Single cell transcriptome of CD8+ T cells in multiple cancers reveals comprehensive exhaustion associated mechanisms

Yunmeng Bai[1], Zixi Chen[1], Xiaoshi Chen[1], Ziqing He[1], Jie Long[2,3,4], Shudai Lin[1], Hongli Du[1*]

[1] School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China

[2] Department of General Surgery, Guangzhou Digestive Disease Center, Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, Guangdong, 510180, China.

[3] Chronic Disease Laboratory, Institutes for Life Sciences and School of Medicine, South China University of Technology, Guangzhou, 510006, China.

[4] Institute of Immunology and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui, 230027, China.

[*]Correspondence: hldu@scut.edu.cn; Tel.: +86-020-3938-0667

Abstract

T cell exhaustion is one of the mechanisms that cancer cells get rid of control from the immune system. Single-cell RNA sequencing has showed superiority on immunity mechanism in recent studies. Here, we collected more than 6000 single CD8+ T cells from three cancers including CRC, HCC and NSCLC, and identified five clusters of each cancer. We obtained 71 and 159 DEGs for pre_exhausted or exhausted vs. effector comparison in all three cancers, respectively. Specially, we found some key genes including the four exhaustion associated genes of *PDCD1*, *HAVCR2*, *TIGIT* and *TOX*, and two vital genes of *CD69* and *JUN* in the interaction network. Additionally, we identified the gene *SAMSN1* which highly expressed in the exhausted cells had a poor overall survival and played a negative role in immunity. We summarized the putative interrelated mechanisms of above key genes identified in this study by integrating the reported knowledge. Furthermore, we explored the heterogeneous and preference of exhausted CD8+ T cells in each patient and found only one exhausted sub-cluster existed in the most of patients, especially in CRC and HCC. As far as we know, this is the first time to study the mechanism of T cell exhaustion with the data of single-cell RNA sequencing of multiple cancers. Our study may facilitate the understanding of the mechanism of T cell exhaustion, and provide a new way for functional research of single-cell RNA sequencing data across cancers.

Keywords: Single-cell RNA sequencing; multiple cancers; T cell exhaustion; tumor heterogeneous

1. Introduction

T cell exhaustion (Tex), a hyporesponsive state of T cells, was originally described in CD8+ T cells during chronic lymphocytic choriomeningitis virus (LCMV) of mice[1]. In recent years, the phenomenon of Tex has also been found in cancer[2, 3]. It has been reported that exhausted T cells in

cancers share many similarities with that in chronic infection[4], and play a significant role in tumorigenesis[5]. The main reasons leading to Tex in chronic infection and cancer are as following[6]: long-term and persistent explosion to antigens[7]; the upregulation of inhibitory receptors (IRs) including PDCD1, CLTA4, HAVCR2, LAG3 and 2B4[8]; complex effects of soluble factors such as cytokines IL-10 and transforming growth factor-β (TGFβ) [9] ; the expression of transcription factor including T-bet and Eomes[10] and corresponding epigenetic regulation[11].However, at present, to reveal molecular mechanisms of exhausted T cells, most researches are still focusing on chronic infected mouse models by microarray or RNA sequencing[8, 12]. Tex is a general trend of tumor, which can be used as one of the main targets of immunosuppression therapy to save T cell from exhaustion and reactivate the cytotoxicity of T cells, providing a new opportunity for clinical immunotherapy[8]. Nevertheless, due to the complexity and heterogeneity of cancer, the concrete mechanism of Tex in cancer is still unclear. Thus, a deeper understanding of this mechanism is urgently needed.

Currently, single-cell RNA sequencing (scRNA-seq) has clearly revealed some new mechanisms and phenomena of cancer with the advantages of high accuracy and reproducibility[13-15]. Using single cell transcriptome profiling, we can identify new types of immune cells which can't be revealed at the original tissue level and can construct a developmental trajectory for immune cells which can reveal the random heterogeneity[16]. These new findings are useful to better understand the immune system and its mechanism of action on tumors. Notably, this technology makes it possible to explore complicated tumor microenvironment including tumor-infiltrating lymphocytes (TILs) in melanoma, head and neck cancer, breast cancer and glioblastoma cancer[17-20]. Thus, using advantage of scRNA-seq to analyze T cells and obtain the hallmarks of exhausted T cells can bring a new therapeutic strategy on clinical cancer treatment.

In the present study, due to the vital role of CD8+ T cells in eliciting antitumor responses[21], we integrated single cell sequencing data from colorectal cancer (CRC), liver cancer (HCC), and non-small cell lung cancer (NSCLC) to analyze CD8+ T cells in various cancers. By classifying cells into different clusters with the unique markers, we identified five clusters of each cancer. Furthermore, we compared the pre_exhausted or exhausted cells with effector cells to make a deeper understanding of exhaustion mechanism. Additionally, the sub-clusters of exhausted cells in each patient revealed the individual differences and preferences. Overall, the findings of molecules changed in pre_exhasuted and exhausted clusters may provide the potential targets to anticancer therapy, and the different exhausted sub-clusters of each patient help the understanding of exhaustion heterogeneity.

2. Materials and methods

2.1 Data Resources and Preprocessing

Single cell transcriptome profiling of human T cells in three cancers including CRC, HCC and NSCLC was obtained from the GEO database (GSE108989, GSE99254, GSE98638), including raw read count and TPM data. According to the annotations, we isolated CD8+ T cells from peripheral blood (PTC), adjacent normal (NTC) and tumor tissues (TTC). To filtered out the low-quality cells, we excluded cells with fewer than 3000 detected genes and the expression of the housekeeping gene ($\log_2(TPM_{ACTB}/10+1)$) below 2.5[22].

2.2 Unsupervised Clustering

For each dataset, we selected the genes with top n highest variance where n changed from 500, 1000, 1500, 2000, 2500, 3000. Then the expression data of these genes was used to perform clustering by single-cell consensus clustering (SC3)[23]. To obtain the reliable result, we set the number of clusters

(k) from 2 to 10. Only the genes with p value < 0.05 and auroc > 0.75 could be considered as markers. For each SC3 run, we focused on three indicators - the silhouette, the consensus matrix and the cluster specific genes, which could help us to fix the suitable n and k. When determined the clusters, we mapped the markers to the CellMarker databases[24] to find out the characteristics of each cluster.

2.3 Trajectory Analysis

To explore the potential functional changes of CD8+ T cell of different clusters for each cancer, we performed development trajectory analysis by Monocle[25] with the top 100 cluster-specific genes of each cluster. The differentially expressed genes (DEGs) between each two cluster pairs were identified by R package edgeR[26] using the criteria as following: (1) the mean of CPM value was greater than 1; (2) false discovery rate (FDR) < 0.05; (3) the absolute value of logFC > 1. We defined the cluster-specific genes in each cluster of each cancer based on the condition that its expression was upregulated when compared with those in the other clusters.

2.4 Differentially Expressed Genes, Functional Annotation and Transcription Network Analysis

To further get known the potential mechanisms that leading to the exhausted CD8+ T cells, we compare genes between pre_exhausted or exhausted cluster in tumor and effector cluster in peripheral blood. The cell numbers of the two groups participating in the comparison were set as m and n, and the smaller one was defined as p. A gene was defined as exhaustion associated DEG if (1) the number of expressed cells > p; (2) FDR < 0.05; (3) the absolute value of logFC > 1. To find the common regulation mechanisms of Tex, we identified the DEGs with the same regulatory status in three cancers. Next, the DEGs were subjected to the Gene Ontology(GO), including biological process (BP), cell component (CC), molecular function (MF), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis by clusterProfiler[27] with adjusted p value < 0.05.

To find out the interaction among these DEGs, Search Tool for the Retrieval of Interacting Genes database (STRING) was used to construct the protein-protein interaction (PPI) network, which was visualized by Cytoscape[28], based on the combined score of 0.4[29]. Additionally, the Molecular Complex Detection (MCODE) as a Cytoscape plugin was performed to explore the closely connected regions with default parameter[30]. Furthermore, to investigate the potential upstream regulators that affect gene differentially expression, we integrated three transcription factor targets databases, including ChIP-X enrichment analysis (ChEA)[31], Encyclopedia of DNA Elements (ENCODE)[32] from Harmonizonme[33] and Transcriptional Regulatory Relationships Unraveled by Sentence-based Text-mining (TRRUST)[34] to reveal regulatory relationships among DEGs.

2.5 Survival analysis

The TCGA data[35] including COAD (colon adenocarcinoma), LIHC (liver hepatocellular carcinoma), LUAD (lung adenocarcinoma), LUSC (lung squamous cell carcinoma) was used to explore the relationship between gene expression levels of the key genes and patients' overall survival. LUAD and LUSC were integrated as NSCLC. The mRNA expression data and corresponding clinical annotation files were obtained from TCGA. To correct the effector of CD8+ T cells[36], the expression of genes were divided by that of the geometric mean of *CD3D*, *CD3E*, *CD3G*, *CD8A* and *CD8B*. The threshold was defined as the median of relative expression. The samples with relative expression below the threshold were classified as ''low expression'' group, and the samples with relative expression above the threshold were classified as "high expression" group. R package survival was used to perform survival analysis.

2.6 Sub-cluster of exhausted cells

We further clustered the exhausted cells into more specific sub-clusters using the same pipeline of

SC3 as mentioned above. Especially, we displayed the distribution of each sub-cluster of each patient. The differential genes expression analysis between exhausted sub-clusters and effector cells was performed by edgeR with the same criteria as mentioned above. To further explore the function of each sub-cluster, we performed gene set variation analysis (GSVA)[37], a sensitive technique that can detect underlying pathways over a sample population in an unsupervised manner, with the DEGs of the exhausted sub-clusters vs. effector comparison and the KEGG pathways in canonical gene sets (C2) from the Molecular Signature Database (MSigDB)[38] as the reference gene set. In addition, comparisons of single-cell enrichment scores were performed using the R package limma[39]. Differentially enriched pathways were defined as the number of genes >10 and FDR adjusted p value < 0.05 as described previously[40].

3. Results

3.1 Single CD8+ T cell transcriptome landscape and processing

We obtained single-cell transcriptome data of human T cells from GEO database, including 12 patients from CRC, 6 patients from HCC and 14 patients from NSCLC. For CD8+ T cells, low-quality cells were filtered out, remaining 2967, 1214, 2053 cells, respectively (Table 1). We noticed that the cells in NSCLC were filtered out more than half.

Table 1. The number of CD8+ T cells of different tissues in three cancers

	N	P	T
CRC	798(1072)	810(1132)	1359(1722)
HCC	262(412)	345(563)	607(777)

NSCLC	351(934)	524(1323)	1178(2182)

Outside the brackets means the number of post-filtered cells, inside means pre-filtered cells

3.2 CD8+ T cell clustering

To reveal the potential functional sub-clusters of CD8+ T cells, we performed unsupervised clustering by SC3. As for CRC, finally we chose the top 2500 genes with highest variance to divide cells into 5 clusters (Figure 1A). Similarly, the top 1000 genes were used to divide into 5 clusters in HCC and NSCLC (Figure 1B, 1C).

To identify each cluster of three cancers, we compared the markers of each 15 clusters with the CellMarker database. For example, the 9 of 10 markers (*CCR7*, *LEF1*, *SELL*, *LDLRAP1*, *NOSIP*, *PRKCQ−AS1*, *TXK*, *TMEM123*, *NELL2*) in CRC_C4 were found in the cell markers of naïve CD8+ T cell, then we identified it as naïve CD8+ T cell. Taken together, we defined the clusters for each cancer with known markers: CRC_C4, HCC_C5, NSCLC_C4 were identified as naïve T cells, CRC_C1, HCC_C3, NSCLC_C1 were identified as effector T cells, CRC_C3, HCC_C2, NSCLC_C3 were identified as exhausted T cells. CRC_C2 was identified as natural killer T (NKT) cells and HCC_C4 was identified as mucosal-associated invariant T (MAIT) cells (Table S1). While in NSCLC, C2 couldn't be identified as a defined cluster according to the CellMarker, in which *ZNF683* was one of the markers belonging to natural killer T (NKT) cell. Thus, we called CRC_C2, HCC_C4, NSCLC_C2 as unconventional T cells[41] temporarily which were excluded in the following analysis. We noted that there was a common cluster with a same marker named *GZMK* that couldn't be well-defined, either. To figure out the potential biological functions of these clusters, we performed trajectory analysis by monocle (Figure 1D, 1E, 1F). Except for the known clusters, we could see that CRC_C5, HCC_C1,

NSCLC_C5 located between effector and exhausted CD8+ T cells and appeared a transition state, thus them were defined as “pre_exhausted” CD8+ T cells. In order to check the classification effect, we used the t-Distributed Stochastic Neighbor Embedding (t-SNE), a non-linear dimensionality reduction technique, by Rtsne package with all the markers of all clusters of each cancer to visualize the cell population clusters in two dimensions (Figure 1G, 1H, 1I). We could see that in CRC, five clusters with different functions were mostly clustered in different parts with obvious boundaries. In addition, the same clustering phenomenon could also be observed in HCC and NSCLC. We combined clusters of three cancers with their all markers and found that the same functional cells from the different cancer types were clustered into the same clusters (Figure 1J).

3.3 Differentially expressed genes associated with exhaustion

The proportion of different clusters varied greatly in different tissues (Figure 2A). Peripheral blood had most naïve and effector CD8+ T cells in all three cancers, tumor tissues had most pre_exhausted and exhausted CD8+ T cells, while there was no regular pattern in normal tissues. The difference in normal tissue may be related to tissue specificity. Since the most of effector cells existed in peripheral blood in three cancers and there showed no difference among peripheral blood, normal and tumor tissues (Figure S1), we compared pre_exhausted or exhausted cells in tumor tissues with effector cells in peripheral blood to explore the exhaustion associated mechanism.

In the pre_exhausted vs. effector comparison, there were 73 DEGs existed in all three cancers, 71 of whom were in the same regulatory status (Figure 2B) including the known exhaustion marker *PDCD1* (programmed cell death□1) and *CD27* (CD27 molecule). The top 15 significant GO and KEGG terms based on the p value were shown as Figure 2D and 2E. We could find that most terms

were related to the production of chemokine and response to tumor necrosis factor which can facilitate the immunoregulation effects[42, 43]. Apart from that, we found other terms associated with immunity. For example, there were 7 upregulated DEGs, including *GPR171* (G protein-coupled receptor 171), *TNFAIP3* (TNF alpha induced protein 3), *CD96* (CD96 molecule), *SAMSN1* (SAM domain, SH3 domain and nuclear localization signals 1, also known as *HACS1*), *ZFP36* (ZFP36 ring finger protein), *DUSP1* (dual specificity phosphatase 1), *PDCD1*, enriched in the negative regulation of immune system process.

In the exhausted vs. effector comparison, there were 159 DEGs existed in all three cancers (Figure 2C) including multiple known exhaustion markers such as *PDCD1*, *HAVCR2* (hepatitis A virus cellular receptor 2, also known as T cell immunoglobulin and mucin-domain containing-3, Tim-3), *TIGIT* (T cell immunoglobulin and ITIM domain), *TOX* (thymocyte selection associated high mobility group box). The top 15 significant GO and KEGG terms were shown as Figure 2F and 2G. Several terms related to glycolytic process which was the main metabolism pathway of T cells to sustain effector function after antigen encounter[44] were enriched. Besides, pathways related to virus infection, NF-kappa B signaling pathway were also enriched. Correspondingly, there were 14 DEGs, including 9 upregulated genes *PDCD1*, *TIGIT*, *HAVCR2*, *CCL3* (C-C Motif Chemokine Ligand 3), *DUSP1*, *TNFAIP3*, *PAG1* (phosphoprotein membrane anchor with glycosphingolipid microdomains 1), *CDK6* (cyclin dependent kinase 6), *SAMSN1*, participated in the negative regulation of immune system process.

Interestingly, we found several common terms between pre-exhausted and exhausted cells, including cell chemotaxis, leukocyte migration, chemokine signaling pathway, cytokine-cytokine receptor interaction and so on. Additionally, the pathways existed only in pre_exhausted cells included

cytolytic granule and interleukin-2 production, which could represent the effector functions of CD8+ T cells, showing the features in transition of pre_exhausted cells. The exhausted cells most enriched in the negative regulation of cytokine production, showing the further dysfunction of T cells. The results of functional analysis of DEGs was shown in Table S2 and S3.

