Single cell transcriptome analysis reveals RGS1 as a new marker and promoting factor for T cell exhaustion in multiple cancers

Yunmeng Bai[1,2†], Meiling Hu[1†], Zixi Chen[1], Jinfen Wei[1*], Hongli Du[1*]

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

[2]Translational Medicine Collaborative Innovation Center, Shenzhen People's Hospital, Shenzhen 518020, China

† These authors have contributed equally to this work

* Corresponding authors

Jinfen Wei,weijinfen@scut.edu.cn

Hongli Du, hldu@scut.edu.cn

Abstract: T cell exhaustion is one of the main reasons of tumor immune escape. Using single cell transcriptome data of CD8+ T cells in multiple cancers, we identified different cell types, in which Pre_exhaust and exhausted T cells participated in negative regulation of immune system process. By analyzing the co-expression network patterns and differentially expressed genes of Pre_exhaust, exhausted and effector T cells, we identified 35 genes related to T cell exhaustion, which high GSVA scores were associated with significantly poor prognosis in various cancers. In the differentially expressed genes, *RGS1* showed the greatest fold change in Pre_exhaust and exhausted cells of three cancers compared with effector T cells, and high expression of *RGS1* was also associated with poor prognosis in various cancers. Additionally, RGS1 protein was upregulated significantly in tumor tissues in the immunohistochemistry verification.

Furthermore, *RGS1* displayed positive correlation with the 35 genes, especially highly correlated with *PDCD1*, *CTLA4*, *HAVCR2* and *TNFRSF9* in CD8+ T cells and cancer tissues, indicating important roles of *RGS1* in CD8+ T cell exhaustion. Considering the GTP-hydrolysis activity of *RGS1* and significantly high mRNA and protein expression in cancer tissues, we speculated that *RGS1* potentially mediate the T cell retention to lead to the persistent antigen stimulation, resulting in T cell exhaustion. In conclusion, our findings suggest that RGS1 is a new marker and promoting factor for CD8+ T cell exhaustion, and provide theoretical basis for research and immunotherapy of exhausted cells.

Keywords: single cell transcriptome; T cell exhaustion; multiple cancers; RGS1; poor prognosis

1. Introduction

T cell exhaustion (Tex), a hyporesponsive state of T cells with increased inhibitory receptors, decreased effector cytokines and impaired cytotoxicity, 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 cancers(2, 3), which is one of the main reasons of tumor immune escape(4). It has been reported that exhausted T cells in cancers share many similarities with that in chronic infection(5), and play a significant role in tumorigenesis(6). Studies show that exhausted T cells 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(7). Nevertheless, due to the complexity and heterogeneity of cancers, the concrete mechanisms and molecules of T cell exhaustion in cancers have not been fully elucidated.

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(8-10). 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

heterogeneity(11). 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(12-15). 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.

Due to the vital role of CD8+ T cells in eliciting antitumor responses(16), we integrated single cell transcriptome data from colorectal cancer (CRC), hepatocellular cancer (HCC), and non-small cell lung cancer (NSCLC) to analyze CD8+ T cells in various cancers in the present study. Focusing on CD8+ T cell exhaustion associated clusters, we identified *RGS1* as a new marker and promoting factor for T cell exhaustion in multiple cancers with poor prognosis, and showed highly positive correlation with the well-known genes associated with T cell exhaustion. Besides, RGS1 protein highly expressed in tumor tissues was also verified in the immunohistochemistry (IHC) experiment. Our findings could facilitate to understand the mechanism during formation and development of T cell exhaustion, and provide theoretical basis for research and immunotherapy of exhausted cells.

2. Materials and Methods

2.1 Data acquisition

The single-cell gene expression matrices including raw count and TPM data were obtained from the GEO database: GSE108989 (CRC), GSE99254(HCC), GSE98638(NSCLC), and we isolated CD8+ T cells from peripheral blood (P), adjacent normal (N) and tumor tissues (T).

2.2 Quality control and data processing

The raw count expression matrices were processed by R package Seurat v3.2.0 (http://satijalab.org/seurat/). To filtered out the low-quality cells, we excluded cells with < 600 and > 10000 detected genes(17). Counts were log-normalized, and scaled by linear regression against the number of reads with function NormalizeData and ScaleData. The highly variable genes (HVGs) were generated with

FindVariableFeatures. Principal component analysis (PCA) was performed on the top 2000 HVGs using functions RunPCA. The appropriate PCs were selected for graph-based clustering with function FindNeighbors and FindClusters. For visualization of clustering analysis, we performed uniform manifold approximation and projection for dimension reduction (UMAP) using RunUMAP function in Seurat. To eliminate the obvious effect from different patients, we performed standard normalization and variable feature selection after acquiring the data. Next, the function FindIntegrationAnchors was performed to find a set of anchors, which used to integrate the data by the function IntegrateData. Besides, the function AddModuleScore was used to calculate scores of gene list in different cell types.

2.3 Cell type annotation

Differentially expressed genes (DEGs) of each cluster were identified based on Wilcoxon rank-sum test using function FindAllMarkers compared with the rest clusters. In brief, for each cluster, only genes that met these criteria were considered as cluster-specific DEGs: 1) log2FC > 0.25; 2) expressed > 25% in either of the two groups of cells; 3) adjusted p value < 0.05. The top DEGs were selected to annotated of each cluster based on the canonical markers from previous studies, besides, the CellMarker databases(18) and R package SingleR v1.4.0(19) were performed to further improve the accuracy of the annotation.

2.4 Trajectory Analysis

To explore the potential functional changes of CD8+ T cell of different clusters for each cancer, we performed development trajectory analysis by R package Monocle v2.18.0(20) with the cluster-specific genes of each cluster. Dimensional reduction and cell ordering were performed using reduceDimension and orderCells function with default parameter.

2.5 Gene ontology (GO) enrichment analysis of DEGs

Biological significance was explored by GO term enrichment analysis by R package clusterProfiler v3.18.0(21), biological process, cellular component, and molecular function included. Adjusted p value < 0.05 was considered statistically

significant. Visualization is realized by R package ggplot2 v3.3.3 and ggalluvial v0.12.3(22).

2.6 Weighted gene co-expression network analysis

In order to identify the highly linked genes in specific clusters, weighted gene co-expression network analysis (WGCNA) was performed with functions in the R package WGCNA(23). To attenuate the effects of noise and outliers, we constructed pseudocells(24) which calculated as averages of 10 cells randomly chosen within each cluster. The function pickSoftThreshold was used to calculate the soft power parameter and blockwiseModules to construct co-expression network. Finally, corPvalueStudent was used to mine modules related to specific cell types.

2.7 Survival analysis

According to the median of gene expression values, cells were divided into high and low groups, then the survival curve was shown using the Kaplan–Meier curve with a log-rank test by GEPIA2 (http://gepia2.cancer-pku.cn/) to illustrate the relationship between differential genes and overall patient survival. The terms with P-value < 0.05 were identified as significant. Additionally, multiple hypothesis testing (FDR) was performed to the significant p value using the function p.adjust.

2.8 Gene set variation analysis

For generated gene set, we performed gene set variation analysis (GSVA) to the data in the present study and validation dataset using R package GSVA(25) with the default parameter, to evaluate the effect distinguishing Tex cells.

2.9 The basic expression of mRNA and protein in normal and cancer tissues

The RNA-seq data were downloaded from TCGA database (https://portal.gdc.cancer.gov/) and then calculated into TPM value. According to the annotation, samples were grouped into normal and four stages of cancer groups. The differential expression of mRNA in the different groups was performed by wilcoxon test with p value<0.05.

The protein expression level in normal and cancer tissues was analyzed using Human Protein Atlas (HPA) database (https://www.proteinatlas.org).

Immunohistochemistry pictures were downloaded from the Tissue Atlas and Pathology Atlas.

2.10 Immunohistochemistry

Immunohistochemistry was carried out on human liver cancer tissue microarray (Shanghai Outdo Biotech Co., Ltd, China). The tissue sections were first dried at 63□ for 1 h, dewaxed and rehydrated before epitope-retrieval by heating at 100 C in 10 mM sodium-citrate (pH 6.0) for 5 min in EDTA solution for 20 min. The sections were cooled down to room temperature for 30 min. The tissue sections were treated with 3% hydrogen peroxide for 20 min to eliminate the endogenous peroxidase and alkaline phosphatase activity in the tissue. After cooling down to room temperature , the sections were treated by blocking agents for 10 min. The sections were incubated with of the individual primary antibody (RGS1, Invitrogen, USA; Product #PA5-86730,) diluted 1:1000 800 overnight at 4□, followed by secondary antibodies at room temperature for 30 min. DAB(3,3'-diaminobenzidine) was then applied as a substrate to reveal the antigen. Hematoxylin was used for counterstaining. The stained images were counted by Image J software(26), and the optical density (OD) value was used for quantification.

3. Results

3.1 Single CD8+ T cell transcriptome landscape

We obtained single cell transcriptome data of human T cells from GEO database, including 12 patients from CRC(27), 6 patients from HCC(28) and 14 patients from NSCLC(29). After strict quality control and filtration, we collected 4010, 1752 and 4439 CD8+ T cells from peripheral blood (P), adjacent normal (N) and tumor tissues (T) (Supplemental Figure 1A, 1B, Supplemental Table 1).

