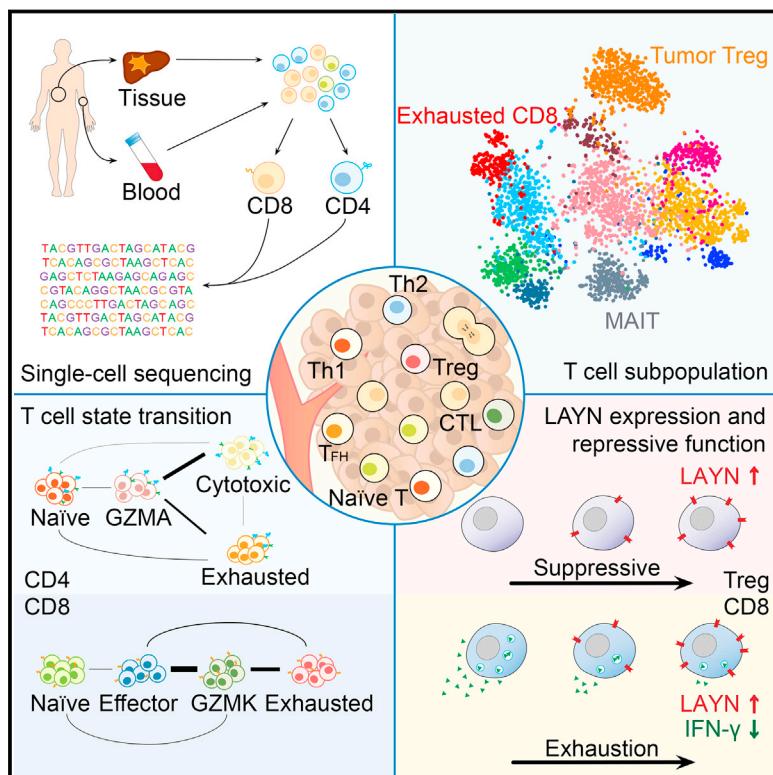


Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing

Graphical Abstract



Authors

Chunhong Zheng, Liangtao Zheng, Jae-Kwang Yoo, ..., Wenjun Ouyang, Jirun Peng, Zemin Zhang

Correspondence

wouyang@amgen.com (W.O.), pengjr@medmail.com.cn (J.P.), zemin@pku.edu.cn (Z.Z.)

In Brief

Analysis of T cell populations from hepatocellular carcinoma patients using single-cell sequencing reveals distinct subtypes and clonal expansion of infiltrating lymphocytes.

Highlights

- Single-cell RNA-seq reveals distinctive functional composition of T cells in HCC
- Clonal enrichment of infiltrating Tregs and exhausted CD8 T cells in HCC
- LAYN is linked to the suppressive function of tumor Treg and exhausted CD8 T cells
- Combined expression and TCR-based analyses show connectivity among T cell subsets

Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing

Chunhong Zheng,^{1,7} Liangtao Zheng,^{2,7} Jae-Kwang Yoo,^{3,7} Huahu Guo,^{4,5,6,7} Yuanyuan Zhang,^{1,7} Xinyi Guo,^{1,7} Boxi Kang,¹ Ruozhen Hu,³ Julie Y. Huang,³ Qiming Zhang,¹ Zhouzerui Liu,¹ Minghui Dong,¹ Xueda Hu,¹ Wenjun Ouyang,^{3,*} Jirun Peng,^{4,5,6,*} and Zemin Zhang^{1,2,8,*}

¹BIOPIC, Beijing Advanced Innovation Center for Genomics, and School of Life Sciences, Peking University, Beijing 100871, China

²Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

³Department of Inflammation and Oncology, Amgen Inc., South San Francisco, CA 94080, USA

⁴Department of Surgery, Beijing Shijitan Hospital, Capital Medical University, Beijing 100038, China

⁵Ninth School of Clinical Medicine, Peking University, Beijing 100038, China

⁶School of Oncology, Capital Medical University, Beijing 100038, China

⁷These authors contributed equally

⁸Lead Contact

*Correspondence: wouyang@amgen.com (W.O.), pengjr@medmail.com.cn (J.P.), zemin@pku.edu.cn (Z.Z.)

<http://dx.doi.org/10.1016/j.cell.2017.05.035>

SUMMARY

Systematic interrogation of tumor-infiltrating lymphocytes is key to the development of immunotherapies and the prediction of their clinical responses in cancers. Here, we perform deep single-cell RNA sequencing on 5,063 single T cells isolated from peripheral blood, tumor, and adjacent normal tissues from six hepatocellular carcinoma patients. The transcriptional profiles of these individual cells, coupled with assembled T cell receptor (TCR) sequences, enable us to identify 11 T cell subsets based on their molecular and functional properties and delineate their developmental trajectory. Specific subsets such as exhausted CD8⁺ T cells and Tregs are preferentially enriched and potentially clonally expanded in hepatocellular carcinoma (HCC), and we identified signature genes for each subset. One of the genes, layilin, is upregulated on activated CD8⁺ T cells and Tregs and represses the CD8⁺ T cell functions *in vitro*. This compendium of transcriptome data provides valuable insights and a rich resource for understanding the immune landscape in cancers.

INTRODUCTION

Cancer immunotherapies have dramatically altered the oncological treatment landscape over the past decades (Sharma and Allison, 2015). While such therapies as checkpoint blockade can lead to remarkable clinical responses (Topalian et al., 2015), their efficacies are not uniform among cancer patients or cancer types. It is, therefore, paramount to identify robust biomarkers that are predictive of therapy responses. Several factors such as mutational loads, tumor-infiltrating lymphocytes (TILs) levels, and the expression of drug targets provide certain correlation with patient responses for anti-CTLA4 or anti-PD1 therapies (Topalian et al., 2016), but these are usually not robust

enough to be uniformly applied as clinical biomarkers. The development of novel cancer immunotherapies and the identification of effective biomarkers require deep understanding of tumor-resident T cells. Transcriptome analyses of infiltrating regulatory T cells (or Tregs) from colon, lung, and breast cancers have revealed their highly suppressive nature (De Simone et al., 2016; Plitas et al., 2016). Although the clinical efficacy of various checkpoint inhibitors hinges at least in part upon their ability to unleash the CD8⁺ cytotoxic T cells in the tumor microenvironment, the cytotoxic activity of these T cells could be rendered ineffective primarily by the suppression of Tregs (Nishikawa and Sakaguchi, 2014) or by reaching a T cell dysfunction state called exhaustion, defined by poor effector function, sustained expression of inhibitory receptors, and a unique transcriptional state (Sen et al., 2016; Wherry and Kurachi, 2015). Therefore, a deep understanding of the mechanisms and pathways leading to the augmented CD8⁺ T cell exhaustion and the accumulation of Tregs in cancer will provide better strategies to orchestrate the immune system to eradicate cancers.

Single-cell analysis of infiltrating lymphocytes allows detailed understanding of these cells in the highly complex tumor microenvironment. Recently, single-cell transcriptome analysis has been applied to cancerous and immune cells from melanoma patients revealing T cell exhaustion signature and their connection to T cell activation (Tirosh et al., 2016). In addition, single-cell RNA sequencing (RNA-seq) provides a powerful tool to define T cell receptor (TCR) sequences for each cell, which are pivotal for recognizing viral antigens (Chisari and Ferrari, 1995) or tumor-specific neoantigens presented by the major histocompatibility complex (MHC) on tumor cells (Coulie et al., 2014). While the TCR repertoire is enormous due to the large amount of TCRs and random recombination, identical TCR sequences can illustrate T cell clonal expansion patterns and T cell lineages (Han et al., 2014).

Hepatocellular carcinoma (HCC) is one of the leading causes of cancer-related death worldwide (El-Serag, 2011) and chronic hepatitis B virus (HBV) infection is the major cause of HCC in China. Yet, there are limited treatment options and lack of clinical success in immunotherapies (Prieto et al., 2015). Although HCC

tumors harbor a significant level of TILs (Kasper et al., 2009), those TILs are by inference incapable of killing tumor cells. Thus, HCC tumors provide an appealing model to characterize the dysregulation of TILs. In this study, the single-cell RNA-seq analyses of >5,000 single T cells isolated from HCC patients enabled us to simultaneously study their complete TCR sequences and transcriptomes. We identified 11 unique T cell subsets with distinct tissue distribution patterns. Combined expression and TCR-based analyses revealed connectivity and potential developmental path of these subsets. Signature genes for exhausted CD8⁺ T cells and tumor-specific Tregs, such as *layilin* (*LAYN*), were examined in detail, and we found that over-expressing *LAYN* in primary CD8⁺ T cells resulted in inhibition of interferon (IFN)- γ production, suggesting a regulatory function of *LAYN*. This unprecedentedly large-scale transcriptome data of T cells can be used as a valuable resource for studying the basic characteristics of TILs and for potentially guiding effective immunotherapy strategies.

RESULTS

Tumor Characteristics and Single T Cell Transcriptome Data Generation

For six treatment-naive HCC patients, we first performed bulk exome- and RNA-sequencing (RNA-seq) using their tumor samples to obtain their basic properties (Figure 1A). Copy-number variation and somatic mutations of these tumors were consistent with known HCC genomic patterns (Figures S1A and S1B) (Kan et al., 2013). The presence of infiltrating lymphocytes in tumor tissues was confirmed using immunohistochemistry (IHC) with CD3, CD4, CD8, and FOXP3 antibodies (Figure S1C), consistent with previous reports (Kasper et al., 2009). We further performed Opal multicolor IHC staining (Figure 1B) to delineate the existence of different T cell subtypes. Notably, we observed that the number of CD8⁺ T cells, but not FOXP3⁺ regulatory T cells, was much lower in the core of tumors than in the outer cortex (Figure S1D), indicating the inefficiency of CD8⁺ T cell infiltration.

We then sorted CD3⁺CD4⁺ and CD3⁺CD8⁺ T cells from single-cell suspension prepared from various tissues (Figures 1A and S1E). In order to enrich Tregs, we further sorted CD4⁺ T cells into CD25⁻ and CD25^{high} populations based on the top 4% gate of blood CD4⁺ T cells (Wing et al., 2002). The percentage of CD4⁺CD25^{high} T cells within CD4⁺ T cells isolated from tumor tissues was much higher than those in normal tissues and peripheral blood (Figure S1F), suggesting potential enrichment of Tregs in the tumor microenvironment. CD8⁺ cytotoxic T cells from peripheral blood, adjacent normal, and tumor tissues were denoted as PTC, NTC, and TTC, respectively. The CD4⁺CD25⁻ T cells from these three locations were enriched for T helpers and denoted as PTH, NTH, and TTH, respectively, and the CD4⁺CD25^{high} T cells were similarly denoted as PTR, NTR, and TTR, respectively.

RNA-seq data were obtained for a total of 5,063 sorted individual cells (Figure 1C), at an average of 1.29 million uniquely mapped read pairs per cell (Table S1). A saturation analysis indicated that at such sequencing depth, we could reliably detect important classes of genes, including cytokines and transcription factors (Figure S2A), and there was no obvious 3' coverage bias for

samples prepared using the Smart-Seq2 method (Picelli et al., 2014) (Figure S2B). Excluding outlier cells (Figures S2C–S2H), we revealed distinct patterns of T cell clustering with the most variable genes (Figure 1D). The clustering of cells from the same location and cell type suggested that there were different properties and intrinsic structures for different groups of lymphocytes within a given patient (Figures S2I and S2J). In some cases, cells from different groups were mixed together (e.g., a subset of TTC cells clustered closely with NTC) indicating their similar functional properties. Meanwhile, PTC cells fell into two apparent clusters, consistent with the presence of both naive and activated CD8⁺ T cells in blood. In particular, TTH cells exhibited high degree of diversity with a small fraction resembling TTR cells. This complex intrinsic composition of T cells from HCC patients highlights the necessity of single-cell technologies for dissecting tumor-related T cells in detail.

T Cell Clustering and Subtype Analysis

To reveal the intrinsic structure and potential functional subtypes of the overall T cell populations, we performed unsupervised clustering of all T cells using the spectral clustering method implemented in SC3 (Kiselev et al., 2017). A total of 11 stable clusters emerged, including 5 clusters for CD8⁺ and 6 clusters for CD4⁺ cells, each with its unique signature genes (Figures 2A–2C; Table S2). Cells of the first CD8⁺ cluster, C1_CD8-LEF1, were dominant in peripheral blood (Figure 2B) and specifically expressed “naïve” marker genes such as *LEF1* and *CCR7* (Förster et al., 2008). The second cluster, C2_CD8-CX3CR1, was characterized by the high expression of the CX3CR1, *FCGR3A*, and *FGFBP2* genes (Figure 2C), commonly associated with T cells with effector functions (Böttcher et al., 2015). The third cluster, C3_CD8-SLC4A10, characterized by specific expression of *SLC4A10*, *ZBTB16*, and *RORC* (Figures 2C and S3A), was largely composed of mucosal-associated invariant T cells (MAIT), which were confirmed by predominantly bearing semi-invariant TCR alpha chains with TRAV1-2/TRAJ33, TRAV1-2/TRAJ20, or TRAV1-2/TRAJ12 (Figure S3B) (Kurioka et al., 2016). The fourth cluster, C4_CD8-LAYN, predominantly composed of cells from tumor tissues (Figure 2B), expressed high levels of exhaustion markers *CTLA4*, *PDCD1*, and *HAVCR2* (Figure 2C), thus representing exhausted CD8⁺ T cells (Chen and Flies, 2013). The remaining CD8⁺ cells, falling into the fifth cluster, C5_CD8-GZMK, shared a few common genes with cluster 4, such as *PDCD1* and *DUSP4*, albeit also with the *GZMK* expression signature that was absent in those exhausted cells (Figure 2C).

The general distribution patterns of CD8⁺ T clusters were comparable among patients (Figure 2D). Naive T cells with *CCR7* expression and effector memory T cells with CX3CR1 were prevalent in peripheral blood, while MAIT cells were prevalent in non-tumor adjacent liver tissue. The percentages of exhausted CD8⁺ T cell were increased significantly in tumor samples (Figure 2E), in agreement with previous findings (Jiang et al., 2015). Meanwhile, MAIT cell fractions were significantly reduced in HCC tumors, compared with adjacent normal tissues. We confirmed this finding in the TCGA cohort, as the expression of the MAIT marker gene *SLC4A10* in HCC tumors was significantly lower ($p = 0.003$, paired Wilcoxon test) than paired normal samples,

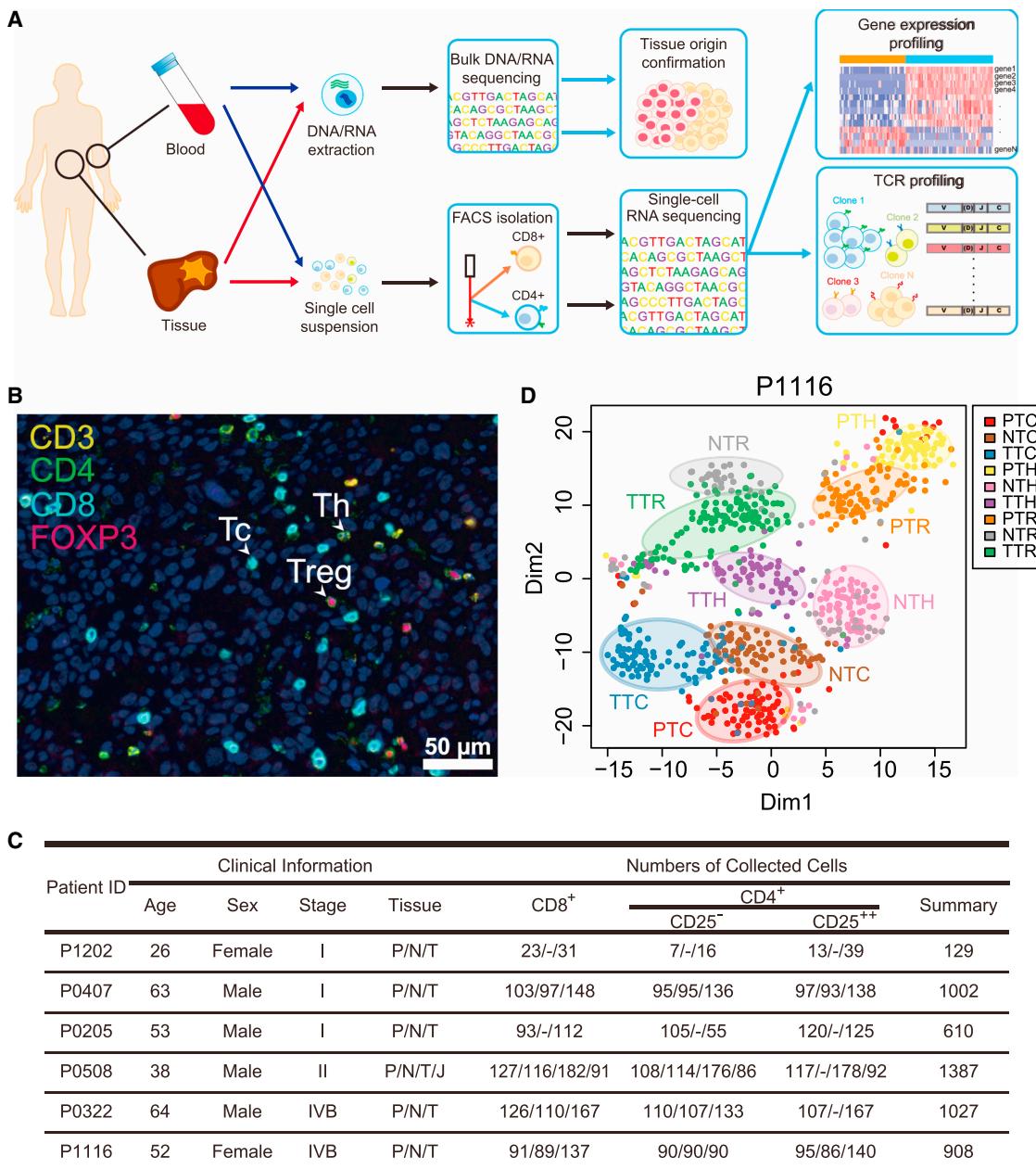


Figure 1. Dissection and Clustering of Tumor-Infiltrating T Cells in HCC

(A) Overview of the study design.

(B) Opal multicolor IHC staining with anti-CD3, CD4, CD8, and FOXP3 antibodies. Tc, CD8⁺ T cells; Th, T helper cells; Treg, regulatory T cells.

(C) The HCC patient information and numbers of T cells sequenced. The tissue types P/N/T/J represent T cells isolated from blood, adjacent normal, tumor tissues, and the joint area between the tumor and adjacent normal tissues.

