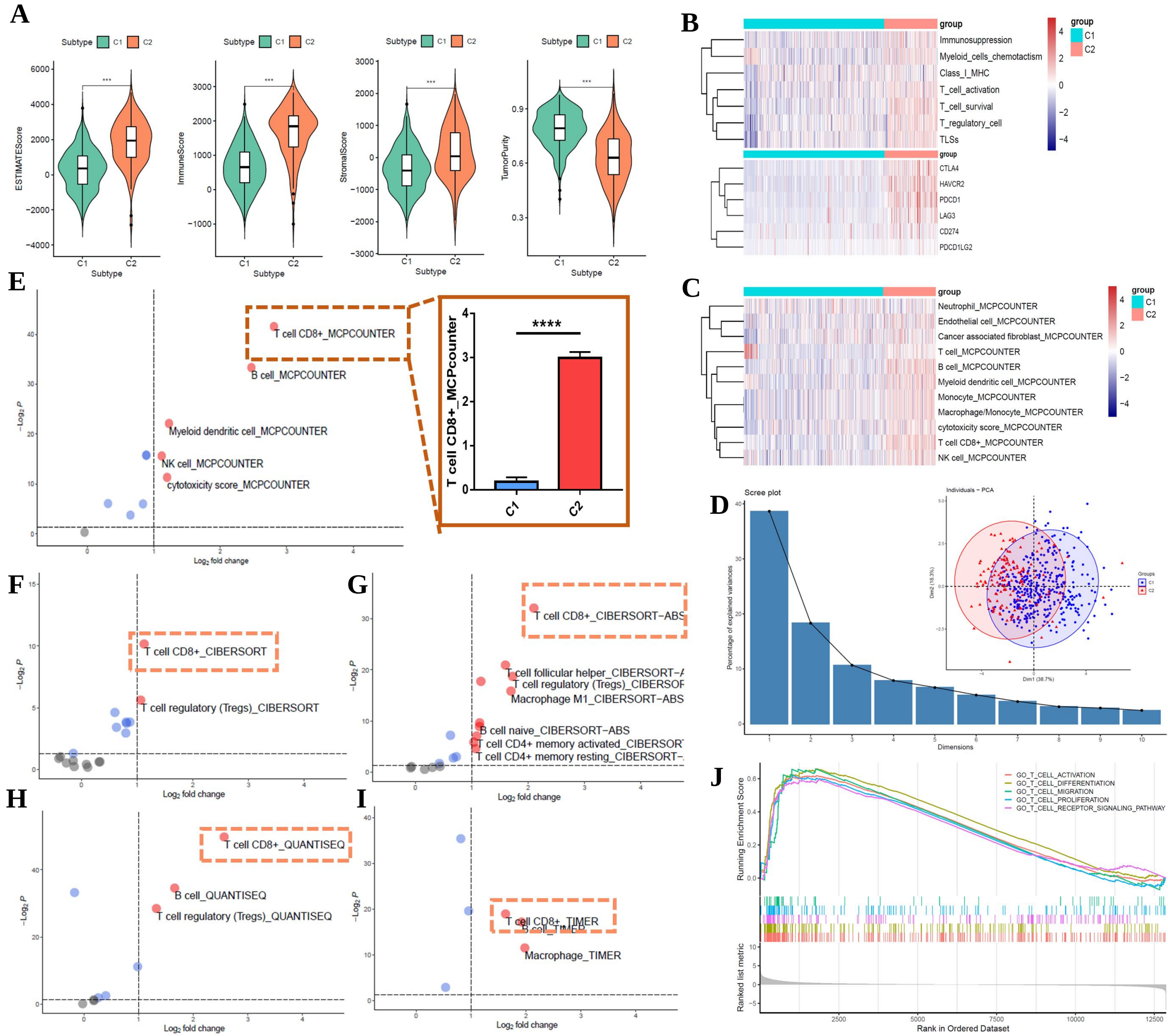
A
Subtype C1 C2
ESTIMATEScore
ImmuneScore
StromalScore
TumorPurity
B
group
Immunosuppression
Myeloid_cells_chemotactism
Class_I_MHC
T_cell_activation
T_cell_survival
T_regulatory_cell
TLSs
CTLA4
HAVCR2
PDCD1
LAG3
CD274
PDCD1LG2
C
Neutrophil_MCPCOUNTER
Endothelial cell_MCPCOUNTER
Cancer associated fibroblast_MCPCOUNTER
T cell_MCPCOUNTER
B cell_MCPCOUNTER
Myeloid dendritic cell_MCPCOUNTER
Monocyte_MCPCOUNTER
Macrophage/Monocyte_MCPCOUNTER
cytotoxicity score_MCPCOUNTER
T cell CD8+_MCPCOUNTER
NK cell_MCPCOUNTER
D
Scree plot
Percentage of explained variances
Dimensions
Individuals - PCA
Dim1 (38.7%)
Dim2 (18.3%)
Groups
E
T cell CD8+_MCPCOUNTER