It was obvious that as the cell state changed from pre-exhausted to exhausted cells, the number of DEGs also increased, especially the exhaustion associated markers. Moreover, the expression level of the common markers increased from pre_exhausted to exhausted cells (Figure S2). Interestingly, we compared the genes existed in both pre_exhausted and exhausted cells and filtered those with greater fold change in the exhausted cells (Figure 3A). We found *SAMSN1* was dysregulated during exhaustion state, which participated in the negative regulation of immune system process in both pre_exhausted and exhausted cells, showed poor survival in the validation based on TCGA data except for CRC (Figure 3B).

Moreover, we performed the PPI network analysis of all 193 DEGs in pre_exhausted or exhausted vs. effector comparison and found seven significant functional modules (Figure 3C and Figure S3). The module with the highest score consisted of 15 nodes and 101 edges, in which *CD69* was the seed gene and linked with 26 genes.

To further explore the factors affecting the expression of DEGs, we integrated the transcription factor (TF) from CHEA, ENCODE and TRRUST databases, and found the TFs with dysregulated expression were ATM (ATM serine/threonine kinase), RORA (RAR Related Orphan Receptor A), TBX21 (T-Box Transcription Factor 21), JUNB (JunB Proto-Oncogene, AP-1 Transcription Factor Subunit), RBPJ (Recombination Signal Binding Protein For Immunoglobulin Kappa J Region), TNFAIP3 and JUN (Jun proto-oncogene, AP-1 transcription factor subunit) (Figure 3D, Figure S4).

ATM, RORA, TBX21, which were the TFs of *DUSP1*, *TUBA1B* (tubulin alpha 1b), *CXCR3* (C-X-C motif chemokine receptor 3) and *IFNG* (Interferon Gamma) respectively, were significantly downregulated in exhausted cells. Among the upregulated factors, JUN regulated more than 100 DEGs and may play a critical role in the occurrence and development of exhausted CD8+ T cells. Therefore, these TFs may lead to the dysregulated expression of DEGs during T cell exhaustion.

3.4 Sub-cluster of exhausted cells

Exhausted T cell populations are heterogeneous[45]. In the further classification of exhausted cells, the top 1500 genes with highest variance were used to divide exhausted cells into 2 sub-clusters in CRC, the top 3000 genes in HCC to 2 sub-clusters, and the top 500 genes in NSCLC to 2 sub-clusters (Figure 5A, 5B, 5C). In the visualization of exhausted sub-clusters, the distribution of exhausted cells of each patient tended to appear in one sub-cluster(Figure 5D, 5E, 5F), for instance, P0413, who was diagnosed at stage III of CRC, mostly exhausted in sub-cluster I, while P0909, who was diagnosed at stage I of CRC, mostly exhausted in sub-cluster II, illustrating the individual differences.

Since different sub-clusters existed in each cancer, we tried to characterize the features of sub-clusters across cancers. We combined the six sub-clusters with re-clustering and found that most cells of sub-cluster I could be distinguished in three cancers (Figure 6A) and the cells of different cancers were clustered relatively closely, illustrating the tumor specificity, while the cells of sub-cluster II were clustered disorderly and couldn’t distinguish the cells from different cancers, showing no tumor specificity.

In order to find out the potential different mechanisms of each sub-cluster, we compared the sub-cluster of exhausted cells from each tumor with effector cells in peripheral blood using the

aforementioned criteria. The numbers of DEGs and intersection of each group were shown by UpSetR[46] (Figure 6B). There were 63 common DEGs among the six sub-clusters, including 27 upregulated and 36 downregulated genes in all three cancers, in which 60 DEGs appeared in the above common DGEs of the exhausted vs. effector comparisons except 3 genes including *PLAC8* (placenta associated 8), *TAPT1-AS1* (transmembrane anterior posterior transformation 1- antisense RNA 1) and *GZMB* (Granzyme B).

In the KEGG pathways enriched by GSVA (Figure 7A-F), there was one common pathway in 6 sub-clusters: cytokine-cytokine receptor interaction, which also existed in the above analysis of pre_exhausted or exhausted clusters compared with effector cluster. Additionally, the widely enriched pathways, including pathways in cancer, natural killer cell mediated cytotoxicity and ribosome, were found in 5 sub-clusters. What's more, the unique pathways belonging to each exhausted sub-cluster might illustrate the specific difference for each exhausted sub-cluster, which were shown in Table S5. For example, Wnt signaling pathway existed only in the sub-cluster I of CRC; apoptosis pathway existed only in the sub-cluster I of NSCLC. These differences may contribute to the various mechanisms of sub-clusters.

4. Discussion

T cell exhaustion is a dysfunctional state and is possible to be used in immunotherapy because of its reversibility[47]. Combining with single-cell RNA sequencing, which can facilitate detecting the transcriptome on the level of single cell[48], we can shed light on the complication of tumor-infiltrating T cells. In order to study the common mechanisms of multiple cancers, we performed transcriptomic analysis of single CD8+ T cells isolated from three cancers including CRC, HCC, NSCLC. As expected,

we identified some known genes related to T cell exhaustion, such as *PDCD1*, *HAVCR2*, *TIGIT*, *TOX*. Moreover, two vital genes *CD69* and *JUN* in the interaction network and the gene *SAMSN1* which highly expressed in the exhausted cells were also identified. Furthermore, the distribution of exhausted T cells for each patient tended to appear in one exhausted sub-cluster, which indicated that the exhaustion mechanism in each individual might have homogeneity.

In the functional enrichment analysis, several DEGs played a negative regulation of immune system process and led to the pre-exhausted or exhausted state of CD8+ T cells. To be specific, the gene *PDCD1* encodes an inhibitory receptor and functions as an immune checkpoint. In the current study, *PDCD1* was significantly upregulated in the pre_exhausted and exhausted cells, which was consistent with previous studies[49, 50]. The upregulation of *PDCD1* played a crucial role in the regulation of CD8+ T cell exhaustion and blockade of the interaction between PDCD1 and its ligand PD-L1 could restore CD8+ effector function[49, 50]. It has been reported that the combination of PDCD1 and PD-L1 can recruit phosphatases[51], dampen TCR signaling[52], inhibit the PI3K signaling pathway[53], suppress proliferation and cytokine production and finally result in CD8+ T cell exhaustion. Besides, the gene *HAVCR2*, which encodes an inhibitory receptor, was upregulated in the exhausted cells in the present study. The previous studies indicated that the binding of its ligand Galectin 9 can mediate the release of BAT3 via tyrosine phosphorylation in the HAVCR2 tail and lead to inhibit the PI3K signaling pathway[54, 55]. In the models of mouse and human tumors, using the antibodies of HAVCR2 can rescue the exhausted T cells[56, 57], revealing that *HAVCR2* was associated with exhaustion of CD8+ T cell. Moreover, the inhibitory receptor encoded by gene *TIGIT* acts as T cell intrinsic inhibitor with its ligand CD155 by restraining effector cell activation and proliferation, leading to the decreased effector function and cytokine production[58], and the knockdown of *TIGIT* can revert

the decreased production of cytokines[59]. *TIGIT* was upregulated in the exhausted cells according to our study, showing its potential key role in the exhaustion of CD8+ T cells. Except for the inhibitory gene, the critical regulator of exhausted CD8+ T cells[60], *TOX*, was upregulated in the exhausted cells in the current study. As previously reported, TOX-knockout cells showed low expression of *PDCD1*, *ENTPD1*, *TIGIT*, *HAVCR2* as well as high expression of the transcription factors genes *TCF7*, *LEF1* and *ID3*[60, 61], while TOX-overexpressed cells resulted in reduced TNF and IFNγ expression as well as increased *PDCD1*, *TIGIT*, *LAG3* levels[62], which supported that *TOX* could drive T cell exhaustion. The genes described above were all upregulated in pre-exhausted or exhausted T cells, which could result in exhaustion of CD8+ T cell.

Apart from that, we found some other crucial genes. *CD69* is an early activation marker of T lymphocytes[63] and acts as a negative regulator of anti-tumor responses[64]. In the PPI network analysis, *CD69* was a hub gene of the module with the highest score and was upregulated in both pre_exhausted and exhausted cells. CD69 can mediate the cell retention via the interaction with S1PR1 (Sphingosine-1-Phosphate Receptor 1) which acts as a central mediator of lymphocyte output[65]. *CD69* was upregulated and *S1PR1* was downregulated significantly in the exhausted cells according to our study (Figure 3A). Thus we deduced that upregulated CD69 binds with downregulated S1PR1, leading to the impairment of lymphocyte output during an immune response[65]. The upregulated CD69 expressed on tumor-infiltrating T cells might interact with unknown ligands in the tumor to lead to the persistent antigen stimulation, which resulted in T cell exhaustion with the overexpression of inhibitory genes such as *PDCD1* and *HAVCR2*[66]. In the murine breast cancer model, *Cd69*$^{-/-}$ mice can reduce tumor growth and metastasis, increase the number of tumor-infiltrating lymphocytes, enhance IFNγ production and attenuate the T cell exhaustion[67]. In light of these researches, the upregulation of *CD69*

and downregulation of *S1PR1* might promote the development of T cell exhaustion. Additionally, *JUN* was a proto-oncogene and involved in the JNK signaling pathway which played a vital role in the proliferation and differentiation of CD8+ T cells[68]. Based on our findings, *JUN* was upregulated in both pre_exhausted and exhausted cells and regulated over 100 DEGs. According to the previous researches, *JUN* was a positive regulator of T cell activation[69, 70] and can promote T cell proliferation and the production of cytokines[71]. It is reported that one of the reasons of T cell exhaustion is the continuous stimulation after the initial cell activation[45, 72], and the upregulation of *JUN* both in the pre_exhausted and exhausted cells showed the lasting role to regulate cell activation, which might lead to T cell exhaustion finally. The ENCODE database showed that JUN is a transcription factor of *PDCD1* and *TIGIT*, thus we speculated that *JUN* may affect T cell exhaustion through transcriptional regulation of the exhaustion markers. However, the mechanisms of T cell exhaustion regulated by *JUN* needed further study. In addition, *SAMSN1* is an immunoinhibitory adaptor and the knockdown of *SAMSN1* can enhance the adaptive immunity in mice[73]. In the comparison of the common DEGs between pre_exhausted or exhausted and effector cells, *SAMSN1* showed greater upregulation in the exhausted cells and was associated with poor prognosis in HCC and NSCLC based on TCGA data. The expression of *SAMSN1* increased in B cells once stimulated by IL-4 through both PI3k/PKC/NF-κB and STAT6 pathway[74, 75]. *SAMSN1* might be associated with T cell exhaustion according to our results, however, further study is needed to explore the exact relationship between *SAMSN1* and T cell exhaustion.

By integrating above previous researches and our results, we summarized the key genes related to T cell exhaustion in Figure 4, including the upregulated genes of *CD69*, *PDCD1*, *HAVCR2*, *TIGIT*, *TOX* and *JUN*, as well as a downregulated gene *S1PR1*. The more specific mechanisms needed to be

explored in the future study.

In the further classification of exhausted cluster to explore the heterogeneity, we found that each patient tended to have one exhausted sub-cluster. Although the sub-clusters of each cancer may be regulated by different hub genes, we found that these sub-clusters were dysregulated in the cytokine-cytokine receptor interaction. Different kinds of cytokine played various and complex effects on T cells, for instance, the decreased expression of effector cytokines such as IL-2 and TNF-α as well as the increased expression of inhibitory cytokines such as IL-10 and TGF-β were reported to be associated with T cell exhaustion[76]. Additionally, the upregulation of pathways in cancer was consistent with the immune evasion of tumor caused by T cell exhaustion[77]. What's more, there were two downregulated pathways including natural killer cell mediated cytotoxicity and ribosome. There is a reciprocal relationship between NK and T cells, and most NK cell receptors are found on T cells[78]. Besides, there are similarities of NK cell and CD8+ T cell on the mechanisms of cell killing. They can release cytotoxic granules to the target cell such as tumor cells, and induce programmed cell death[79]. The dysfunctional cytotoxicity might attenuate the killing effector of CD8+ T cells on tumor cells. Additionally, the pathway of ribosome was also downregulated. Ribosome were essential for production in all living cells and regulated the cell growth and proliferation[80, 81]. During T cell exhaustion, the cells were characterized by the dysfunctional activity and bioenergetic deficiencies[82], which might be related to the damaged functions of ribosome. In addition to the universal pathways, there existed different pathways owing to each sub-cluster. In the CRC_S1 pathways, wnt signaling pathway was upregulated in our study. It has been reported that wnt signaling pathway contributes to tumor progression[83] and its blockade can improve the generation of effector cells[84], thus the upregulation of wnt signaling pathway might contribute to T cell exhaustion and facilitate the growth

of tumor; In the NSCLC_S1 pathways, the upregulated apoptosis may lead to the impaired function of T cells[85].

In the current study, we integrated three cancers and found some key genes associated with T cell exhaustion, including the known markers *PDCD1*, *HAVCR2*, *TIGIT*, *TOX*. In the gene interaction network, we found significant genes *CD69* and *JUN*. Besides, we firstly identified a potential exhaustion associated gene, *SAMSN1*, which enriched in the negative regulation of immunity and showed poor survival. We summarized the putative interrelated mechanisms of above key genes identified in this study by integrating the previous knowledge. Furthermore, we explored the heterogeneous and preference of exhausted CD8+ T cells in each patient and found only one exhausted sub-cluster existed in most of patients, especially in CRC and HCC. Our study provides a universal molecular mechanism on T cell exhaustion and a new way for functional research of single-cell RNA sequencing data, which might promote the further realization of immunotherapy.

Acknowledgements

This work was supported by the National Key R&D Program of China (2018YFC0910200), the Key R&D Program of Guangdong Province (2019B020226001), and the Science and the Technology Planning Project of Guangzhou (201704020176).

References

[1] MOSKOPHIDIS D, LECHNER F, PIRCHER H, et al. Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells [J]. Nature, 1993, 362(6434): 758-61.

[2] CATAKOVIC K, KLIESER E, NEUREITER D, et al. T cell exhaustion: from pathophysiological basics to tumor immunotherapy [J]. Cell Communication & Signaling Ccs, 2017, 15(1): 1.

[3] LEE P P, YEE C, SAVAGE P A, et al. Characterization of circulating T cells specific for tumor-associated antigens in melanoma patients [J]. Nat Med, 1999, 5(6): 677-85.

[4] PAUKEN K E, WHERRY E J. Overcoming T cell exhaustion in infection and cancer [J]. Trends Immunol, 2015, 36(4): 265-76.

[5] SCHIETINGER A, PHILIP M, KRISNAWAN V E, et al. Tumor-Specific T Cell Dysfunction Is a Dynamic Antigen-Driven Differentiation Program Initiated Early during Tumorigenesis [J]. Immunity, 2016, 45(2): 389-401.

[6] PAUKEN K E, WHERRY E J. Overcoming T cell exhaustion in infection and cancer [J]. Trends in Immunology, 2015, 36(4): 265-76.

[7] BUCKS C M, NORTON J A, BOESTEANU A C, et al. Chronic antigen stimulation alone is sufficient to drive CD8+ T cell exhaustion [J]. J Immunol, 2009, 182(11): 6697-708.

[8] BLACKBURN S D, SHIN H, HAINING W N, et al. Coregulation of CD8+ T cell exhaustion by multiple inhibitory receptors during chronic viral infection [J]. Nature immunology, 2009, 10(1): 29.

[9] BAITSCH L, BAUMGAERTNER P, DEVEVRE E, et al. Exhaustion of tumor-specific CD8(+) T cells in metastases from melanoma patients [J]. J Clin Invest, 2011, 121(6): 2350-60.

[10] MCLANE L M, ABDEL-HAKEEM M S, WHERRY E J. CD8 T Cell Exhaustion During Chronic Viral Infection and Cancer [J]. Annual Review of Immunology, 2019, 37(1): 457-95.

[11] SEN D R, KAMINSKI J, BARNITZ R A, et al. The epigenetic landscape of T cell exhaustion [J]. Science, 2016, 354(6316): 1165-9.

[12] CRAWFORD A, ANGELOSANTO J M, KAO C, et al. Molecular and transcriptional basis of $CD4^+$ T cell dysfunction during chronic infection [J]. Immunity, 2014, 40(2): 289-302.

[13] WAGNER A, REGEV A, YOSEF N. Revealing the vectors of cellular identity with single-cell genomics [J]. Nature Biotechnology, 2016, 34(11): 1145.