We divided the cells of each cancer into different cell types annotated with cluster-specific genes expression (Figure 1A, Supplemental Table 2). Specifically, in CRC, cells were identified as naïve T cells (Tn, cluster 4, marked with selectin L (*SELL*), lymphoid enhancer binding factor 1(*LEF1*), C-C motif chemokine receptor 7(*CCR7*)), transcription Factor 7 (*TCF7*), effector T cells (Teff, cluster 2 and cluster 5, marked with C-X3-C motif chemokine receptor 1(*CX3CR1*), killer cell lectin like

receptor F1(*KLRF1*)), fibroblast growth factor binding protein 2 (*FGFBP2*), Fc fragment of IgG receptor IIIa (*FCGR3A*), exhausted T cells (Tex, cluster 3 and cluster 6, marked with hepatitis A virus cellular receptor 2(*HAVCR2*), programmed cell death 1(*PDCD1*), lymphocyte activating 3 (*LAG3*), *TOX*, *CXCL13*, tissue-resident memory T cells (Trm, cluster 1, marked with *CD69*, integrin subunit alpha E (*ITGAE*)), mucosal-associated invariant T cells (MAIT, cluster 7, marked with solute carrier family 4 member 10(*SLC4A10*)), RAR related orphan receptor C (*RORC*) (Figure 1B, 1C). Cluster 0 was located between cluster 2 and cluster 3, which represented as Teff and Tex cells, besides, the genes marked Teff and Tex cells showed relative high expression levels in this cluster, therefore, we identified it as Pre_exhaust T cells. Similarly, cells in HCC were identified as Tn (cluster 4), Teff (cluster 0), Pre_exhaust (cluster 1), Tex (cluster 3 and cluster5), MAIT (cluster 2), cells in NSCLC were identified as Tn (cluster 4), Teff (cluster 0 and cluster 6), Pre_exhaust (cluster 1 and cluster 7), Tex (cluster 3), Trm (cluster 2), MAIT (cluster 5). Different cell types showed preference in different tissues. In general, Tn and Teff cells were mainly enriched in peripheral blood, Pre_exhaust and Tex cells were mainly enriched in tumor tissues, Trm cells existed more in adjacent normal tissues of CRC and HCC, and in tumor tissues of NSCLC, which may be related to tissue specificity. Besides, the number of MAIT cells was relative less than the others. Additionally, we performed exhausted cell scoring by the function AddModuleScore in R package Seurat using the exhaustion gene list including *HAVCR2*, T cell immunoreceptor with Ig and ITIM domains (*TIGIT*), *LAG3*, *PDCD1*, *CXCL13*, layilin (*LAYN*), *TOX*, cytotoxic T-lymphocyte associated protein 4 (*CTLA4*), B and T lymphocyte associated (*BTLA*) and visualized in Supplemental Figure 1C, showing the accuracy of Tex cell classification (Supplemental Figure 1C).

To further explore the function of each cluster, we performed GO enrichment analysis using the cluster-specific genes (Figure 1D). In all three cancers, Tn cells were enriched in T cell activation, T cell differentiation, ribosome biogenesis; Teff cells were enriched in cellular defense response, positive regulation of cytokine production, T cell mediated cytotoxicity; Tex cells were enriched in negative regulation of immune system process, T cell apoptotic process, response to hypoxia; Trm cells were enriched in antigen receptor-mediated signaling pathway, leukocyte chemotaxis; MAIT cells were enriched in interferon-gamma production; and

Pre_exhaust cells were enriched in terms associated with Teff and Tex cells, including cellular defense response and negative regulation of immune system process, showing the transitional characteristics during T cell exhaustion.

3.2 Establishment of co-expression network

Tex cells, as shown above, played a negative role in immune system process, and it was suggested that Pre_exhaust cells was a transitional stage from Teff to Tex cells. To find out the highly linked genes associated with T cell exhaustion, we used the R package WGCNA to construct the weighted co-expression network in Teff, Pre_exhaust and Tex cells (Figure 2A). The blue module (R-value 0.96, P-value 2e-145), turquoise module (R-value 0.96, P-value 7e-67) and blue module (R-value 0.96, P-value 2e-172) represented Tex cells in CRC, HCC and NSCLC, receptively. Combining with the differentially expressed genes (DEGs) up-regulated in the Tex vs. Teff cells comparison in three cancers (Supplemental Table 3), there were 35 genes were found in three Tex cells-related modules and overexpressed in Tex cells (Figure 2B, Table 1), which were defined as "Candidate" gene set, including exhaustion marker *PDCD1*, *CTLA4*, *HAVCR2*, *TOX*, *TIGIT*. It is noteworthy that the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), which can catalyze an important energy-yielding step in glycolysis metabolism, was also upregulated and enriched in the highly linked DEGs, thus we speculated that the glycolysis progress was disordered in Tex cells.

As shown in Figure 2C, poor overall survival was correlated with higher Candidate gene set expression in multiple cancers including liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), skin cutaneous melanoma (SKCM), thymoma (THYM). Besides, the GSVA scores of Candidate gene set in Tex cells were much higher than those of the other clusters in CRC, HCC and NSCLC, so were in the other independent validation datasets (Figure 2D). GSE116390 dataset(30) was composed of four types of tumor-infiltrating CD8+ T cells, which were exhausted, memory-like, naïve, and effector memory-like (EM-like) subsets, from B16 melanoma

tumor-bearing mice, and GSE123813 dataset(31) contained six types of tumor-infiltrating CD8+ T cells, which were from 11 patients with advanced basal cell carcinoma, showing higher score in the exhausted related clusters.

As indicated above, the Candidate gene set was enriched in exhausted CD8+ T cells with poor prognosis and was able to distinguishing Tex cells from the other CD8+ T cells in different cancers, indicating the GSVA score of these 35 genes might be an effective prognostic marker or a marker to identify Tex cells.

3.3 Differentially expressed genes associated with T cell exhaustion

The exhaustion of T cells was gradually formed, so we performed differential expression analysis of Pre_exhaust and Tex cells compared with Teff cells, in order to further explore the critical genes associated with the formation and development of T cell exhaustion. In the Pre_exhaust vs. Teff cells comparison, there were 119 DEGs existed in all three cancers, 57 of whom were up-regulated and the others were down-regulated. Correspondingly, there were 162 up-regulated and 88 down-regulated DEGs in the Tex vs. Teff cells comparison. Furthermore, there were 40 DEGs overexpressed in both Pre_exhaust and Tex cells compared with Teff cells (Figure 3A), which may contribute to the origin of T cell exhaustion, including the canonical exhaustion marker *PDCD1* which encodes an inhibitory receptor(32), *CD69* with the capability to mediate the cell retention(33), Cbl Proto-Oncogene B(*CBLB*) whose deletion can inhibit CD8+ T cell exhaustion(34), hypoxia inducible factor-1 (*HIF1A*) which can stably expressed in hypoxia condition, consistent with the biological process of Tex cells (Figure 1D).

Obviously, regulator of G protein signaling 1(*RGS1*) showed the greatest fold change in Tex cells across three cancers and also across different patients, showed the greatest fold change in Pre_exhaust cell in HCC and NSCLC (Figure 3A, Supplemental Figure 2), illustrating its potential roles during T cell exhaustion. Besides, *RGS1* was expressed highly in whole Tex cells compared with Teff cells (Figure 3B), eliminating that the high fold change of *RGS1* in Pre_exhaust and Tex cells was not caused by partial cells with abnormally high expression value but the high expression in whole cells. In order to evaluate the role of *RGS1* in tumorigenesis, we analyzed the expression levels of *RGS1* between tumor and normal tissues in TCGA database

(Figure 3C). *RGS1* expression was significantly upregulated in multiple cancers including BRCA, cholangiocarcinoma (CHOL), esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), HNSC, kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), LUAD, rectal carcinoma (READ) and stomach adenocarcinoma (STAD) when compared with the normal samples. Besides, we added and compared the expression levels of *RGS1* across different cancer stages, and found *RGS1* expression was associated with stage in some cancer types, such as, *RGS1* expression was higher in Stage II, III and IV vs Stage I in STAD. This result revealed that *RGS1* was likely a key tumorigenesis regulator in multiple cancers and may associated with prognosis. Conspicuously, the prognosis analysis was analyzed in 33 TCGA cancer types (Supplemental Table 4). RGS1 expression was significantly correlated to poor prognosis in seven cancers, including LIHC, adrenocortical carcinoma (ACC), pancreatic adenocarcinoma (PAAD), sarcoma (SARC), SKCM, STAD and THYM (Figure 3D), suggesting that *RGS1* was a potential prognostic factor in survival of above cancers. Apart from that, the protein level of RGS1 in HPA database showed the immunohistochemical (IHC) staining of RGS1 was negative staining in normal tissues and positive in liver cancer tissues, demonstrating that RGS1 was significantly expressed in cancer tissues than in normal liver tissues (Figure 3E). Additionally, we performed IHC verification to quantify the protein expression of RGS1 using the local clinical samples of liver cancer and normal tissues (Figure 3F, 3G), similarly, RGS1 protein displayed stronger staining in hepatocarcinoma, line with the statistical result (p=7.1e-5). The clinicopathological information of the patient samples and protein expression value were provided in Supplemental Table 5. These results showed that *RGS1* was highly expressed in Tex cells in cancers, upregulated in tumor tissues in mRNA and protein level, and with poor prognosis in multiple cancers, which indicated its potential key role in T cells exhaustion or cancer progress, and RGS1 might be an effective prognostic marker or a marker to identify Tex cells.