B cell_MCPCOUNTER
Myeloid dendritic cell_MCPCOUNTER
NK cell_MCPCOUNTER
cytotoxicity score_MCPCOUNTER
T cell CD8+_MCPcounter
****
F
T cell CD8+_CIBERSORT
T cell regulatory (Tregs)_CIBERSORT
G
T cell CD8+_CIBERSORT-ABS
T cell follicular helper_CIBERSORT-A
T cell regulatory (Tregs)_CIBERSOR
Macrophage M1_CIBERSORT-ABS
B cell naive_CIBERSORT-ABS
T cell CD4+ memory activated_CIBERSOR
T cell CD4+ memory resting_CIBERSORT-
H
T cell CD8+_QUANTISEQ
B cell_QUANTISEQ
T cell regulatory (Tregs)_QUANTISEQ
I
T cell CD8+_TIMER
Macrophage_TIMER
Log2 fold change
-Log2 P
J
GO_T_CELL_ACTIVATION
GO_T_CELL_DIFFERENTIATION
GO_T_CELL_MIGRATION
GO_T_CELL_PROLIFERATION
GO_T_CELL_RECEPTOR_SIGNALING_PATHWAY
Running Enrichment Score
Ranked list metric
Rank in Ordered Dataset

Figure 3

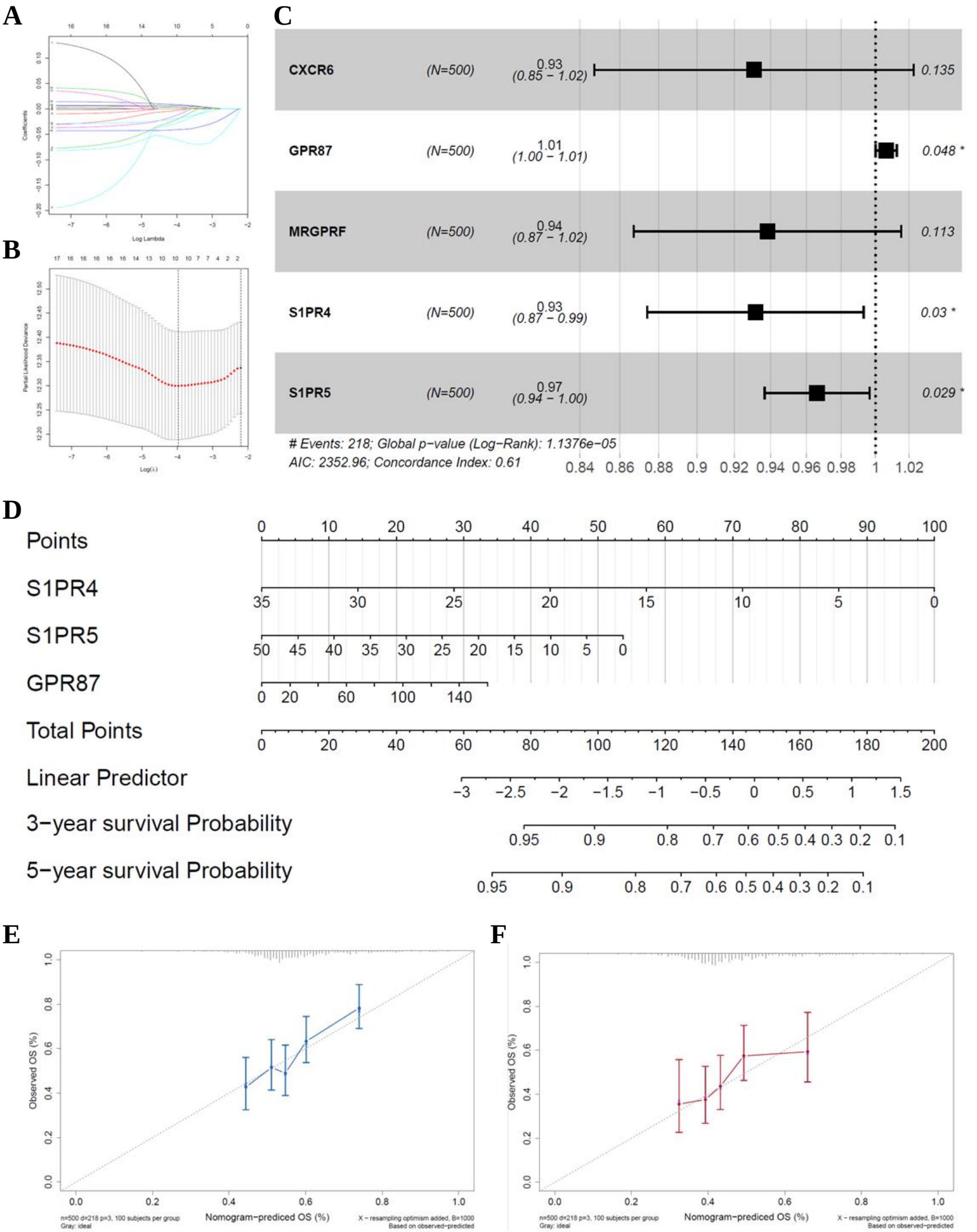
A
Coefficients
Log Lambda
B
Partial Likelihood Deviance
Log(λ)
C
CXCR6 (N=500) 0.93 (0.85 − 1.02) 0.135
GPR87 (N=500) 1.01 (1.00 − 1.01) 0.048 *
MRGPRF (N=500) 0.94 (0.87 − 1.02) 0.113
S1PR4 (N=500) 0.93 (0.87 − 0.99) 0.03 *
S1PR5 (N=500) 0.97 (0.94 − 1.00) 0.029 *
Events: 218; Global p−value (Log−Rank): 1.1376e−05
AIC: 2352.96; Concordance Index: 0.61
0.84 0.86 0.88 0.9 0.92 0.94 0.96 0.98 1 1.02
D
Points
S1PR4
S1PR5
GPR87
Total Points
Linear Predictor
3−year survival Probability
5−year survival Probability
E
Observed OS (%)