[14] WINTERHOFF B J, MAILE M, MITRA A K, et al. Single cell sequencing reveals heterogeneity within ovarian cancer epithelium and cancer associated stromal cells [J]. Gynecol Oncol, 2017, 144(3): 598-606.

[15] WU H, YU J, LI Y, et al. Single-cell RNA sequencing reveals diverse intratumoral heterogeneities and gene signatures of two types of esophageal cancers [J]. Cancer Lett, 2018, 438(133-43.

[16] PAPALEXI E, SATIJA R. Single-cell RNA sequencing to explore immune cell heterogeneity [J]. Nature Reviews Immunology, 2017, 18(1): 35.

[17] AZIZI E, CARR A J, PLITAS G, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment [J]. Cell, 2018, 174(S0092867418307232-.

[18] DARMANIS S, SLOAN S A, CROOTE D, et al. Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma [J]. Cell Reports, 2017, 21(5): 1399.

[19] PURAM S V, TIROSH I, PARIKH A S, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer [J]. Cell, 2017, 171(7): 1611-24 e24.

[20] TIROSH I, IZAR B, PRAKADAN S M, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq [J]. Science, 2016, 352(6282): 189-96.

[21] HUANG A C, POSTOW M A, ORLOWSKI R J, et al. T-cell invigoration to tumour burden ratio associated with anti-PD-1 response [J]. Nature, 2017, 545(7652): 60-5.

[22] VENTEICHER A S, TIROSH I, HEBERT C, et al. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq [J]. Science, 2017, 355(6332):

[23] KISELEV V Y, KIRSCHNER K, SCHAUB M T, et al. SC3: consensus clustering of single-cell RNA-seq data [J]. Nature Methods, 2017, 14(483-6.

[24] ZHANG X, LAN Y, XU J, et al. CellMarker: a manually curated resource of cell markers in human and mouse [J]. Nucleic Acids Research, 2018, 47(D1): D721-D8.

[25] TRAPNELL C, CACCHIARELLI D, GRIMSBY J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells [J]. Nat Biotechnol, 2014, 32(4): 381-6.

[26] ROBINSON M D, MCCARTHY D J, SMYTH G K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data [J]. Bioinformatics, 2010, 26(1): 139-40.

[27] YU G, WANG L G, HAN Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters [J]. OMICS, 2012, 16(5): 284-7.

[28] SHANNON P, MARKIEL A, OZIER O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks [J]. Genome Res, 2003, 13(11): 2498-504.

[29] VON MERING C, HUYNEN M, JAEGGI D, et al. STRING: a database of predicted functional associations between proteins [J]. Nucleic Acids Res, 2003, 31(1): 258-61.

[30] BADER G D, HOGUE C W. An automated method for finding molecular complexes in large protein interaction networks [J]. BMC Bioinformatics, 2003, 4(2.

[31] LACHMANN A, XU H, KRISHNAN J, et al. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments [J]. Bioinformatics, 2010, 26(19): 2438-44.

[32] CONSORTIUM E P. The ENCODE (ENCyclopedia Of DNA Elements) Project [J]. Science, 2004, 306(5696): 636-40.

[33] ROUILLARD A D, GUNDERSEN G W, FERNANDEZ N F, et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins [J]. Database (Oxford), 2016, 2016(

[34] HAN H, CHO J W, LEE S, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions [J]. Nucleic Acids Res, 2018, 46(D1): D380-D6.

[35] TOMCZAK K, CZERWINSKA P, WIZNEROWICZ M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge [J]. Contemp Oncol (Pozn), 2015, 19(1A): A68-77.

[36] ZHENG C, ZHENG L, YOO J K, et al. Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing [J]. Cell, 2017, 169(7): 1342-56 e16.

[37] HANZELMANN S, CASTELO R, GUINNEY J. GSVA: gene set variation analysis for microarray and RNA-seq data [J]. BMC Bioinformatics, 2013, 14(7.

[38] LIBERZON A, BIRGER C, THORVALDSDOTTIR H, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection [J]. Cell Syst, 2015, 1(6): 417-25.

481 [39] RITCHIE M E, PHIPSON B, WU D, et al. limma powers differential expression analyses
482 for RNA-sequencing and microarray studies [J]. Nucleic Acids Res, 2015, 43(7): e47.
483 [40] LI R, QIAN J, WANG Y Y, et al. Long noncoding RNA profiles reveal three molecular
484 subtypes in glioma [J]. CNS Neurosci Ther, 2014, 20(4): 339-43.
485 [41] XIYING F, ALEXANDER.Y R. Hallmarks of Tissue-Resident Lymphocytes [J]. Cell, 2016,
486 164(1198-211.
487 [42] MEHTA A K, GRACIAS D T, CROFT M. TNF activity and T cells [J]. Cytokine, 2018,
488 101(14-8.
489 [43] HARLIN H, MENG Y, PETERSON A C, et al. Chemokine expression in melanoma
490 metastases associated with CD8+ T-cell recruitment [J]. Cancer Res, 2009, 69(7): 3077-85.
491 [44] SUKUMAR M, LIU J, JI Y, et al. Inhibiting glycolytic metabolism enhances CD8+ T cell
492 memory and antitumor function [J]. J Clin Invest, 2013, 123(10): 4479-88.
493 [45] WHERRY E J K, M. Molecular and cellular insights into T cell exhaustion [J]. Nat Rev
494 Immunol, 2015, 15(486–99.
495 [46] CONWAY J R, LEX A, GEHLENBORG N. UpSetR: an R package for the visualization of
496 intersecting sets and their properties [J]. Bioinformatics, 2017, 33(18): 2938-40.
497 [47] KALLIES A, ZEHN D, UTZSCHNEIDER D T. Precursor exhausted T cells: key to
498 successful immunotherapy? [J]. Nat Rev Immunol, 2019,
499 [48] POLLEN A A, NOWAKOWSKI T J, SHUGA J, et al. Low-coverage single-cell mRNA
500 sequencing reveals cellular heterogeneity and activated signaling pathways in developing
501 cerebral cortex [J]. Nat Biotechnol, 2014, 32(10): 1053-8.
502 [49] WHERRY E J. T cell exhaustion [J]. Nat Immunol, 2011, 12(6): 492-9.

503 [50] BARBER D L, WHERRY E J, MASOPUST D, et al. Restoring function in exhausted CD8
504 T cells during chronic viral infection [J]. Nature, 2006, 439(7077): 682-7.
505 [51] CHEMNITZ J M, PARRY R V, NICHOLS K E, et al. SHP-1 and SHP-2 associate with
506 immunoreceptor tyrosine-based switch motif of programmed death 1 upon primary human T
507 cell stimulation, but only receptor ligation prevents T cell activation [J]. J Immunol, 2004,
508 173(2): 945-54.
509 [52] RIELLA L V, PATERSON A M, SHARPE A H, et al. Role of the PD-1 pathway in the
510 immune response [J]. American journal of transplantation : official journal of the American
511 Society of Transplantation and the American Society of Transplant Surgeons, 2012, 12(10):
512 2575-87.
513 [53] PATSOUKIS N, BROWN J, PETKOVA V, et al. Selective effects of PD-1 on Akt and Ras
514 pathways regulate molecular components of the cell cycle and inhibit T cell proliferation [J]. Sci
515 Signal, 2012, 5(230): ra46.
516 [54] KOGUCHI K, ANDERSON D E, YANG L, et al. Dysregulated T cell expression of TIM3 in
517 multiple sclerosis [J]. J Exp Med, 2006, 203(6): 1413-8.
518 [55] DAS M, ZHU C, KUCHROO V K. Tim-3 and its role in regulating anti-tumor immunity [J].
519 Immunol Rev, 2017, 276(1): 97-111.
520 [56] NGIOW S F, VON SCHEIDT B, AKIBA H, et al. Anti-TIM3 antibody promotes T cell
521 IFN-gamma-mediated antitumor immunity and suppresses established tumors [J]. Cancer Res,
522 2011, 71(10): 3540-51.
523 [57] FOURCADE J, SUN Z, BENALLAOUA M, et al. Upregulation of Tim-3 and PD-1
524 expression is associated with tumor antigen-specific CD8+ T cell dysfunction in melanoma

525 patients [J]. J Exp Med, 2010, 207(10): 2175-86.

526 [58] JOLLER N, KUCHROO V K. Tim-3, Lag-3, and TIGIT [J]. Curr Top Microbiol Immunol,

527 2017, 410(127-56.

528 [59] KONG Y, ZHU L, SCHELL T D, et al. T-Cell Immunoglobulin and ITIM Domain (TIGIT)

529 Associates with CD8+ T-Cell Exhaustion and Poor Clinical Outcome in AML Patients [J]. Clin

530 Cancer Res, 2016, 22(12): 3057-66.

531 [60] SCOTT A C, DUNDAR F, ZUMBO P, et al. TOX is a critical regulator of tumour-specific T

532 cell differentiation [J]. Nature, 2019, 571(7764): 270-4.

533 [61] KHAN O, GILES J R, MCDONALD S, et al. TOX transcriptionally and epigenetically

534 programs CD8(+) T cell exhaustion [J]. Nature, 2019, 571(7764): 211-8.

535 [62] ALFEI F, KANEV K, HOFMANN M, et al. TOX reinforces the phenotype and longevity of

536 exhausted T cells in chronic viral infection [J]. Nature, 2019, 571(7764): 265-9.

537 [63] ZIEGLER S F, RAMSDELL F, HJERRILD K A, et al. Molecular characterization of the

538 early activation antigen CD69: a type II membrane glycoprotein related to a family of natural

539 killer cell activation antigens [J]. Eur J Immunol, 1993, 23(7): 1643-8.

540 [64] ESPLUGUES E, SANCHO D, VEGA-RAMOS J, et al. Enhanced antitumor immunity in

541 mice deficient in CD69 [J]. The Journal of experimental medicine, 2003, 197(9): 1093-106.

542 [65] CYSTER J G, SCHWAB S R. Sphingosine-1-phosphate and lymphocyte egress from

543 lymphoid organs [J]. Annu Rev Immunol, 2012, 30(69-94.

544 [66] KIMURA M Y, HAYASHIZAKI K, TOKOYODA K, et al. Crucial role for CD69 in allergic

545 inflammatory responses: CD69-Myl9 system in the pathogenesis of airway inflammation [J].

546 Immunol Rev, 2017, 278(1): 87-100.

547 [67] MITA Y, KIMURA M Y, HAYASHIZAKI K, et al. Crucial role of CD69 in anti-tumor
548 immunity through regulating the exhaustion of tumor-infiltrating T cells [J]. Int Immunol, 2018,
549 30(12): 559-67.
550 [68] ARBOUR, N. c-Jun NH2-Terminal Kinase (JNK)1 and JNK2 Signaling Pathways Have
551 Divergent Roles in CD8+ T Cell-mediated Antiviral Immunity [J]. Journal of Experimental
552 Medicine, 195(7): 801-10.
553 [69] BEIQING L, CHEN M, WHISLER R L. Sublethal levels of oxidative stress stimulate
554 transcriptional activation of c-jun and suppress IL-2 promoter activation in Jurkat T cells [J]. J
555 Immunol, 1996, 157(1): 160-9.
556 [70] QIAO X, PHAM D N T, LUO H, et al. Ran Overexpression Leads to Diminished T Cell
557 Responses and Selectively Modulates Nuclear Levels of c-Jun and c-Fos [J]. Journal of
558 Biological Chemistry, 285(8): 5488-96.
559 [71] UCHIDA M, SAWA H. [The SWI/SNF chromatin-remodeling complex regulates
560 asymmetric cell division in C. elegans] [J]. Tanpakushitsu Kakusan Koso Protein Nucleic Acid
561 Enzyme, 2005, 50(6 Suppl): 569-74.
562 [72] MOGNOL G P, SPREAFICO R, WONG V, et al. Exhaustion-associated regulatory
563 regions in CD8(+) tumor-infiltrating T cells [J]. Proc Natl Acad Sci U S A, 2017, 114(13):
564 E2776-E85.
565 [73] WANG D, STEWART A K, ZHUANG L, et al. Enhanced adaptive immunity in mice lacking
566 the immunoinhibitory adaptor Hacs1 [J]. FASEB J, 2010, 24(3): 947-56.
567 [74] YAN Y, ZHANG L, XU T, et al. SAMSN1 is highly expressed and associated with a poor
568 survival in glioblastoma multiforme [J]. PLoS One, 2013, 8(11): e81905.

569 [75] ZHU Y X, BENN S, LI Z H, et al. The SH3-SAM adaptor HACS1 is up-regulated in B cell
570 activation signaling cascades [J]. J Exp Med, 2004, 200(6): 737-47.
571 [76] HE Q F, XU Y, LI J, et al. CD8+ T-cell exhaustion in cancer: mechanisms and new area
572 for cancer immunotherapy [J]. Briefings in functional genomics, 2019, 18(2): 99-106.
573 [77] CATAKOVIC K, KLIESER E, NEUREITER D, et al. T cell exhaustion: from
574 pathophysiological basics to tumor immunotherapy [J]. Cell Commun Signal, 2017, 15(1): 1.
575 [78] VIVIER E, NUNES J A, VELY F. Natural killer cell signaling pathways [J]. Science, 2004,
576 306(5701): 1517-9.
577 [79] GALVIN J P, SPAENY-DEKKING L H, WANG B, et al. Apoptosis induced by granzyme
578 B-glycosaminoglycan complexes: implications for granule-mediated apoptosis in vivo [J].
579 Journal of immunology (Baltimore, Md : 1950), 1999, 162(9): 5345-50.
580 [80] CATEZ F, DALLA VENEZIA N, MARCEL V, et al. Ribosome biogenesis: An emerging
581 druggable pathway for cancer therapeutics [J]. Biochem Pharmacol, 2019, 159(74-81.
582 [81] TAN T C J, KNIGHT J, SBARRATO T, et al. Suboptimal T-cell receptor signaling
583 compromises protein translation, ribosome biogenesis, and proliferation of mouse CD8 T cells
584 [J]. Proc Natl Acad Sci U S A, 2017, 114(30): E6117-E26.
585 [82] WHERRY E J, HA S J, KAECH S M, et al. Molecular signature of CD8+ T cell exhaustion
586 during chronic viral infection [J]. Immunity, 2007, 27(4): 670-84.
587 [83] POLAKIS P. Wnt signaling in cancer [J]. Cold Spring Harbor perspectives in biology, 2012,
588 4(5):
589 [84] SCHINZARI V, TIMPERI E, PECORA G, et al. Wnt3a/beta-Catenin Signaling Conditions
590 Differentiation of Partially Exhausted T-effector Cells in Human Cancers [J]. Cancer Immunol

Res, 2018, 6(8): 941-52.

[85] BARATHAN M, GOPAL K, MOHAMED R, et al. Chronic hepatitis C virus infection triggers spontaneous differential expression of biosignatures associated with T cell exhaustion and apoptosis signaling in peripheral blood mononucleocytes [J]. Apoptosis, 2015, 20(4): 466-80.

Figure legends

Figure 1. Clustering of tumor-infiltrating CD8+ T cells.

A-C. Cell clustering using the selected genes of the top n ones with highest variance in CRC (A, n=2500), HCC (B, n=1000), NSCLC (C, n=1000). The colors of the upper bar highlight the different clusters.

D-F. The trajectory analysis of CD8+ T cell (excluding conventional cells) state transition inferred by Monocle in CRC (D), HCC (E), NSCLC (F). Each dot represents an individual cell, colored according to the clusters.

G-I. t-SNE plots of CD8+ T cells show distinct clusters predominantly determined by cell types in CRC (G), HCC (H), NSCLC (I). Each dot represents an individual cell, colored according to the clusters.

J. t-SNE plots of the total CD8+ T cells of three cancers show the uniformity according to the cell types, regardless of the type of different cancers.

Figure 2. The Differentially expressed genes associated with exhausted state.

A. The fractions of five clusters of CD8+ T cells across peripheral blood (P), adjacent normal (N), and

tumor tissues (T). Bars were colored according to the clusters.

B-C. Venn diagram of DEGs in pre_exhausted (B) or exhausted (C) vs. effector comparison.

D-E. GO and KEGG enrichment analysis of DEGs in pre_exhausted vs. effector comparison. The top 15 significantly enriched GO terms (D) and KEGG pathways (E).

F-G. GO and KEGG enrichment analysis of DEGs in exhausted vs. effector comparison. The top 15 significantly enriched GO terms (F) and KEGG pathways (G).

Figure 3. The key DEGs associated with exhausted state.

