

RESEARCH ARTICLE SUMMARY

CANCER IMMUNOLOGY

Pan-cancer single cell landscape of tumor-infiltrating T cells

Liangtao Zheng†, Shishang Qin†, Wen Si†, Anqiang Wang, Baocai Xing, Ranran Gao, Xianwen Ren, Li Wang, Xiaojiang Wu, Ji Zhang, Nan Wu, Ning Zhang, Hong Zheng, Hanqiang Ouyang, Keyuan Chen, Zhaode Bu*, Xueda Hu*, Jiafu Ji*, Zemin Zhang*

INTRODUCTION: Cancer immunotherapies that target tumor-specific T cells have benefited many cancer patients, but the clinical efficacy varies greatly among different cancer types. Tumor-infiltrating T cells often enter a dysfunctional state, widely known as T cell exhaustion, and the antitumor functions of effector T cells are regulated by multiple factors, including the presence of regulatory T cells (T_{reg} cells). The states and abundances of T cells vary across tumor microenvironments (TMEs) of different cancer types, which may fundamentally influence different clinical parameters such as drug response to immunotherapies.

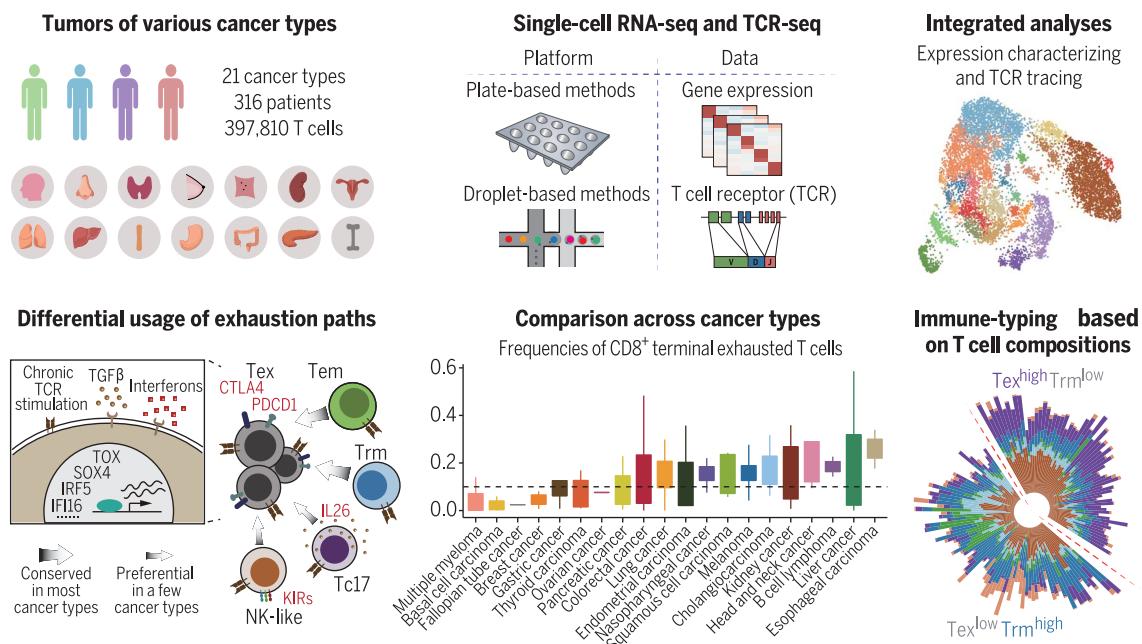
RATIONALE: To build a high-resolution pan-cancer T cell atlas, we performed single-cell RNA sequencing (scRNA-seq) on tumors, paracancerous tissues, and blood samples from patients of various cancer types and collected additional published scRNA-seq datasets. The diverse data were integrated after correcting confounding factors and batch effects. This atlas was composed of scRNA-seq data from 316 patients across 21 cancer types. T cell receptor (TCR) sequences of individual

T cells with gene expression profiles were assembled to characterize the expansion and dynamics of T cells. Various computational methods were applied to investigate the features and abundance of T cells across cancer types.