Nomogram−prediced OS (%)
n=500 d=218 p=3, 100 subjects per group
Gray: ideal
X − resampling optimism added, B=1000
Based on observed−predicted
F
Observed OS (%)
Nomogram−prediced OS (%)
n=500 d=218 p=3, 100 subjects per group
Gray: ideal
X − resampling optimism added, B=1000
Based on observed−predicted

Figure 4

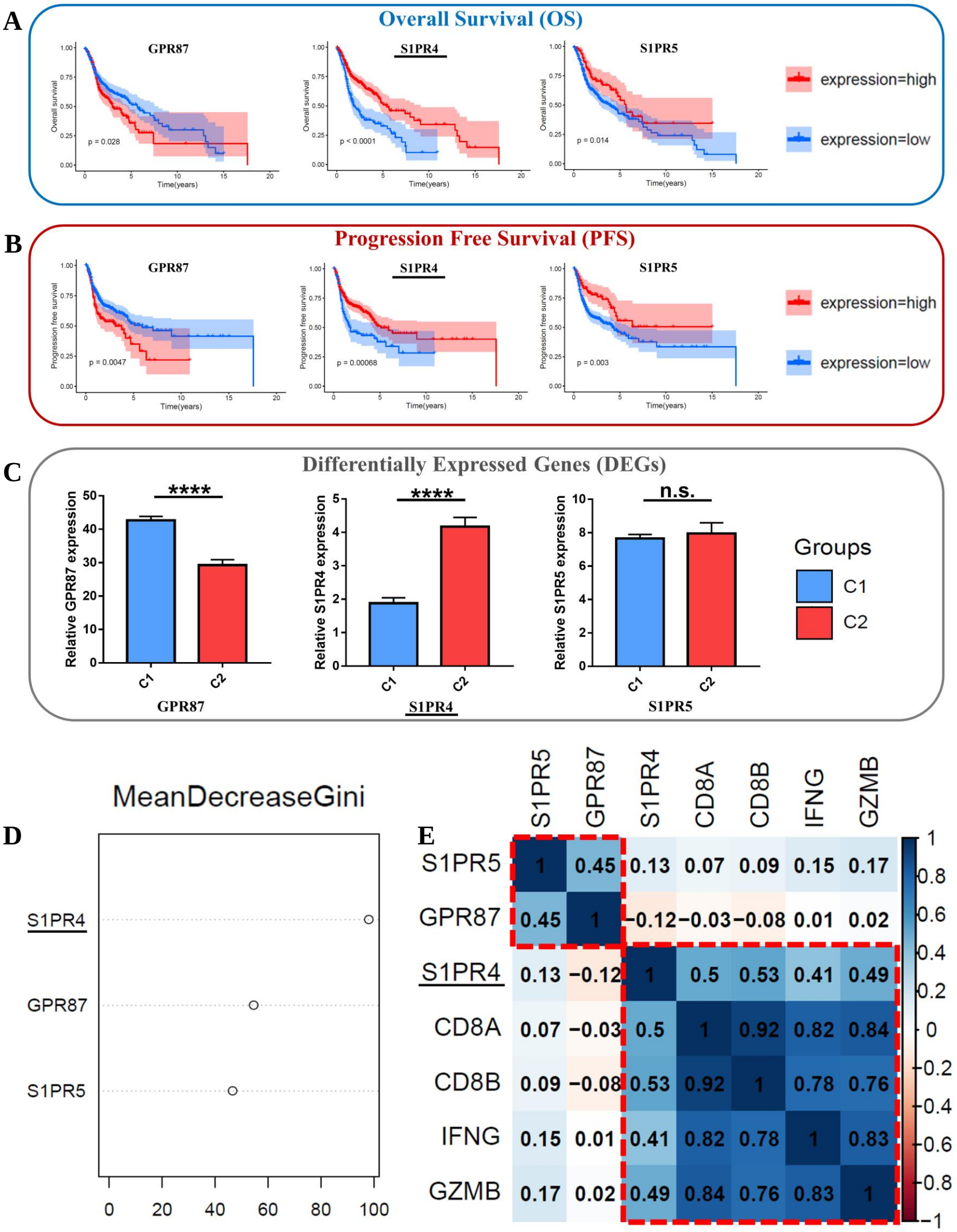
A
Overall Survival (OS)
GPR87
S1PR4
S1PR5
Overall survival
Time(years)
p = 0.028
p < 0.0001
p = 0.014
expression=high
expression=low
B
Progression Free Survival (PFS)
Progression free survival
p = 0.0047
p = 0.00068
p = 0.003
C
Differentially Expressed Genes (DEGs)
Relative GPR87 expression
Relative S1PR4 expression
Relative S1PR5 expression