A. The DEGs with higher fold change in the exhausted cells. The number represents the logFC value.

B. The survival analysis of *SAMSN1* based on TCGA data. Red corresponds to the group of high expression, blue corresponds to the group of low expression.

C. The PPI network module with the highest score. Colors correspond to state of DEGs (red: up, blue: down), shapes correspond to the types of DEGs (triangle: the specific genes existed in only pre_exhausted cells, ellipse: the common genes existed in both exhausted and pre_exhausted cells, diamond: the specific genes existed in only exhausted cells).

D. The number of DEGs regulated by their corresponding TFs. Colors correspond to state of TFs (red: up, blue: down), shapes correspond to the types (ellipse: the dysregulated TFs, diamond: the number DEGs regulated by corresponding TFs).

Figure 4. The putative interrelated mechanisms of key genes related to CD8+ T cell exhaustion.

A. CD69, which can combine with S1PR1 and inhibit the function of S1PR1 to regulate cell retention, expressed on tumor-infiltrating T cells and may interact with unknown ligands in the tumor cell to lead to the persistent antigen stimulation.

B-D. The mechanism of genes encoding inhibitory receptor related to CD8+ T cell exhaustion. The upregulation of all three inhibitory receptors, *PDCD1*, *HAVCR2* and *TIGIT*, would result in the inhibition of PI3K/Akt signaling pathway and lead to reduced survival and proliferation.

E-F. The transcription factor related to CD8+ T cell exhaustion. (E)*TOX* can promote the expression of inhibitory receptors including *PDCD1*, *HAVCR2*, *TIGIT* to mediate T cell exhaustion program. (F) *JUN* is proposed to regulate the expression of exhaustion associated *PDCD1* and *TIGIT* to play vital role in T cell exhaustion at present study.

The red fonts represent upregulated expression, while the blue fonts represent downregulated expression.

Figure 5. Clustering of exhausted cells.

A-C. Cell clustering using the selected genes of the top n ones with highest variance in the exhausted cluster of CRC (A, n=1500), HCC (B, n=3000), NSCLC (C, n=500). The colors of the upper bar highlight the different sub-clusters.

D-F. t-SNE plots of the exhausted sub-clusters show heterogeneous and preference in each patient of CRC (D), HCC (E) and NSCLC (F). Each dot represents an individual cell, colored according to the sub-clusters, gray means the non-exhausted cells.

Figure 6. The DEGs of exhausted sub-clusters.

A. Heatmap of six sub-clusters across three cancers. The colors of the upper bar highlight the different sub-clusters.

B. Upset plot of DEGs between exhausted sub-clusters and effector cells. The numbers of exhausted

sub-clusters are shown on the left of the bar. The red line linked with dots show the common genes across six sub-clusters.

Figure 7. The differential pathways between exhausted sub-clusters and effector cells by GSVA with more than 10 DEGs and FDR < 0.05.

Supplements

Table S1. The number of markers in each cluster of each cancer and corresponding genes in CellMarker database.

Table S2. DEGs of pre_exhausted or exhausted vs. effector comparison.

Table S3. GO and KEGG pathways in pre_exhausted or exhausted vs. effector comparison.

Table S4. DEGs of exhausted sub-clusters vs. effector comparison.

Table S5. GSVA pathway analysis results in exhausted sub-clusters vs. effector comparison.

Figure S1. The distribution of effector cells among peripheral blood, normal and tumor tissues in CRC (A), HCC (B), NSCLC (C).

Figure S2. Violin plots of the key genes in pre_exhausted or exhausted vs. effector comparison.

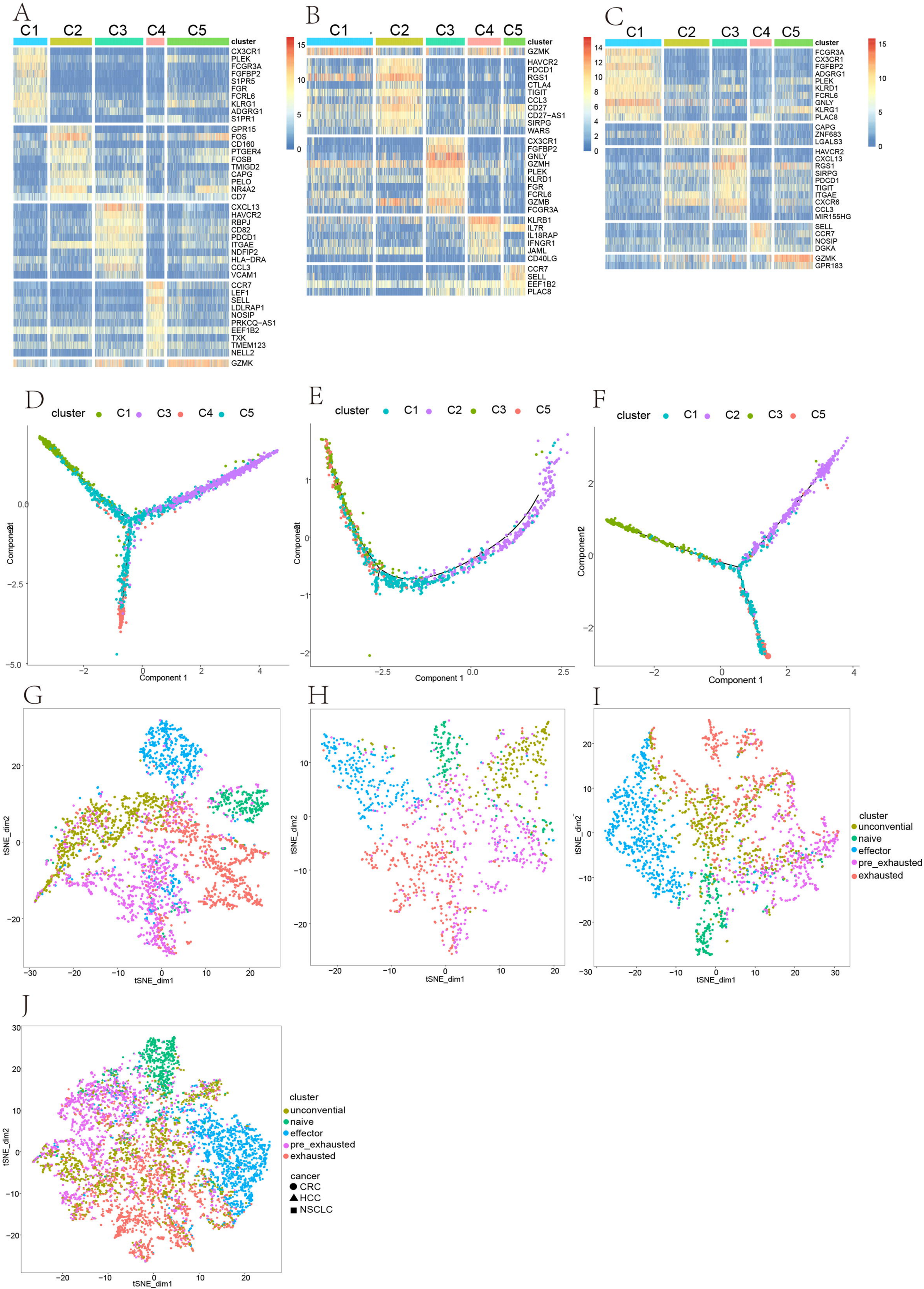
A. The expression of key genes in the effector, pre_exhausted and exhausted cells of different cancers.

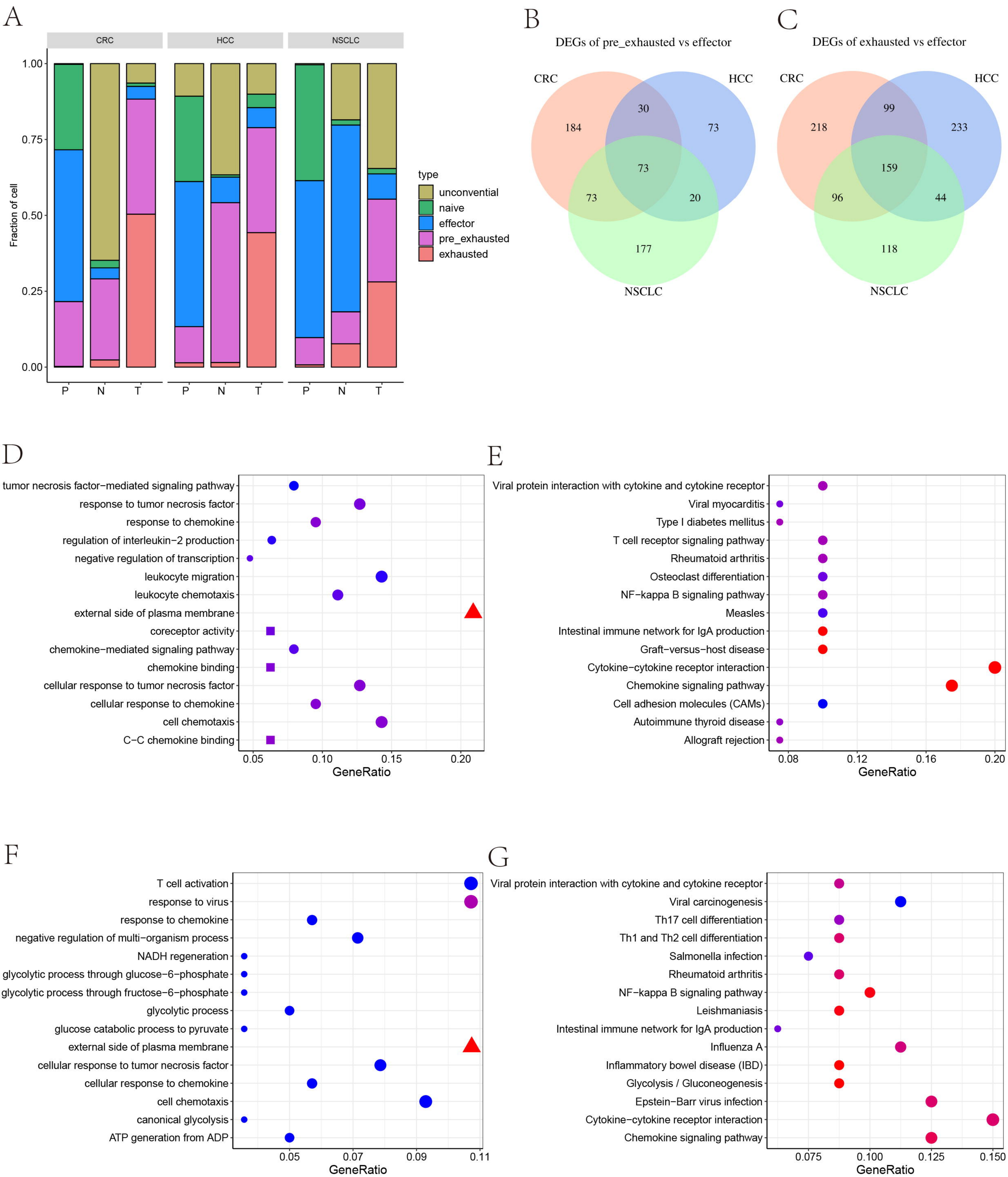
The black dot represents the median of expression value of genes (log_2(TPM/10+1)).

B-D. The expression of key genes in each patient of CRC (B), HCC (C) and NSCLC (D).

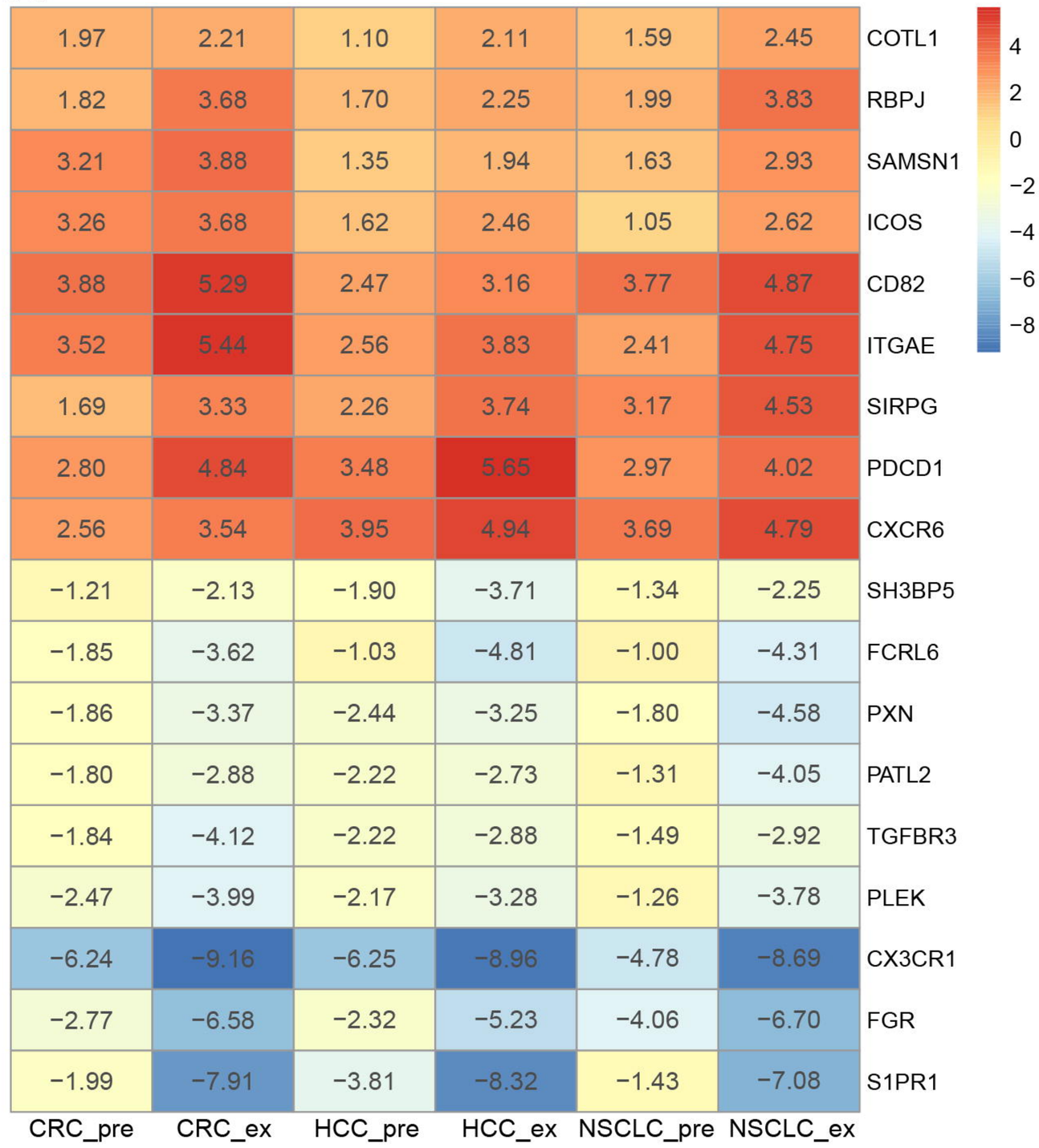
Figure S3. The PPI network of DEGs in pre_exhausted or exhausted vs. effector comparison. The top network consists of overall DEGs, and the remaining six networks are the module identified by plugin MCODE of Cytoscape. Colors correspond to state of DEGs (red: up, blue: down), shapes correspond to the types of DEGs (triangle: the specific genes existed in only pre_exhausted cells, ellipse: the common genes existed in both exhausted and pre_exhausted cells, diamond: the specific genes existed in only exhausted cells).

Figure S4. The network of DEGs and their corresponding TFs. Colors correspond to state of TFs (red: up, blue: down), shapes correspond to the types (ellipse: the dysregulated TFs, diamond: the target genes of the TFs).