RESULTS: We identified multiple potentially tumor-reactive T cell (pTRT) populations in cancer patients. The states of the pTRTs varied dramatically in the tumor microenvironment of different cancer types. For $CD8^+$ T cells, the major pTRTs were exhausted T cells and exhibited high heterogeneity. We computationally inferred two major developmental paths to T cell exhaustion, through effector memory T cells and tissue-resident memory T cells, respectively, and both were prevalent among cancer types. We also noted the state transitions between terminal exhausted T cells and cells such as natural killer (NK)-like T cells, Type 17 $CD8^+$ T cells (Tc17 cells), and $CD8^+$ T_{reg} cells, but such transitions tend to occur in specific cancer types. For $CD4^+$ T cells, follicular helper T cell (T_{FH})/T helper 1 (T_{H1}) dual-functional T cells, which appeared to originate from T_{FH} cells, were also notable

pTRTs and correlated with the tumor mutation burden. We also found that the transcriptional programs of pTRTs could be affected by transforming growth factor- β (TGF- β) and interferons in the TMEs. The abundances of T cell states vary dramatically depending on cancer types. On the basis of tumor-infiltrating T cell compositions, cancer patients could be immune-typed as a group with high frequencies of terminal exhausted $CD8^+$ T cells and another group with high frequencies of tissue-resident memory $CD8^+$ T cells, and the immune types were associated with clinical traits such as patient survival and responses to immune checkpoint blockade.

CONCLUSION: We depicted the pan-cancer landscape of T cell heterogeneity and dynamics in the TME and established a baseline reference for future temporal or spatial studies associated with cancer treatments. The systematic comparison across cancer types revealed the commonalities and differences of T cell states in different TMEs. Our detailed signature, dynamics, and regulations of tumor-infiltrating T cells will facilitate the development of immunotherapies, and our proposed immune-typing can aid the therapeutic and diagnostic strategies that target T cells. ■



RESEARCH ARTICLE

CANCER IMMUNOLOGY

Pan-cancer single-cell landscape of tumor-infiltrating T cells

Liangtao Zheng^{1†}, Shishang Qin^{2†}, Wen Si^{1†}, Anqiang Wang³, Baocai Xing⁴, Ranran Gao², Xianwen Ren², Li Wang², Xiaojiang Wu³, Ji Zhang³, Nan Wu⁵, Ning Zhang⁶, Hong Zheng⁷, Hanqiang Ouyang^{8,9}, Keyuan Chen^{8,9}, Zhaode Bu^{2,10*}, Xueda Hu^{2,10*}, Jiafu Ji^{3,11*}, Zemin Zhang^{1,2*}

T cells play a central role in cancer immunotherapy, but we lack systematic comparison of the heterogeneity and dynamics of tumor-infiltrating T cells across cancer types. We built a single-cell RNA-sequencing pan-cancer atlas of T cells for 316 donors across 21 cancer types and revealed distinct T cell composition patterns. We found multiple state-transition paths in the exhaustion of CD8⁺ T cells and the preference of those paths among different tumor types. Certain T cell populations showed specific correlation with patient properties such as mutation burden, shedding light on the possible determinants of the tumor microenvironment. T cell compositions within tumors alone could classify cancer patients into groups with clinical trait specificity, providing new insights into T cell immunity and precision immunotherapy targeting T cells.

Tumor-infiltrating lymphocytes (TILs) are central players in the tumor microenvironment (TME), shaping fundamental clinical properties such as responses to immunotherapies. Immune checkpoint blockade (ICB) has shown tremendous clinical success, but its efficacy varies dramatically across cancer types, suggesting underlying differences of tumor immunity. Within the TME, effector T cells tend to exhibit high expression levels of multiple inhibitory receptors such as PD-1 (programmed cell death 1), TIM3 (T cell immunoglobulin

and mucin domain-containing protein 3), TIGIT (T cell immunoreceptor with Ig and ITIM domains), and LAG3 (lymphocyte activating 3) (1), which are considered to be hallmarks of a dysfunctional state, widely known as T cell exhaustion. The varied ICB efficacies could be logically linked to the tumor-infiltrating T cell state differences among cancer types, especially the exhaustion differences. In melanoma patients, CD8⁺ tumor-infiltrating T cells exhibit a linear and continuous progression from predysfunctional cell state to dysfunction (2), but in lung cancer patients, there are two pre-exhaustion states that could develop to exhaustion (3). Thus, the exhaustion dynamics may differ among TMEs of various cancers. Intrinsically, T cell exhaustion appears to be tightly regulated by several transcription factors (TFs), including TOX (thymocyte selection-associated high mobility group box) (4, 5) and TCF7 (transcription factor 7) (6), as well as epigenetic regulators that shape the specific state observed in dysfunctional CD8⁺ T cells (7). In addition, multiple TME factors contribute to the exhaustion phenotype (8), and distinct regulatory processes dictating the phenotypes and abundance of T cells may exist within the TMEs of various cancer types. Distinguishable T cell features have been observed in different cancer types. For example, liver and colon cancers have higher fractions of exhausted T cells than that of lung cancer (9), and cancer types such as multiple myeloma do not show notable exhausted T cell populations (10). However, direct comparative studies have been restricted to only three or four isolated cancer types (11, 12).