3.4 Correlation between RGS1 and Candidate gene set

To discover the relationship between *RGS1* and Candidate gene set as their potential roles in exhausted T cells, we calculated the correlation coefficient in single cell and tissue level (Figure 4). In three cancer single cell datasets, it was obvious that *RGS1* showed positive correlation with 35 genes (up to 0.3~0.8), indicating that

consistency of coexpression patterns between these genes. In the bulk RNA sequencing datasets, the correlation coefficients were almost positive, especially high with *PDCD1*, *CTLA4*, TNF receptor superfamily member 9 (*TNFRSF9*), *HAVCR2*, *TOX* and *TIGIT*, which further confirmed the potential key roles of *RGS1* in Tex cells. The negative correlations were mainly with the genes involved in cell cycle and DNA replication, including minichromosome maintenance complex component 7 (*MCM7*), enolase superfamily member 1 (*ENOSF1*), minichromosome maintenance complex component 5 (*MCM5*), minichromosome maintenance complex component 3 (*MCM3*), proliferating cell nuclear antigen (*PCNA*), which pointed out the different results observed between single cell and bulk RNA sequencing analysis, suggesting we could obtain deeper understanding about T cell exhaustion by single cell sequencing. One interpretation of above inconsistent result is that the bulk transcriptome change (which mixes many immune and non-immune cells together) likely reflects an overall expression in these cells and does not discriminate specific T cell states, further highlights the advantages in defining and studying T cell exhaustion state by using single cell sequencing data.

4. Discussion

T cell exhaustion is characterized by loss of effector functions, continuously high expression of numerous inhibitory receptors, epigenetic and transcription profile changes, and dysregulated metabolism. Exhaustion CD8+ T cells are associated with suppressive immune microenvironment and poor overall survival in various cancer types, such as, in invasive bladder cancer(35) and clear-cell renal cell carcinoma(36). Besides, increased exhausted CD8+ T cell subpopulations predict PD-1 blockade resistance response in melanoma(37). Accumulating evidences support exhausted T cells are possible to be rescued in cancer immunotherapy. Anti-PD1 antibodies, including atezolizumab and nivolumab, can renew the activity of exhausted CD8 T cells through preventing PD-1-mediated attenuation of proximal TCR cascades(38, 39) and can effect metabolic reprogramming to reinvigorate T cells(40). However, one study found an association between increased accumulation of one CD8+ T cell exhaustion phenotype and clinical benefit, suggesting exhausted T cells may comprise heterogenous cell population with distinct responsiveness to intervention and the standard definition of exhaustion cells is unclear in the context of treatment(41). Thus,

understanding molecular mechanism of T cell exhaustion and comprehensively exploring potential markers associated with T cell exhaustion is essential to precisely define T cell exhaustion and establish rational immunotherapeutic interventions.

In this study, combining with single-cell RNA sequencing, which can facilitate to detect he transcriptome on the level of single cell(42), we can shed light on the complication of tumor-infiltrating T cells. In order to explore the key genes associated with T cell exhaustion in multiple cancers, we performed transcriptomic analysis of single CD8+ T cells isolated from three cancers, including CRC, HCC and NSCLC, and identified different cell types, thereinto, Pre_exhaust and Tex cells overexpressed exhaustion markers and enriched in the negative regulation of immune progress. In the comparison with Teff cells, *RGS1* showed almost the greatest fold change in Pre_exhaust and Tex cells of three cancers with poor prognosis, and displayed highly positive correlation with the well-known genes associated with T cell exhaustion.

In the WGCNA analysis, we identified a Candidate gene set consisting of 35 DEGs, including exhaustion markers such as *PDCD1*, *CTLA4*, *HAVCR2*, *TOX*, *TIGIT*. Apart from that, genes involved in cell cycle and DNA replication also included, such as *MCM7*, *MCM5*, *MCM3*, *PCNA*, stathmin 1 (*STMN1*), titin (*TTN*), TBC1 domain family member 4 (*TBC1D4*), suggesting that exhausted cells still retained the ability of proliferation, which was also observed in chronically infected models(43). Functionally, the Candidate gene set was able to distinguish Tex cells from the other subtypes of CD8+ T cells in different cancers and higher GSVA scores of Candidate gene set showed poor prognosis in multiple cancers.

Besides, several DEGs appeared in Pre_exhaust and Tex cells, suggesting the role in the formation and development of T cell exhaustion, in which *RGS1* showed almost the greatest fold change. *RGS1* encodes a member of the regulator of G-protein signaling family, can act as a GTPase activating protein (GAP), increasing the rate of conversion of the GTP to GDP, driving G-protein into its inactive GDP-bound form, hence attenuating or turning off G-protein-coupled receptors signaling(44). *RGS1* is highly expressed in immune cells including T cells(45), B cells(46), natural killer (NK) cells(47), dendritic cells(48) and monocytes(49), suggesting a role for RGS1 in immune cell regulation. RGS1 inhibits the chemokine-induced lymphocyte migration(50) because chemokine-dependent activation of G-protein-coupled

receptors can cause the activation of heterotrimeric G-protein subunits resulting in enhanced cell migration and adhesion(51), which has been found in Treg cells(45). In the present study, *RGS1* was highly expressed in tumor tissues and correlated with shorter overall survival, which also appeared in several previous studies, including multiple myeloma(52), melanoma(53, 54), non-small cell lung cancer(55), gastric cancer(56), diffuse large B cell lymphoma(57) and so on. However, the role of *RGS1* in CD8+T cells especially in Tex cells has not been reported. In addition, RGS1 protein, locating at cytoplasm and membrane, enriched in tumor tissues compared with normal tissues according to the IHC staining in HPA database and verification experiment, further verifying its pathogenicity. Considering the ability to block cell migration of *RGS1*, we speculate that *RGS1* can mediate the cell retention to lead to the persistent antigen stimulation of T cells, which resulted in T cell exhaustion with the overexpression of inhibitory genes such as *PDCD1* and *HAVCR2*(58). Additionally, *RGS1* was identified as a HIF-dependent hypoxia target that dampens cell migration and signal transduction(59), indicating its role in exhausted T cells might be caused by hypoxia condition (Figure 1B).

RGS1, the most up-regulated gene in Pre_exhaust and Tex cells and a potential marker for T cell exhaustion, was excluded from the Candidate gene set. This phenomenon happened in other T cell exhaustion-related genes as well, such as *CD69* and *CBLB*. *CD69*, upregulated in Pre_exhaust and Tex cells, is an early activation marker of T cells(60). It can mediate the cell retention via the interaction with sphingosine-1-phosphate receptor 1 (*S1PR1*) which acts as a central mediator of lymphocyte output(61), leading to the persistent antigen stimulation of T cells, which resulted in T cell exhaustion by overexpression of *PDCD1* and *HAVCR2* (62). *CBLB*, also upregulated in Pre_exhausted and Tex cells, deletion can inhibit CD8+ T cell exhaustion and promote chimeric antigen receptor T-cell function(34). Considering the positive correlation between *RGS1* with Candidate gene set in single cells and tissues and its high expression in Tex cells, it is important and the necessity to further study *RGS1* mechanism in T cell exhaustion.

In summary, our findings suggest that the GSVA score of the 35 Candidate gene set could be an effective prognostic marker or a marker to identify Tex cells. *RGS1,* as the most up-regulated gene in Pre_exhaust and Tex cells, might play key roles in T

cells exhaustion or cancer progress. As a HIF-dependent hypoxia target, *RGS1* might be upregulated by hypoxia, and further mediate the cell retention by inhibiting chemokine-induced lymphocyte migration. The current study could provide theoretical basis for research and immunotherapy of exhausted cells, while further studies are essential to fully elucidate the concrete mechanism of *RGS1* during CD8+ T cell exhaustion.

Author Contributions

All authors contributed significantly to the work and the preparation of the manuscript. HD conceived the study. YB analyzed the data with assistance from ZC, wrote the manuscript. MH collected the data and performed IHC experiment. JW reviewed and revised the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

This research was funded by the National Key R&D Program of China (2018YFC0910200), the Key R&D Program of Guangdong Province (2019B020226001)

Conflicts of Interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data Availability Statement

The original contributions presented in the study are included in the article/Supplementary Material, the original IHC figures were uploaded in https://www.jianguoyun.com/p/DdToBqoQ4ISAChiXhJkE and further inquiries can be directed to the corresponding author/s.