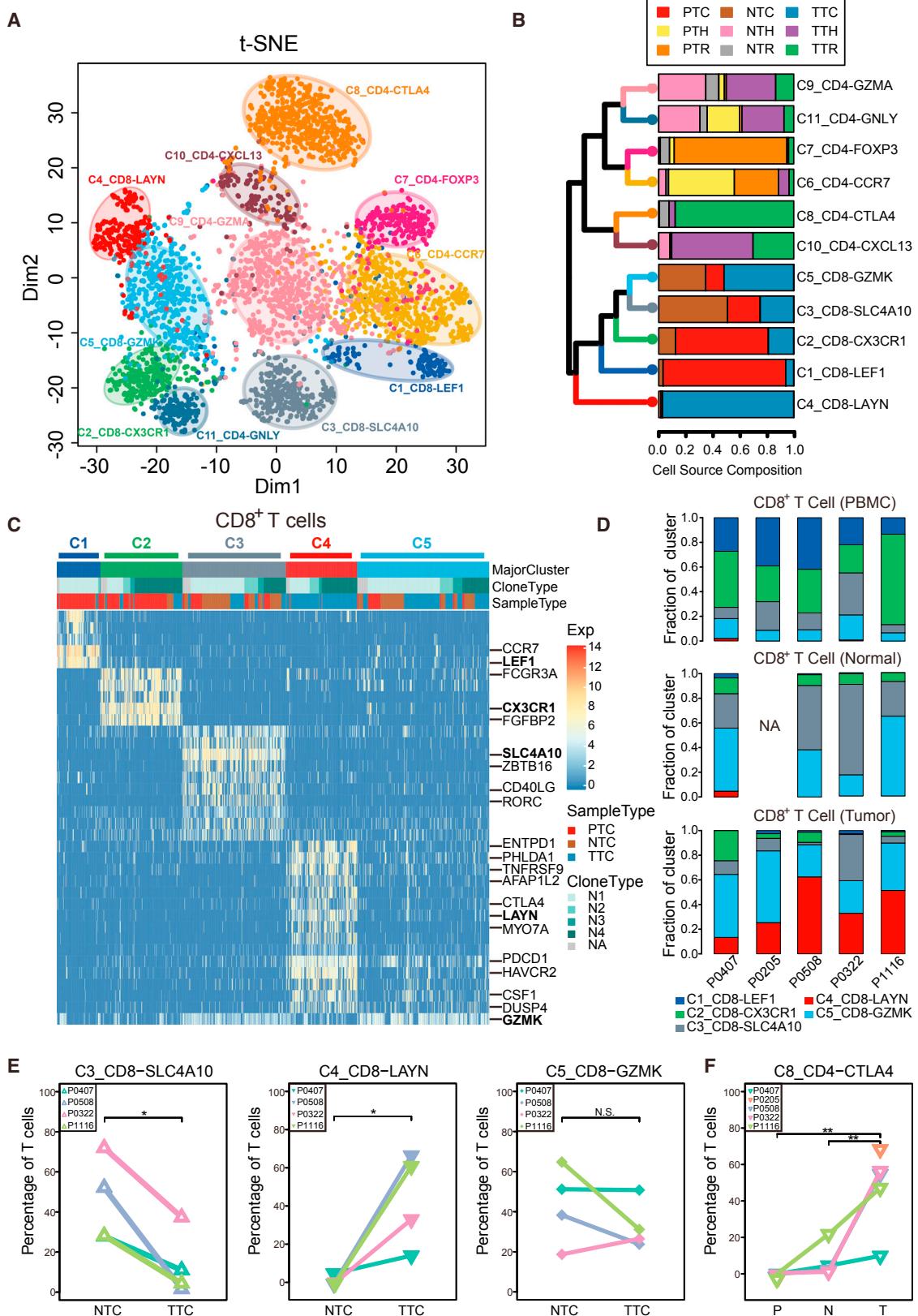
(D) 2D visualization of single-cell clusters in patient P1116 by t-SNE. Each dot corresponds to one single cell, colored according to cell type and location. PTC, NTC, TTC: CD8⁺ T cells from peripheral blood, adjacent normal, and tumor tissues, respectively; PTH, NTR, TTH: CD4⁺CD25⁻ T cells from these three locations; PTR, NTR, TTR: CD4⁺CD25^{high} T cells from the same three locations.

See also Figures S1 and S2 and Table S1.

confirming that the MAIT cell reduction is common in liver cancer (Figure S3C). In addition, lower *SLC4A10* expression in HCC correlates with poor prognosis in the TCGA cohort ($p = 0.037$, log rank test; Figure S3D). Thus, our results provide evidence for

the differential distribution of CD8⁺ T cells as a feature in the tumor microenvironment of HCC.

Similarly, we identified six major CD4⁺ T cell clusters (Figure S4A). C6_CD4-CCR7 cells, mostly derived from blood,



(legend on next page)

comprised CD4⁺ T cells with high expression of naive marker genes, such as *SELL*, *TCF7*, and *CCR7*. Both C7_CD4-FOXP3 and C8_CD4-CTLA4 were FOXP3⁺ Tregs, but C8_CD4-CTLA4 had higher *CTLA4* expression (Figure S4B). While C7_CD4-FOXP3 mainly comprised of sorted CD4⁺CD25^{high} cells from peripheral blood (PTR), C8_CD4-CTLA4 cells were primarily sorted CD4⁺CD25^{high} cells from tumors (TTR). Although we used CD4⁺CD25^{high} sorting to enrich Tregs, some of the sorted cells fell into other cell types based on detailed transcriptome analysis (Figure 2B). Thus, in the subsequent analyses, we only relied on Tregs that were digitally defined by the Treg signature genes expression. For example, in the C8_CD4-CTLA4, in addition to the expression of *FOXP3*, other well-defined Treg genes such as *TNFRSF9*, *TIGIT*, and *CTLA4*, were also present (Bhairavabhotla et al., 2016). These cells were preferentially enriched in tumors (Figure 2F). Cells from the remaining clusters were primarily T helper cells from normal and tumor tissues and were characterized by the high expression of *GZMA* and *CCL5*. Among these cells, C10_CD4-CXCL13 specifically expressed CXCL13, *PDCD1*, *CTLA4*, and *TIGIT*, suggestive of the identity of exhausted CD4 T cells (Crawford et al., 2014). C11_CD4-GNLY exhibited high expression of cytotoxic molecules, including *NKG7*, *GNLY*, and *GZMB*, indicative of the status of cytotoxic CD4 T cells (Marshall and Swain, 2011).

Identification and Confirmation of Genes Uniquely Associated with HCC-Infiltrating Tregs and Exhausted CD8⁺ T Cells

Our data demonstrated the preferential enrichment of CTLA4^{high} Tregs and exhausted CD8 cells in HCC (Figures 2E and 2F). Because co-inhibitory receptors, such as *PDCD1* and *TIGIT*, expressed by these two T cell subsets are targets for cancer immunotherapies, we focused our further analyses on these cell types. We identified a total of 401 genes (adjusted p value < 0.01, limma moderated t test, and fold change ≥ 2) specifically expressed in tumor-infiltrating Tregs (Table S3), including *FOXP3*, *CTLA4*, *TNFRSF18*, *TNFRSF4*, and *CCR8*. These HCC Treg-specific genes significantly overlapped with those identified by three previous studies in melanoma, breast, colon, and lung cancers (p value < 2e-85, hypergeometric test) (Figure 3A) (De Simone et al., 2016; Plitas et al., 2016; Tirosh et al., 2016). Among the 33 common Treg signature genes from these 3 studies, 31 genes were also identified in our data (Figures 3B, 3C, and S4C). Two missed genes from the shared gene list, *CRADD* and *IL32*, also

showed significantly higher expression (adjusted p value < 6e-12, limma moderated t test) in tumor-specific infiltrating Tregs, albeit with slightly lower log₂FC values (0.99 and 0.96, just under the cutoff of 1.0). Therefore, our single-cell strategy based on relatively limited patient samples could robustly provide detailed portraits of these tumor-infiltrating lymphocytes. Our study also unveiled 146 genes that uniquely enriched in our dataset (Figures 3D and S4C), including *RGS1*, *STAT3*, and *PTPN22* that were previously reported to be associated with Tregs or regulate their functions (Fousteri et al., 2014; Goode et al., 2013; Pallandre et al., 2007).

Next, we analyzed tumor-infiltrating exhausted CD8⁺ T cells and obtained a list of 82 exhaustion-specific genes by comparing exhausted and non-exhausted TTC cells (Figure 3E; Table S4). Multiple known exhaustion markers, such as *HAVCR2*, *PDCD1*, *ENTPD1*, *CTLA4*, *TIGIT*, *TNFRSF9*, and *CD27*, were among the top-ranked genes. When evaluating the exhaustion status with such exhaustion-specific gene signature, we found that patients at the late stage (P0322 and P1116) exhibited higher exhaustion level than other patients (p = 0.02, Student's t test) (Figure S5A). The 82-gene list also contained several less-described genes, such as *MYO7A*, *WARS*, and *CXCL13* (Tirosh et al., 2016), as well as novel exhaustion makers such as *LAYN*, *PHLDA1*, and *SNAP47* (Figure S5B). Notably, 22 of these exhaustion-related genes were also highly expressed by tumor-infiltrating Tregs (Figure 3F), including *CTLA4* and *TIGIT*. Based on the TCGA HCC data, higher expression of *PHLDA1* and *SNAP47* were both associated with poor prognosis, while such a trend was not observed for the Treg-associated *CCR8* gene (Figure S5C) despite a recent report showing its association with poor prognosis in colon and lung cancer (De Simone et al., 2016). Thus, our data not only confirmed previously identified genes associated with tumor-infiltrating exhausted CD8⁺ T cells and Tregs, but also discovered additional markers for these cell types.

LAYN Is Induced in Blood-Isolated CD8⁺ T Cells and Tregs upon Activation

LAYN, encoding layilin, was recently reported to be highly expressed in Tregs isolated from lung and colon tumors (De Simone et al., 2016) but it is under-characterized functionally. Based on our data, *LAYN* exhibited a highly specific expression pattern in both tumor Tregs (C8_CD4-CTLA4) and exhausted TTC cells (C4_CD8-LAYN) (Figure 4A). In addition, by

Figure 2. T Cell Subtype Analysis Based on Single-Cell Gene Expression

(A) The t-SNE projection of T cells from HBV⁺ patients, showing the formation of 11 main clusters shown in different colors. The functional description of each cluster is determined by the gene expression characteristics of each cluster. Cluster C1_CD8-LEF1: naive CD8⁺ T cells; C2_CD8-CX3CR1: effector memory CD8⁺ T cells; C3_CD8-SLC4A10: MAIT cells; C4_CD8-LAYN: exhausted CD8⁺ T cells; C6_CD4-CCR7: naive CD4⁺ T cells; C7_CD4-FOXP3: peripheral Tregs; C8_CD4-CTLA4: tumor Tregs; C10_CD4-CXCL13: exhausted CD4⁺ T cells; C11_CD4-GNLY: cytotoxic CD4 cells.

(B) Unsupervised hierarchical clustering of the 11 clusters based on the average gene expression of cells in each cluster. The proportions of different cell sources are shown.

(C) Heatmap of five CD8⁺ T cell clusters with unique signature genes. Information of clonal status and tissue-of-origin are colored for each cell. Selective specifically expressed genes are marked alongside. Clone type N1/N2/N3/N4 indicates that the TCR sequence was observed 1/2/3/ ≥ 4 times across the CD8⁺ T cells, respectively. NA indicates that no dominant TCR alpha/beta was observed in the cell.

(D) The fractions of five clusters defined in CD8⁺ T cells in each patient across peripheral blood, adjacent normal, and tumor tissues.

(E) The percentages of cells in different clusters in adjacent normal and tumor tissues. *p < 0.05, N.S., not significant, Student's t test.

(F) The percentages of tumor Tregs in C8_CD4-CTLA4 cluster in peripheral blood, adjacent normal, and tumor tissues. **p < 0.01, Student's t test.

See also Figures S3 and S4 and Table S2.

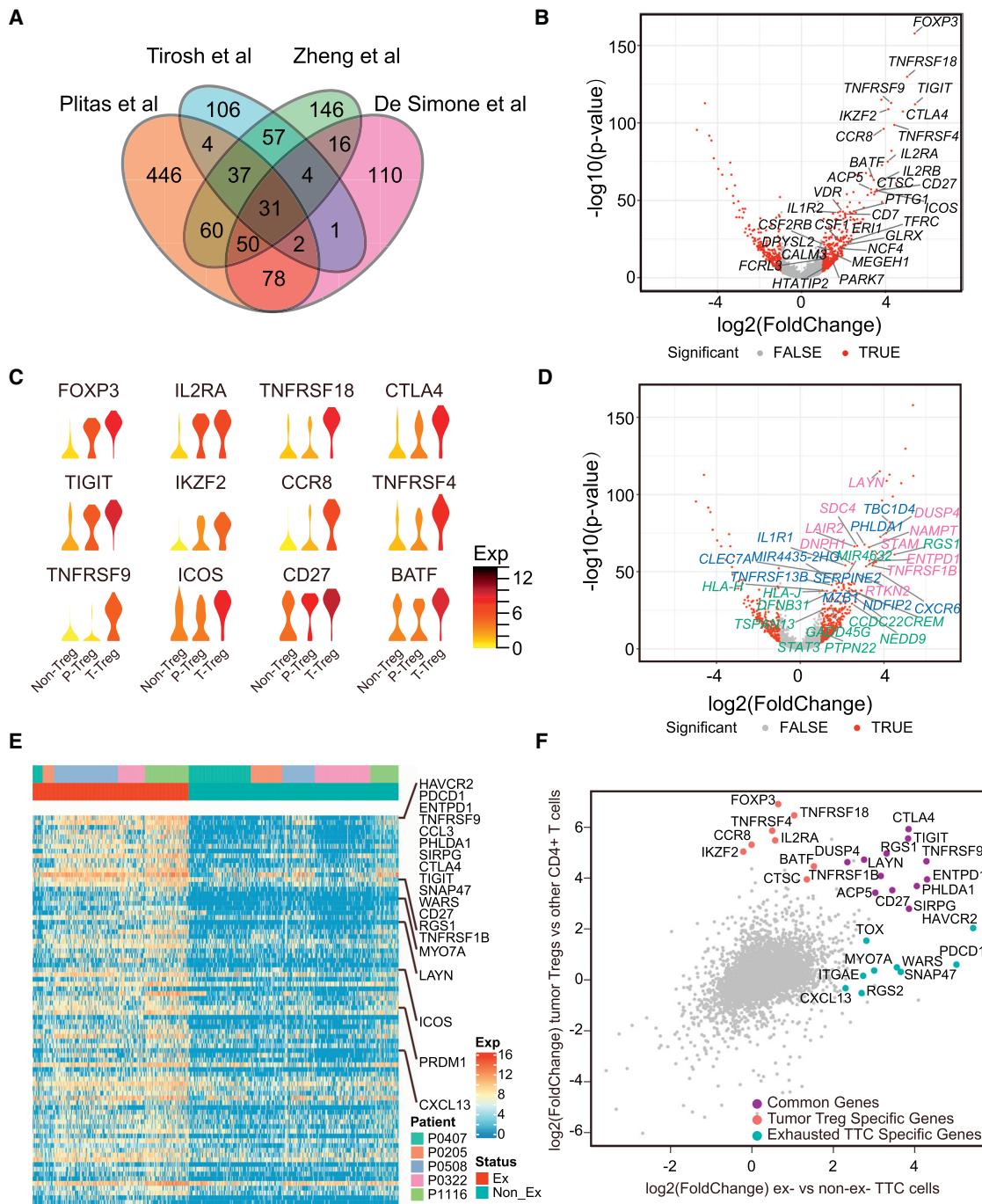


Figure 3. Gene Expression Characterization Associated with HCC-Infiltrating Tregs and Exhausted CD8⁺ T Cells

- (A) The Venn graph showing the overlap of tumor Treg genes identified in this study with those from previous studies by De Simone et al. (2016) ($p = 2e-85$), Plitas et al. (2016) ($p = 6e-129$), and Tirosh et al. (2016) ($p = 2.1e-141$), determined by hypergeometric test.
- (B) Volcano plot showing differentially expressed genes in tumor-infiltrating Tregs. Each red dot denotes an individual gene passing our p value and fold difference thresholds. Only those 31 genes found by all previous three studies are marked with gene names.
- (C) Violin plots showing the expression profile of those genes recurrently identified in HCC data and previous studies across peripheral blood Tregs (C7_CD4-FOXP3), tumor-infiltrating Tregs (C8_CD4-CTLA4), and non-Tregs (other CD4 clusters). The expression is measured as the $\log_2(\text{TPM}+1)$.
- (D) The same volcano plot as in (B) but with those unique genes found by this study labeled in green. Those identified only once (blue) and twice (purple) previously are also marked.
- (E) Heatmap showing differentially expressed genes between exhausted and non-exhausted TTCs. The top bar indicates patient origins. The lower bar indicates the exhaustion states. Selected known T cell exhaustion markers are denoted to the right.

(legend continued on next page)

interrogating the TCGA HCC survival data with respect to *LAYN* expression after normalizing the effect of T cell infiltration levels by total CD3 expression, we found that higher expression of *LAYN* was associated with reduced disease-free survival rate ($p = 0.002$, log rank test, Figure 4B). Because *LAYN* encodes a cell surface protein, we validated its expression at the protein level on both CD8⁺ and Tregs isolated from human peripheral blood mononuclear cells (PBMCs) by fluorescence-activated cell sorting (FACS) analysis. At the resting stage, *LAYN* was not expressed on CD14⁺ myeloid cells, B cells, CD4⁺ T cells, Tregs, or CD8⁺ T cells (Figure S5D). However, after T cells were activated by anti-CD3 and anti-CD28 antibodies for 2 days, we readily detected *LAYN* presence on >30% Tregs and CD8⁺ T cells, although on CD4⁺ T cells *LAYN* was only minimally upregulated (Figure 4C). These data were consistent with the observed upregulation of *LAYN* mRNA in both tumor-infiltrating CD8⁺ and Tregs. Although FOXP3 is the master transcription factor for Treg differentiation and function, in human T cells, FOXP3 can also be transiently upregulated in conventional CD4⁺ T cells (Wang et al., 2007). Another transcription factor, Helios, was recently reported to stabilize FOXP3 expression and the repressive function of Tregs (Kim et al., 2015). Therefore, we evaluated *LAYN* expression in either FOXP3⁺ or FOXP3⁺Helios⁺ Tregs. Notably, *LAYN* was preferentially upregulated in FOXP3⁺ Helios⁺ double-positive Tregs (Figure 4D) suggesting that *LAYN* expression was associated with more repressive and stable Tregs.