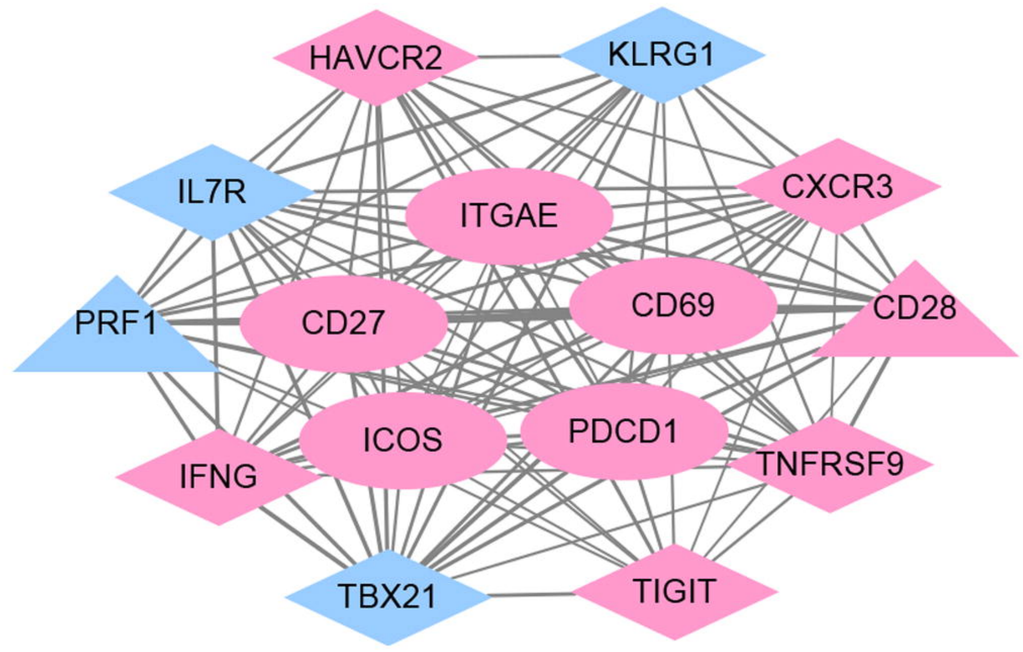

A
C1 C2 C3 C4 C5
cluster
CX3CR1 PLEK FCGR3A FGFBP2 S1PR5 FGR FCRL6 KLRG1 ADGRG1 S1PR1
GPR15 FOS CD160 PTGER4 FOSB TMIGD2 CAPG PELO NR4A2 CD7
CXCL13 HAVCR2 RBPJ CD82 PDCD1 ITGAE NDFIP2 HLA-DRA CCL3 VCAM1
CCR7 LEF1 SELL LDLRAP1 NOSIP PRKCQ-AS1 EEF1B2 TXK TMEM123 NELL2
GZMK
B
C1 C2 C3 C4 C5
cluster
GZMK
HAVCR2 PDCD1 RGS1 CTLA4 TIGIT CCL3 CD27 CD27-AS1 SIRPG WARS
CX3CR1 FGFBP2 GNLY GZMH PLEK KLRD1 FGR FCRL6 GZMB FCGR3A
KLRB1 IL7R IL18RAP IFNGR1 JAML CD40LG
CCR7 SELL EEF1B2 PLAC8
C
C1 C2 C3 C4 C5
cluster
FCGR3A CX3CR1 FGFBP2 ADGRG1 PLEK KLRD1 FCRL6 GNLY KLRG1 PLAC8
CAPG ZNF683 LGALS3
HAVCR2 CXCL13 RGS1 SIRPG PDCD1 TIGIT ITGAE CXCR6 CCL3 MIR155HG
SELL CCR7 NOSIP DGKA
GZMK GPR183
D
cluster C1 C3 C4 C5
Component 1
Component 2
E
cluster C1 C2 C3 C5
Component 1
Component 2
F
cluster C1 C2 C3 C5
Component 1
Component 2
G
tSNE_dim1
tSNE_dim2
H
tSNE_dim1
tSNE_dim2
I
tSNE_dim1
tSNE_dim2
cluster
unconvential
naive
effector
pre_exhausted
exhausted
J
tSNE_dim1
tSNE_dim2
cluster
unconvential
naive
effector
pre_exhausted
exhausted
cancer
CRC
HCC
NSCLC


A
CRC
HCC
NSCLC
Fraction of cell
1.00
0.75
0.50
0.25
0.00
P N T P N T P N T
type
unconvential
naive
effector
pre_exhausted
exhausted
B
DEGs of pre_exhausted vs effector
CRC
HCC
NSCLC
184
30
73
73
73
20
177
C
DEGs of exhausted vs effector
CRC
HCC
NSCLC
218
99
233
159
96
44
118
D
tumor necrosis factor−mediated signaling pathway
response to tumor necrosis factor
response to chemokine
regulation of interleukin−2 production
negative regulation of transcription
leukocyte migration
leukocyte chemotaxis
external side of plasma membrane
coreceptor activity
chemokine−mediated signaling pathway
chemokine binding
cellular response to tumor necrosis factor
cellular response to chemokine
cell chemotaxis
C−C chemokine binding
0.05
0.10
0.15
0.20
GeneRatio
E
Viral protein interaction with cytokine and cytokine receptor
Viral myocarditis
Type I diabetes mellitus
T cell receptor signaling pathway
Rheumatoid arthritis
Osteoclast differentiation
NF−kappa B signaling pathway
Measles
Intestinal immune network for IgA production
Graft−versus−host disease
Cytokine−cytokine receptor interaction
Chemokine signaling pathway
Cell adhesion molecules (CAMs)
Autoimmune thyroid disease
Allograft rejection
0.08
0.12
0.16
0.20
GeneRatio
F
T cell activation
response to virus
response to chemokine
negative regulation of multi−organism process
NADH regeneration
glycolytic process through glucose−6−phosphate
glycolytic process through fructose−6−phosphate
glycolytic process
glucose catabolic process to pyruvate
external side of plasma membrane
cellular response to tumor necrosis factor
cellular response to chemokine
cell chemotaxis
canonical glycolysis
ATP generation from ADP
0.05
0.07
0.09
0.11
GeneRatio
G
Viral protein interaction with cytokine and cytokine receptor
Viral carcinogenesis
Th17 cell differentiation
Th1 and Th2 cell differentiation
Salmonella infection
Rheumatoid arthritis
NF−kappa B signaling pathway
Leishmaniasis
Intestinal immune network for IgA production
Influenza A
Inflammatory bowel disease (IBD)
Glycolysis / Gluconeogenesis
Epstein−Barr virus infection
Cytokine−cytokine receptor interaction
Chemokine signaling pathway
0.075
0.100
0.125
0.150
GeneRatio