References

1. Moskophidis D, Lechner F, Pircher H, Zinkernagel RM. Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells. *Nature* (1993) 362(6422):758-61. Epub 1993/04/22. doi: 10.1038/362758a0. PubMed PMID: 8469287.
2. Lee PP, Yee C, Savage PA, Fong L, Brockstedt D, Weber JS, et al. Characterization of circulating T cells specific for tumor-associated antigens in melanoma patients. *Nat Med* (1999) 5(6):677-85. Epub 1999/06/17. doi: 10.1038/9525. PubMed PMID: 10371507.
3. Catakovic K, Klieser E, Neureiter D, Geisberger R. T cell exhaustion: from pathophysiological basics to tumor immunotherapy. *Cell communication and signaling : CCS* (2017) 15(1):1. Epub 2017/01/12. doi: 10.1186/s12964-016-0160-z. PubMed PMID: 28073373; PubMed Central PMCID: PMCPMC5225559.
4. Dunn GP, Old LJ, Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity* (2004) 21(2):137-48. doi: 10.1016/j.immuni.2004.07.017. PubMed PMID: 15308095.
5. Pauken KE, Wherry EJ. Overcoming T cell exhaustion in infection and cancer. *Trends Immunol* (2015) 36(4):265-76. doi: 10.1016/j.it.2015.02.008. PubMed PMID: 25797516; PubMed Central PMCID: PMCPMC4393798.
6. Schietinger A, Philip M, Krisnawan VE, Chiu EY, Delrow JJ, Basom RS, et al. Tumor-Specific T Cell Dysfunction Is a Dynamic Antigen-Driven Differentiation Program Initiated Early during Tumorigenesis. *Immunity* (2016) 45(2):389-401. doi: 10.1016/j.immuni.2016.07.011. PubMed PMID: 27521269; PubMed Central PMCID: PMCPMC5119632.
7. Blackburn SD, Shin H, Haining WN, Zou T, Workman CJ, Polley A, et al. Coregulation of CD8 + T cell exhaustion by multiple inhibitory receptors during chronic viral infection. (2008). doi: 10.1038/ni.1679.
8. Wagner A, Regev A, Yosef N. Revealing the vectors of cellular identity with single-cell genomics. *Nature Biotechnology* (2016) 34(11):1145-60. doi: 10.1038/nbt.3711.
9. Winterhoff BJ, Maile M, Mitra AK, Sebe A, Bazzaro M, Geller MA, et al. Single cell sequencing reveals heterogeneity within ovarian cancer epithelium and cancer associated stromal cells. *Gynecologic oncology* (2017) 144(3):598-606. Epub 2017/01/24. doi: 10.1016/j.ygyno.2017.01.015. PubMed PMID: 28111004; PubMed Central PMCID: PMCPMC5316302.
10. Wu H, Yu J, Li Y, Hou Q, Zhou R, Zhang N, et al. Single-cell RNA sequencing reveals diverse intratumoral heterogeneities and gene signatures of two types of esophageal cancers. *Cancer letters* (2018) 438:133-43. Epub 2018/09/18. doi: 10.1016/j.canlet.2018.09.017. PubMed PMID: 30223068.
11. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nature reviews Immunology* (2018) 18(1):35-45. Epub 2017/08/09. doi: 10.1038/nri.2017.76. PubMed PMID: 28787399.
12. Darmanis S, Sloan SA, Croote D, Mignardi M, Chernikova S, Samghababi P, et al. Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell reports* (2017) 21(5):1399-410. Epub 2017/11/02. doi: 10.1016/j.celrep.2017.10.030. PubMed PMID: 29091775; PubMed Central PMCID: PMCPMC5810554.
13. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (New York, NY)* (2016) 352(6282):189-96. Epub 2016/04/29. doi: 10.1126/science.aad0501. PubMed PMID: 27124452;

PubMed Central PMCID: PMCPMC4944528.

14. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* (2017) 171(7):1611-24.e24. Epub 2017/12/05. doi: 10.1016/j.cell.2017.10.044. PubMed PMID: 29198524; PubMed Central PMCID: PMCPMC5878932.

15. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* (2018) 174(5):1293-308.e36. Epub 2018/07/03. doi: 10.1016/j.cell.2018.05.060. PubMed PMID: 29961579; PubMed Central PMCID: PMCPMC6348010.

16. Huang AC, Postow MA, Orlowski RJ, Mick R, Bengsch B, Manne S, et al. T-cell invigoration to tumour burden ratio associated with anti-PD-1 response. *Nature* (2017) 545(7652):60-5. doi: 10.1038/nature22079. PubMed PMID: 28397821; PubMed Central PMCID: PMCPMC5554367.

17. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol Cell* (2017) 65(4):631-43 e4. doi: 10.1016/j.molcel.2017.01.023. PubMed PMID: 28212749.

18. Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Research* (2018) 47(D1):D721-D8. doi: 10.1093/nar/gky900.

19. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature immunology* (2019) 20(2):163-72. Epub 2019/01/16. doi: 10.1038/s41590-018-0276-y. PubMed PMID: 30643263; PubMed Central PMCID: PMCPMC6340744.

20. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* (2014) 32(4):381-6. doi: 10.1038/nbt.2859. PubMed PMID: 24658644; PubMed Central PMCID: PMCPMC4122333.

21. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology* (2012) 16(5):284-7. Epub 2012/03/30. doi: 10.1089/omi.2011.0118. PubMed PMID: 22455463; PubMed Central PMCID: PMCPMC3339379.

22. JC B. ggalluvial: Layered Grammar for Alluvial Plots. *Journal of Open Source Software* (202) 5(49):2017-23. Epub 2020/5/21. doi: https://doi.org/10.21105/joss.02017.

23. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* (2008) 9:559. Epub 2008/12/31. doi: 10.1186/1471-2105-9-559. PubMed PMID: 19114008; PubMed Central PMCID: PMCPMC2631488.

24. Tosches MA, Yamawaki TM, Naumann RK, Jacobi AA, Tushev G, Laurent G. Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* (2018) 360(6391):881-8. doi: 10.1126/science.aar4237. PubMed PMID: 29724907.

25. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics* (2013) 14:7. Epub 2013/01/18. doi: 10.1186/1471-2105-14-7. PubMed PMID: 23323831; PubMed Central PMCID: PMCPMC3618321.

26. Jensen EC. Quantitative analysis of histological staining and fluorescence using ImageJ. *Anatomical record (Hoboken, NJ : 2007)* (2013) 296(3):378-81. Epub 2013/02/06. doi: 10.1002/ar.22641. PubMed PMID: 23382140.

27. Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, et al. Lineage tracking reveals dynamic

relationships of T cells in colorectal cancer. *Nature* (2018) 564(7735):268-72. doi: 10.1038/s41586-018-0694-x. PubMed PMID: 30479382.

28. Zheng C, Zheng L, Yoo JK, Guo H, Zhang Y, Guo X, et al. Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell* (2017) 169(7):1342-56 e16. doi: 10.1016/j.cell.2017.05.035. PubMed PMID: 28622514.

29. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med* (2018) 24(7):978-85. doi: 10.1038/s41591-018-0045-3. PubMed PMID: 29942094.

30. Carmona SJ, Siddiqui I, Bilous M, Held W, Gfeller D. Deciphering the transcriptomic landscape of tumor-infiltrating CD8 lymphocytes in B16 melanoma tumors with single-cell RNA-Seq. *Oncoimmunology* (2020) 9(1):1737369. doi: 10.1080/2162402X.2020.1737369. PubMed PMID: 32313720; PubMed Central PMCID: PMCPMC7153840.

31. Yost KE, Satpathy AT, Wells DK, Qi Y, Wang C, Kageyama R, et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat Med* (2019) 25(8):1251-9. doi: 10.1038/s41591-019-0522-3. PubMed PMID: 31359002; PubMed Central PMCID: PMCPMC6689255.

32. Wherry EJ. T cell exhaustion. *Nat Immunol* (2011) 12(6):492-9. doi: 10.1038/ni.2035. PubMed PMID: 21739672.

33. Mita Y, Kimura MY, Hayashizaki K, Koyama-Nasu R, Ito T, Motohashi S, et al. Crucial role of CD69 in anti-tumor immunity through regulating the exhaustion of tumor-infiltrating T cells. *Int Immunol* (2018) 30(12):559-67. doi: 10.1093/intimm/dxy050. PubMed PMID: 30085193.

34. Kumar J, Kumar R, Kumar Singh A, Tsakem EL, Kathania M, Riese MJ, et al. Deletion of Cbl-b inhibits CD8(+) T-cell exhaustion and promotes CAR T-cell function. *J Immunother Cancer* (2021) 9(1):e001688. Epub 2021/01/20. doi: 10.1136/jitc-2020-001688. PubMed PMID: 33462140; PubMed Central PMCID: PMCPMC7813298.

35. Liu Z, Zhou Q, Wang Z, Zhang H, Zeng H, Huang Q, et al. Intratumoral TIGIT(+) CD8(+) T-cell infiltration determines poor prognosis and immune evasion in patients with muscle-invasive bladder cancer. *J Immunother Cancer* (2020) 8(2):e000978. Epub 2020/08/21. doi: 10.1136/jitc-2020-000978. PubMed PMID: 32817209; PubMed Central PMCID: PMCPMC7430558.