Because the association of *LAYN* with tumor-infiltrating exhausted CD8⁺ T cells in HCC was not previously reported, we further characterized its expression and regulation in CD8⁺ T cells in vitro. The *LAYN* protein was induced on CD8 T cells 1 day after T cell activation and its presence could still be detected 6 days after activation (Figure 4E). *LAYN* could be detected on both PD-1⁺ and PD-1⁻ CD8⁺ T cells under this in vitro stimulation condition. Notably, *LAYN* was exclusively induced on LAG-3 negative CD8⁺ T cells (Figure S5E). The LAG3⁺CD8⁺ T cells have been suggested to represent a unique subset of exhausted CD8⁺ T cells (Grosso et al., 2009). The mutually exclusive expression pattern of *LAYN* and LAG3 suggests that *LAYN* might serve as a marker for a different subset of CD8⁺ T cells. To better understand the potential function of *LAYN* on CD8⁺ T cells, we used retroviral-mediated overexpression of *LAYN* in these cells to mimic the high expression of *LAYN* on tumor-infiltrating CD8⁺ T cells (Rutz et al., 2011). Increased *LAYN* expression on CD8⁺ T cells was observed on day 3 after retroviral infection (Figure 4F). Interestingly, the *LAYN*-overexpressed CD8⁺ T cells produced significantly less IFN- γ than that by control virus-infected cells (Figure 4G), supporting an inhibitory role of *LAYN* in CD8⁺ T cells. In summary, these data corroborated with the transcriptome data that *LAYN* protein could be induced on both Tregs and CD8 T cells and provided insight into its potential functions.

Enrichment of Clonal CD8 T Cells and Tregs in HCC Revealed by TCR Identification

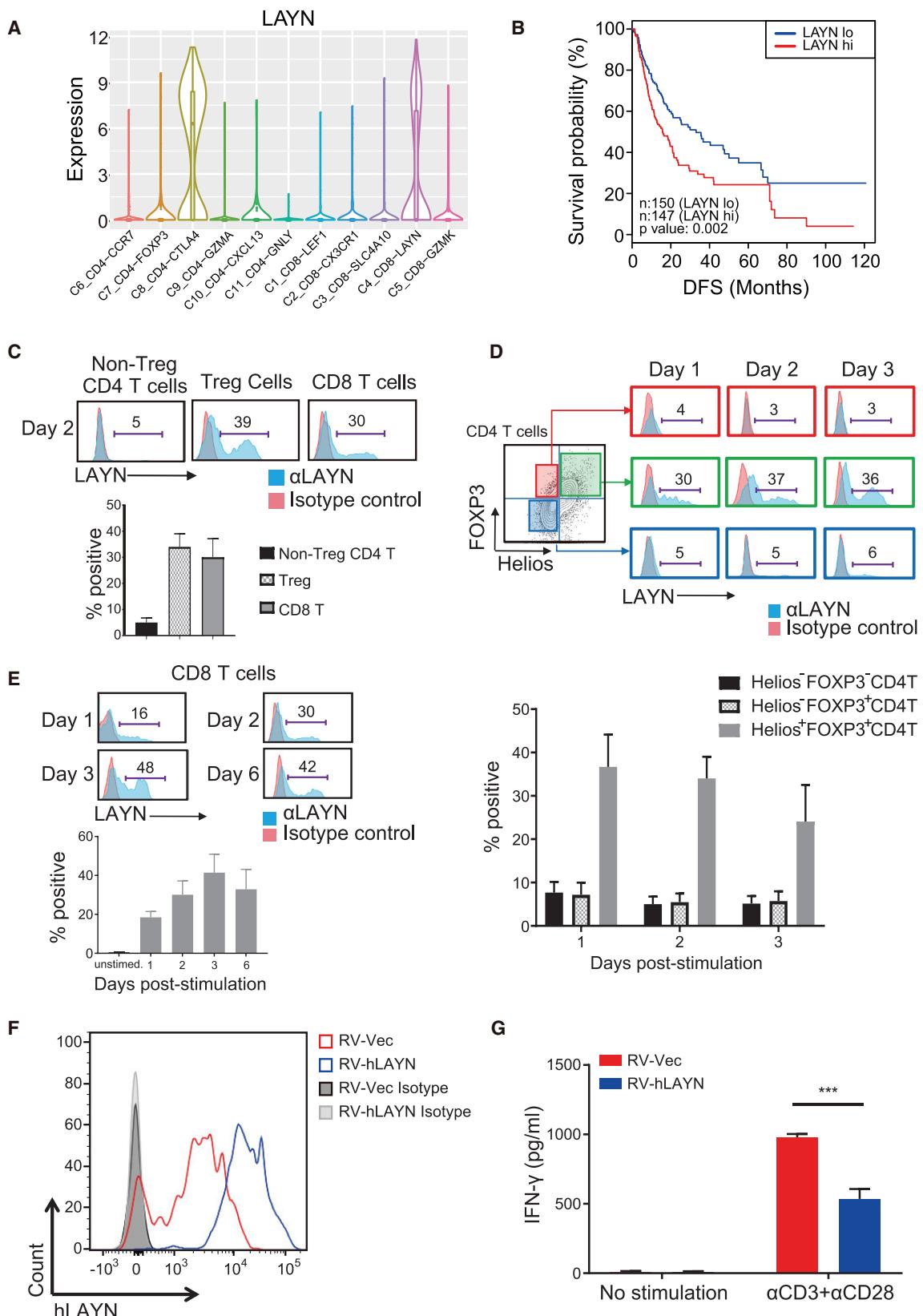
Because TCRs are often used as unique identifiers of T cell ancestries (Han et al., 2014), we took advantage of our single-cell RNA-seq data to track the lineage of each single T cell based on their full-length TCR α and β sequences assembled by the TraCeR method (Stubbington et al., 2016) for cells from five HBV-positive HCC patients. In total, we detected full TCR sequences for 94% (3,792/4,032) T cells, with at least one paired productive TCR α - β chain for subsequent analyses (Table S5). While most cells expressed unique TCR α and β alleles, non-unique α and/or β could be detected in a fraction of T cells. After eliminating non-productive alleles (e.g., out-of-frame transcripts) or low-abundance TCRs (Figure S6A), we found that 84% (3,174/3,792) contained unique and productive α chains and 94% (3,559/3,792) unique and productive β chains (Figure S6B), in agreement with previous reports (Stubbington et al., 2016). Therefore, to adequately define T cell clonality, we strictly defined T cells with at least one pair of identical paired α - β chains to be one clone from the same ancestry, and the expanded clones were defined as those whose α and β TCR pairs were shared by at least three cells in a given cell population. A strong correlation was observed between the recurring frequency of various TCR α chains and that of β chains among T cells, suggesting a common ancestral cell of origin (Figure S6C). We further investigated TCR rearrangements to identify V-D-J usage preference and found the expected TCR segment usage bias (Figures S6D and S6E) (Freeman et al., 2009).

The available TCR sequences for each cell provide an opportunity to study the relationship among different T cells. While most cells contained unique TCRs, patterns of repeat uses were observed at varying degrees in different cell types, especially in tumor-infiltrating CD8⁺ T cells and Tregs (Figure 5A and 5B). While only ~10% CD8⁺ T cells harbored clonal TCRs in blood and normal liver tissues, tumor tissues showed a much higher percentage at ~30% ($p < 0.05$, Student's t test) (Figures 5C and S6F). The number of non-unique TCR clones in tumor tissues was 2-fold higher than that in blood and adjacent normal tissues, or 4- to 5-fold higher when only considering clonal TCRs ($p < 0.01$, Student's t test) (Figure S6G).

Similar patterns were identified in tumor-infiltrating Tregs (Figure 5B). While PTR cells showed little clonal enrichment, much more TTR cells were clonally enriched (Figures 5D and S6H), especially for cluster C8_CD4-CTLA4. By contrast, similar clonal enrichment of tumor-infiltrating T helper cells were not observed (Figure 5E), with only patient P0508 showing a high proportion of clonal TCRs in TTH cells (Figure S6I).

Given the low clonal enrichments of T cells in blood and normal tissues, it was likely that the clonal accumulation of both Tregs and CD8 T cells was a result of local T cell proliferation and activation in the tumor environment, as suggested by previous studies (Ferradini et al., 1993; Tanchot et al., 2013). However, we could not formally rule out the possibility that these clonal

(F) Scatterplot showing genes with high expression in tumor Tregs (C8_CD4-CTLA4) and/or exhausted tumor CD8⁺ T cells (C4_CD8-LAYN). The x axis represents the relative expression of a given gene between exhausted and non-exhausted tumor CD8⁺ T cells, and the y axis represents the relative expression of a given gene between tumor Tregs and the remaining CD4⁺ T cells. Each dot represents a gene, with color annotation inside.
See also Figures S4 and S5 and Tables S3 and S4.



(legend on next page)

T cells were amplified in the peripheral lymph organs and then migrated into tumors. Because these patients were HBV-positive, clonal T cells could also be a result of T cell activation in response to viral antigens during chronic infection. However, in this case, one would expect similar clonal enrichment in the adjacent HBV⁺ normal tissues, which was not observed in our study (Figure 5C). Altogether, our data enabled us to link TCR information to various states of T cells and supported the potential clonal expansion of both CD8⁺ T cells and Tregs in HCC.

Mapping Subset Connectivity by Pseudotime State Transition and Clonal TCRs

The complete transcriptome data along with TCR information for a large number of T cells allowed us to gain insights into the functional states of and relationship among these cells. We applied the Monocle 2 algorithm (Trapnell et al., 2014) to order CD8⁺ T cells or CD4⁺ T helper cells in pseudotime to indicate their developmental trajectories (Figures 6A and 6B). Such pseudotime analysis is a measure of progress through biological processes based on transcriptional similarities. We also applied other algorithms including EMBEDDR (Campbell et al., 2015), SCORPIUS (Cannoodt et al., 2016), and TSCAN (Ji and Ji, 2016) to the same dataset and obtained comparable results (Figure S7A).

For CD8⁺ T cell analysis, we removed those MAIT cells (C3_CD8-SLC4A10) due to their different TCR characteristics. Most cells from each cluster aggregated based on expression similarities, and different clusters formed into a relative process in pseudotime that began with the C1_CD8-LEF1 cells (naïve CD8⁺ T cells), followed by C2_CD8-CX3CR1 (effector memory CD8⁺ T cells), C5_CD8-GZMK, and ended with C4_CD8-LAYN cells (exhausted CD8⁺ T cells) (Figure 6A). Thus, those exhausted T cells were highly enriched at the late period of the pseudotime, demonstrating the T cell state transition from activation to exhaustion.

C5_CD8-GZMK appeared to be an intermediate state between effector and exhausted T cells based on such trajectory analysis. This finding was also independently validated by the clonal analysis based on those identical TCRs from common ancestors but falling into different cell clusters (Figure 6C). Of the total of 61 TCR pairs shared by cells from at least two different clusters, 30 were shared between C2_CD8-CX3CR1 and C5_CD8-GZMK and 20 were shared between C5_CD8-GZMK and C4_CD8-LAYN. By contrast, no more than five TCR

pairs were shared between other clusters, supporting the development connections between the transitional C5_CD8-GZMK and effector cells or exhausted cells.

Similarly, we analyzed CD4⁺ T helper cells to determine the differentiation trajectory. Based on the pseudotime result (Figure 6B), the C6_CD4-CCR7 naïve T cells and C9_CD4-GZMA T helper cells primarily aggregated on the pseudotime backbone. The C10_CD4-CXCL13 (exhausted CD4⁺ T cells) and C11_CD4-GNLY (cytotoxic CD4⁺ T cells) cells were located in different directions in the pseudotime trajectory plot, indicating functional divergence of these cells. We again examined shared TCR sequences among these groups of cells and found that most shared TCRs occurred between C9_CD4-GZMA and C10_CD4-CXCL13 or between C9_CD4-GZMA and C11_CD4-GNLY (Figure 6D). Almost no TCRs were shared between C10_CD4-CXCL13 and C11_CD4-GNLY indicating their exclusive developmental fate and supporting the conclusion from the pseudotime analyses. Based on both the trajectory and TCR analyses, we inferred that exhausted CD4 and CD8 T cells were more closely linked to intermediate populations marked by GZMA and GZMK signatures, respectively, than to the effector populations. It remains to be seen if therapeutic strategies targeting these intermediate populations could prevent their transition to exhausted cells and instead promote them to effector cells.

Preferential and Clonal Enrichment of Exhausted CD8⁺ T Cells and Tregs in HCC Microenvironment

It has been recognized previously that T cells within the tumor microenvironment are prone to either exhaustion or Treg suppression, thus preventing such CD8 cells from eliciting sufficient T cell-mediated killing of tumor cells (Wherry and Kurachi, 2015). Focusing our analysis on tumor-infiltrating T cells in HCC, we observed much fewer effector CD8⁺ T cells in the tumor environment but more pronounced exhaustion phenotypes (Figures 2D and 7A). We also noticed a trend of increased exhausted CD8⁺ T cells in HCC from late stage patients compared with early stages ($p = 0.03$, Student's t test) (Figure 7B). Furthermore, clonal CD8⁺ T cells were more likely to exhibit the exhaustion phenotypes, especially among late-stage HCC tumors (Figure 7C), and high clonal population (greater than four cells per clone) was much more likely to be exhausted than non-clonal population ($p < 0.01$, Student's t test) (Figure 7D) in HCC.

Figure 4. LAYN Expression Is Induced on Activated Human Treg and CD8⁺ T Cells

- (A) Violin plots comparing the expression of LAYN in 11 T cell clusters.
- (B) The disease-free survival curve based on TCGA HCC data showing patients with higher expression of LAYN in tumor had poor prognosis. $p = 0.002$, log rank tests, hazard ratio: 1.62 (1.19–2.21, 95% confidence interval [CI]).
- (C) Human PBMCs were stimulated in vitro with anti-CD3 and anti-CD28 monoclonal antibodies (mAbs) conjugated microbeads for 2 days. Gated on each indicated population (non-Treg CD4 T cells: CD3⁺CD4⁺Helios⁻Foxp3⁻; Tregs: CD3⁺CD4⁺Helios⁺Foxp3⁺; CD8 T cells: CD3⁺CD8⁺), LAYN expression was determined by FACS. From (C) to (E), data represent mean \pm SEM $n \geq 3$.
- (D) Human PBMCs were similarly stimulated as in (C). Gated on each indicated CD4 T cell sub-populations (Helios⁻Foxp3⁻CD4 T: CD3⁺CD4⁺Helios⁻Foxp3⁻; Helios⁻Foxp3⁺CD4 T: CD3⁺CD4⁺Helios⁻Foxp3⁺; Helios⁺Foxp3⁺CD4 T: CD3⁺CD4⁺Helios⁺Foxp3⁺), the LAYN presence was determined by FACS.
- (E) Human PBMCs were similarly stimulated as in (C) for the indicated time. Gated on CD8 T cells, the LAYN presence was determined by FACS.
- (F) Flow cytometry plots showing human LAYN expression on CD8⁺GFP⁺ T cells.
- (G) ELISA of human IFN- γ in the cell supernatant of sorted control and LAYN-infected CD8⁺GFP⁺ T cells after 24 hr re-stimulation. * $p < 0.05$, Student's t test. Data represent mean \pm SEM $n \geq 2$.

See also Figure S5.

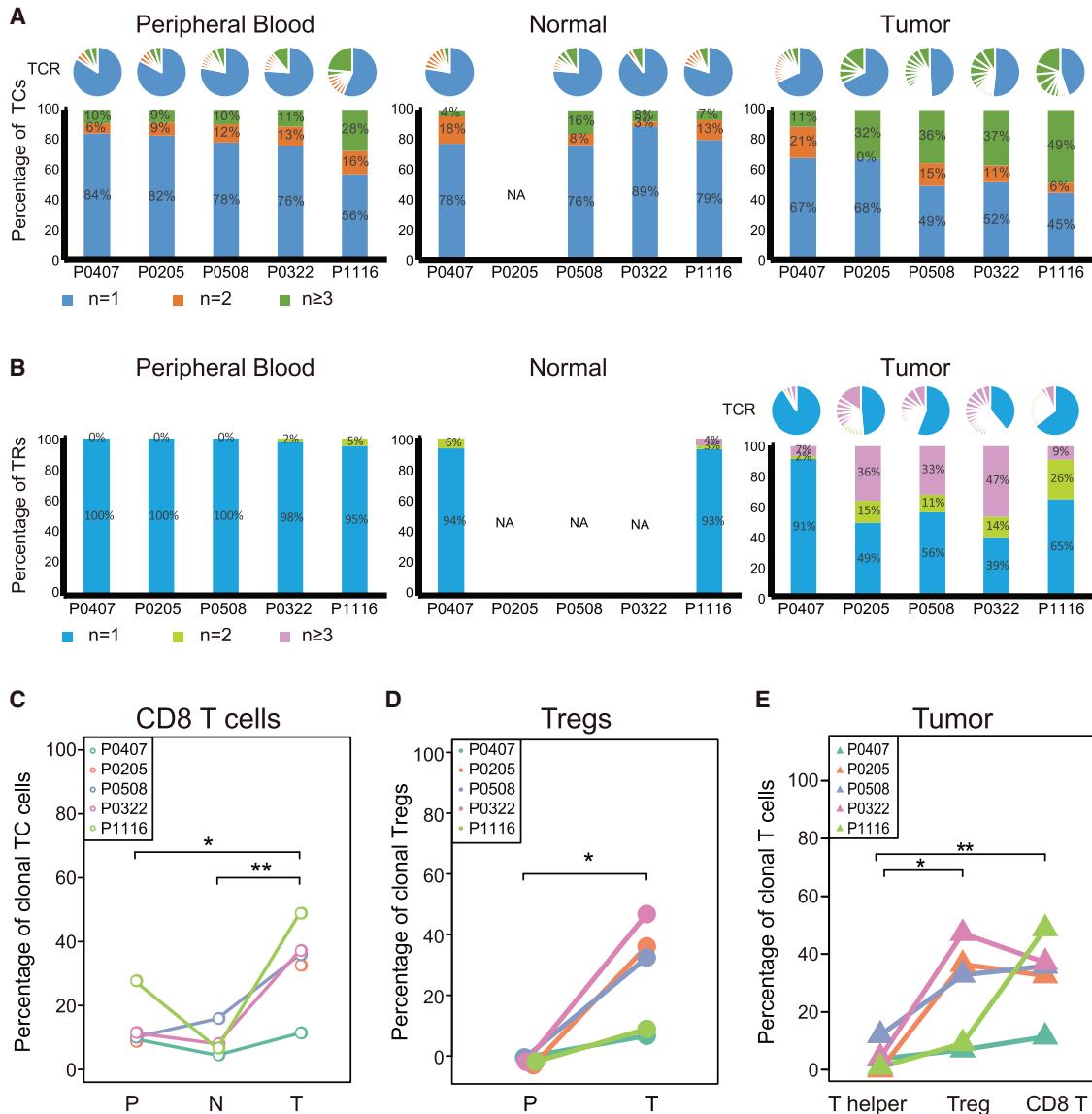


Figure 5. T Cell Receptor Distribution and Clonality Analysis

(A) The TCR distribution of CD8⁺ T cells across different tissues. Unique (n = 1), duplicated (n = 2), and clonal (n ≥ 3) TCRs are labeled with different colors. Pie charts above each bar illustrate the composition of every individual TCR. NA indicates that no such cell type was analyzed for the given patient.