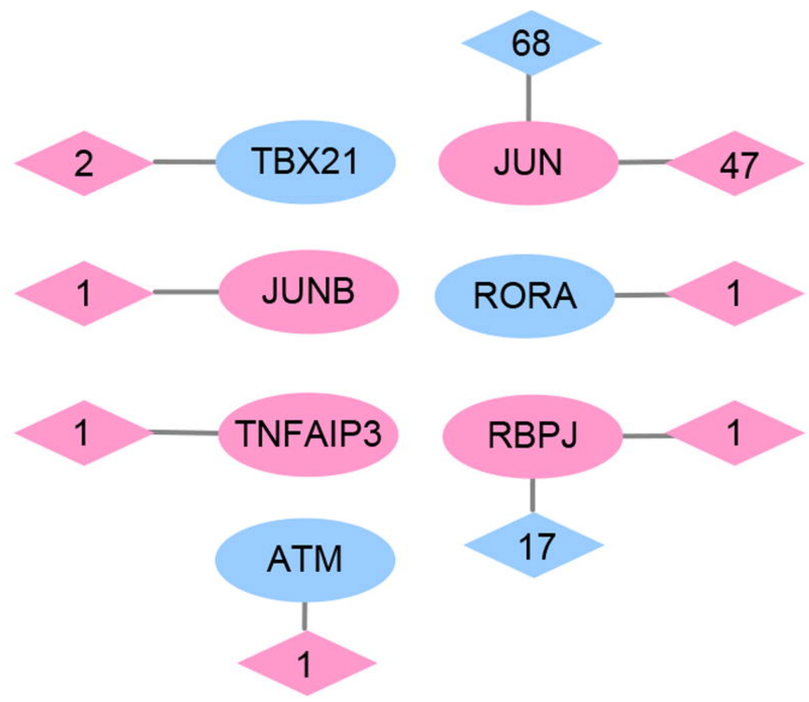
A



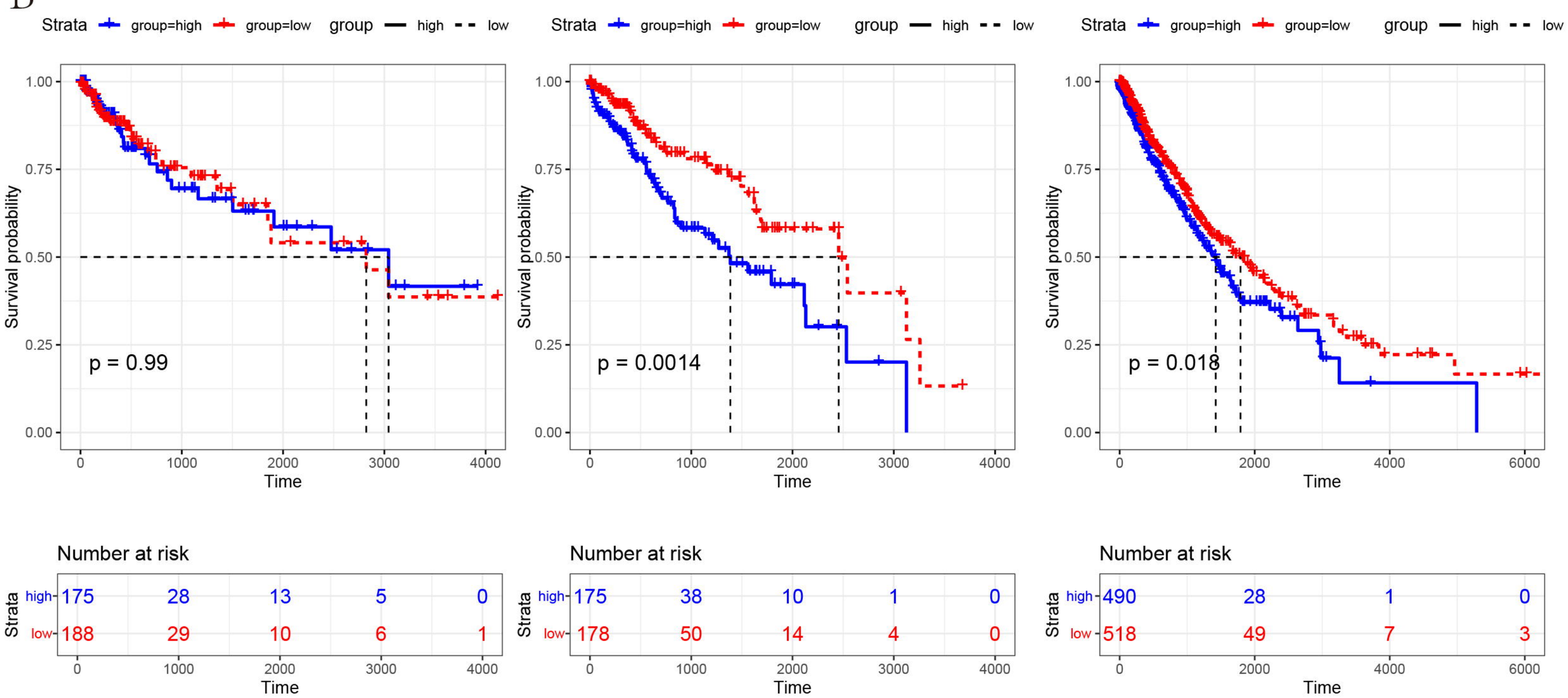
C



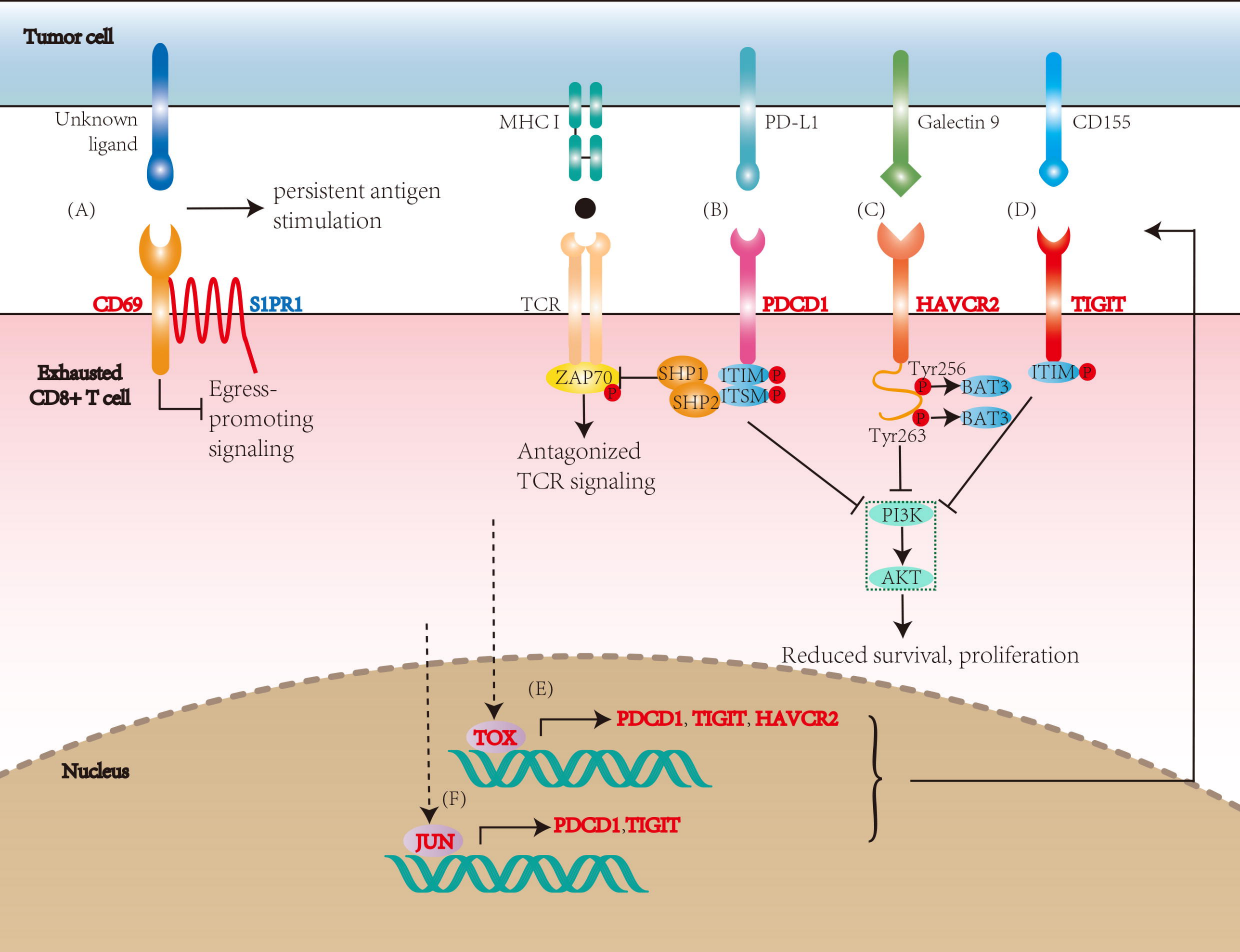
D

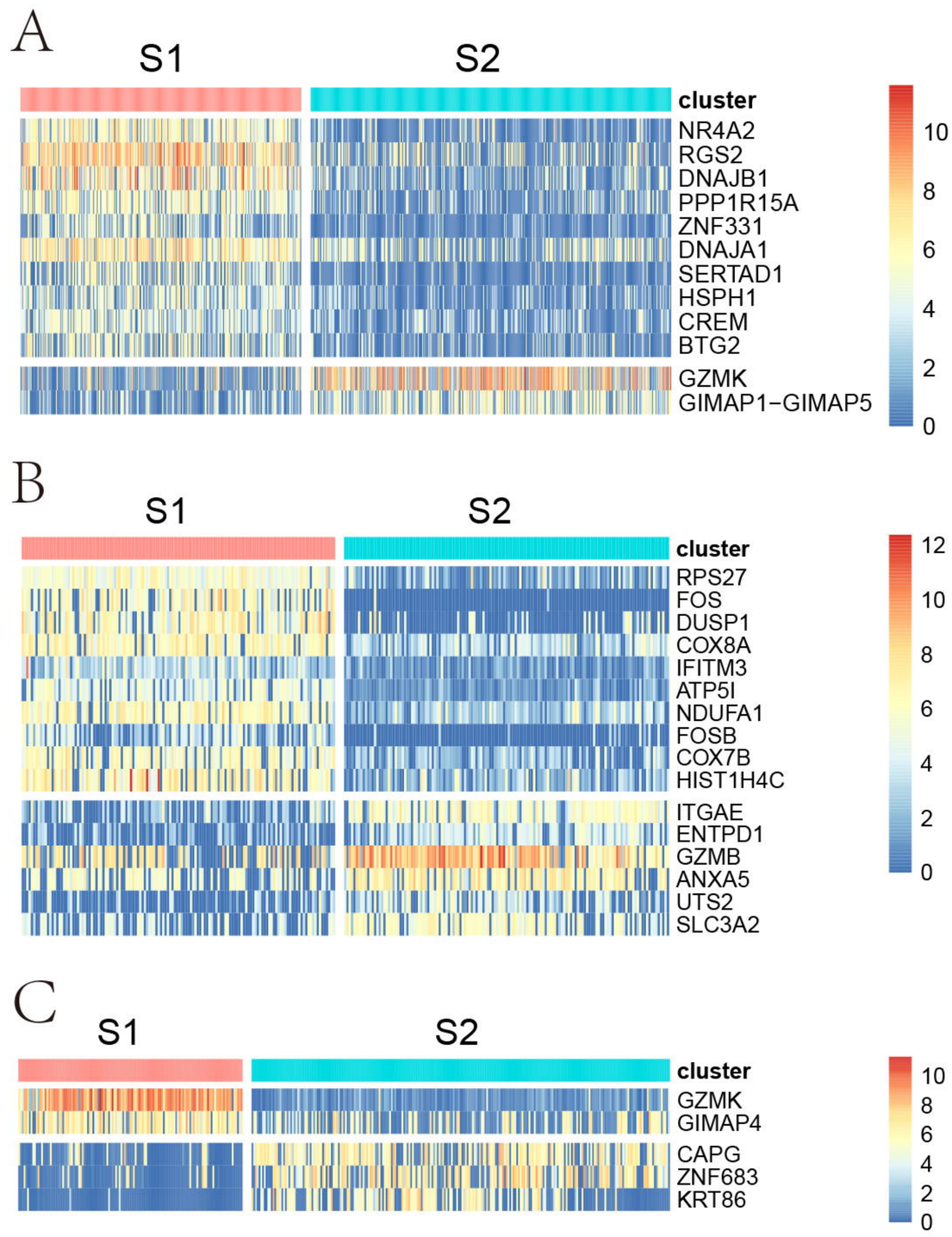


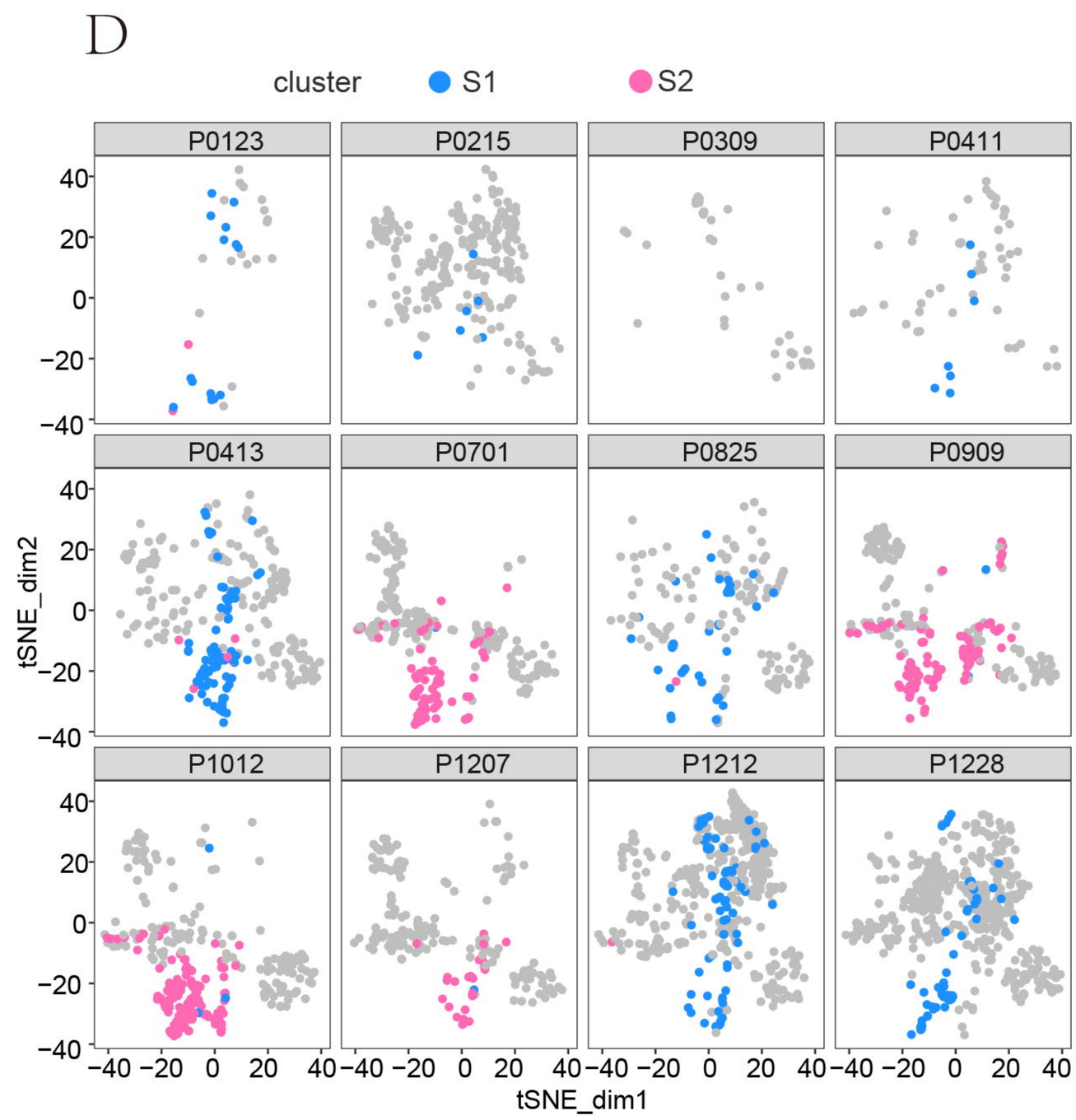
B

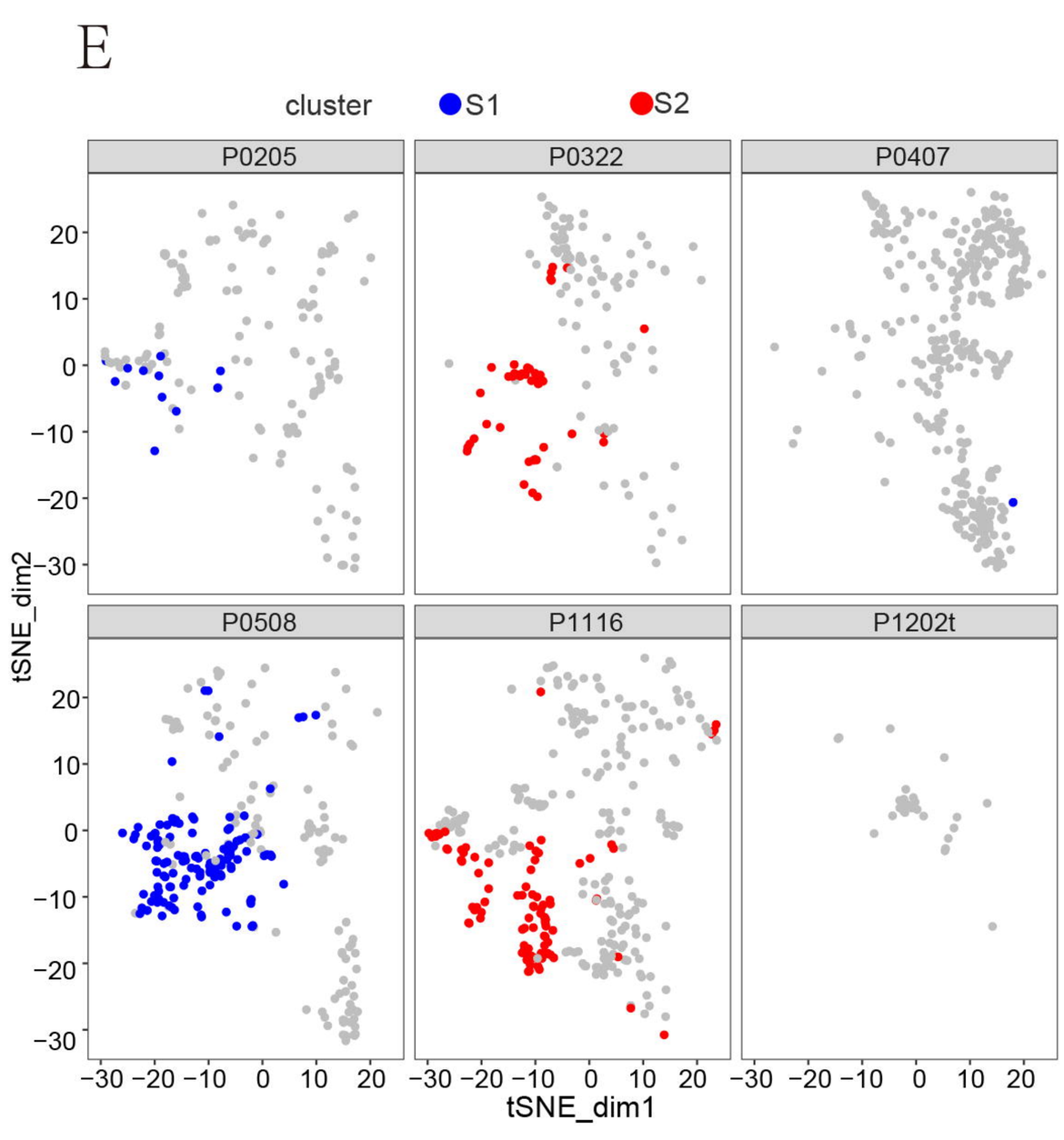


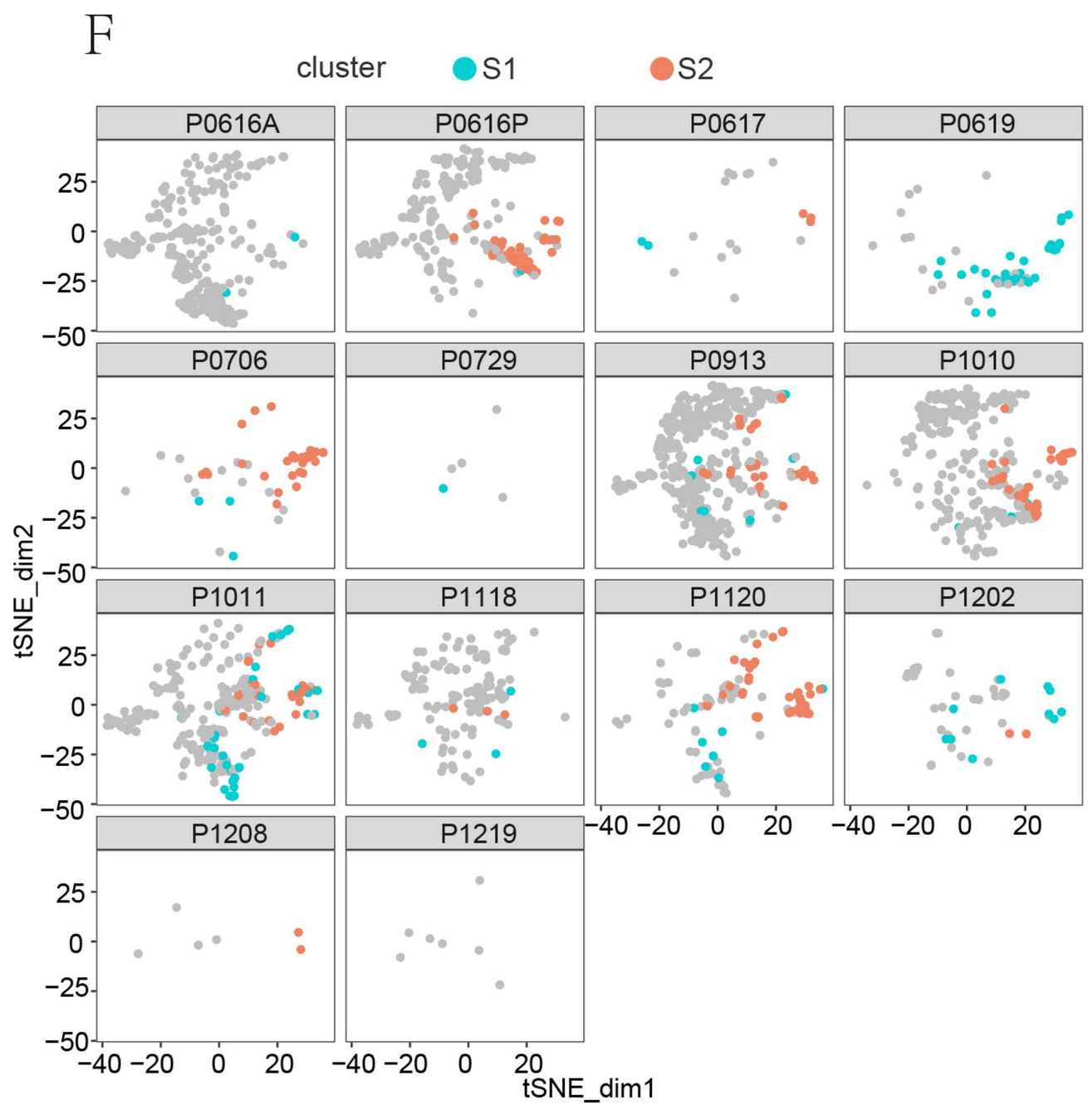
TCGA_COAD TCGA_LIHC TCGA_NSCLC


Tumor cell
Unknown ligand
MHC I
PD-L1
Galectin 9
CD155
(A)
persistent antigen stimulation
(B)
(C)
(D)
CD69
S1PR1
TCR
PDCD1
HAVCR2
TIGIT
Exhausted CD8+ T cell
Egress-promoting signaling
ZAP70
P
SHP1
SHP2
ITIM
ITSM
Tyr256
BAT3
Tyr263
Antagonized TCR signaling
PI3K
AKT
Reduced survival, proliferation
Nucleus
(E)
TOX
PDCD1, TIGIT, HAVCR2
(F)
JUN
PDCD1,TIGIT