36. Hu J, Chen Z, Bao L, Zhou L, Hou Y, Liu L, et al. Single-Cell Transcriptome Analysis Reveals Intratumoral Heterogeneity in ccRCC, which Results in Different Clinical Outcomes. *Mol Ther* (2020) 28(7):1658-72. doi: 10.1016/j.ymthe.2020.04.023. PubMed PMID: 32396851; PubMed Central PMCID: PMCPMC7335756.

37. Peng DH, Rodriguez BL, Diao L, Chen L, Wang J, Byers LA, et al. Collagen promotes anti-PD-1/PD-L1 resistance in cancer through LAIR1-dependent CD8+ T cell exhaustion. *Nature Communications* (2020) 11(1):4520. doi: 10.1038/s41467-020-18298-8.

38. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* (2014) 515(7528):568-71. doi: 10.1038/nature13954. PubMed PMID: 25428505; PubMed Central PMCID: PMCPMC4246418.

39. Herbst RS, Soria JC, Kowanetz M, Fine GD, Hamid O, Gordon MS, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* (2014) 515(7528):563-7. doi: 10.1038/nature14011. PubMed PMID: 25428504; PubMed Central PMCID: PMCPMC4836193.

40. Bengsch B, Johnson AL, Kurachi M, Odorizzi PM, Pauken KE, Attanasio J, et al. Bioenergetic Insufficiencies Due to Metabolic Alterations Regulated by the Inhibitory Receptor PD-1 Are an Early

Driver of CD8(+) T Cell Exhaustion. *Immunity* (2016) 45(2):358-73. Epub 2016/08/09. doi: 10.1016/j.immuni.2016.07.008. PubMed PMID: 27496729; PubMed Central PMCID: PMCPMC4988919.

41. Huang AC, Orlowski RJ, Xu X, Mick R, George SM, Yan PK, et al. A single dose of neoadjuvant PD-1 blockade predicts clinical outcomes in resectable melanoma. *Nat Med* (2019) 25(3):454-61. doi: 10.1038/s41591-019-0357-y. PubMed PMID: 30804515; PubMed Central PMCID: PMCPMC6699626.

42. Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, et al. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol* (2014) 32(10):1053-8. doi: 10.1038/nbt.2967. PubMed PMID: 25086649; PubMed Central PMCID: PMCPMC4191988.

43. Paley MA, Kroy DC, Odorizzi PM, Johnnidis JB, Dolfi DV, Barnett BE, et al. Progenitor and terminal subsets of CD8+ T cells cooperate to contain chronic viral infection. *Science* (2012) 338(6111):1220-5. doi: 10.1126/science.1229620. PubMed PMID: 23197535; PubMed Central PMCID: PMCPMC3653769.

44. Watson N, Linder ME, Druey KM, Kehrl JH, Blumer KJ. RGS family members: GTPase-activating proteins for heterotrimeric G-protein alpha-subunits. *Nature* (1996) 383(6596):172-5. doi: 10.1038/383172a0. PubMed PMID: 8774882.

45. Agenes F, Bosco N, Mascarell L, Fritah S, Ceredig R. Differential expression of regulator of G-protein signalling transcripts and in vivo migration of CD4+ naive and regulatory T cells. *Immunology* (2005) 115(2):179-88. doi: 10.1111/j.1365-2567.2005.02146.x. PubMed PMID: 15885123; PubMed Central PMCID: PMCPMC1782143.

46. Moratz C, Kang VH, Druey KM, Shi CS, Scheschonka A, Murphy PM, et al. Regulator of G protein signaling 1 (RGS1) markedly impairs Gi alpha signaling responses of B lymphocytes. *J Immunol* (2000) 164(4):1829-38. doi: 10.4049/jimmunol.164.4.1829. PubMed PMID: 10657631.

47. Kveberg L, Ryan JC, Rolstad B, Inngjerdingen M. Expression of regulator of G protein signalling proteins in natural killer cells, and their modulation by Ly49A and Ly49D. *Immunology* (2005) 115(3):358-65. doi: 10.1111/j.1365-2567.2005.02174.x. PubMed PMID: 15946253; PubMed Central PMCID: PMCPMC1782169.

48. Shi GX, Harrison K, Han SB, Moratz C, Kehrl JH. Toll-like receptor signaling alters the expression of regulator of G protein signaling proteins in dendritic cells: implications for G protein-coupled receptor signaling. *J Immunol* (2004) 172(9):5175-84. doi: 10.4049/jimmunol.172.9.5175. PubMed PMID: 15100254.

49. Denecke B, Meyerdierks A, Bottger EC. RGS1 is expressed in monocytes and acts as a GTPase-activating protein for G-protein-coupled chemoattractant receptors. *The Journal of biological chemistry* (1999) 274(38):26860-8. doi: 10.1074/jbc.274.38.26860. PubMed PMID: 10480894.

50. Han JI, Huang NN, Kim DU, Kehrl JH. RGS1 and RGS13 mRNA silencing in a human B lymphoma line enhances responsiveness to chemoattractants and impairs desensitization. *J Leukoc Biol* (2006) 79(6):1357-68. doi: 10.1189/jlb.1105693. PubMed PMID: 16565322.

51. Offermanns S, Simon MI. Organization of transmembrane signalling by heterotrimeric G proteins. *Cancer surveys* (1996) 27(27):177-98. doi: 10.3109/07357909609076911. PubMed Central PMCID: PMC8909801.

52. Roh J, Shin SJ, Lee AN, Yoon DH, Suh C, Park CJ, et al. RGS1 expression is associated with poor prognosis in multiple myeloma. *J Clin Pathol* (2017) 70(3):202-7. doi: 10.1136/jclinpath-2016-203713.

PubMed PMID: 27445341.
53. Kashani-Sabet M, Nosrati M, Miller JR, Sagebiel RW, Leong SPL, Lesniak A, et al. Prospective Validation of Molecular Prognostic Markers in Cutaneous Melanoma: A Correlative Analysis of E1690. *Clinical Cancer Research* (2017) 23(22):6888-92. doi: 10.1158/1078-0432.ccr-17-1317.
54. Rangel J, Nosrati M, Leong SP, Haqq C, Miller JR, 3rd, Sagebiel RW, et al. Novel role for RGS1 in melanoma progression. *Am J Surg Pathol* (2008) 32(8):1207-12. doi: 10.1097/PAS.0b013e31816fd53c. PubMed PMID: 18580492.
55. Dai J, Gu J, Lu C, Lin J, Stewart D, Chang D, et al. Genetic variations in the regulator of G-protein signaling genes are associated with survival in late-stage non-small cell lung cancer. *PLoS One* (2011) 6(6):e21120. doi: 10.1371/journal.pone.0021120. PubMed PMID: 21698121; PubMed Central PMCID: PMCPMC3117866.
56. Li S, Yang H, Li S, Zhao Z, Wang D, Fu W. High expression of regulator of G-protein signalling 1 is associated with the poor differentiation and prognosis of gastric cancer. *Oncology letters* (2021) 21(4):322. Epub 2021/03/12. doi: 10.3892/ol.2021.12584. PubMed PMID: 33692854; PubMed Central PMCID: PMCPMC7933750.
57. Carreras J, Kikuti YY, Beà S, Miyaoka M, Hiraiwa S, Ikoma H, et al. Clinicopathological characteristics and genomic profile of primary sinonasal tract diffuse large B cell lymphoma (DLBCL) reveals gain at 1q31 and RGS1 encoding protein; high RGS1 immunohistochemical expression associates with poor overall survival in DLBCL not otherwise specified (NOS). *Histopathology* (2017) 70(4):595-621. Epub 2016/10/25. doi: 10.1111/his.13106. PubMed PMID: 27775850.
58. Blank CU, Haining WN, Held W, Hogan PG, Kallies A, Lugli E, et al. Defining 'T cell exhaustion'. *Nature reviews Immunology* (2019) 19(11):665-74. Epub 2019/10/02. doi: 10.1038/s41577-019-0221-9. PubMed PMID: 31570879; PubMed Central PMCID: PMCPMC7286441.
59. Wierenga AT, Vellenga E, Schuringa JJ. Convergence of hypoxia and TGFβ pathways on cell cycle regulation in human hematopoietic stem/progenitor cells. *PLoS One* (2014) 9(3):e93494. Epub 2014/04/02. doi: 10.1371/journal.pone.0093494. PubMed PMID: 24686421; PubMed Central PMCID: PMCPMC3970968.
60. Ziegler SF, Ramsdell F, Hjerrild KA, Armitage RJ, Grabstein KH, Hennen KB, et al. Molecular characterization of the early activation antigen CD69: a type II membrane glycoprotein related to a family of natural killer cell activation antigens. *Eur J Immunol* (1993) 23(7):1643-8. doi: 10.1002/eji.1830230737. PubMed PMID: 8100776.
61. Cyster JG, Schwab SR. Sphingosine-1-phosphate and lymphocyte egress from lymphoid organs. *Annu Rev Immunol* (2012) 30:69-94. doi: 10.1146/annurev-immunol-020711-075011. PubMed PMID: 22149932.
62. Kimura MY, Hayashizaki K, Tokoyoda K, Takamura S, Motohashi S, Nakayama T. Crucial role for CD69 in allergic inflammatory responses: CD69-Myl9 system in the pathogenesis of airway inflammation. *Immunol Rev* (2017) 278(1):87-100. doi: 10.1111/imr.12559. PubMed PMID: 28658550.