We constructed a comprehensive tumor-infiltrating T cell compendium across 21 distinct cancer types through single-cell RNA-

sequencing (scRNA-seq). By finding the commonalities and differences of tumor-infiltrating T cells, we aim to reveal the “pan-cancer” features of the T cell states, dynamics, and regulation.

Results

Construction of a pan-cancer single-cell transcriptome atlas of T cells

We compiled a single-cell transcriptome atlas of T cells across 21 cancer types (Fig. 1A). After stringent quality-control filtering, this atlas contained data for 397,810 T cells from 316 donors—derived from their tumors, adjacent normal tissues, and peripheral blood—of which 46.4% cells were newly sequenced in this study, whereas others were from previously published datasets (table S1). We integrated the diverse data generated from multiple technologies (fig. S1) on the basis of “minicluster” and the batch effect correction algorithm Harmony (figs. S2, A to D, and S3, A to D) (13, 14). Both visual and quantitative evaluations showed that cells were well mixed in the integrated data (figs. S2E and S3E). T cell receptor (TCR) sequences from individual cells were assembled for data generated through 10x VDJ (15) and Smart-Seq2 (16) protocols. A total of 168,901 cells from 92,533 clonotypes spanning 87 donors from 15 cancer types harbored at least one pair of productive TCR α chain and β chain, of which 53.9% were clonal cells (with identical TCR pairs found in at least two cells), corresponding to 14,631 expanded clonotypes (fig. S4).

A total of 17 CD8⁺ and 24 CD4⁺ metaclusters were identified, all of which were shared by at least 80% of cancer types (Fig. 1, B and C, and figs. S2F and S3F). Analysis of expression signatures of these metaclusters revealed the presence of both previously described T cell subtypes and new groups, including granzyme K-positive ($GZMK^+$) effector memory cells (T_{em} cells), terminally differentiated effector memory or effector cells (T_{emra} cells), and interferon-stimulated genes (ISG)-positive T cells in both CD4⁺ and CD8⁺ compartments; killer cell immunoglobulin-like receptor (KIR)-positive natural killer (NK)-like T cells, $ZNP683^+CXCR6^+$ tissue-resident memory T cells (T_{rm} cells), and four exhausted CD8⁺ T cell (T_{ex} cell) populations in the CD8⁺ compartment; and three follicular helper T cell (T_{FH} cell)-related populations [C-X-C motif chemokine receptor 5-positive (CXCR5⁺) pre- T_{FH} , classical $IL21^+T_{FH}$, and $IFNG^+T_{FH}/T$ helper 1 (T_{H1}) dual-functional T cells] and four regulatory T cell (T_{reg} cell) populations in the CD4⁺ compartment (fig. S5 and table S2). For the CD8⁺ metacluster c16, nearly half of the cells harbored the semi-invariant TCR α chains of mucosal-associated invariant T cells (MAIT) (fig. S6A), and cells with or without such TCR α chains both highly expressed genes related to Type 17 CD8⁺ T cells (Tc17 cells) (fig. S6B) (17, 18),

¹Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China. ²BIOPIC, Beijing Advanced Innovation Center for Genomics, School of Life Sciences, Peking University, Beijing 100871, China. ³Gastrointestinal Cancer Center, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing 100142, China. ⁴Department of Hepatopancreatobiliary Surgery I, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing 100142, China. ⁵Department of Thoracic Surgery II, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing 100142, China. ⁶Department of Urology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing 100142, China. ⁷Department of Gynecologic Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing 100142, China. ⁸Department of Orthopaedics, Peking University Third Hospital, Beijing 100191, China. ⁹Beijing Key Laboratory of Spinal Disease Research, Peking University Third Hospital, Beijing 100191, China. ¹⁰Analytical Biosciences Limited, Beijing 100084, China. ¹¹Department of Biobank, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing 100142, China.

*Corresponding author. Email: zemin@pku.edu.cn (Z.Z.); jjiafu@hsc.pku.edu.cn (J.J.); huxueda@abiosciences.com (X.H.); buzhaode@cjrcn.org (Z.B.).
†These authors contributed equally to this work.