(B) The same TCR distribution plot as in (A) for Tregs.

(C) The proportions of clonal CD8⁺ T cells in tumors (T) are higher than peripheral blood (P) and adjacent normal tissues (N). **p < 0.01; *p < 0.05, Student's t test.

(D) The proportions of clonal Tregs in tumor tissue (T) are higher than peripheral blood (P). *p < 0.05.

(E) Within HCC tumors, the proportions of clonal tumor-infiltrating CD8⁺ T and Tregs are higher than those of T helpers. **p < 0.01; *p < 0.05.

See also Figure S6 and Table S5.

In addition to exhausted CD8⁺ T cells, a previous study also identified CD8⁺FOXP3⁺ regulatory T cells in colorectal and prostate cancers (Chaput et al., 2009). The scale of our data allowed us to examine whether these cells also existed in HCC. Indeed, there was a small number of exhausted CD8⁺ T cells expressing the Treg marker FOXP3. Some of these FOXP3⁺ cells shared the same TCRs with other FOXP3⁻-exhausted CD8⁺ T cells based on TCR analysis (Figure 7E), suggesting those FOXP3⁺ cells were developmentally connected with typical FOXP3⁻-ex-

hausted CD8⁺ T cells. We further confirmed the presence of these CD8⁺FOXP3⁺ regulatory T cells in HCC by multi-color IHC (Figure 7F). These cells shared special expression properties resembling both Treg and cytolytic characteristics (Figure 7G) with expression of Treg hallmarks, such as FOXP3, CTLA4, TNFRSF18, and TNFRSF9 and cytolytic-related genes PRF1, GZMA, and NKG7 indicating the presence of both suppressive and cell-killing characteristics. Thus, HCC microenvironment promoted the transition of infiltrating CD8⁺ T cells into exhausted

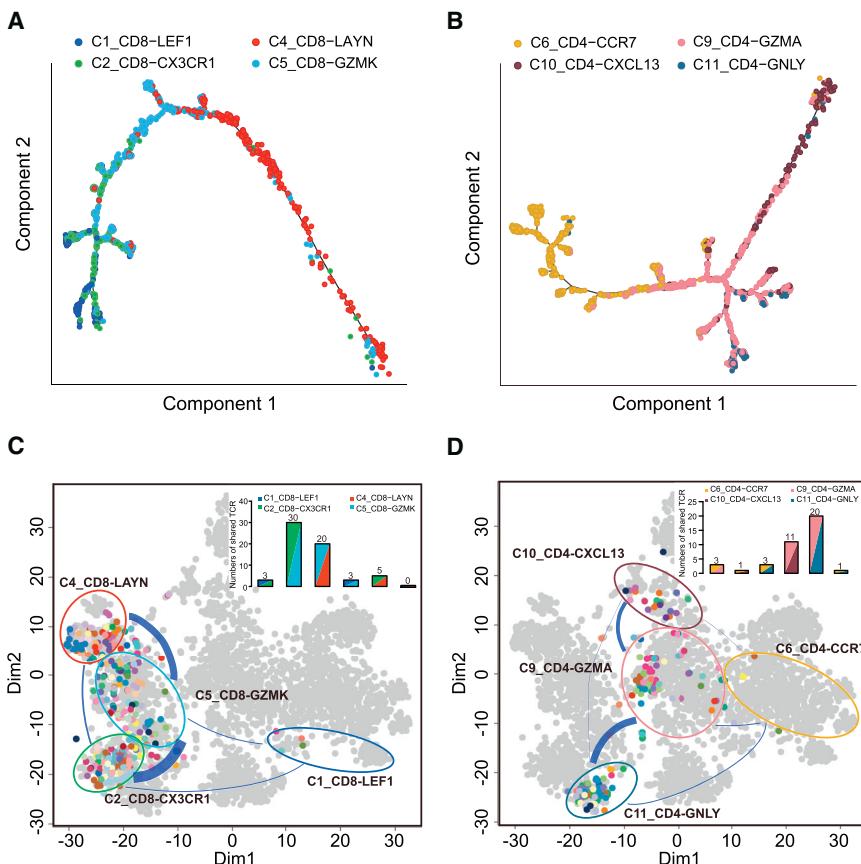


Figure 6. The CD8⁺ and CD4⁺ T Cell State Transition Analysis Based on Integrated Expression and TCR Clonality

(A) The ordering of CD8⁺ T cells along pseudotime in a two-dimensional state-space defined by Monocle2. Cell orders are inferred from the expression of most dispersed genes across CD8⁺ T cell populations sans MAIT. Each point corresponds to a single cell, and each color represents a T cell cluster.

(B) The same pseudotime plot as in (A) for four clusters of CD4⁺ T helper cells.

(C) Cell state transition of CD8⁺ T cell clusters inferred by shared TCRs. The space distribution of cells is defined by Figure 2A, with only those cells sharing TCRs across different CD8⁺ T cell clusters shown in colors. Each color represents a distinct TCR clonotype. Lines connecting different clusters are based on the degree to TCR sharing, with the thickness of lines representing the number of shared TCRs. The insert picture shows the numbers of TCRs shared by two clusters. The two colors in every bar represent the particular two clusters sharing TCRs.

(D) Similar to (C), the cell states transition of CD4⁺ T helper cell clusters inferred by shared TCRs. See also Figure S7.

DISCUSSION

The deep transcriptome data along with the complete TCR information for >5,000 individual T cells provided a comprehensive resource for understanding the multi-dimensional characterization of T cells, especially TILs. The higher resolution provided by our dataset was exemplified by the identification of 11 large subsets as well as unique subpopulations such as CD8⁺FOXP3⁺ regulatory-like cells. The high quantity and quality of single-cell data allowed us to map their developmental trajectory. In addition, the ability to identify clonal TCRs at single-cell level permitted us not only to confirm their developmental relationships in various clusters, but also to deduce their activation status in the cancer microenvironment.

The status of T cell infiltration and their characteristics are usually associated with different prognostic outcomes (Sharma and Allison, 2015). In our study, we identified CCR8 and LAYN as HCC-associated Treg marker genes, although LAYN, but not CCR8, is associated with tumor-infiltrating exhausted CD8⁺ T cells and poor prognosis. Previous study with small interfering RNA (siRNA)-mediated knockdown in vitro revealed a regulatory role of LAYN in Treg function (Bhairavabhotla et al., 2016). We demonstrated its induction in both CD8⁺ T cells and Tregs upon activation. In addition, overexpression of LAYN on CD8⁺ T cells from human blood resulted in significantly reduced production of IFN- γ , a key cytokine involved in tumor killing activity, supporting LAYN as a negative regulator. Further studies are needed to generate mechanistic insights into the function of LAYN as well as other genes associated with HCC-infiltrating CD8⁺ T cells and Tregs.

state and even the occasional acquirement of the suppressive function, because these CD8⁺FOXP3⁺ T cells were absent in either blood or adjacent normal tissues.

The manifestation of a large fraction of clonally expanded Tregs in the HCC microenvironment might represent another mechanism to suppress the killing mediated by effector CD8⁺ T cells. By using both pseudotime and TCR-based methods, we also traced lineage relationship of tumor-infiltrating CD4⁺ T cells, especially Tregs (Figures 7H and 7I). Although tumor-infiltrating Tregs, as defined as C8_CD4-CTLA4, shared close developmental trajectory to exhausted CD4 T cells (C10_CD4-CXCL13), these two clusters had only a limited number of shared TCRs (Figure 7I). The vast majority (71/87) of the tumor-specific Treg clones were exclusive to themselves, suggesting an independent developmental path from other tumor-infiltrating CD4 T helper cells. This was in contrast to tumor-associated exhausted CD8⁺ T cells that share a higher number of common TCRs with other clusters of CD8 T cells in tumor (Figure 7J). The few common T cell clones in both Tregs and exhausted CD4 T cell clusters indicated the potential conversion from exhausted CD4 T cells to inducible Tregs (Figure 7I and 7K). The enrichment of many Treg clones suggested that their TCRs might recognize HCC antigens, and the amplification of these Treg clones most likely augmented their immune repressive functions, which further constrained the effector functions of tumor-infiltrating T cells.

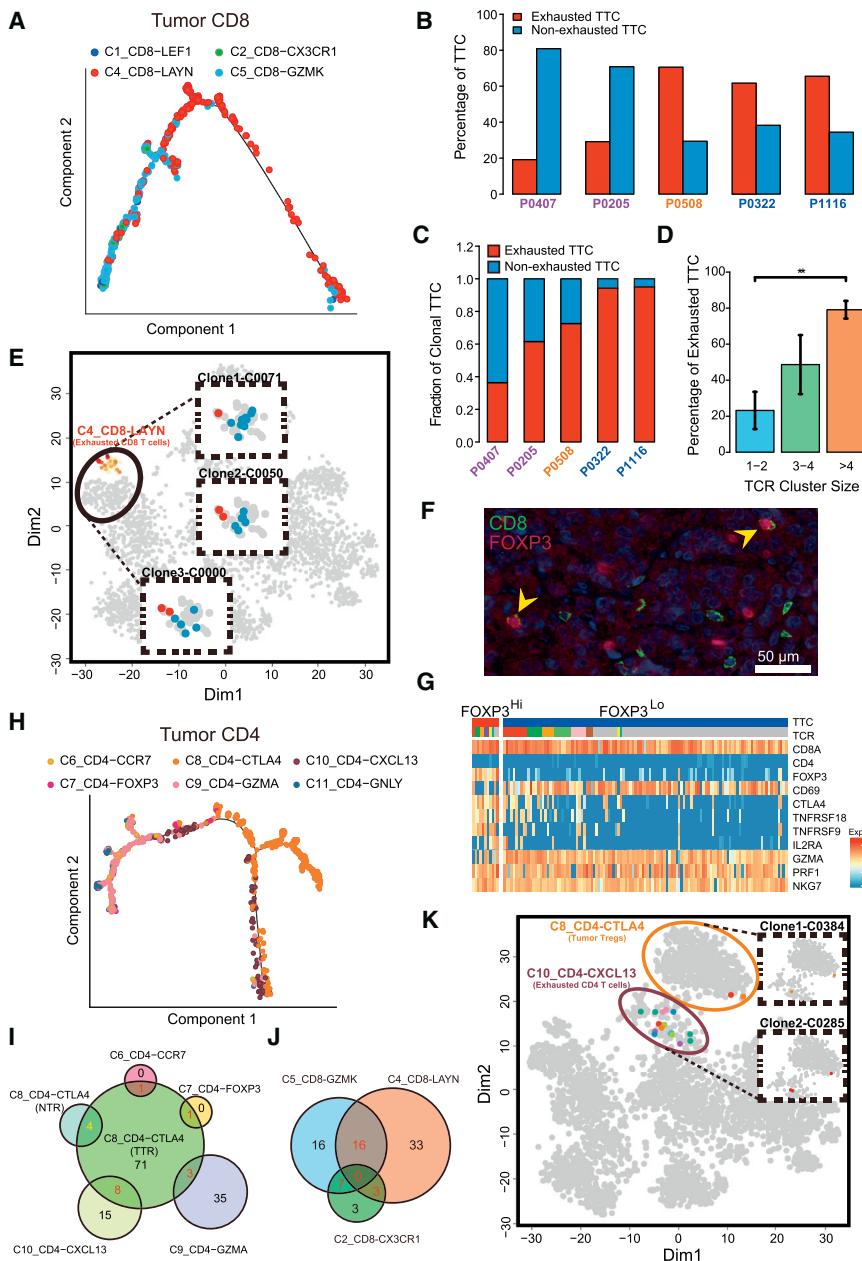


Figure 7. The Tumor-Specific CD8⁺ T Cell and Treg Analysis Based on Integrated Expression and TCR Clonality

(A) The ordering of tumor-derived CD8⁺ cells along pseudotime in a two-dimensional state-space. Each point corresponds to a single cell, and each color represents a T cell cluster.

(B) The percentages of exhausted and non-exhausted TTCs in each patient, based on T cell classification in Figure 2. The purple color marks patients in stage I, orange in stage II, and blue stage IVB.

(C) The fractions of exhausted and non-exhausted cells among clonal TTCs in each patient.

(D) Bar plot showing the percentages of exhausted TTCs in varying TCR-cluster sizes in each patient. Error bars represent \pm SEM.

(E) Among exhausted TTCs, a small number of cells show high FOXP3 expression. The space distribution of cells is defined by Figure 2A, with only those FOXP3⁺CD8⁺ T cells and other CD8⁺ T cells sharing the same TCRs shown in colors. The inserts show three different groups of cells with identical TCR clonotypes, with red dots denoting FOXP3⁺CD8⁺ cells and blue dots denoting FOXP3⁻CD8⁺ cells.

(F) Opal multicolor IHC staining showing the presence of FOXP3⁺CD8⁺ T cells, labeled with yellow arrow, using CD8 and FOXP3 antibodies.

(G) Expression characteristics and TCR clonal status of CD8⁺ FOXP3[±] T cells in patient P0322. The same color (except gray) in the TCR annotation bar represents the identical TCR clonotypes. (H) The same pseudotime plot as in (A) for tumor CD4⁺ T cells.

(I) The number of shared TCRs between tumor Tregs (C8_CD4-CTLA4) and the other CD4⁺ T cell clusters. The red numbers represent shared TCRs, the black numbers the unique TCRs detected within each CD4⁺ T cell cluster, and the yellow numbers the shared TCRs between NTR and TTR within cluster C8_CD4-CTLA4.

(J) The number of shared TCRs between exhausted CD8⁺ T cells (C4_CD8-LAYN) and other CD8⁺ T cell clusters in tumor.

(K) Similar to (E), multiple TCRs are shared between exhausted T helper cells (C10_CD4-CXCL13) and Tregs (C8_CD4-CTLA4) in tumor.

The preferential accumulation of both Tregs and exhausted CD8⁺ T cells in HCC might be a result of local expansion of these cells as suggested by previous studies (Ferradini et al., 1993; Hindley et al., 2011). A previous TCR β chain-based study suggested that breast cancer-associated Tregs were mainly recruited from periphery, but less likely evolved from conventional CD4⁺ T cells in tumor or expansion of Tregs from adjacent normal tissue into the tumor environment (Plitas et al., 2016). Our data, based on clonal TCRs, corroborated with this finding because the majority clonal HCC-infiltrating Tregs (82%) were unique (Figure 7) and only a small portion of them shared common TCRs with CD4⁺ T helper cells in the tumor or adjacent normal tissue Tregs.

Similar TCR analysis, however, painted a different picture for the origins of tumor-infiltrating exhausted CD8⁺ T cells. Although exhausted CD8⁺ T cells represented the largest cluster (42.6% on average) in tumor-infiltrating CD8⁺ T cells, the CX3CR1⁺ effector cells (7.75%) and GZMK-expressing CD8⁺ T cells (35.17%) could also be detected. Among all exhausted CD8⁺ T cell clones identified in tumor, 37% were shared with other CD8 T cell clusters, especially with C5_CD8-GZMK cluster (31%), suggesting exhausted CD8 T cells were more likely evolved from other types of CD8 T cells inside the tumor.

C5_CD8-GZMK cluster appears to be distinct from conventional T cell subtypes. These cells express lower levels of cytotoxic markers (GZMB, GNLY, and KLRG1) compared to effector

T cells and also certain levels of exhaustion markers (*PDCD1* and *TIGIT*) and thus could possibly represent cells in a transition state from effector to exhausted T cells. Indeed, these cells also shared high percentage of common TCRs with CX3CR1⁺ effector T cells (18%). Therefore, promoting cells in C5_CD8-GZMK cluster to effector-like cells and preventing them from exhaustion might be a possible therapeutic strategy. In fact, antibody blockade of PD1 pathway has been shown to reinvigorate exhausted CD8⁺ T cells with intermediate expression of *PDCD1*, but not those with high *PDCD1* expression (Blackburn et al., 2008). In addition to CX3CR1⁺ CD8 effector T cells, we also identified GNLY⁺ CD4 effector T cells, which shared similar gene expression characteristics to CD8 effector T cells, implicating their cytotoxic functions. These cells are more closely linked to CD4⁺ Cluster 9 with GZMA expression, which appear to be in a mixed state among known subtypes of T helper cells but are more similar to the Th1 cells due to the expression of CXCR6, *TBX21*, and CXCR3 (Di Cesare et al., 2009). Therefore, promoting these cells to more closely resemble cytolytic CD4⁺ T cells (Soghoian and Streeck, 2010) might be another strategy for cancer immunotherapy.

Another cluster of special interest in HCC consists of the MAIT cells with specific semi-invariant TCR alpha chains. MAIT cells are activated during bacterial or viral infection (van Wilgenburg et al., 2016) but have not been linked to HBV infection previously. MAIT cells are considered as first line of defense in liver, whereas their role in cancers is unclear (Kurioka et al., 2016). We observed markedly decreased representation of MAIT cells in tumors compared to adjacent normal liver tissues, consistent with the previous notion that MAIT had the potential to kill cancer cells, and TCGA HCC patients with low expression of the MAIT marker gene *SLC4A10* showed poor prognosis. These findings invite further investigations on role of MAIT cells in HCC and immunotherapies.

In conclusion, our comprehensive single T cell database revealed detailed characteristics of HCC-infiltrating T cells, in term of their clustering, dynamic, and developmental trajectory, as well as unique signatures. To facilitate the usage of our data for the wide research community, we also developed an interactive, web-based tool (<http://hcc.cancer-pku.cn>) for analyzing, visualizing, and downloading our single-cell data for single or multiple user-input genes (Figure S7B). The transcriptome data, coupled with detailed TCR-based lineage information, can serve as a rich resource for deeper understanding of tumor-infiltrating lymphocytes in general.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Human Specimens
- METHOD DETAILS
 - Single cell collection
 - Single cell sorting, RT and amplification

- Bulk DNA and RNA isolation and sequencing
- Immunohistochemistry
- In vitro T cell activation and FACS analysis
- Overexpressing LAYN in primary CD8⁺ T cells

- QUANTIFICATION AND STATISTICAL ANALYSIS

- Single-cell RNA-Seq data processing
- Bulk RNA- and exome-seq data processing
- TCR analysis
- Unsupervised clustering
- T cell exhaustion state analysis
- TCGA data analysis

- DATA AND SOFTWARE AVAILABILITY

- Data resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and five tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2017.05.035>.