Figure and Table legends

Figure 1 Clustering of CD8+ T cells in three cancers.

A. UMAP of single cells to visualize cell-type clusters (left), Pseudo-time trajectory graph (middle) and the proportion of different cell types in different sources (right).

B. Heatmap showing marker genes for CD8+ cell types.

C. The top 10 DEGs in each cell type in three cancers.

D. The GO enrichment analysis of different cell types of CD8+ T cells in three cancers.

Figure 2 Candidate gene set associated with CD8+ Tex cells.

A. The gene co-expression network modules of CD8+ T cells with correlation coefficient and P-value.

B. The number of genes with differential expression (left) and co-expression (right).

C. The Kaplan-Meier overall survival curves of TCGA patients grouped by the middle expression value of Candidate gene set. The red and blue lines denote higher and lower expression group, respectively.

D. Distinguishing Tex cells from the other CD8+ T cells effectly in different cancers by GSVA score of Candidate gene set.

Figure 3 DEGs in Pre_exhasuted and Tex cells compared with Teff cells

A. The common up-regulated genes in Pre_exhasuted and Tex cells compared with Teff cells.

B. The mRNA expression value of *RGS1* in single cell dataset and TCGA database (**C**).

D. The Kaplan-Meier overall survival curves of TCGA patients grouped by the middle expression value of *RGS1*. The red and blue lines denote higher and lower expression group, respectively.

E. Representative IHC images of RGS1 protein in tumor and normal tissues of liver derived from the HPA database and verification experiment (**F**, scale bar 100 μm, magnification ×20).

G. The protein expression value of RGS1 in hepatocarcinoma and adjacent noncancerous tissues in the IHC verification experiment.

Figure 4 The correlation coefficient between *RGS1* and Candidate gene set of Tex cells in CD8+ T cells (left) and TCGA database(right).

Table 1 The logarithm of fold change of Candidate gene set in three cancers

Supplemental Information titles and legends

Supplemental Figure 1: Clustering of CD8+ T cells in three cancers.

A. UMAP visualization of T cells in three cancers.

B. The number of detected genes in each cell before filter in three cancers.

C. The exhaustion scores of different cell types using the exhaustion gene list including *HAVCR2*, *TIGIT*, *LAG3*, *PDCD1*, *CXCL13*, *LAYN*, *TOX*, *CTLA4*, *BTLA*.

Supplemental Figure 2: The correlation coefficient between *RGS1* and Candidate gene set of Tex cells in CD8+ T cells of different patients.

Supplemental Figure 3: The mRNA expression value of *RGS1* of different stages in TCGA database.

Supplemental Table 1: The patient ID and numbers of cell in three cancers

Supplemental Table 2: Differentially expressed genes of each cluster in three cancers.

Supplemental Table 3: Differentially expressed genes of Pre_exhaust and Tex cells compared with Teff cells in three cancers.

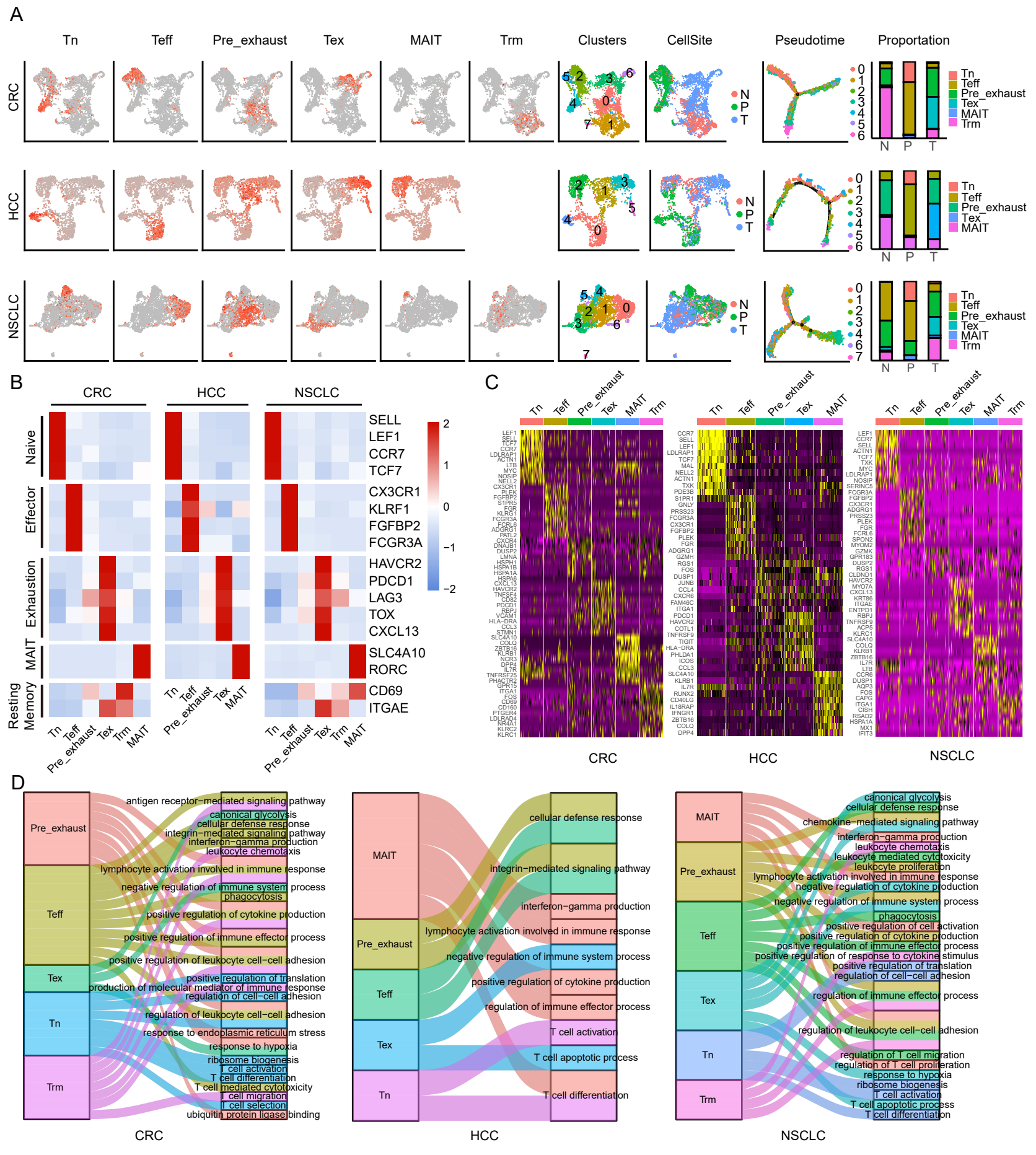
Supplemental Table 4: Survival analysis results of .RGS1 in different cancer types. Adj.p value referred to the adjusted p value corrected by multiple hypothesis testing (FDR).