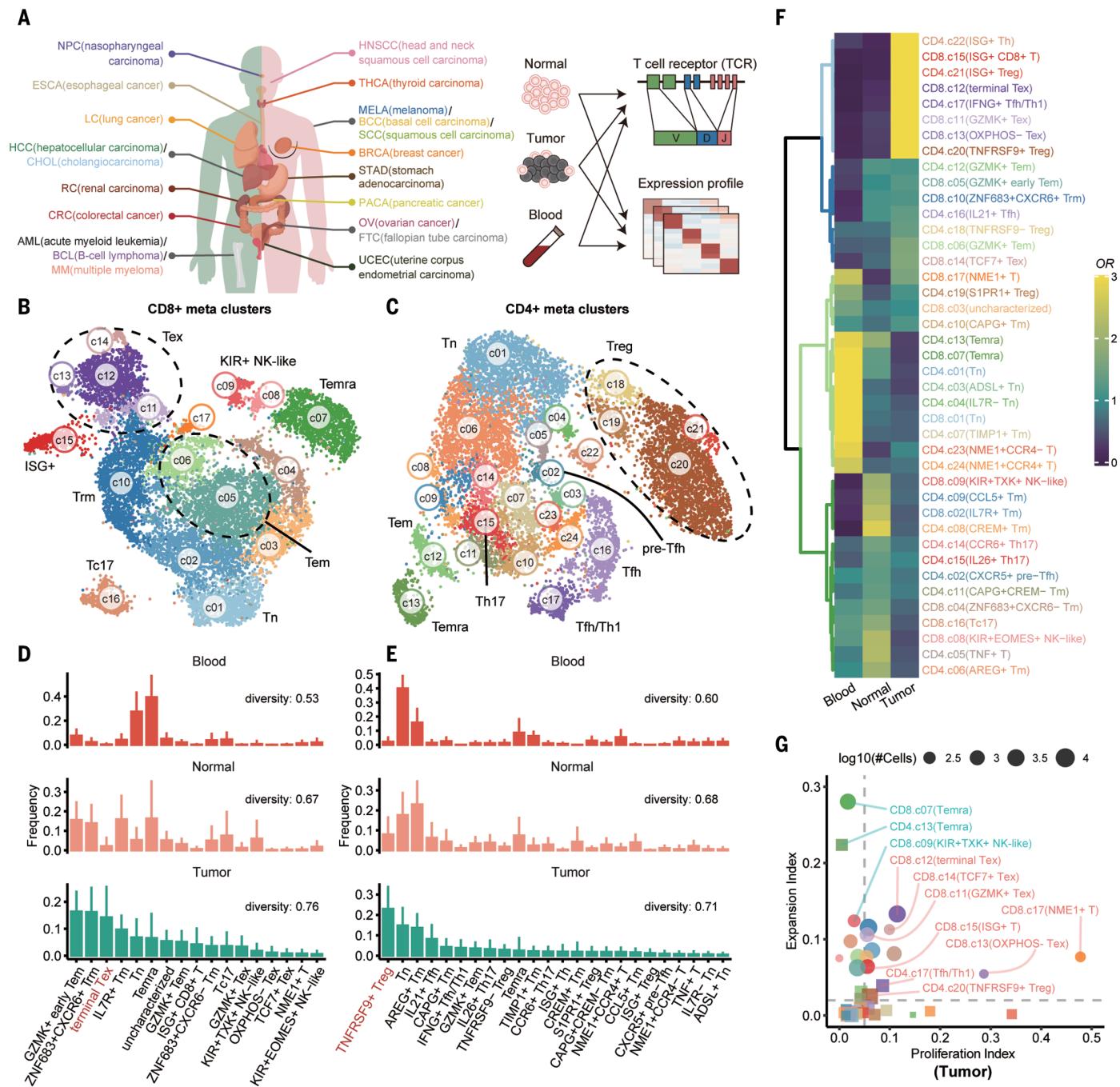


Fig. 1. Pan-cancer T cell profile at the single-cell resolution.