AUTHOR CONTRIBUTIONS

Z.Z., J.P., W.O., and C.Z. designed experiments. C.Z., X.G., J.-K.Y., R.H., J.Y.H., Q.Z., Z.L., B.K., and X.H. performed the experiments. C.Z., L.Z., Y.Z., and M.D. analyzed sequencing data. H.G. performed immunofluorescence staining. C.Z., W.O., and Z.Z. wrote the manuscript, with all authors contributing to writing and providing feedback.

ACKNOWLEDGMENTS

We thank Xianwen Ren and Shuang Geng for discussions, the contribution by the BIOPIC high-throughput sequencing facility, and the Computing Platform of the Center for Life Science. We thank National Center for Protein Sciences Beijing (Peking University) and Z. Fu, X. Zhang, S. Wang, and F. Wang for assistance with FACS. This project was supported by grants from Beijing Advanced Innovation Center for Genomics at Peking University, Key Technologies R&D Program (2016YFC0900100) and the National Natural Science Foundation of China (81573022, 31530036, 81372604). C.Z. was supported in part by the Postdoctoral Fellowship of Peking-Tsinghua Center for Life Sciences. W.O., J.-K.Y., R.H., and J.Y.H. are employees of Amgen.

Received: December 19, 2016

Revised: April 9, 2017

Accepted: May 23, 2017

Published: June 15, 2017

REFERENCES

- Bhairavabhotla, R., Kim, Y.C., Glass, D.D., Escobar, T.M., Patel, M.C., Zahr, R., Nguyen, C.K., Kilaru, G.K., Muljo, S.A., and Shevach, E.M. (2016). Transcriptome profiling of human FoxP3⁺ regulatory T cells. *Hum. Immunol.* 77, 201–213.
- Blackburn, S.D., Shin, H., Freeman, G.J., and Wherry, E.J. (2008). Selective expansion of a subset of exhausted CD8 T cells by alphaPD-L1 blockade. *Proc. Natl. Acad. Sci. USA* 105, 15016–15021.
- Böttcher, J.P., Beyer, M., Meissner, F., Abdullah, Z., Sander, J., Höchst, B., Eickhoff, S., Rieckmann, J.C., Russo, C., Bauer, T., et al. (2015). Functional classification of memory CD8(+) T cells by CX3CR1 expression. *Nat. Commun.* 6, 8306.
- Buettner, F., Natarajan, K.N., Casale, F.P., Prosperio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* 33, 155–160.

- Campbell, K., Ponting, C.P., and Webber, C. (2015). Laplacian eigenmaps and principal curves for high resolution pseudotemporal ordering of single-cell RNA-seq profiles. *bioRxiv*. <http://dx.doi.org/10.1101/027219>.
- Cannoodt, R., Saelens, W., Sichien, D., Tavernier, S., Janssens, S., Guiliams, M., Lambrecht, B.N., Preter, K.D., and Saeys, Y. (2016). SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. *bioRxiv*. <http://dx.doi.org/10.1101/079509>.
- Chaput, N., Louafi, S., Bardier, A., Charlotte, F., Vaillant, J.-C., Ménegaux, F., Rosenzwaig, M., Lemoine, F., Klatzmann, D., and Taieb, J. (2009). Identification of CD8+CD25+Foxp3+ suppressive T cells in colorectal cancer tissue. *Gut* 58, 520–529.
- Chen, L., and Flies, D.B. (2013). Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat. Rev. Immunol.* 13, 227–242.
- Chisari, F.V., and Ferrari, C. (1995). Hepatitis B virus immunopathogenesis. *Annu. Rev. Immunol.* 13, 29–60.
- Coulie, P.G., Van den Eynde, B.J., van der Bruggen, P., and Boon, T. (2014). Tumour antigens recognized by T lymphocytes: at the core of cancer immunotherapy. *Nat. Rev. Cancer* 14, 135–146.
- Crawford, A., Angelosanto, J.M., Kao, C., Doering, T.A., Odorizzi, P.M., Barnett, B.E., and Wherry, E.J. (2014). Molecular and transcriptional basis of CD4⁺ T cell dysfunction during chronic infection. *Immunity* 40, 289–302.
- De Simone, M., Arrigoni, A., Rossetti, G., Gruarin, P., Ranzani, V., Politano, C., Bonnal, R.J.P., Provasi, E., Sarnicola, M.L., Panzeri, I., et al. (2016). Transcriptional landscape of human tissue lymphocytes unveils uniqueness of tumor-infiltrating T regulatory cells. *Immunity* 45, 1135–1147.
- Di Cesare, A., Di Meglio, P., and Nestle, F.O. (2009). The IL-23/Th17 axis in the immunopathogenesis of psoriasis. *J. Invest. Dermatol.* 129, 1339–1350.
- El-Serag, H.B. (2011). Hepatocellular carcinoma. *N. Engl. J. Med.* 365, 1118–1127.
- Ferradini, L., Mackensen, A., Genevée, C., Bosq, J., Duvillard, P., Avril, M.F., and Hercend, T. (1993). Analysis of T cell receptor variability in tumor-infiltrating lymphocytes from a human regressive melanoma. Evidence for in situ T cell clonal expansion. *J. Clin. Invest.* 91, 1183–1190.
- Förster, R., Davalos-Misslitz, A.C., and Rot, A. (2008). CCR7 and its ligands: balancing immunity and tolerance. *Nat. Rev. Immunol.* 8, 362–371.
- Fousteri, G., Jofra, T., Debernardis, I., Stanford, S.M., Laurenzi, A., Bottini, N., and Battaglia, M. (2014). The protein tyrosine phosphatase PTPN22 controls forkhead box protein 3 T regulatory cell induction but is dispensable for T helper type 1 cell polarization. *Clin. Exp. Immunol.* 178, 178–189.
- Freeman, J.D., Warren, R.L., Webb, J.R., Nelson, B.H., and Holt, R.A. (2009). Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res.* 19, 1817–1824.
- Goode, E.L., DeRycke, M., Kalli, K.R., Oberg, A.L., Cunningham, J.M., Maurer, M.J., Fridley, B.L., Armasu, S.M., Serie, D.J., Ramar, P., et al. (2013). Inherited variants in regulatory T cell genes and outcome of ovarian cancer. *PLoS ONE* 8, e53903.
- Grosso, J.F., Goldberg, M.V., Getnet, D., Bruno, T.C., Yen, H.-R., Pyle, K.J., Hipkiss, E., Vignali, D.A.A., Pardoll, D.M., and Drake, C.G. (2009). Functionally distinct LAG-3 and PD-1 subsets on activated and chronically stimulated CD8 T cells. *J. Immunol.* 182, 6659–6669.
- Han, A., Glanville, J., Hansmann, L., and Davis, M.M. (2014). Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat. Biotechnol.* 32, 684–692.
- Hindley, J.P., Ferreira, C., Jones, E., Lauder, S.N., Ladell, K., Wynn, K.K., Betts, G.J., Singh, Y., Price, D.A., Godkin, A.J., et al. (2011). Analysis of the T-cell receptor repertoires of tumor-infiltrating conventional and regulatory T cells reveals no evidence for conversion in carcinogen-induced tumors. *Cancer Res.* 71, 736–746.
- Ji, Z., and Ji, H. (2016). TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* 44, e117.
- Jiang, Y., Li, Y., and Zhu, B. (2015). T-cell exhaustion in the tumor microenvironment. *Cell Death Dis.* 6, e1792.
- Kan, Z., Zheng, H., Liu, X., Li, S., Barber, T.D., Gong, Z., Gao, H., Hao, K., Willard, M.D., Xu, J., et al. (2013). Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res.* 23, 1422–1433.
- Kasper, H.-U., Drebber, U., Stippel, D.L., Dienes, H.P., and Gillessen, A. (2009). Liver tumor infiltrating lymphocytes: comparison of hepatocellular and cholangiolar carcinoma. *World J. Gastroenterol.* 15, 5053–5057.
- Kim, H.-J., Barnitz, R.A., Kreslavsky, T., Brown, F.D., Moffett, H., Lemieux, M.E., Kaygusuz, Y., Meissner, T., Holderried, T.A.W., Chan, S., et al. (2015). Stable inhibitory activity of regulatory T cells requires the transcription factor Helios. *Science* 350, 334–339.
- Kiselev, V.Y., Kirschner, K., Schaub, M.T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K.N., Reik, W., Barahona, M., Green, A.R., and Hemberg, M. (2017). SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* 14, 483–486.
- Kurioka, A., Walker, L.J., Kleinerman, P., and Willberg, C.B. (2016). MAIT cells: new guardians of the liver. *Clin. Transl. Immunology* 5, e98.
- Marshall, N.B., and Swain, S.L. (2011). Cytotoxic CD4 T cells in antiviral immunity. *J. Biomed. Biotechnol.* 2011, 954602.
- Nishikawa, H., and Sakaguchi, S. (2014). Regulatory T cells in cancer immunotherapy. *Curr. Opin. Immunol.* 27, 1–7.
- Pallandré, J.-R., Brillard, E., Créhange, G., Radlovic, A., Remy-Martin, J.-P., Saas, P., Rohrlich, P.-S., Pivot, X., Ling, X., Tibergien, P., and Borg, C. (2007). Role of STAT3 in CD4+CD25+FOXP3+ regulatory lymphocyte generation: implications in graft-versus-host disease and antitumor immunity. *J. Immunol.* 179, 7593–7604.
- Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181.
- Plitas, G., Konopacki, C., Wu, K., Bos, P.D., Morrow, M., Putintseva, E.V., Chudakov, D.M., and Rudensky, A.Y. (2016). Regulatory T cells exhibit distinct features in human breast cancer. *Immunity* 45, 1122–1134.
- Prieto, J., Melero, I., and Sangro, B. (2015). Immunological landscape and immunotherapy of hepatocellular carcinoma. *Nat. Rev. Gastroenterol. Hepatol.* 12, 681–700.
- Rutz, S., Noubade, R., Eidenschenk, C., Ota, N., Zeng, W., Zheng, Y., Hackney, J., Ding, J., Singh, H., and Ouyang, W. (2011). Transcription factor c-Maf mediates the TGF-β-dependent suppression of IL-22 production in T(H)17 cells. *Nat. Immunol.* 12, 1238–1245.
- Sen, D.R., Kaminski, J., Barnitz, R.A., Kurachi, M., Gerdemann, U., Yates, K.B., Tsao, H.-W., Godec, J., LaFleur, M.W., Brown, F.D., et al. (2016). The epigenetic landscape of T cell exhaustion. *Science* 354, 1165–1169.
- Sharma, P., and Allison, J.P. (2015). Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential. *Cell* 161, 205–214.
- Soghoian, D.Z., and Streeck, H. (2010). Cytolytic CD4(+) T cells in viral immunity. *Expert Rev. Vaccines* 9, 1453–1463.
- Stubbington, M.J.T., Lönnberg, T., Proserpio, V., Clare, S., Speak, A.O., Dougan, G., and Teichmann, S.A. (2016). T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* 13, 329–332.
- Tanchot, C., Terme, M., Pere, H., Tran, T., Benhamouda, N., Strioga, M., Banni, C., Galluzzi, L., Kroemer, G., and Tartour, E. (2013). Tumor-infiltrating regulatory T cells: phenotype, role, mechanism of expansion in situ and clinical significance. *Cancer Microenviron.* 6, 147–157.
- Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V.I., Lao, K., and Surani, M.A. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat. Protoc.* 5, 516–535.
- Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196.

- Topalian, S.L., Drake, C.G., and Pardoll, D.M. (2015). Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell* 27, 450–461.
- Topalian, S.L., Taube, J.M., Anders, R.A., and Pardoll, D.M. (2016). Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. *Nat. Rev. Cancer* 16, 275–287.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386.
- van Wilgenburg, B., Scherwitzl, I., Hutchinson, E.C., Leng, T., Kurioka, A., Kulicke, C., de Lara, C., Cole, S., Vasanawathana, S., Limpitikul, W., et al.; STOP-HCV consortium (2016). MAIT cells are activated during human viral infections. *Nat. Commun.* 7, 11653.
- Wang, J., Ioan-Facsinay, A., van der Voort, E.I.H., Huizinga, T.W.J., and Toes, R.E.M. (2007). Transient expression of FOXP3 in human activated nonregulatory CD4+ T cells. *Eur. J. Immunol.* 37, 129–138.
- Wherry, E.J., and Kurachi, M. (2015). Molecular and cellular insights into T cell exhaustion. *Nat. Rev. Immunol.* 15, 486–499.
- Wing, K., Ekmark, A., Karlsson, H., Rudin, A., and Suri-Payer, E. (2002). Characterization of human CD25+ CD4+ T cells in thymus, cord and adult blood. *Immunology* 106, 190–199.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Human CD3 eFluor 450 (FACS)	eBioscience	Cat#48-0037-41
Anti-Human CD4 FITC (FACS)	eBioscience	Cat#11-0048-41
Anti-Human CD8a APC (FACS)	eBioscience	Cat#17-0086-41
Anti-Human CD25 PE (FACS)	eBioscience	Cat#12-0259-42
Human Layilin Antibody (FACS)	Sino Biological	Cat#10208-MM02
7-AAD Viability Staining Solution (FACS)	eBioscience	Cat#00-6993-50
Anti-CD3 antibody (IHC)	Abcam	Cat#ab16669
Anti-CD4 antibody (IHC)	Abcam	Cat#ab846
Anti-CD8 antibody (IHC)	Abcam	Cat#ab17147
Anti-FOXP3 antibody (IHC)	Abcam	Cat#ab22510
CD3 Functional Grade Monoclonal Antibody	eBioscience	Cat#16-0037-85
CD28 Functional Grade Monoclonal Antibody	eBioscience	Cat#16-0289-85
Biological Samples		
Human PBMC	AllCells	https://www.allcells.com
Critical Commercial Assays		
Live/Dead Fixable Blue Dead Cell Stain Kit	Invitrogen	Cat#L34962
Alexa Fluor 647 Conjugation Kit	Molecular Probes	Cat#A20186
IFN Gamma Human Uncoated ELISA Kit	eBioscience	Cat#88-7316-88
Dynabeads Human T-activator CD3/CD28 for T Cell Expansion and Activation	ThermoFisher Scientific	Cat#11131D
Retro-X Universal Packaging System	Clontech	Cat#631530
Pan T cell Isolation Kit	Miltenyi Biotec	Cat#130-096-535
Human T cell Activation/Expansion Kit	Miltenyi Biotec	Cat#130-091-441
NEBNext Ultra RNA Library Prep Kit for Illumina Paired-end Multiplexed Sequencing Library	NEB	Cat#E7530
SureSelectXT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library kit	Agilent	Cat#G9701
TruePrep DNA Library Prep Kit V2 for Illumina	Vazyme Biotech	Cat#TD503
Hiseq 3000/4000 SBS kit	Illumina	Cat#FC-410-1003
Hiseq 3000/4000 PE cluster kit	Illumina	Cat#PE-410-1001
Deposited Data		
Data files for single-cell RNA sequencing (raw data)	This paper	EGAS00001002072
Data files for bulk RNA sequencing (raw data)	This paper	EGAS00001002072
Data files for bulk exome sequencing (raw data)	This paper	EGAS00001002072
Data files for single-cell RNA sequencing (processed data)	This paper	GSE98638
Oligonucleotides		
Primer: CD3D Forward: TCATTGCCACTCTGCTCC	This paper	N/A
Primer: CD3D Reverse: GTTCACTTGTCCGAGCC	This paper	N/A
Software and Algorithms		
SC3	Kiselev et al., 2017	https://github.com/hemberg-lab/SC3
Monocle 2.0	Trapnell et al., 2014	http://monocle-bio.sourceforge.net/
ScLVM	Buettner et al., 2015	https://github.com/PMBio/scLVM
TraCeR	Stubbington et al., 2016	https://github.com/teichlab/tracer

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Zemin Zhang (zemin@pku.edu.cn).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Specimens

Four male and two female patients who were pathologically diagnosed with hepatocellular carcinoma (HCC) were enrolled in this study. Their ages ranged from 26 to 64, with a median age of 49. Among those patients, three were diagnosed as stage I, one as stage II, and two as stage IVB. All patients, with the exception of P1202, were HBV-positive based on the HBsAg test. None of the patients was treated with chemotherapy or radiation prior to tumor resection. The available clinical characteristics of these patients are summarized in [Figure 1C](#). For patients P0508, P1116, P0322 and P0407, their peripheral blood and paired fresh HCC tumor and adjacent normal liver tissues were obtained for the subsequent lymphocyte isolation. The adjacent normal tissues were at least 2 cm from the matched tumor tissue. For patients P1202 and P0205, only fresh liver tumor tissues and matched peripheral blood were collected. For patient P0508, we also collected cells from the joint area between the tumor and the adjacent normal tissue. This study was approved by the Ethics Committee of Beijing Shijitan Hospital, Capital Medical University. All patients in this study provided written informed consent for sample collection and data analyses. For the LAYN study, frozen human PBMCs were obtained from AllCells Inc (Alameda, CA). PBMCs from only normal healthy donors (*i.e.* BMI score < 30, no virus infections including HIV/HCV/HBV, no medication within 2 wks and non-smokers) were used in this study.

METHOD DETAILS

Single cell collection

Peripheral blood mononuclear cells (PBMCs) were isolated using HISTOPAQUE-1077 (Sigma-Aldrich) solution according to the manufacturer's instructions. Briefly, 3 mL of fresh peripheral blood was collected prior to surgery in EDTA anticoagulant tubes and subsequently layered onto HISTOPAQUE-1077. After centrifugation, lymphocyte cells remained at the plasma-HISTOPAQUE-1077 interface and were carefully transferred to a new tube and washed twice with 1x PBS (Invitrogen). These lymphocytes were re-suspended with sorting buffer (PBS supplemented with 1% fetal bovine serum (FBS, ScienCell)).

Fresh tumor and adjacent normal tissue samples were cut into approximately 1-mm³ pieces and gently triturated with a 20 mL syringe plunger on a 40 µm Cell-Strainer (BD) in the RPMI-1640 medium (Invitrogen) with 10% FBS until uniform cell suspensions were obtained. The suspended cells were subsequently passed through cell strainers and centrifuged at 400 g for 10 min. After the supernatant was removed, the pelleted cells were suspended in red blood cell lysis buffer (Solarbio) and incubated on ice for 2 min to lyse red blood cells. After washing twice with 1x PBS, the cell pellets were re-suspended in sorting buffer.