****
n.s.
C1
C2
Groups
D
MeanDecreaseGini
E
S1PR5
GPR87
S1PR4
CD8A
CD8B
IFNG
GZMB

Figure 5

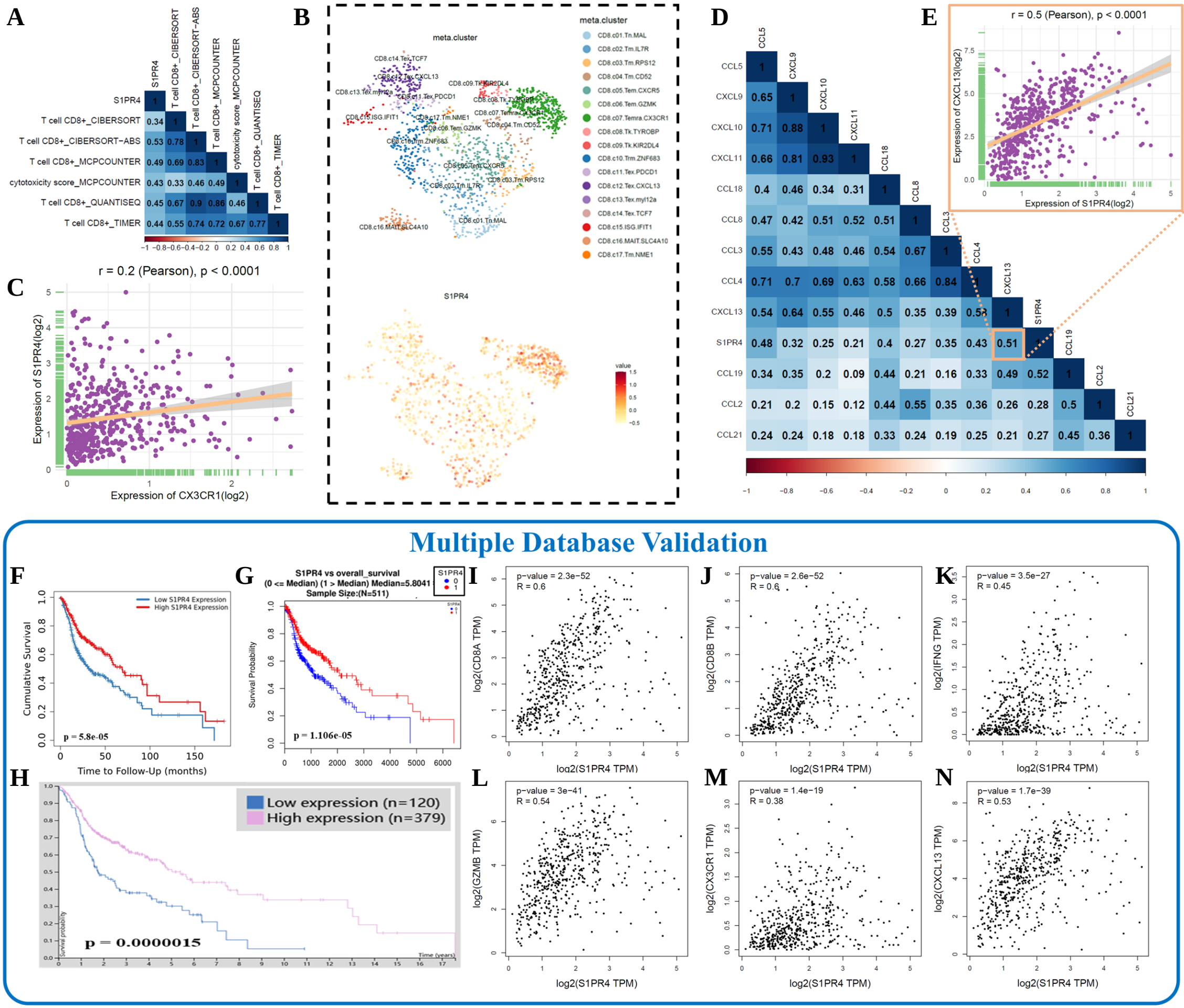
A
B
C
D
E
r = 0.2 (Pearson), p < 0.0001
Expression of S1PR4(log2)
Expression of CX3CR1(log2)
meta.cluster
S1PR4
r = 0.5 (Pearson), p < 0.0001
Expression of CXCL13(log2)
Expression of S1PR4(log2)
Multiple Database Validation
F
G
H
I
J
K
L
M
N
Low S1PR4 Expression
High S1PR4 Expression
Cumulative Survival
Time to Follow-Up (months)
p = 5.8e-05
S1PR4 vs overall_survival
(0 <= Median) (1 > Median) Median=5.8041
Sample Size:(N=511)
Survival Probability
p = 1.106e-05
Low expression (n=120)
High expression (n=379)
p = 0.0000015