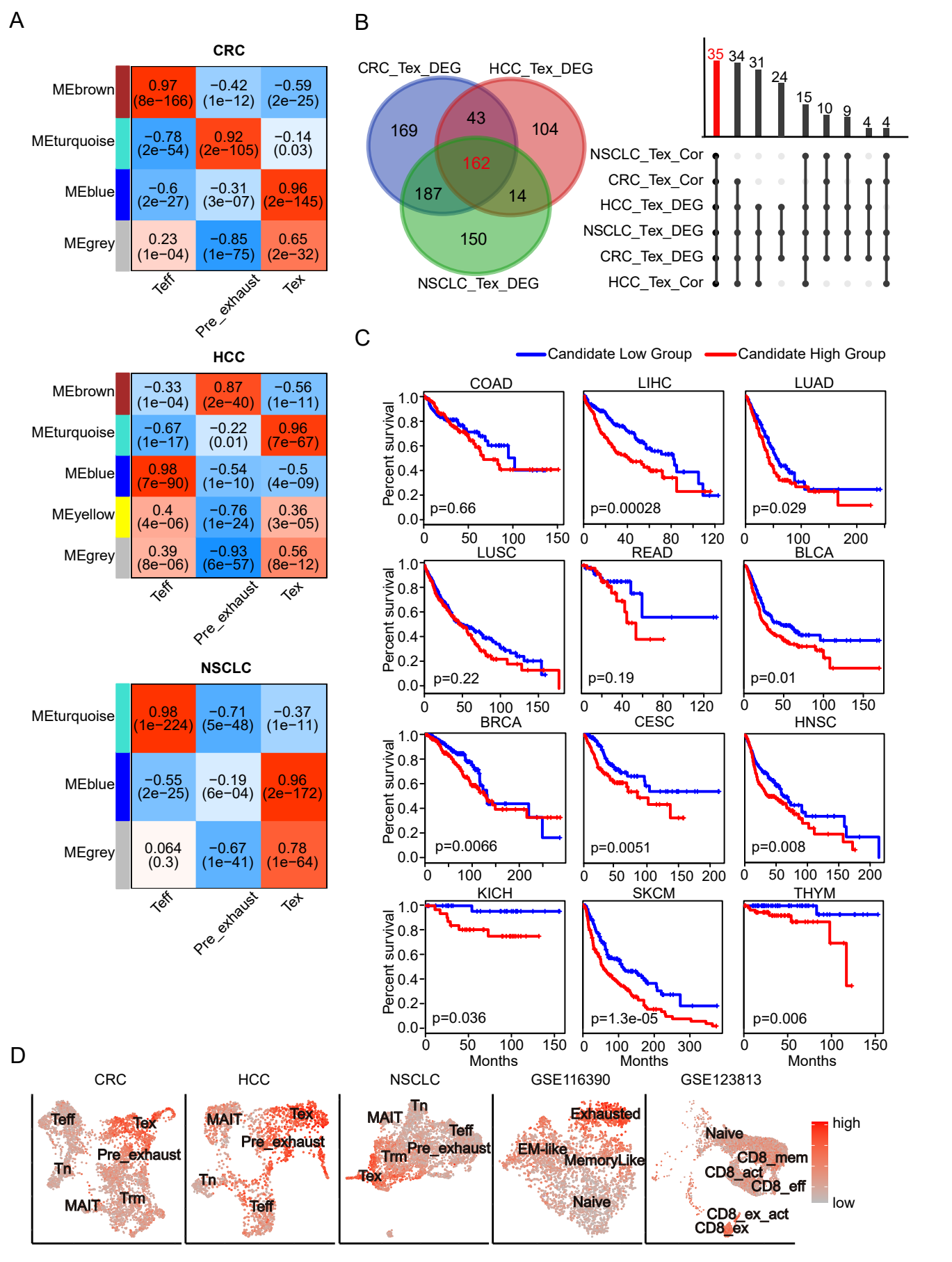
Supplemental Table 5: The clinicopathological information of the patient samples and RGS1 protein expression value.

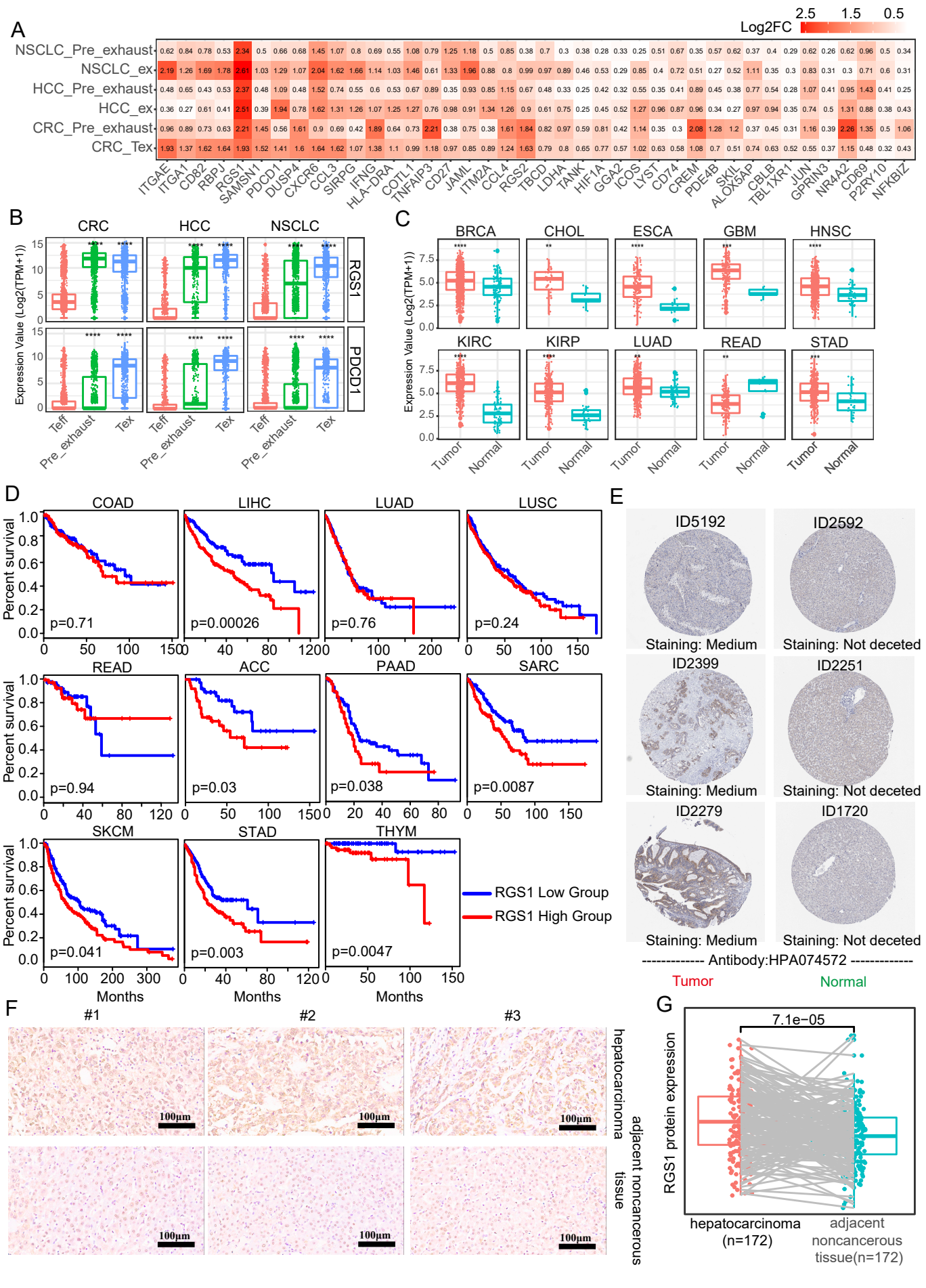
706

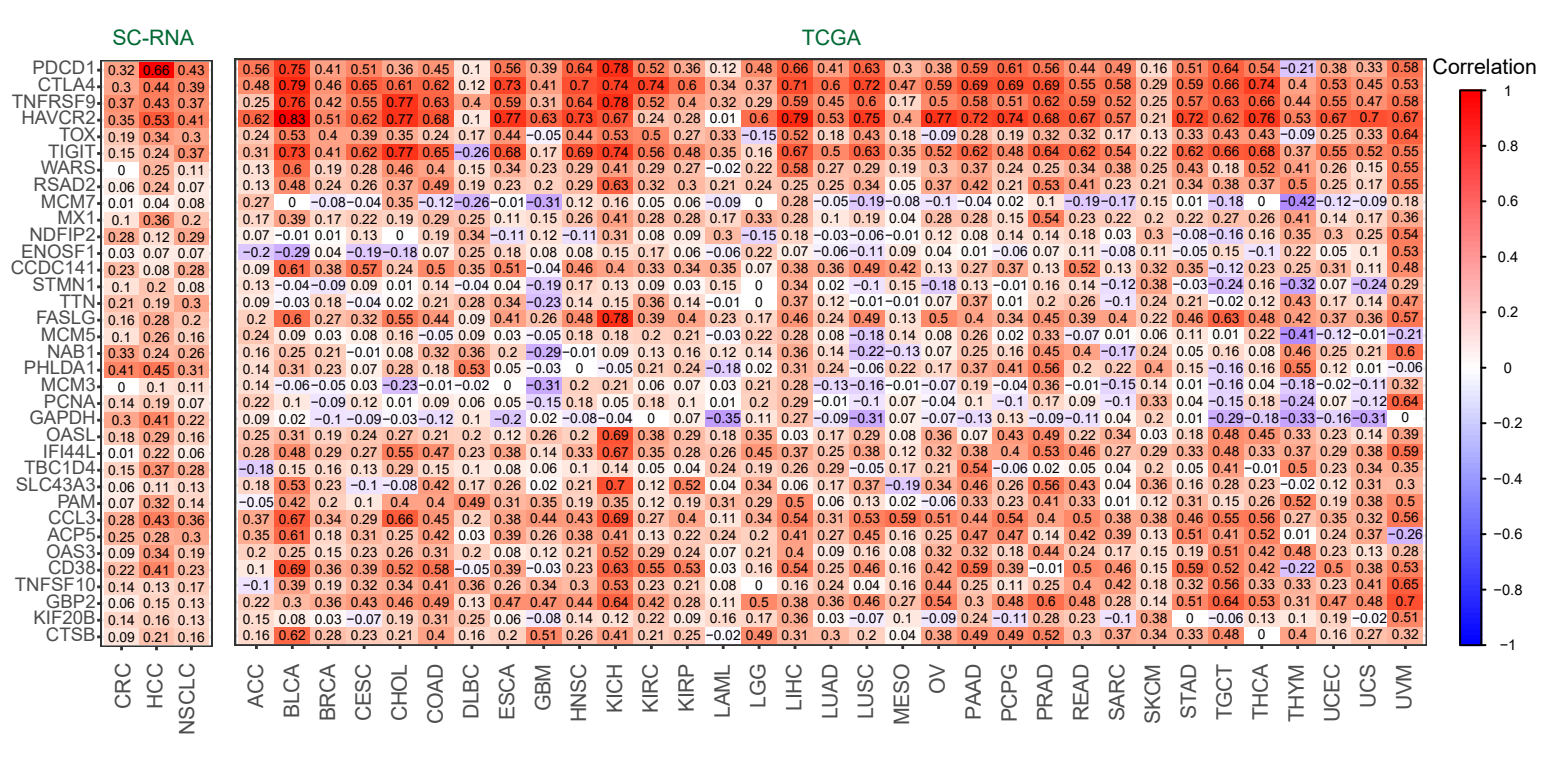
Table 1 The logarithm of fold change of Candidate gene set in three cancers