Single cell sorting, RT and amplification

Based on FACS analysis, single cells of different subtypes, including CD8⁺ T cells (CD3⁺ and CD8⁺), T helper cells (CD3⁺, CD4⁺ and CD25⁺), and regulatory T cells (CD3⁺, CD4⁺ and CD25^{high}), were sorted into wells of 96-well plates (Axygen) with lysis buffer, which contained 1 µl 10 mM dNTP mix (Fermentas), 1 µl 10 µM Oligo dT primer, 1.9 µl 1% Triton X-100 (Sigma) plus 0.1 µl 40 U/µl RNase Inhibitor (Takara). The sealed plates were stored frozen at -80°C. For patients P0322 and P0407, the External RNA Controls Consortium (ERCC, Ambion) was added into each well as a known quantity of exogenous spike-in before the reverse transcription. Single cell transcriptome amplifications were performed following the Smart-Seq2 protocol. The amplified cDNA products were purified with 1x Agencourt XP DNA beads (Beckman). Following the first round of beads purification, the cDNA of each single cell was quantified with qPCR of CD3D, and fragment analysis (using fragment analyzer AATI).

For those single cell samples with high quality after this step, the DNA products were further cleaned with 0.5x Agencourt XP DNA beads (Beckman) to eliminate short fragments (less than 500 bp). At this step, the concentration of each sample was quantified with Qubit HsDNA kits (Invitrogen), and libraries were then constructed with the TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme Biotech). Constructed libraries derived from patients P0205, P0508 and P0322 were analyzed by an Illumina HiSeq2500 sequencer with 100 bp pair-end reads, and for patients P0407 and P1116, libraries were analyzed by an Illumina Hiseq 4000 sequencer with 150 bp pair-end reads. For patient P1202, single cells were manually picked into each well with mouth pipet and the single cell transcriptome amplifications were performed following the Tang2010 protocol ([Tang et al., 2010](#)). TCRs of those cells could not be assembled with their single cell RNA-seq data due to the obvious 3' bias and bulk exome and RNA sequencing were not performed further for this patient.

Bulk DNA and RNA isolation and sequencing

Genomic DNA of peripheral blood and tissue samples were extracted using the QIAamp DNA Mini Kit (QIAGEN) according to the manufacturer's specification. The concentrations of DNA were quantified using the Qubit HsDNA Kits (Invitrogen) and the qualities of DNA were evaluated with agarose gel electrophoresis. Exon libraries were constructed using the SureSelectXT Target Enrichment

System for Illumina Paired-End Multiplexed Sequencing Library kit (Agilent). Samples were sequenced on the Illumina Hiseq 4000 sequencer with 150 bp paired-end reads.

For bulk RNA analysis, small fragments of tumor tissues and adjacent normal tissues were first stored in RNAlater RNA stabilization reagent (QIAGEN) after surgical resection and kept on ice to avoid RNA degradation. RNA of tumor and adjacent normal tissue samples were extracted using the RNeasy Mini Kit (QIAGEN) according to the manufacturer's specification. The concentrations of RNA were quantified using the NanoDrop instrument (Thermo) and the qualities of RNA were evaluated with fragment analyzer (AATI). Libraries were constructed using NEBNext Poly(A) mRNA Magnetic Isolation Module kit (NEB) and NEBNext Ultra RNA Library Prep Kit for Illumina Paired-end Multiplexed Sequencing Library (NEB). Samples were sequenced on the Illumina Hiseq 4000 sequencer with 150 bp paired-end reads.

Immunohistochemistry

Human tissue specimens were provided by Beijing Shijitan Hospital under an approved Institutional Review Board protocol. The specimens were collected within 30 min after the tumor resection and fixed in formalin for 48 hr. Dehydration and embedding in paraffin was performed following routine methods. These paraffin blocks were cut into 5 μ m slides and adhered on the slides glass. Then the paraffin sections were placed in the 70°C paraffin oven for 1 hr before deparaffinized in xylene and then rehydrated in 100%, 90%, 70% alcohol successively. Antigen was retrieved by citric acid buffer (pH 6.0) in the 95°C water bath for 20 min. Endogenous peroxidase was inactivated by incubation in 3% H₂O₂ for 15 min. Following a preincubation with 10% normal goat serum to block nonspecific sites for 30 min, the sections were incubated with primary antibodies in a humidified chamber at 4°C overnight. (Rabbit anti-CD3 antibody were diluted with PBS at 1:50, and rabbit anti-CD4 antibody at 1:150, mouse anti-CD8 antibody at 1:50, mouse anti-FOXP3 antibody at 1:50, Abcam) After the sections were washed with PBS twice for 5 min, the antigenic binding sites were visualized using the GTVisionTMIDetection System/Mo&Rb according to the manufacturer's protocol.

In vitro T cell activation and FACS analysis

Human PBMCs were cultured in complete media X-vivo (Lonza) supplemented with 10% human serum (Sigma), Glutamax, non-essential amino acid and pen/strep in the presence of anti-human CD3 and CD28 mAbs conjugated microbeads. At the indicated time points post-stimulation, cells were harvested and stained with indicated mAbs (anti-human CD3 (SK7, BD), CD4 (SK3, BD), CD8a (SK1, BD), CD20 (2H7, BioLegend), CD14 (M5E2, BD), Helios (22F6, BioLegend), Foxp3 (206D, BioLegend), PD-1 (EH12.1, BD), TIM-3 (F38-2E2, BioLegend), LAG-3 (3DS223H, eBioscience)). Samples were acquired with LSR-II analyzer. Using the Flowjo software, LAYN and T cell activation/exhaustion markers (*i.e.* PD-1 and LAG-3) surface expressions were analyzed within each live/singlet gated sub-population. All data are from at least three independent experiments with more than 6 different donor PBMCs.

Overexpressing LAYN in primary CD8⁺ T cells

Human layilin was cloned into MSCV-IRES-EGFP retroviral vector (RV-Vec). Retrovirus was generated by using pAmpho packaging system before infecting T cells. For T cell infection, pan T cells were isolated from human PBMCs by using human Pan T cell isolation kit and activated with Dynabeads Human T-Activator CD3/CD28 at 1:1 ratio for 72 hr. 72 hr later, Dynabeads were removed and activated T cells were spin-infected with retrovirus at 2000 rpm for 1 hr under 32°C. 3 days post spin-infection, GFP+ human layilin overexpressing CD8⁺ T cells were FACS sorted and rested with human IL2 for 2 days. Equal numbers of control or human layilin overexpressing CD8 T cells were seeded into 96 round-bottom wells and re-stimulated with plate-bound anti-CD3 (1 μ g/ml) and anti-CD28 (1 μ g/ml). Cell supernatant was collected 24 hr after re-stimulation and human IFN- γ was measured by ELISA.

QUANTIFICATION AND STATISTICAL ANALYSIS

Single-cell RNA-Seq data processing

Sequencing data from Illumina Hiseq2500 or Hiseq4000 sequencer were first processed to filter out low quality reads which were defined as: 1) "N" bases account for 10% reads' length; 2) bases with quality < 5 account for 50% reads' length; 3) contain adaptor sequence. The reads passing the above filters usually accounted for about 90% of original reads. These reads were then aligned to ribosomal RNA sequences download from RFam using the GSNAP alignment tool. Next the unmapped reads were mapped with GSNAP to reference sequences, containing human genome reference sequence version 19 (hg19) downloaded from UCSC and 92 ERCC spike-in RNA sequences obtained from the datasheet of the External RNA Controls Consortium kit. In addition, a gene model file named knownGene.txt was downloaded from UCSC and used to define known exon-exon junctions with the adjustable parameters "-novellicing 1 -n 10 -i 1 -M 2" except other default parameters. After retrieving reads overlapping with genes defined in the gene model file with R function findOverlaps, the read counts of each gene were obtained. We further normalized the read counts with the DESeq method. Basically, the geometric mean of read counts across all cells was first calculated from the same patient for each gene, and then the ratios between read count and the corresponding geometric mean across cells with no-zero read counts were derived. Finally, the median value of the ratios was obtained as the size factor for each cell. Next the normalized read counts ($RC_{normalized}$) was defined as $RC / size_factor$, and the value $\log_2(RC_{normalized} + 1)$ was used for downstream analyses such as PCA and clustering.

We also filtered out low-quality cells following these steps: 1) to discard cells with the number of uniquely mapped read pairs less than 250,000; 2) to discard cells with RPKM of *CD3D* less than 1.0; 3) to discard FACS sorted CD8⁺ T cells with RPKM of *CD8A* less than 1.0 or with RPKM of *CD4* larger than 10; 4) to discard FACS sorted CD4⁺ T cells with RPKM of *CD4* less than 1.0 or with RPKM of *CD8A* larger than 10; 5) to discard cells that failed the outlier analysis. In the outlier analysis, we performed PCA analysis with $\log_2(RC_{\text{normalized}}+1)$ values and fitting analysis between the squared coefficient of variance (CV^2) and the mean of $RC_{\text{normalized}}$ among each FACS-sorted T cell subtype, and then we manually examined the result to identify outlier cells. Outlier cells typically were located far from other cells and had unusually high contribution to principal components. In addition, those cells also exhibited low size factor, indicative of low cellular RNA amount, and fitted curves would improve upon removing these cells. After these filtering and discarding outlier cells, total 4128 cells from 6 patients were remained, including 4070 cells from 5 HBV positive patients.

Bulk RNA- and exome-seq data processing

The Bulk RNA-seq data were processed using the same GSNAp pipeline as single cell RNA-seq data. For bulk exome sequencing data, we first filtered out low-quality sequencing reads with the same procedure as bulk RNA-seq data processing. The high quality reads were mapped to the human reference sequence (version b37 with decoy sequence added, download from Broad) with bwa-0.7.12. The post-alignment processing, including bam file sorting, duplicate reads marking, reads' local re-aligning around candidate INDEL and base quality re-calibration, was performed using GATK according the suggested best practice. Somatic mutations, including SNV and INDEL, were called using Strelka with default parameters. Those called mutations, especially low-frequency mutations, were manually inspected further using the IGV tool. Somatic copy number alterations and LOH (loss of heterogeneity) were obtained by ADTEx. Based on the overlap with the driver gene list proposed previous, we further obtained candidate driver mutations for these patients. Among these, a few occurred at known mutational hotspots: TP53 G245C (patient P0205), PIK3CA H1047R (patient P0508), PTEN C136R (patient P1116), TP53 R249S (patient P0322) and TP53 I195F (patient P0427). The late stage patient P0322 had two different TP53 mutations.

TCR analysis

The TCR sequences for each single T cell were assembled by the TraCeR method from single cell RNA-Seq data, leading to the identification of the CDR3 sequence, the rearranged TCR genes, and their expression abundance (transcripts per million, TPM). First, we discard those cells with no obvious TCR forms. Then we arrange TCR alpha and beta chain respectively with the following steps. The first TCR alpha (beta) chain was defined as follows: 1) keep all single T cells in which only one productive TCR alpha and beta chain was present. 2) if more than one TCR alpha or beta chain were identified in one T cell, we kept only the cells in which a dominant form of alpha and beta was detected. Often, one alpha/beta chain was productive and the other chain was non-productive, or the expression level of one was far higher than the alternative allele, and the productive or dominant form was identified. Next, we filtered out the second TCR alpha chains with TPM less than 10 and beta chains with TPM less than 15 to eliminate the biological and bioinformatics error based on the histogram analysis for the expression distribution (Figure S6A). From a total 4032 cells with successfully assembled TCR sequences, we identified the TCR alpha/beta pairs for 3792 cells.

Unsupervised clustering

For each patient, the relative expression measurement of a given gene was calculated as $Y-\text{mean}(Y)$, where Y is the normalized expression measurement $\log_2(RC_{\text{normalized}}+1)$, and $\text{mean}(Y)$ is the mean of Y across all cells of one patient. This procedure removed the unwanted variation caused by patient differences. Such centered data were subsequently combined across all patients. Genes with the top n highest standard deviation were considered highly variable genes, and we experimented different n values of 1500, 2000, 2500 and 3000. The resulting centered expression data across all genes were used for a modified SC3 (Kiselev et al., 2017) clustering pipeline. Basically, the distance matrixes based on Spearman correlation were first constructed and then transformed by calculating the eigenvectors of the graph Laplacian. Next, the k-means algorithm was applied to the first d eigenvectors multiple times where d was chosen different values range from 4% to 7% of the total number of cells. A consensus matrix could be subsequently obtained by averaging the results from different runs. The consensus matrix indicate how often two cells are clustered together into the same group in different runs. Finally, hierarchical clustering with complete agglomeration was performed on the consensus matrix and the k clusters were inferred. The SC3's parameters k , which was used in the k-means and hierarchical clustering, was chosen from 2 to 10 iteratively. For each SC3 run, the silhouette was calculated, the consensus matrix plotted, and cluster specific genes identified. All these three aspects aid us to empirically determine the optimal k and n . Once the stable clusters were determined, the above procedure was iteratively applied to each of these clusters to reveal highly variable genes among cells in each cluster, and then use these variable genes to identify sub-clusters.

After clustering results were obtained, to identify differential expressed genes ANOVA was performed using the R function *aov*, and the difference between each cluster pair was tested using Tukey's range test implemented in R function *TukeyHSD*. Then differential expressed genes of each cluster were identified as those met these criteria: 1) FDR adjusted p value of F test < 0.01; 2) absolute log2 fold change between the cluster under consideration and others were larger than 1; 3) p values of the Tukey's range test comparing the cluster under consideration and others < 0.01. t-SNE method was used for visualization of cells' distance in the reduced 2D space.

In addition, cell cycle was considered as one important confounding factor of cell population analysis. In our study we examined the contribution of cell cycle by ScLVM (Buettnner et al., 2015) and found only minimal effect on our clustering results. The ScLVM results revealed that the variance explained by cell cycle factor was only 4.5~13.6% of that by biological factors for 4 patients (P0508, P0322, P0407, P1116), and was 37.7% for patient P0205, which was still much lower than those reported previously (85%). Indeed, we identified clusters consisting of a few cells expressing cell cycle related genes, such as *TUBB4B*, *TUBA1B*, *MCM7*, *STMN1* and *MKI67*, but all these clusters (containing 62 cells) were small sub-clusters of the major clusters “C4_CD8-LAYN,” “C8_CD4-CTLA4” and “C10_CD4-CXCL13,” and excluded for tSNE visualization and differential gene expression analysis.

T cell exhaustion state analysis

To reveal characteristics of exhausted T cells, we first removed tumor infiltrating T cells that carried the same TCR sequences observed in T cells from non-cancer tissues. MAIT cells were also excluded from the exhaustion analysis due to their special characteristics. Based on the SC3 cluster analysis, tumor-infiltrating CD8⁺ T cells in C4_CD8-LAYN cluster were defined as exhausted T cells, while others as non-exhausted T cells. Differentially expressed genes (DEGs) between exhausted and non-exhausted T cells were detected using the linear model and the empirical Bayes method implemented by the R package limma, with stringent significance thresholds for adjusted p value (Benjamini-Hochberg multiple testing correction) < 0.01 and fold change ≥ 4 . The stringent threshold was designed to reduce false positives and find reliable exhaustion markers. We used the limma method because it leverages the highly parallel nature of genomic data, borrowing information between the gene-wise models. The relative expression (log2 fold-change) of each DEG within each patient was calculated and the mean of log2 fold change of all DEG for each patient were calculated to give the per-patient exhaustion score. The exhaustion score trended with the clinical stages. Tumor Treg-specific genes were also detected with the same method but different threshold (adjusted p value < 0.01 and fold change ≥ 2).

TCGA data analysis

The TCGA data were used to test the correlation between selected genes and patient survival. RNaseq v2 mRNA expression data and clinical parameters were retrieved from cBioportal. To correct the effect of T cell levels within each sample, the expression of selected genes in tumor were divided by that of the geometric mean of CD3 genes (*CD3D*, *CD3E* and *CD3G*). The relative expression was further discretized as follow: the median of relative expression minus (plus) 10% MAD (median absolute deviation) defined the lower (upper) threshold; and samples with relative expression below the lower threshold were categorized as the “low expression” group, and those with relative expression above the upper threshold were categorized as the “high expression” group. Cases with the relative expression between the two thresholds were excluded from survival analyses. The statistical analysis was performed by the R package “survival,” and survival curves were fitted by the *survfit* function and the difference between high and low expression group was test by *survdiff*.

DATA AND SOFTWARE AVAILABILITY

Data resources

HCC sequencing data

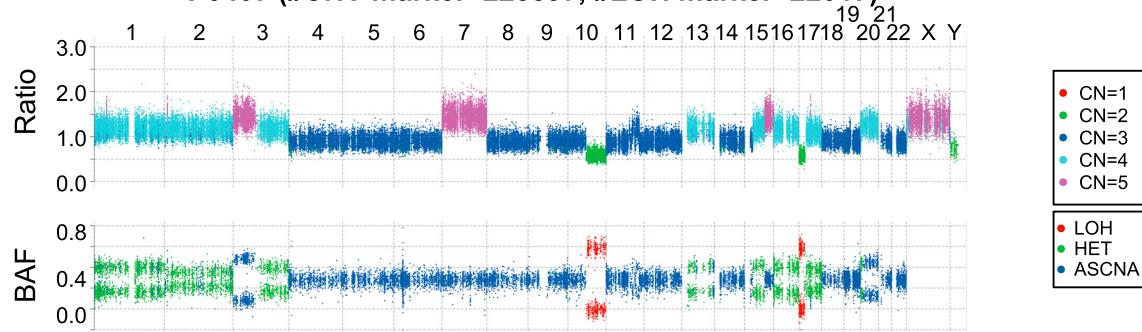
The accession numbers for the sequencing raw data and processed data in this paper are EGA: EGAS00001002072 and GEO: GSE98638. Analysis of such HCC data can also be found at <http://hcc.cancer-pku.cn>.