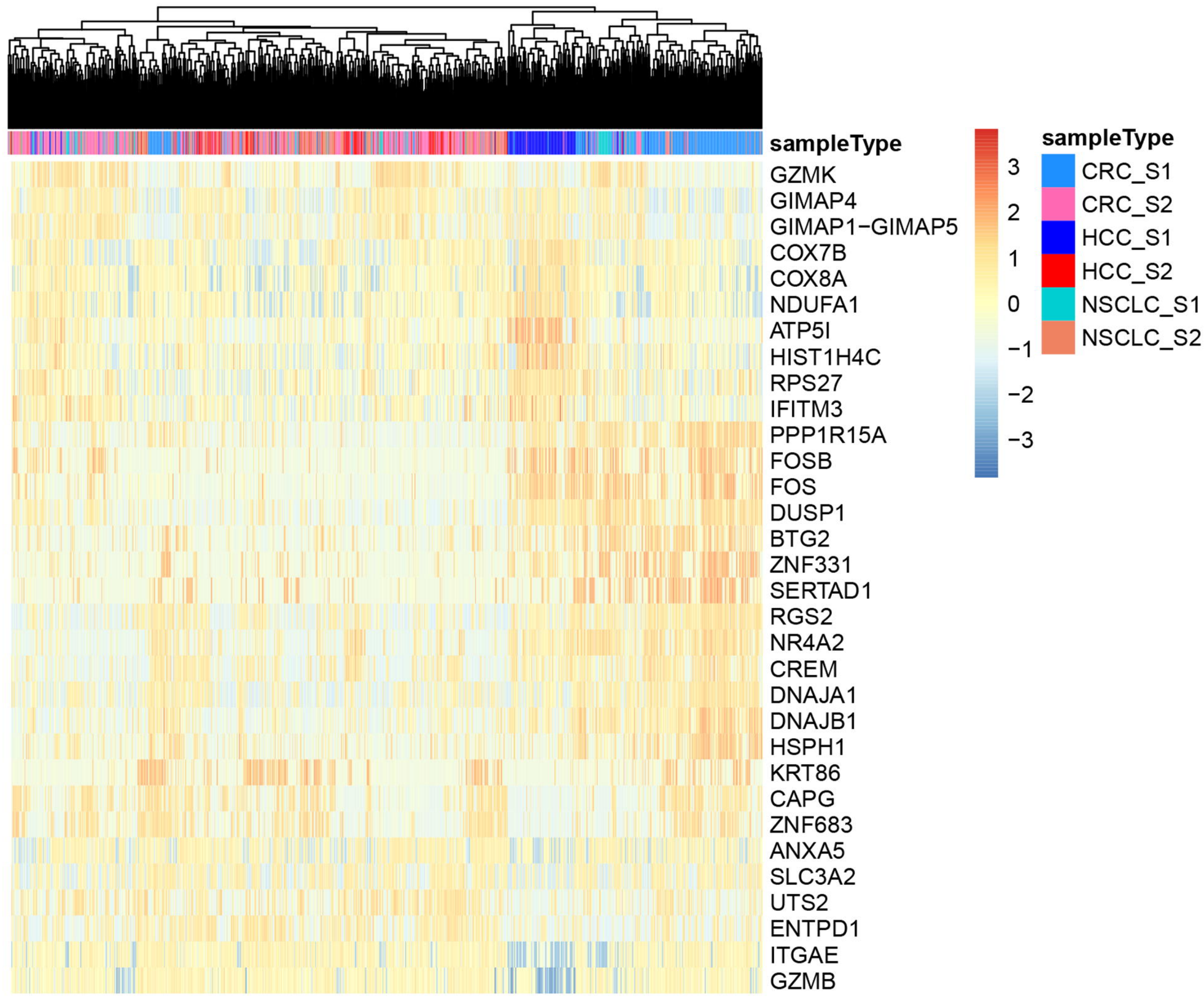

A
S1
S2
cluster
NR4A2
RGS2
DNAJB1
PPP1R15A
ZNF331
DNAJA1
SERTAD1
HSPH1
CREM
BTG2
GZMK
GIMAP1−GIMAP5
B
RPS27
FOS
DUSP1
COX8A
IFITM3
ATP5I
NDUFA1
FOSB
COX7B
HIST1H4C
ITGAE
ENTPD1
GZMB
ANXA5
UTS2
SLC3A2
C
GZMK
GIMAP4
CAPG
ZNF683
KRT86


D
cluster
S1
S2
P0123
P0215
P0309
P0411
P0413
P0701
P0825
P0909
P1012
P1207
P1212
P1228
tSNE_dim2
tSNE_dim1

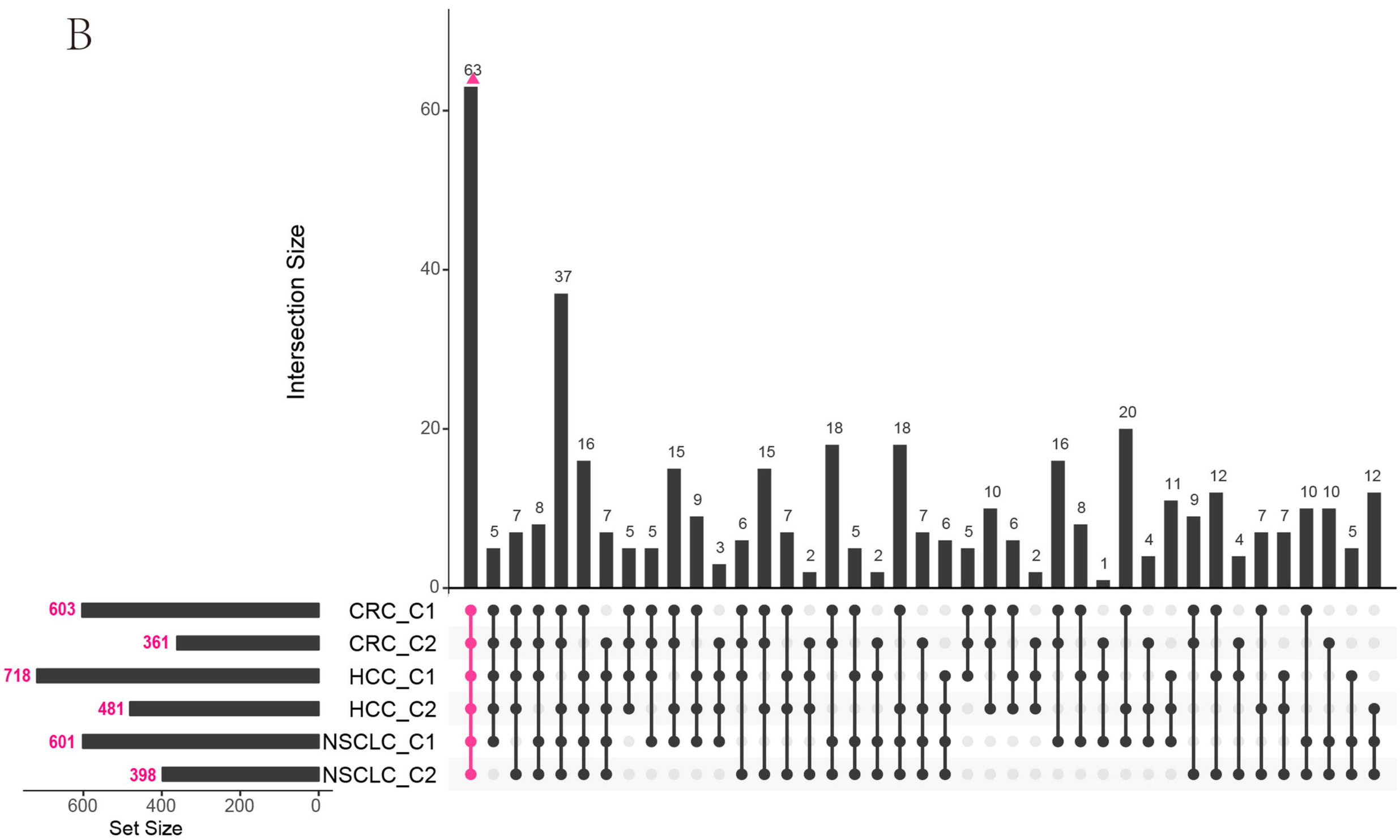

E
cluster
S1
S2
P0205
P0322
P0407
P0508
P1116
P1202t
tSNE_dim2
tSNE_dim1

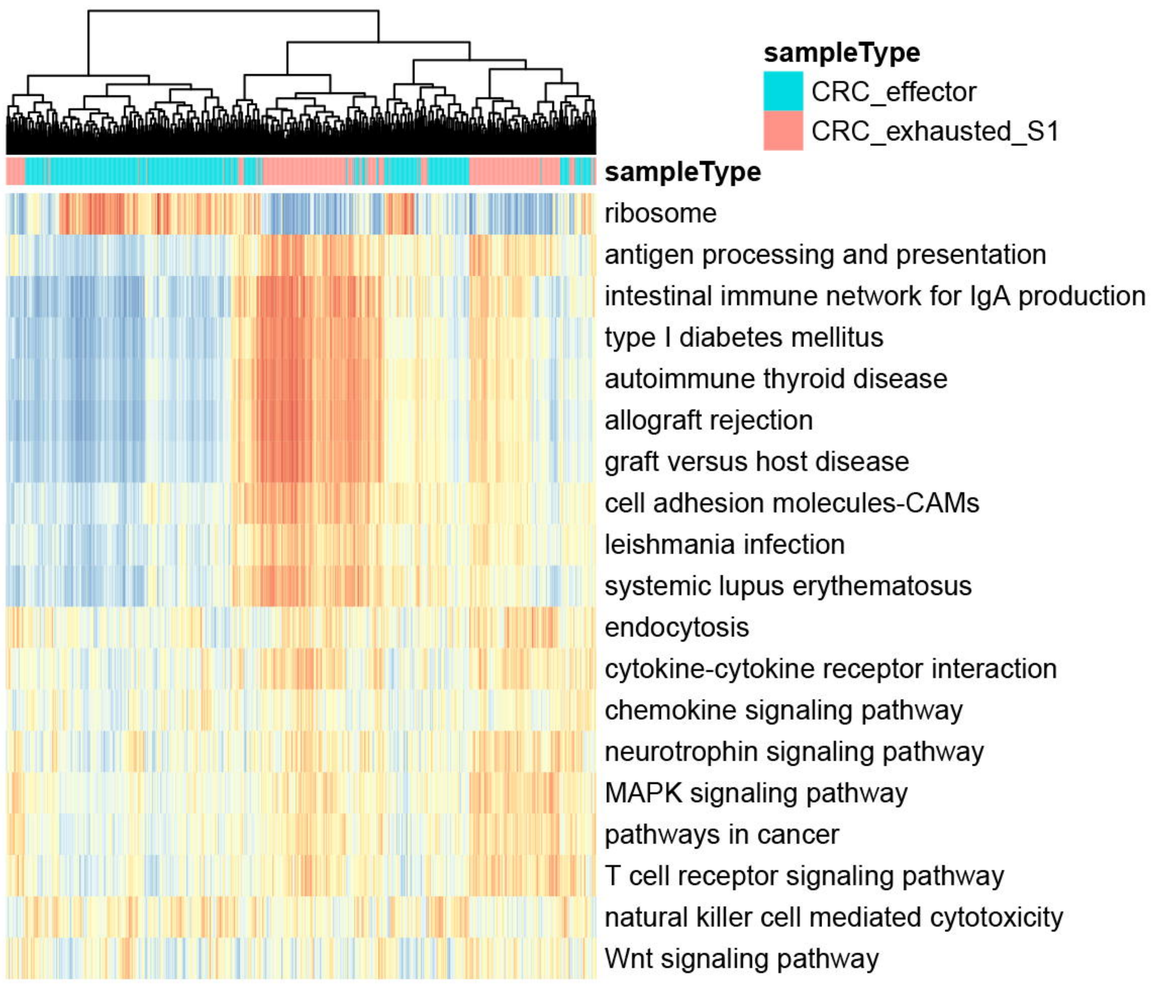

F
cluster
S1
S2
P0616A
P0616P
P0617
P0619
P0706
P0729
P0913
P1010
P1011
P1118
P1120
P1202
P1208
P1219
tSNE_dim2
tSNE_dim1

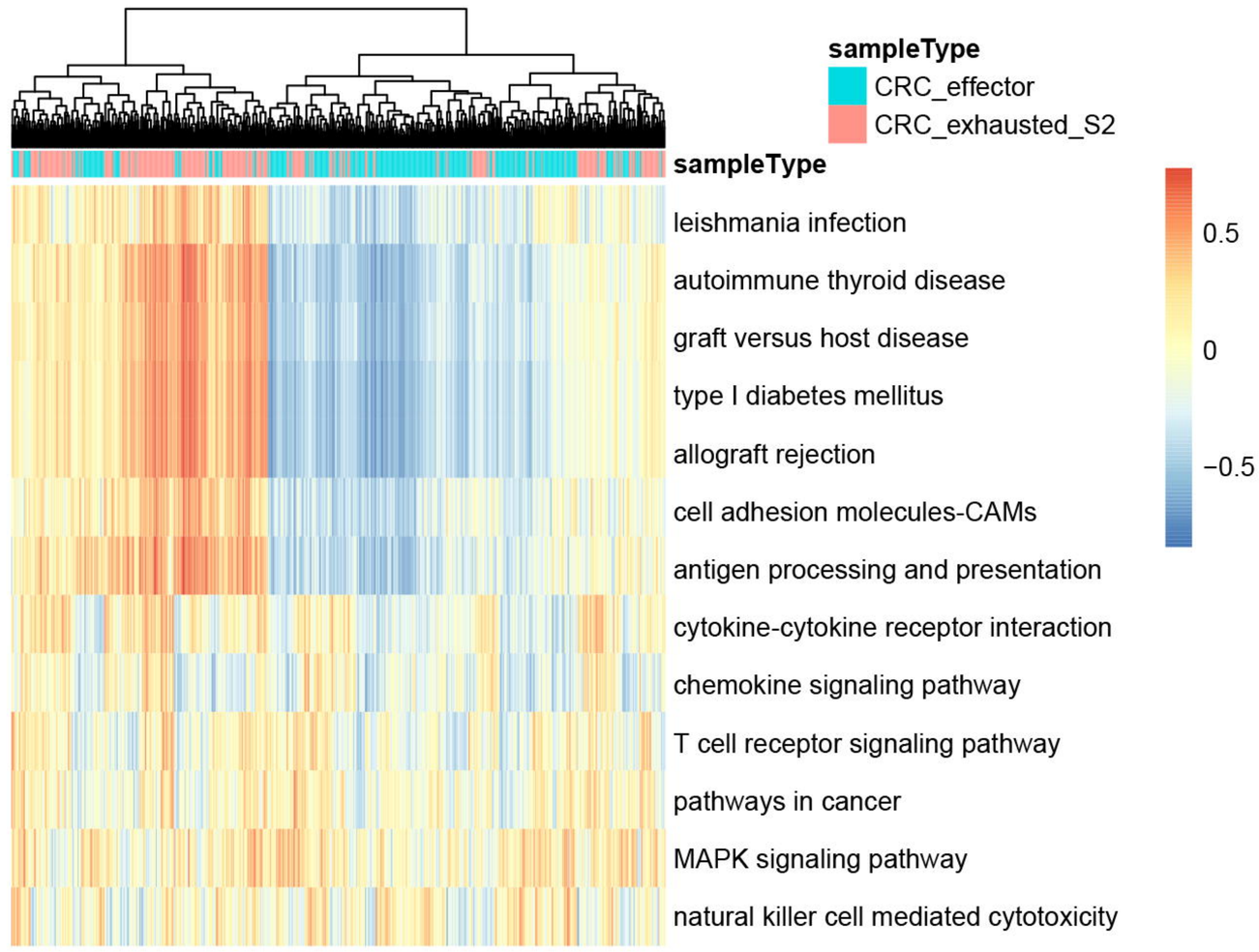
A

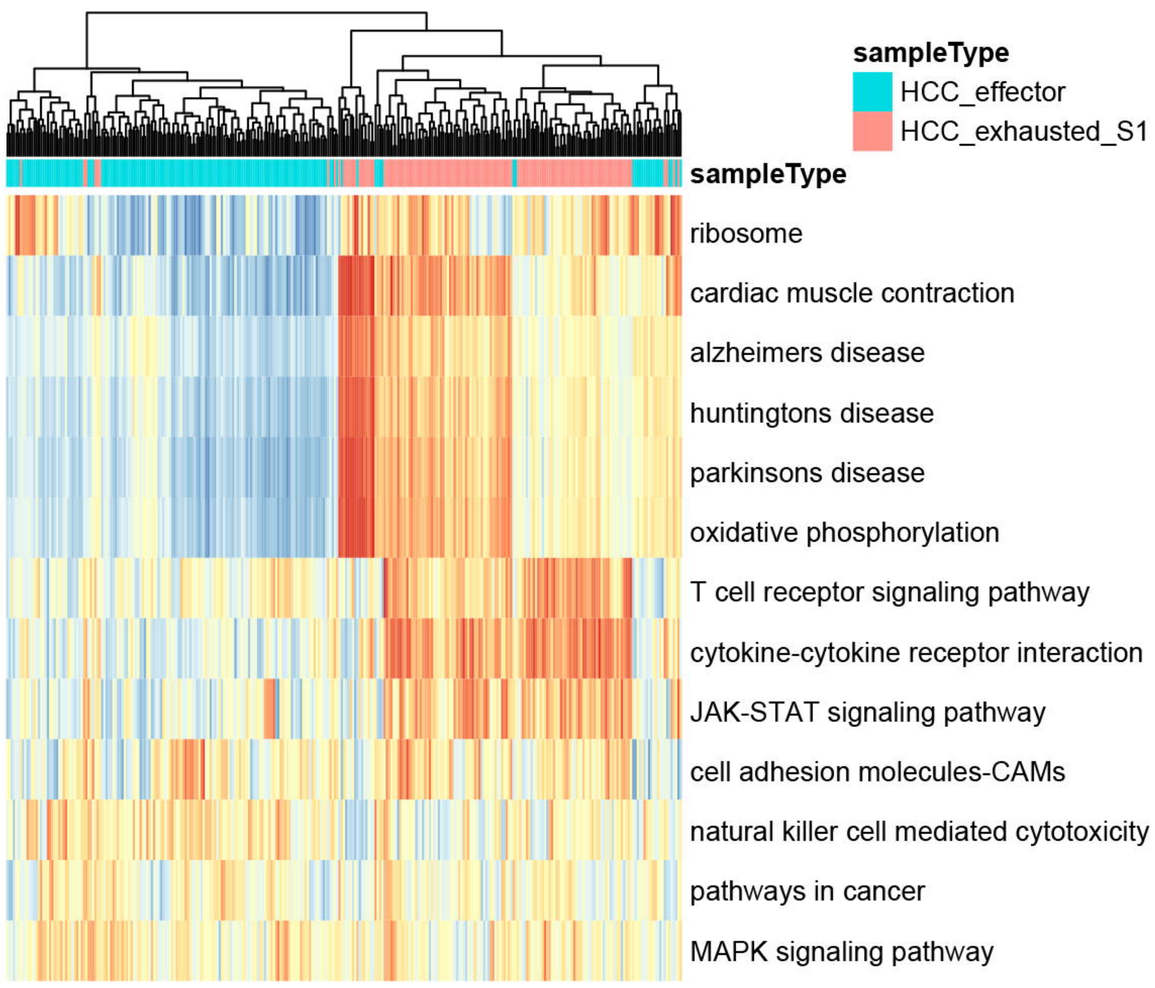

sampleType
GZMK
GIMAP4
GIMAP1−GIMAP5
COX7B
COX8A
NDUFA1
ATP5I
HIST1H4C
RPS27
IFITM3
PPP1R15A
FOSB
FOS
DUSP1
BTG2
ZNF331
SERTAD1
RGS2
NR4A2
CREM
DNAJA1
DNAJB1
HSPH1
KRT86
CAPG
ZNF683
ANXA5
SLC3A2
UTS2
ENTPD1
ITGAE
GZMB
3
2
1
0
−1
−2
−3
sampleType
CRC_S1
CRC_S2
HCC_S1
HCC_S2
NSCLC_S1
NSCLC_S2