	CRC	HCC	NSCLC		CRC	HCC	NSCLC
PDCD1	1.41	1.94	1.29	*PHLDA1*	1.52	1.31	1.05
CTLA4	1.02	0.63	0.74	*MCM3*	0.36	0.39	0.51
TNFRSF9	1.53	1.53	1.58	*PCNA*	0.59	0.49	0.52
HAVCR2	2.15	1.75	2.05	*GAPDH*	1.06	1.00	1.10
TOX	0.71	0.92	0.74	*OASL*	0.65	0.55	0.65
TIGIT	0.59	0.79	1.24	*IFI44L*	0.83	0.37	0.74
WARS	0.41	0.99	0.34	*TBC1D4*	0.66	0.82	0.74
RSAD2	0.62	0.31	0.43	*SLC43A3*	0.33	0.27	0.45
MCM7	0.69	0.41	0.62	*PAM*	0.27	0.68	0.30
MX1	0.71	0.81	1.06	*CCL3*	1.62	1.31	1.62
NDFIP2	1.33	0.46	1.21	*ACP5*	1.17	0.75	1.36
ENOSF1	0.55	0.49	0.55	*OAS3*	0.58	0.57	0.63
CCDC141	0.89	0.36	1.03	*CD38*	0.73	0.79	0.51
STMN1	1.30	1.04	0.95	*TNFSF10*	0.64	0.36	0.71
TTN	0.81	0.45	1.08	*GBP2*	0.38	0.49	0.35
FASLG	0.88	0.87	0.82	*KIF20B*	0.37	0.35	0.44
MCM5	0.71	0.75	0.68	*CTSB*	0.43	0.31	0.35
NAB1	0.76	0.54	0.56				


A
Tn
Teff
Pre_exhaust
Tex
MAIT
Trm
Clusters
CellSite
Pseudotime
Proportation
CRC
HCC
NSCLC
N
P
T
B
Naive
Effector
Exhaustion
MAIT
Resting Memory
SELL
LEF1
CCR7
TCF7
CX3CR1
KLRF1
FGFBP2
FCGR3A
HAVCR2
PDCD1
LAG3
TOX
CXCL13
SLC4A10
RORC
CD69
ITGAE
C
D
antigen receptor−mediated signaling pathway
canonical glycolysis
cellular defense response
integrin−mediated signaling pathway
interferon−gamma production
leukocyte chemotaxis
lymphocyte activation involved in immune response
negative regulation of immune system process
phagocytosis
positive regulation of cytokine production
positive regulation of immune effector process
positive regulation of leukocyte cell−cell adhesion
positive regulation of translation
production of molecular mediator of immune response
regulation of cell−cell adhesion
regulation of leukocyte cell−cell adhesion
response to endoplasmic reticulum stress
response to hypoxia
ribosome biogenesis
T cell activation
T cell differentiation
T cell mediated cytotoxicity
T cell migration
T cell selection
ubiquitin protein ligase binding
T cell apoptotic process
chemokine−mediated signaling pathway
leukocyte mediated cytotoxicity
leukocyte proliferation
negative regulation of cytokine production
positive regulation of response to cytokine stimulus
regulation of T cell migration
regulation of T cell proliferation


A
CRC
MEbrown
MEturquoise
MEblue
MEgrey
0.97 (8e−166)
−0.42 (1e−12)
−0.59 (2e−25)
−0.78 (2e−54)
0.92 (2e−105)
−0.14 (0.03)
−0.6 (2e−27)
−0.31 (3e−07)
0.96 (2e−145)
0.23 (1e−04)
−0.85 (1e−75)
0.65 (2e−32)
Teff
Pre_exhaust
Tex
HCC
MEbrown
MEturquoise
MEblue
MEyellow
MEgrey
−0.33 (1e−04)
0.87 (2e−40)
−0.56 (1e−11)
−0.67 (1e−17)
−0.22 (0.01)
0.96 (7e−67)
0.98 (7e−90)
−0.54 (1e−10)
−0.5 (4e−09)
0.4 (4e−06)
−0.76 (1e−24)
0.36 (3e−05)
0.39 (8e−06)
−0.93 (6e−57)
0.56 (8e−12)
Teff
Pre_exhaust
Tex
NSCLC
MEturquoise
MEblue
MEgrey
0.98 (1e−224)
−0.71 (5e−48)
−0.37 (1e−11)
−0.55 (2e−25)
−0.19 (6e−04)
0.96 (2e−172)
0.064 (0.3)
−0.67 (1e−41)
0.78 (1e−64)
Teff
Pre_exhaust
Tex
B
CRC_Tex_DEG
HCC_Tex_DEG
NSCLC_Tex_DEG
169
43
104
162
187
14
150
35
34
31
24
15
10
9
4
4
NSCLC_Tex_Cor
CRC_Tex_Cor
HCC_Tex_DEG
NSCLC_Tex_DEG
CRC_Tex_DEG
HCC_Tex_Cor
C
Candidate Low Group
Candidate High Group
COAD
p=0.66
LIHC
p=0.00028
LUAD
p=0.029
LUSC
p=0.22
READ
p=0.19
BLCA
p=0.01
BRCA
p=0.0066
CESC
p=0.0051
HNSC
p=0.008
KICH
p=0.036
SKCM
p=1.3e-05
THYM
p=0.006
Percent survival
Months
D
CRC
Teff
Tex
Pre_exhaust
Tn
MAIT
Trm
HCC
MAIT
Tex
Pre_exhaust
Tn
Teff
NSCLC
Tn
MAIT
Teff
Pre_exhaust
Trm
Tex
GSE116390
Exhausted
EM-like
MemoryLike
Naive
GSE123813
Naive
CD8_mem
CD8_act
CD8_eff
CD8_ex_act
CD8_ex
high
low


A
Log2FC
2.5
1.5
0.5
NSCLC_Pre_exhaust
NSCLC_ex
HCC_Pre_exhaust
HCC_ex
CRC_Pre_exhaust
CRC_Tex
ITGAE ITGA1 CD82 RBPJ RGS1 SAMSN1 PDCD1 DUSP4 CXCR6 CCL3 SIRPG IFNG HLA-DRA COTL1 TNFAIP3 CD27 JAML ITM2A CCL4 RGS2 TBCD LDHA TANK HIF1A GGA2 ICOS LYST CD74 CREM PDE4B SKIL ALOX5AP CBLB TBL1XR1 JUN GPRIN3 NR4A2 CD69 P2RY10 NFKBIZ
B
CRC
HCC
NSCLC
RGS1
PDCD1
Expression Value (Log2(TPM+1))
Teff
Pre_exhaust
Tex
C
BRCA
CHOL
ESCA
GBM
HNSC
KIRC
KIRP
LUAD
READ
STAD
Tumor
Normal
D
COAD
LIHC
LUAD
LUSC
READ
ACC
PAAD
SARC
SKCM
STAD
THYM
Percent survival
p=0.71
p=0.00026
p=0.76
p=0.24
p=0.94
p=0.03
p=0.038
p=0.0087
p=0.041
p=0.003
p=0.0047
Months
RGS1 Low Group
RGS1 High Group
E
ID5192
ID2592
ID2399
ID2251
ID2279
ID1720
Staining: Medium
Staining: Not deceted
Antibody:HPA074572
Tumor
Normal
F
#1
#2
#3
100μm
hepatocarcinoma
adjacent noncancerous tissue
G
7.1e−05
RGS1 protein expression
hepatocarcinoma (n=172)
adjacent noncancerous tissue(n=172)


SC-RNA
TCGA
Correlation
PDCD1
CTLA4
TNFRSF9
HAVCR2
TOX
TIGIT
WARS
RSAD2
MCM7
MX1
NDFIP2
ENOSF1
CCDC141
STMN1
TTN
FASLG
MCM5
NAB1
PHLDA1
MCM3
PCNA
GAPDH
OASL
IFI44L
TBC1D4
SLC43A3
PAM
CCL3
ACP5
OAS3
CD38
TNFSF10
GBP2
KIF20B
CTSB
CRC
HCC
NSCLC
ACC
BLCA
BRCA
CESC
CHOL
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
PCPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
UVM
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1