Supplemental Figures

Cell

A

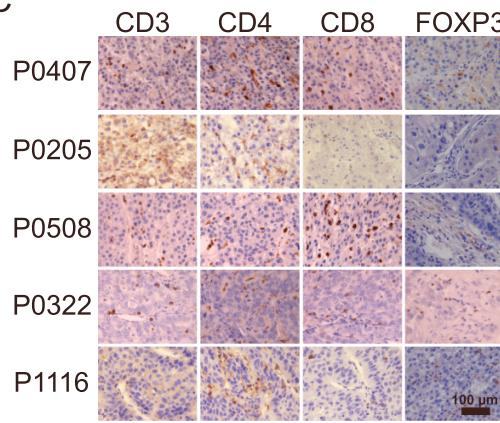
P0407 (#CNV marker=228897, #LOH marker=22017)



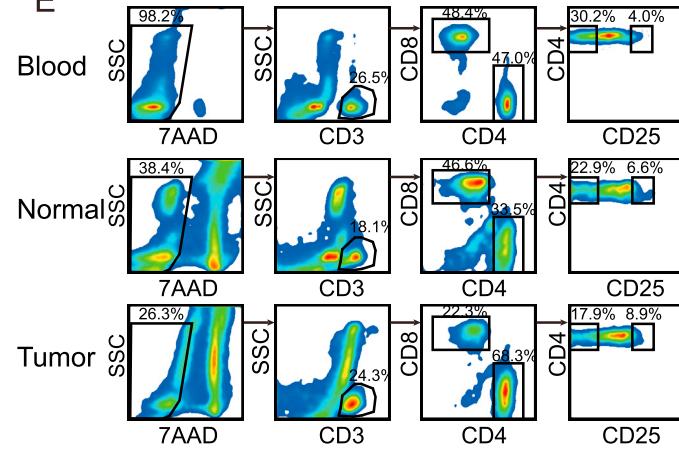
B

Patient ID	Genomic mutation	Variant frequency in tumor	Exonic Function	Gene	cDNA mutation	Protein mutation	Mutation annotation	Gene annotation
P0407	3:52651373,T>A	0.1825	missense SNV	PBRM1	c.A1768T	p.I590F	-	TSG
P0407	12:52379069,T>A	0.4085	missense SNV	ACVR1B	c.T1073A	p.V358D	-	TSG
P0407	17:7578266,T>A	0.5762	missense SNV	TP53	c.A583T	p.I195F	hotspot	TSG
P0407	17:70120050,A>T	0.3182	missense SNV	SOX9	c.A1052T	p.Q351L	-	TSG
P0205	2:47637405,A>T	0.2063	missense SNV	MSH2	c.A539T	p.D180V	-	TSG
P0205	10:89725053, TACTTCACAAAAACAGTAGAGGGAGCCGTCAA>T	0.0578	nonframeshift_deletion	PTEN	c.1037_1066del	p.346_356del	-	TSG
P0205	17:7577548,C>A	0.2044	missense SNV	TP53	c.G733T	p.G245C	hotspot	TSG
P0205	17:37873620,C>A	0.2015	missense SNV	ERBB2	c.C1785A	p.F595L	-	Oncogene
P0205	17:37879858,C>T	0.0536	missense SNV	ERBB2	c.C2153T	p.T718M	-	Oncogene
P0508	3:178952085,A>G	0.1005	missense SNV	PIK3CA	c.A3140G	p.H1047R	hotspot	Oncogene
P0508	6:157525129,A>T	0.2037	missense SNV	ARID1B	c.A5144T	p.Q1715L	-	TSG
P0322	17:7577534,C>A	0.3126	missense SNV	TP53	c.G747T	p.R249S	hotspot	TSG
P0322	17:7579854,GA>G	0.3465	frameshift_deletion	TP53	c.58delT	p.S20fs	-	TSG
P1116	10:89692922,T>C	0.1048	missense SNV	PTEN	c.T406C	p.C136R	hotspot	TSG

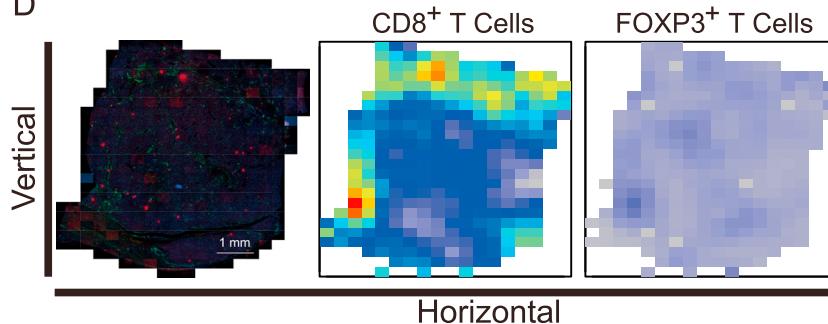
C



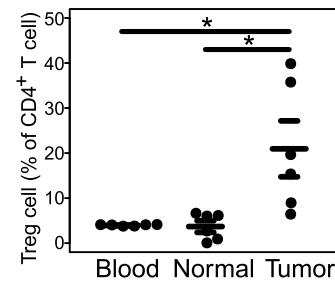
E



D



F



(legend on next page)

Figure S1. Characteristics of HCC Samples Used and the Presence of Multiple Types of Tumor-Infiltrating T Cells, Related to Figure 1

(A) DNA copy number profiles of a representative HCC tumor. The copy number information was obtained by ADTex and depicted in bin count plots across chromosomes. The read count ratios (Ratio, “1” in y axis means baseline copy number) and B allele frequencies (BAF) were shown. Various colored dots in the Ratio graph represents different copy number status of each segments. The red dots in BAF graph represents region of loss-of heterozygosity (LOH), the green dots heterozygosity (HET), and the blue dots allele-specific copy number alterations (ASCNA).

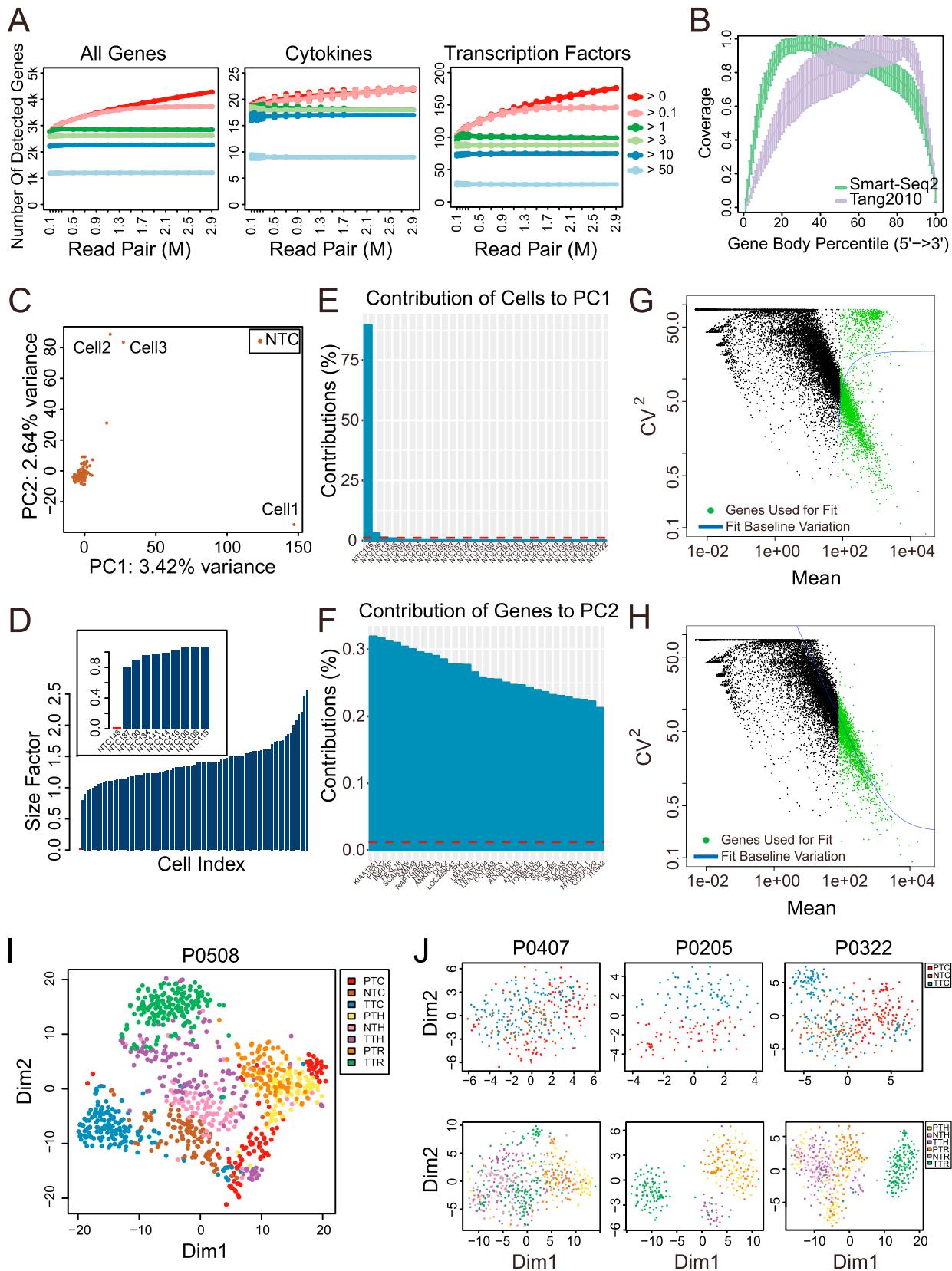
(B) Selected cancer associated somatic mutations detected in HCC tumors. The mutation annotations refer to COSMIC database version 70, and Gene Annotations refer to the classification scheme used by Vogelstein B et al. *Science* 2013;339(6127):1548-58.

(C) Immunohistochemistry images of tumor infiltrating T cells that express CD3, CD4, CD8 and FOXP3 in all five patients.

(D) Opal multicolor IHC staining images (left panel), with grids showing the distribution of CD8⁺ T cells (middle panel) and FOXP3⁺ T cells (right panel). Each grid represents one corresponding site in the left original picture, and the color of the grid represents the cell density.

(E) CD8⁺ T cells, T helpers and CD4⁺CD25^{high} T cells were sorted using flow cytometry according to their cell-surface markers. Antibodies for CD3, CD4, CD8, CD25 and 7AAD dye were used to enrich living target cells.

(F) Representative flow cytometric analysis of lymphocytes from tumor tissue (Tumor), normal tissue (Normal) and peripheral blood (Blood). Scatterplots indicating the frequency of Tregs among all CD4⁺ T cells in different tissues respectively. Error bars represent \pm SEM; *p < 0.05, Student's t test.



(legend on next page)

Figure S2. Coverage Assessment and Detection for Outlier Samples of Single-Cell RNA-Seq Data, Related to Figure 1

For outlier detection (C-H), PCA was performed on each type of cells from every patient to exclude abnormal cell data. The PCA analysis of all NTCs in patient P1116 is shown here as a representative.

(A) Saturation curves for the number of genes detected at different expression levels. Each point on the curve is derived from calculations based on the random selection of a fraction of raw reads from each sample, representing the average of 100 replicate sub-samplings. Standard error bars are depicted. Each line with different color shows how fast a gene can reach detection saturation at different expression levels, represented by a particular RPKM value. For genes with $\text{RPKM} > 1$, 0.5 M reads are sufficient for detecting the vast majority of such genes. When including genes with lower expression levels ($\text{RPKM} > 0.1$), it takes 1.3 million reads before reaching saturation, which is particularly true for genes with lower expression levels, such as cytokines and transcription factor genes.

(B) While data based on the [Tang et al. \(2010\)](#) instructions show a 3' bias, data based on Smart-Seq2 ([Picelli et al., 2014](#)) are more evenly distributed throughout the gene body and thus are more suitable for full-length gene assembly.

(C) Three highlighted cells are potential outliers and far away from the majority of NTC cells in P1116. The size factor of one PC1 outlier (NTC146-1116) is unusually small, compared with the others (D), but the apparent contribution is large (E).

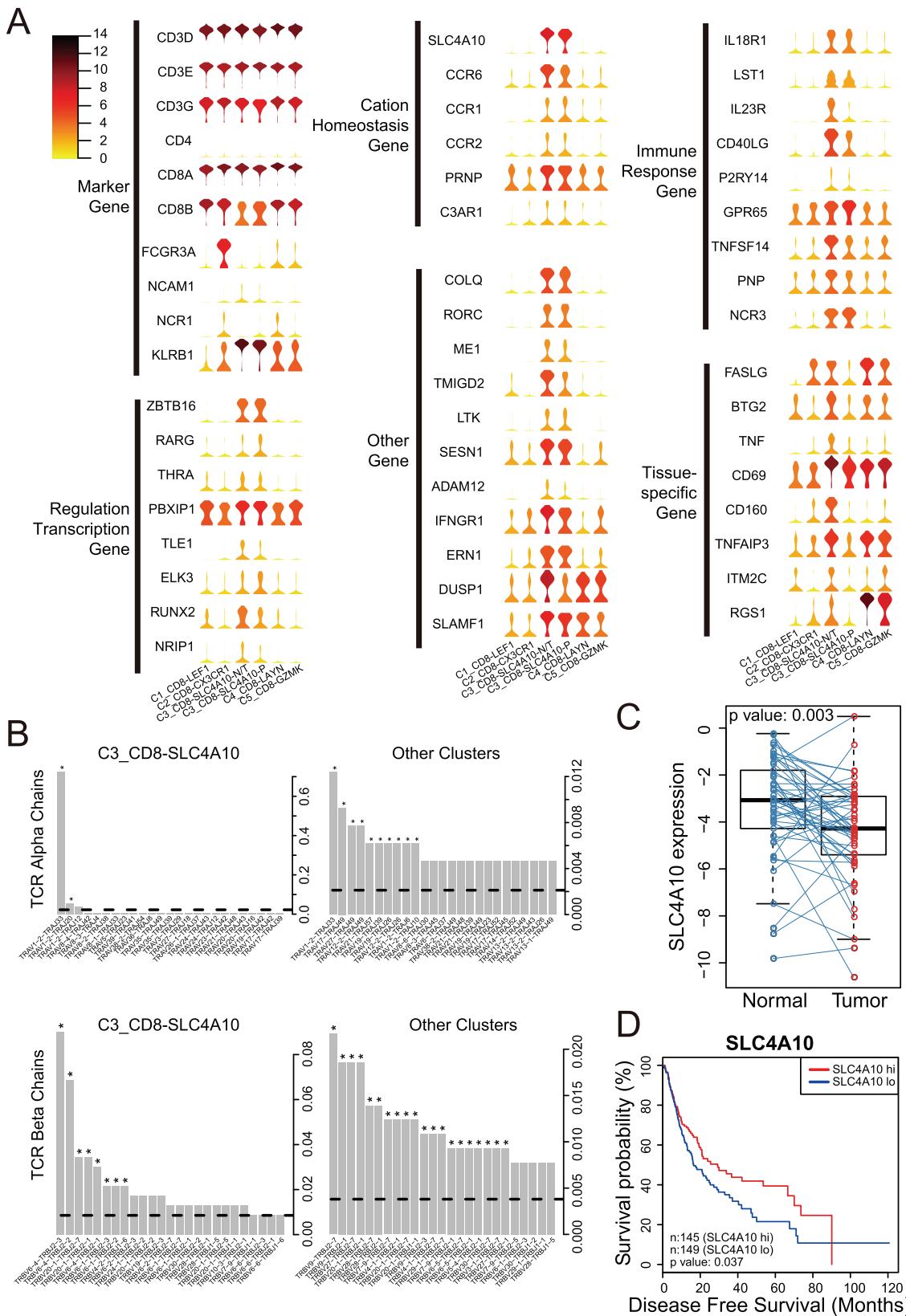
(F) The PC2 variance from cell 2 and 3 is primarily due to genes involved in the cell cycle; thus, these two cells are retained.

(G) The inclusion of the PC1 outliers would lead to aberrant squared coefficient of variation, based on expression fitting.

(H) The removal of PC1 outliers makes the fitting curve normal. For (G) and (H), each dot shows one gene, and the x axis shows the average expression levels across all cells in a population. The y axis shows the squared coefficient of variation. Green dots are genes used for fitting, and the blue line indicates the fitted relationship between technical noise and gene expression strength.

(I) 2D visualization of single-cell clusters using the t-SNE method in patient P0508. Each dot corresponds to one single cell and is colored according to its type and location. The cell type description is provided in [Figure 1](#).

(J) 2D visualization of single-cell clusters using the t-SNE method in patients P0407, P0205 and P0322. CD4⁺ and CD8⁺ T cells were projected separately.



(legend on next page)

- Figure S3. Highly Expressed Genes in CD8 Cells, TCR V/J Gene Usage and Association of SLC4A10 Expression Pattern with HCC Prognosis in the C3_CD8-SLC4A10 Cluster, Related to Figure 2**
- (A) The color of each violin represents the average expression value of a given gene in every one of the six cell clusters (four CD8⁺ T clusters and C3_CD8-SLC4A10 cells further divided into non-tissue origin (C3_CD8-SLC4A10-P) and tissue origin (C3_CD8-SLC4A10-N/T)). Cells with SLC4A10 expression showed high level of gene *CD3* and *CD8* but with low level expression of NK cell marker genes *FCGR3A* (*CD16*), *NCAM1* (*CD56*) and *NCR1* (*NKP46*), confirming their T cell identity but not NK cells. They also express high level of *ZBTB16*, *RORC*, *KLRB1* (*CD161*), *IL18R1*, *CCR2* and *CCR6*, consisting with gene expression profile of mucosal-associated invariant T (MAIT) cells. The expression is measured as the log₂(TPM+1).
- (B) The bar graphs show gene expression of TCR alpha and beta chains distribution in CD8⁺ SLC4A10 cells and other cells. The dashed lines represent the mean frequency without any bias in TCR allele expression, and the stars denote transcripts of certain TCR alleles that occur more frequently than the dashed line frequency. Only the top 25 genes with the highest frequency are shown in each panel.
- (C) HCC tumor samples from the TCGA cohort showed significantly ($p = 0.003$, Paired Wilcoxon Test) lower *SLC4A10* relative expression compared with paired normal samples in the 50 TCGA cases.
- (D) The TCGA HCC patients with lower *SLC4A10* relative expression in tumor had significantly ($p = 0.037$, Log-Rank Test) lower disease free survival. The hazard ratio was 1.39 (1.02 - 1.91, 95% CI).