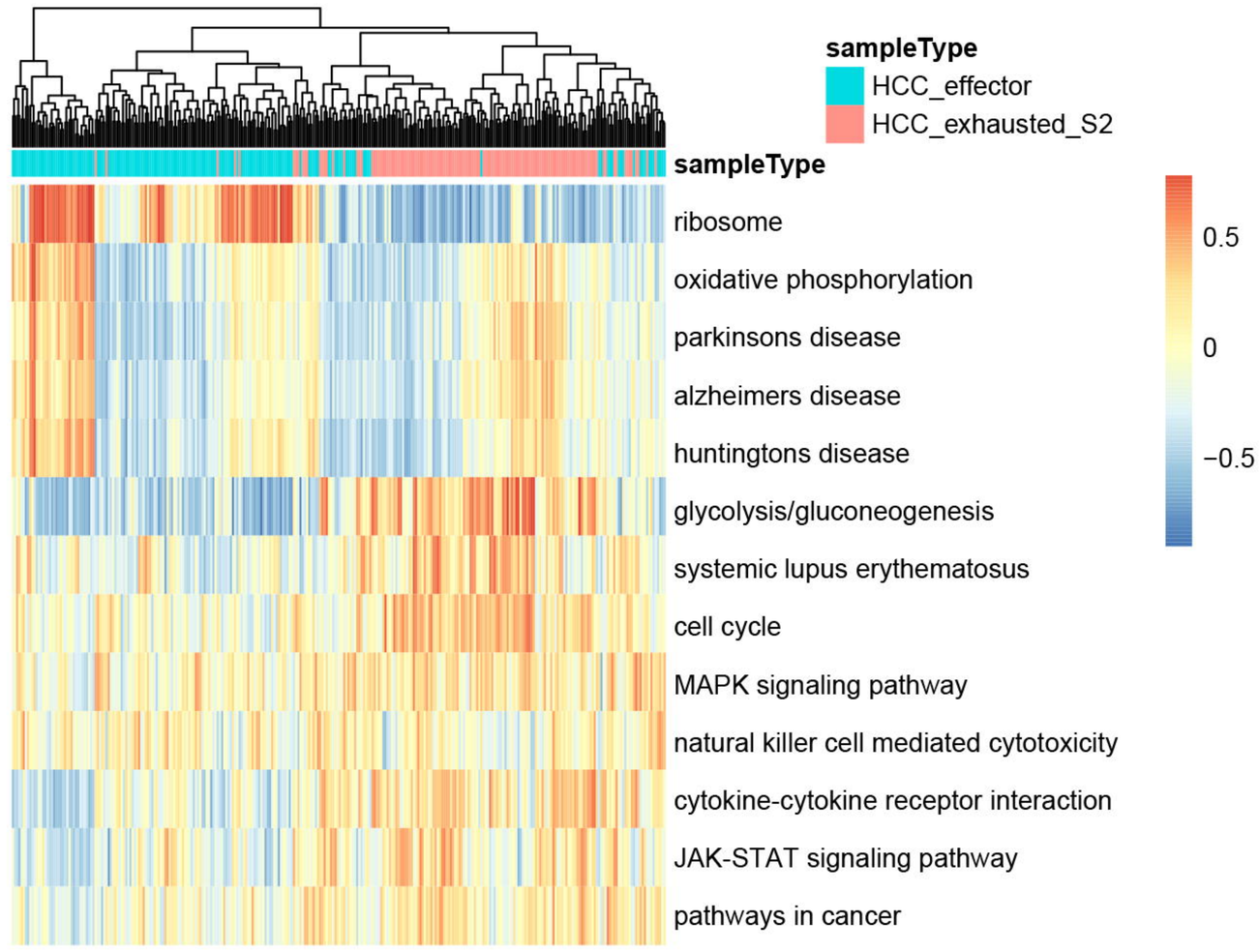
B

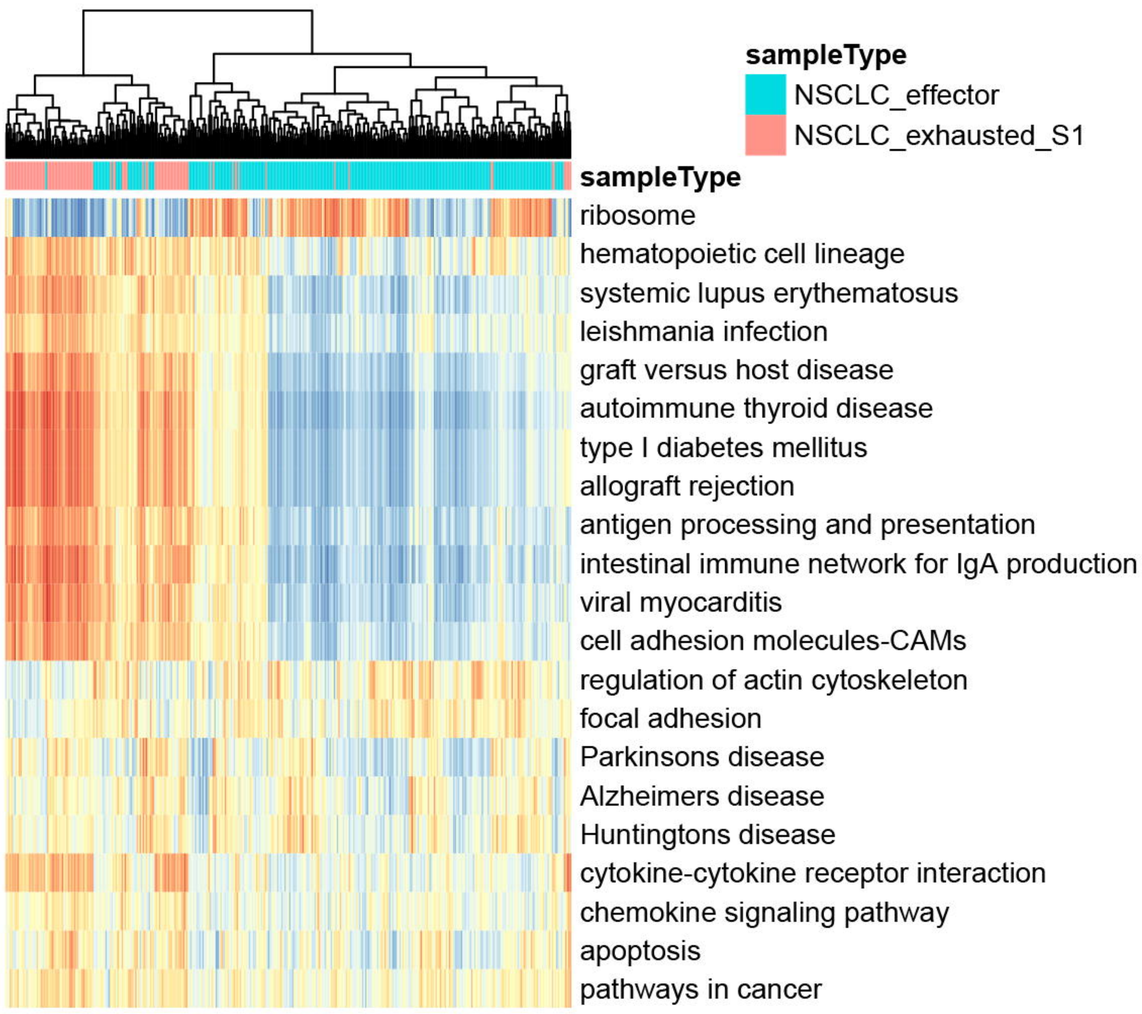

Intersection Size
0
20
40
60
63
5
7
8
37
16
7
5
5
15
9
3
6
15
7
2
18
5
2
18
7
6
5
10
6
2
16
8
1
20
4
11
9
12
4
7
7
10
10
5
12
603
361
718
481
601
398
CRC_C1
CRC_C2
HCC_C1
HCC_C2
NSCLC_C1
NSCLC_C2
600
400
200
0
Set Size

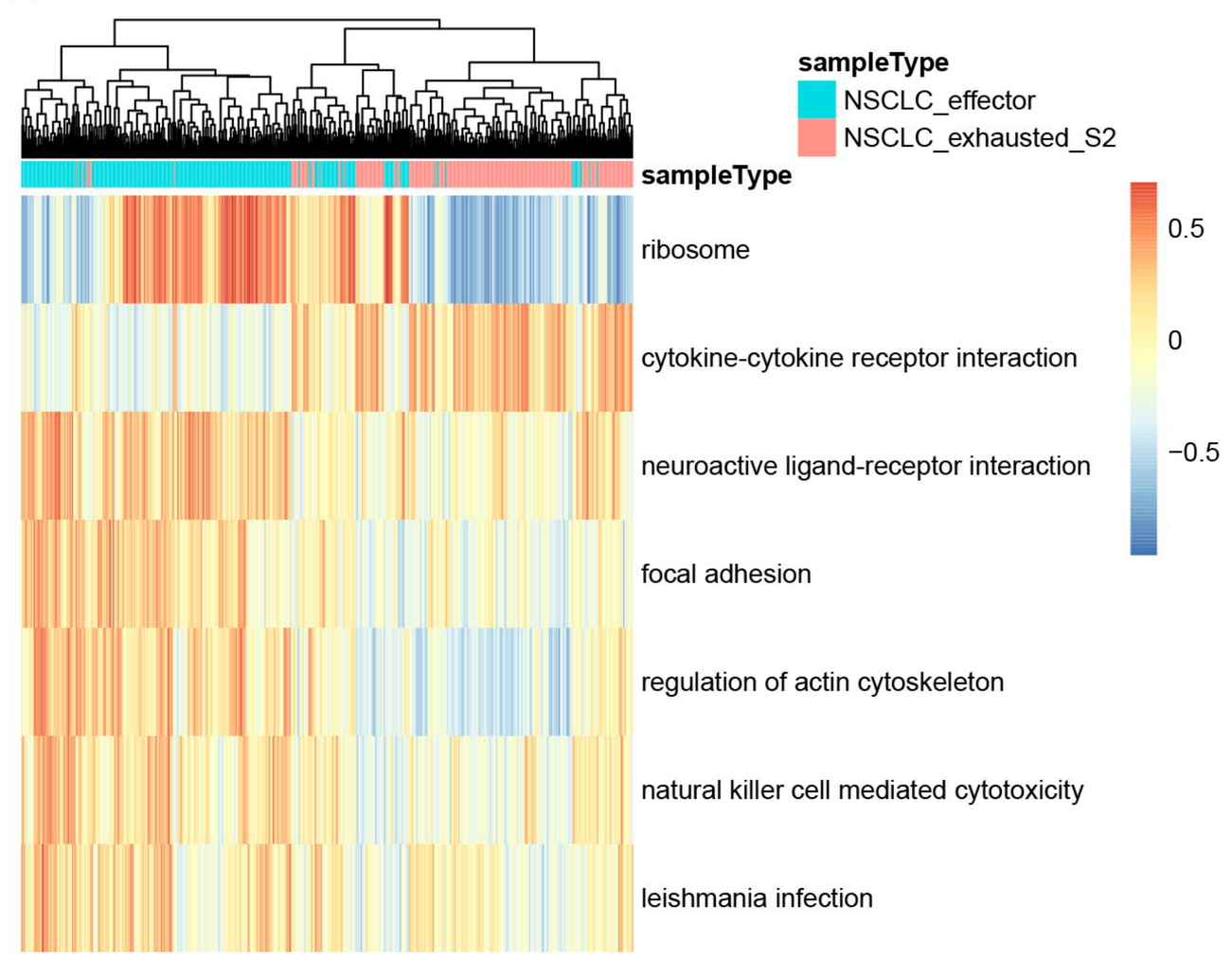

A
sampleType
CRC_effector
CRC_exhausted_S1
sampleType
ribosome
antigen processing and presentation
intestinal immune network for IgA production
type I diabetes mellitus
autoimmune thyroid disease
allograft rejection
graft versus host disease
cell adhesion molecules-CAMs
leishmania infection
systemic lupus erythematosus
endocytosis
cytokine-cytokine receptor interaction
chemokine signaling pathway
neurotrophin signaling pathway
MAPK signaling pathway
pathways in cancer
T cell receptor signaling pathway
natural killer cell mediated cytotoxicity
Wnt signaling pathway


B
sampleType
CRC_effector
CRC_exhausted_S2
sampleType
leishmania infection
autoimmune thyroid disease
graft versus host disease
type I diabetes mellitus
allograft rejection
cell adhesion molecules-CAMs
antigen processing and presentation
cytokine-cytokine receptor interaction
chemokine signaling pathway
T cell receptor signaling pathway
pathways in cancer
MAPK signaling pathway
natural killer cell mediated cytotoxicity
0.5
0
−0.5


C
sampleType
HCC_effector
HCC_exhausted_S1
sampleType
ribosome
cardiac muscle contraction
alzheimers disease
huntingtons disease
parkinsons disease
oxidative phosphorylation
T cell receptor signaling pathway
cytokine-cytokine receptor interaction
JAK-STAT signaling pathway
cell adhesion molecules-CAMs
natural killer cell mediated cytotoxicity
pathways in cancer
MAPK signaling pathway


D
sampleType
HCC_effector
HCC_exhausted_S2
sampleType
ribosome
oxidative phosphorylation
parkinsons disease
alzheimers disease
huntingtons disease
glycolysis/gluconeogenesis
systemic lupus erythematosus
cell cycle
MAPK signaling pathway
natural killer cell mediated cytotoxicity
cytokine-cytokine receptor interaction
JAK-STAT signaling pathway
pathways in cancer
0.5
0
−0.5


E
sampleType
NSCLC_effector
NSCLC_exhausted_S1
sampleType
ribosome
hematopoietic cell lineage
systemic lupus erythematosus
leishmania infection
graft versus host disease
autoimmune thyroid disease
type I diabetes mellitus
allograft rejection
antigen processing and presentation
intestinal immune network for IgA production
viral myocarditis
cell adhesion molecules-CAMs
regulation of actin cytoskeleton
focal adhesion
Parkinsons disease
Alzheimers disease
Huntingtons disease
cytokine-cytokine receptor interaction
chemokine signaling pathway
apoptosis
pathways in cancer


F
sampleType
NSCLC_effector
NSCLC_exhausted_S2
sampleType
ribosome
cytokine-cytokine receptor interaction
neuroactive ligand-receptor interaction
focal adhesion
regulation of actin cytoskeleton
natural killer cell mediated cytotoxicity
leishmania infection
0.5
0
−0.5