p-value = 2.3e-52
R = 0.6
log2(CD8A TPM)
log2(S1PR4 TPM)
p-value = 2.6e-52
R = 0.6
log2(CD8B TPM)
p-value = 3.5e-27
R = 0.45
log2(IFNG TPM)
p-value = 3e-41
R = 0.54
log2(GZMB TPM)
p-value = 1.4e-19
R = 0.38
log2(CX3CR1 TPM)
p-value = 1.7e-39
R = 0.53
log2(CXCL13 TPM)

Figure 6

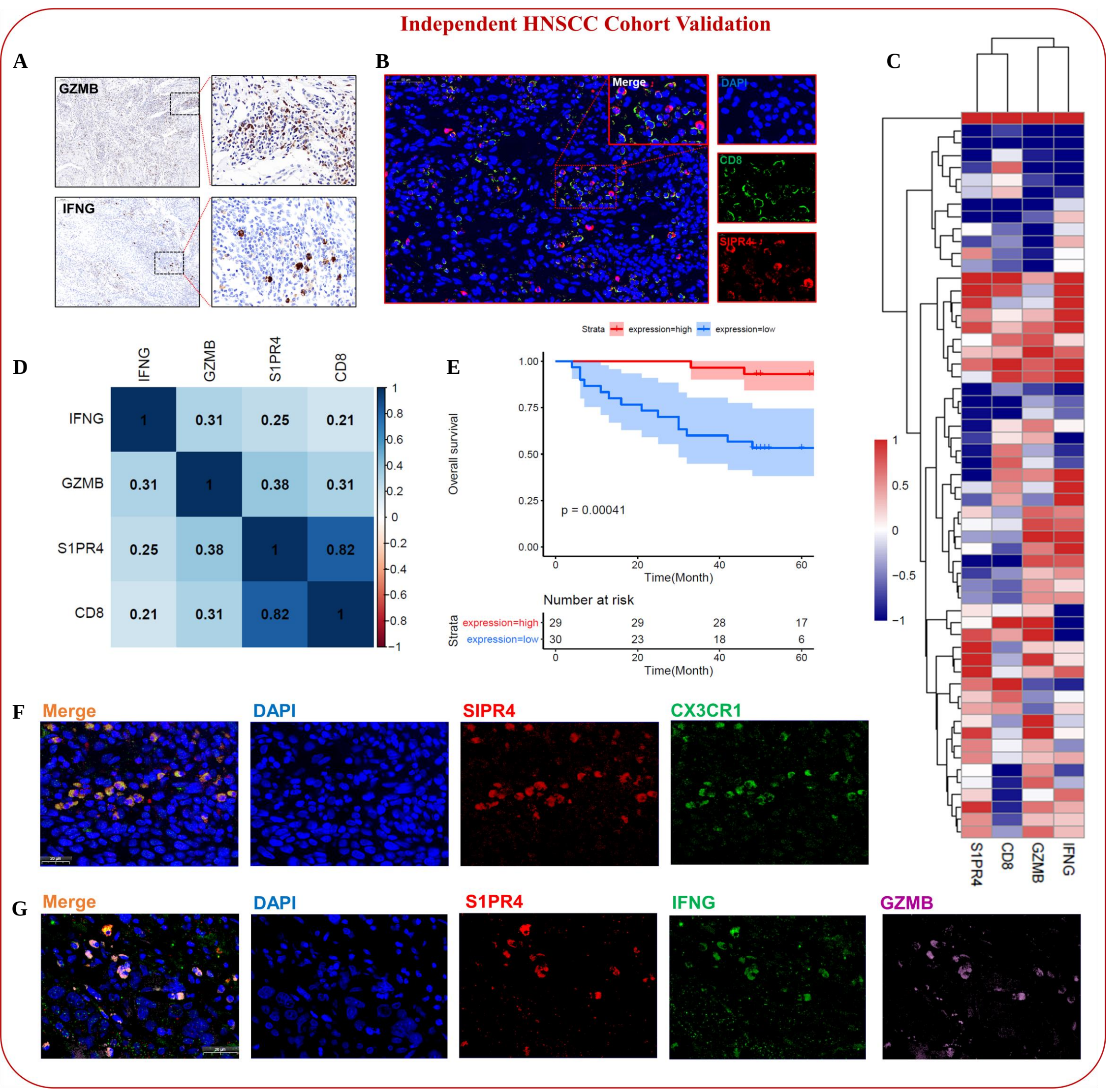
Independent HNSCC Cohort Validation
A
GZMB
IFNG
B
Merge
DAPI
CD8
SIPR4
C
S1PR4
CD8
GZMB
IFNG
D
IFNG
GZMB
S1PR4
CD8
1
0.31
0.25
0.21
0.31
1
0.38
0.31
0.25
0.38
1
0.82
0.21
0.31
0.82
1
E
Strata
expression=high
expression=low
Overall survival
p = 0.00041
Time(Month)
Number at risk
29
29
28
17
30
23
18
6
F
Merge
DAPI
SIPR4
CX3CR1
G
Merge
DAPI
S1PR4
IFNG
GZMB

Figure 7