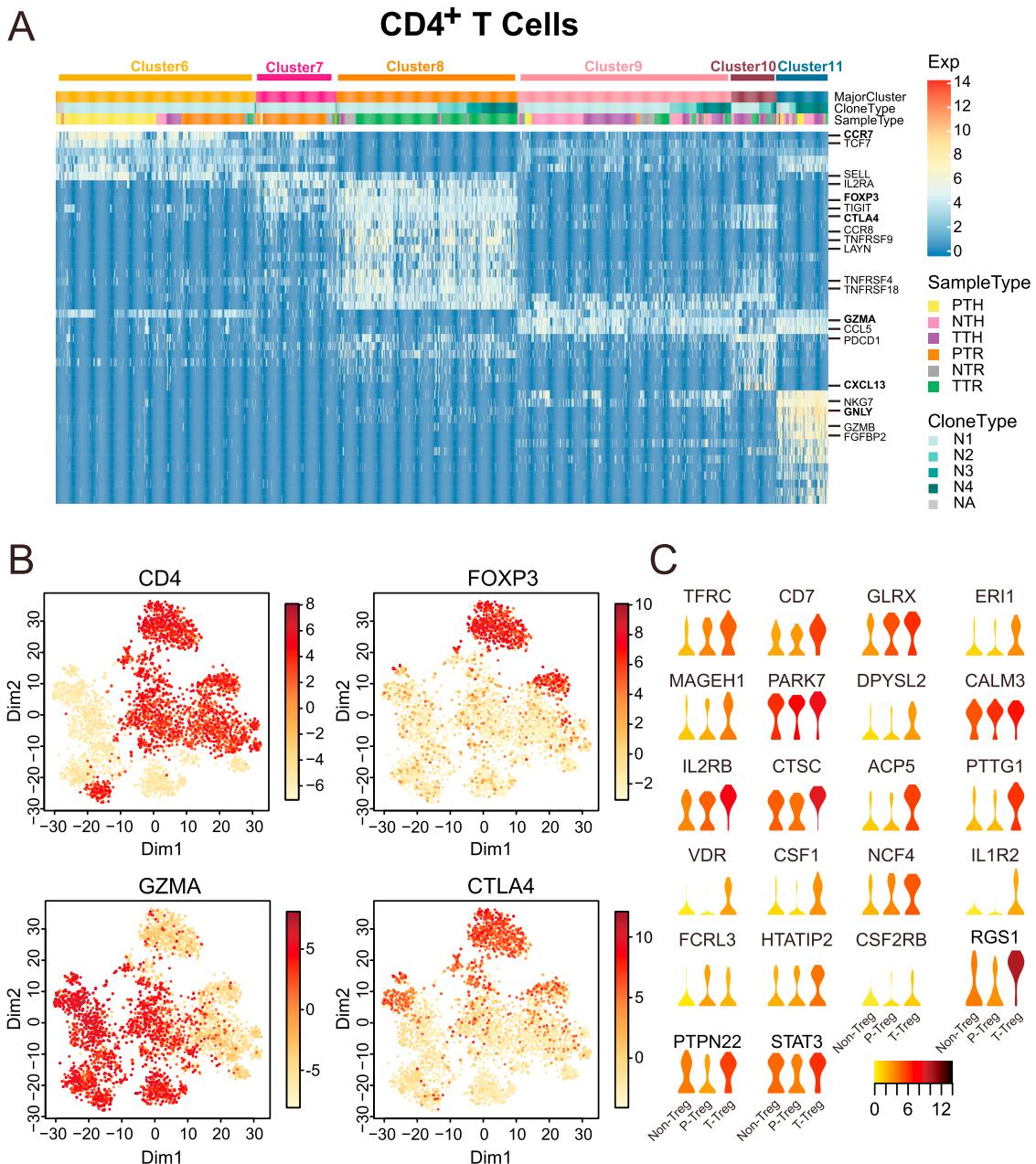


Figure S4. Clustering of CD4⁺ T Cells and Distribution of Selective Functional-Related Genes Expression, Related to Figures 2 and 3

(A) Heatmap of CD4⁺ T cells, with 6 main clusters identified, each containing a unique set of signature genes. Information regarding clonal status, tissue-of-origin and cell classification are also colored for each cell. Selective functional-related genes are marked to the right. Clone type N1/2/3 indicates that the cell TCR sequences were observed 1/2/3 times across the CD4⁺ T cells, whereas N4 indicates that it was observed ≥ 4 times, and NA indicates that no dominant TCR alpha/beta was observed in the cell.

(B) t-SNE projection of T cells, with each cell colored based on the relative normalized expression of CD4, FOXP3, GZMA and CTLA4.

(C) Violin plots showing expression comparison of selected Treg markers in this dataset. The expression is measured as the $\log_2(\text{TPM}+1)$.

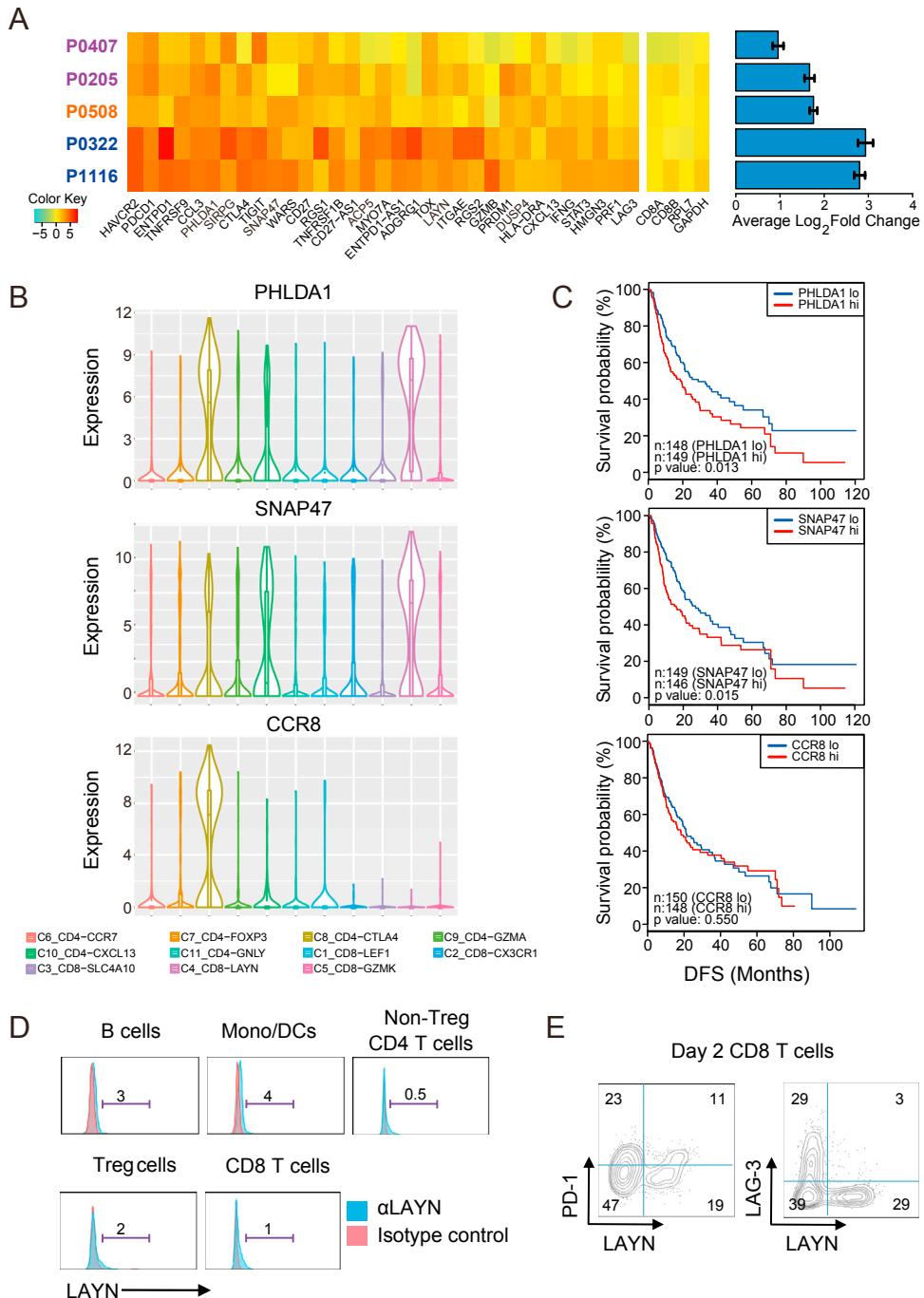


Figure S5. The Expression Pattern of Exhaustion-Related Genes in Different Tumors and Clinical Implications of PHLDA1, SNAP47 and CCR8 Expression in Liver Cancer, Related to Figures 3 and 4

(A) The expression pattern of exhaustion-related genes and cell-type marker genes in different tumors. The bar chart on the right shows the relative expression (log2FC) of the overall exhaustion-related genes between exhausted and non-exhausted TTCs in each tumor. Error bars represent \pm SEM.

(B) Violin plots comparing the expression distribution of PHLDA1, SNAP47 and CCR8 across 11 T cell clusters.

(C) Disease-free-survival (DFS) curve comparing the high and low expression of PHLDA1, SNAP47 and CCR8 based on the TCGA HCC cohort.

(D) FACS analyses showing the absence of LAYN expression in multiple resting immune cell populations isolated from human PBMCs. B cells: CD3⁻CD20; Monocyte/DCs: CD3⁻CD20⁻CD14; Non-Treg CD4 T cells: CD3⁺CD4CD25⁻; Tregs: CD3⁺CD4CD25⁺CD127⁻; CD8 T cells: CD3⁺CD8⁺CD20⁻. Representative data from three independent experiments with 6 different donors are shown.

(E) Human PBMCs were similarly stimulated as in Figure 4C for 2 days. Gated on CD8 T cells, LAYN, PD-1 and LAG-3 expressions were determined by FACS. Representative data from three independent experiments with 6 different donors are shown.

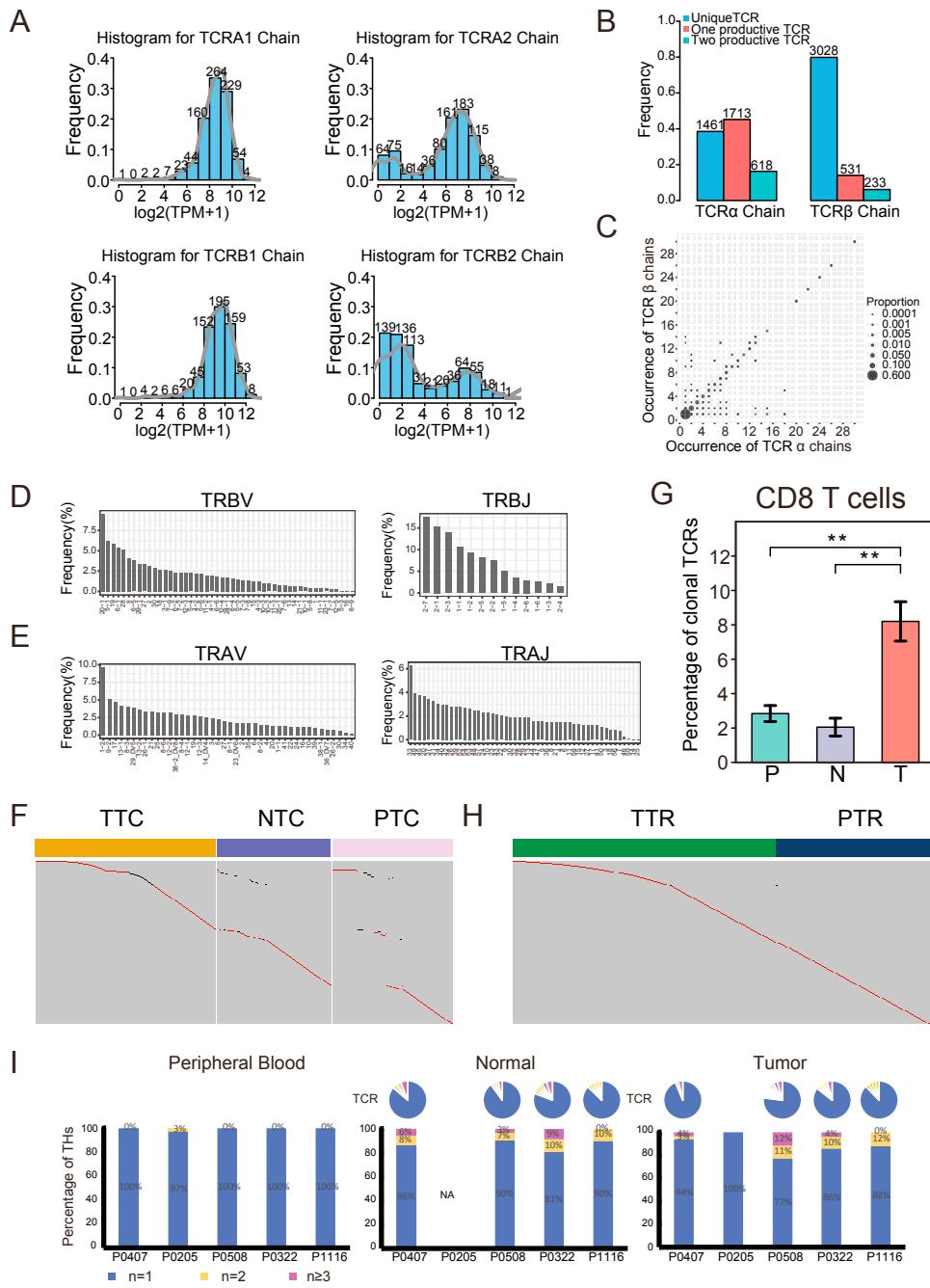


Figure S6. The Profile of TCR Usage in HCC-Related T Cells, Related to Figure 5

- (A) The distribution of abundance of TCR α - or β -chain expressed in single T cells. The gray lines represent the fitting values.
- (B) Bar plots showing the number of cells expressing unique or dual TCR alleles of α - and β -chains.
- (C) The correlation between the degree of recurrent usage of various TCR α chains with that of β -chains among T cells. Each dot represents a group of TCR alpha/beta allele expressed in a given number of cells. Dot size represents the proportion of such group in all TCR chains detected.
- (D) Bar plots show usage of expressed V- and J- alleles of TCR beta chain in all T cells.
- (E) Bar plots show usage of expressed V- and J- alleles of TCR alpha chain in all T cells.
- (F) TCR distribution of CD8 $^{+}$ T cells from different locations in patient P1116. Each column represents an individual cell, and each row represents a distinct TCR type. Most cells have unique TCR sequences, but some cells showed identical TCRs (multiple cells in one row, marked with black color).
- (G) The proportions of clonal TCRs of CD8 $^{+}$ T cells in different locations and patients. ** $p < 0.01$, Student's t test. P: peripheral blood; N: adjacent normal tissue; T: tumor tissue. Error bars represent \pm SEM.
- (H) TCR distribution in Tregs from different tissues in patient P0322.

(legend continued on next page)

(I) Bar graphs showing the fractions of unique and non-unique TCRs expression in T helper cells across peripheral blood, normal liver tissues and tumor tissues. Unique TCRs ($n = 1$), duplicated TCRs ($n = 2$) and the clonal TCRs (shared by at least 3 cells in a given cell population, $n \geq 3$) are labeled with different colors. Pie charts above each bar illustrate the composition of every individual TCR. NA indicates that no such cell type was analyzed for the given patient.

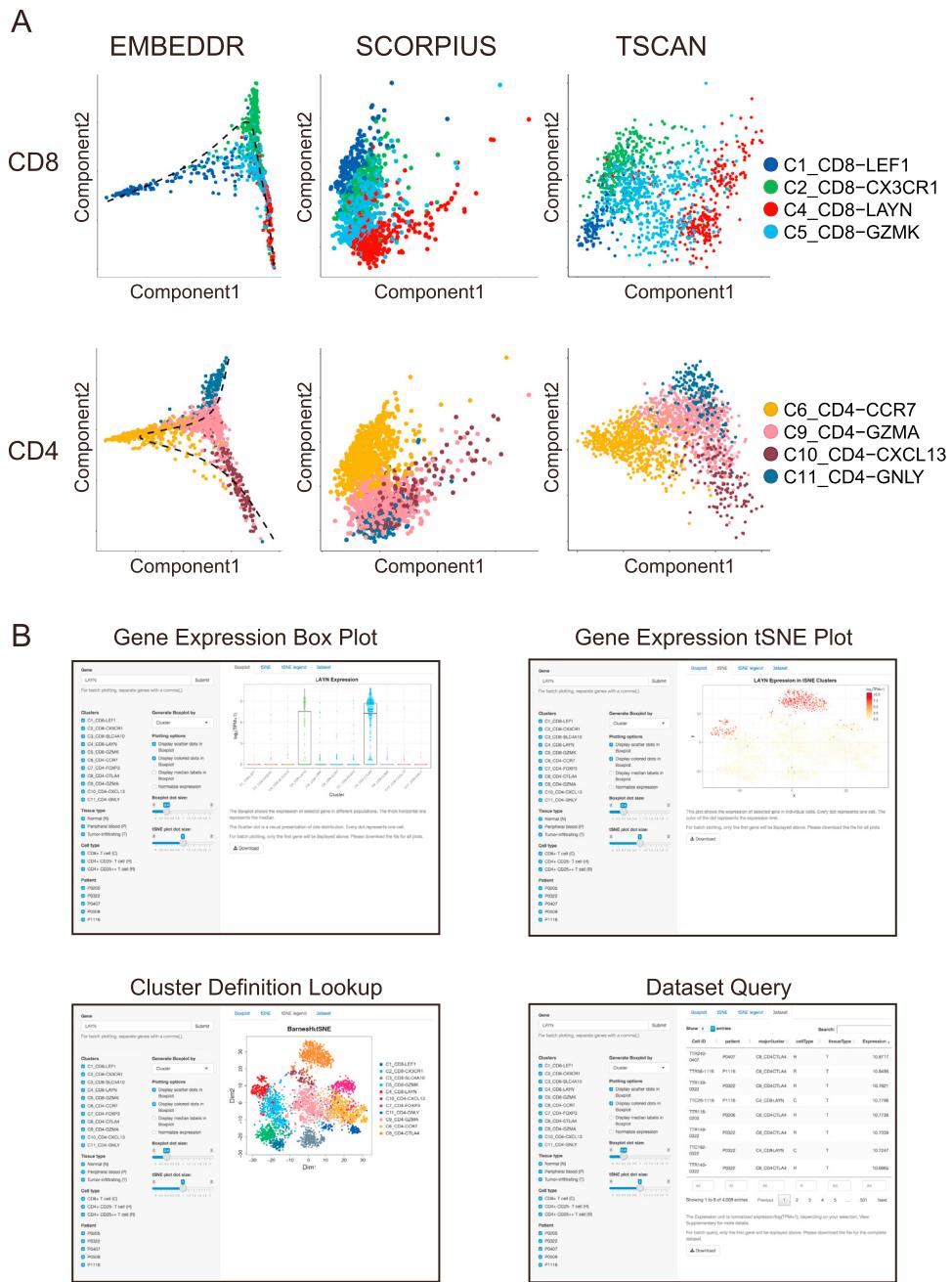


Figure S7. Additional Data Visualization of Single-Cell Expression Patterns, Related to Figure 6

(A) The trajectory analyses of CD8⁺ and CD4⁺ T cells by other trajectory inference methods EMBEDDR, SCORPIUS and TSCAN, with the same datasets and dispersed gene set as those used in Monocle 2.0. Each dot corresponds to a single cell, and each color represents a T cell cluster. The black dotted lines show the state transitional trajectories.

(B) We developed a web-based tool for interactive online data analysis, visualization, and downloading (<http://hcc.cancer-pku.cn>) for user-inputted single or multiple genes, based on the underlying HCC infiltrating T cell data.