(A) Schematics of pan-cancer single-cell transcriptome and TCR profiling of T cells. (B) UMAP visualization of CD8⁺ T cell metaclusters. Selective metaclusters are highlighted by using their functional annotation: T_{ex}, exhausted T cells; ISG, interferon-stimulated genes; T_{emra}, terminally differentiated effector memory or effector; T_{emra}, effector memory T cells; T_{rm}, tissue-resident memory T cells; T_n, naïve T cells; and KIR, killer cell immunoglobulin-like receptors. (C) The same plot as in (B) applied to CD4⁺ T cells. (D) Bar plots showing the CD8⁺ T cell compositions in the blood ($n = 46$ patients), normal ($n = 82$ patients), and tumor ($n = 197$ patients) from treatment-naïve patients (mean \pm SD). Average diversity measured with the Shannon equitability index for each tissue is shown. A high index indicated

similar frequencies across different metaclusters. (E) The same plots as in (D) applied to CD4⁺ T cells from blood ($n = 15$ patients), normal ($n = 51$ patients), and tumor ($n = 163$ patients). (F) Heatmap showing the ORs of metaclusters occurring in each tissue. OR > 1.5 indicates that the metacluster is preferred to distribute in the corresponding tissue. Hierarchical clustering based on cosine distance is applied for rows. The naming, numbering, and colors of the metaclusters are in accordance with (B) and (C). (G) Scatter plot showing the expansion index and the proliferation index of metaclusters in the tumor. Metaclusters of high expansion (expansion index > 0.1 , $P < 0.01$) are highlighted in blue, and tumor-enriched metaclusters with significant expansion ($P < 0.01$) and proliferation (index > 0.05) are highlighted in red. Size indicates the number of cells (in log10 scale).

indicating that this metacluster contained both MAIT and non-MAIT Tc17 cells (designated Tc17 hereafter).

The compositions of T cells from different tissue-of-origin of treatment-naïve patients displayed prominent differences because the diversities measured with Shannon equitability index (19) in normal tissues and tumors were significantly higher than that in the blood ($P < 0.01$, two-sided Wilcoxon tests), and for CD8⁺ T cells, the diversity was even higher in tumors than that in normal tissues (Fig. 1, D and E, and fig. S7 A and B). CD8⁺ cells in the tumor were featured by the emergence of exhausted T cells (Fig. 1D and fig. S7C), whereas among CD4⁺ T cell populations, the most abundant population was the TNFRSF9⁺ T_{reg} cell, which showed significantly lower frequencies in both blood and normal tissues ($P < 0.01$, two-sided Wilcoxon tests) (Fig. 1E and fig. S7D).

To distinguish the T cells reacting to tumors from bystander T cells, we jointly analyzed the features of tissue distribution, transcriptional phenotypes, proliferation, and clonal expansion. The characteristics of proliferation and clonal expansion of tumor-enriched T cells have been viewed as evidence of their tumor reactivity (20). From the odds ratio (OR) analysis, the naïve T cells and T_{emra} cells (both CD4⁺ and CD8⁺) showed a strong distribution preference in blood, whereas TNFRSF9⁺ T_{reg} cells, T_{FH}/T_{H1} cells, CD8⁺ISG⁺ T cells, and four CD8⁺ T_{ex} cell groups appeared to be tumor-enriched (Fig. 1F and table S2). On the basis of the STARTRAC (single T cell analysis by RNA-seq and TCR tracking) analysis (9), such tumor-enriched metaclusters exhibited expansion and ongoing proliferation (Fig. 1G), implying their clonal expansion in response to tumor antigens. In addition, most of the tumor-enriched metaclusters with high expansion and proliferation indices (except for TCF7⁺ T_{ex} cells and OXPHOS-T_{ex} cells) tended to exhibit high activities in TCR signaling, confirming their high antigen reactivities (fig. S8). At the individual tumor level, we identified groups of cells that share clonotypes with specific expanded TILs that had high TCR signaling or proliferation, and these cells were collectively considered as potentially tumor-reactive T cells (pTRTs). For pTRTs in tumors, the most frequently observed cell states were terminal T_{ex} cells and TNFRSF9⁺ T_{reg} cells for CD8⁺ and CD4⁺ T cells, respectively (fig. S9A), although their occurrence varied among different cancer types (fig. S9C). Meanwhile, the T_{emra} cells of both CD4⁺ and CD8⁺ compartments exhibited significant expansion ($P < 0.01$, permutation test) but low proliferation (<2%) in all tissues (Fig. 1G and fig. S10). The STARTRAC migration indices, which quantify the extent of tissue migration, revealed that both CD8⁺ and CD4⁺ T_{emra} cells had the highest

mobility between blood and normal or tumor tissues in most tested cancer types (fig. S11, A and B). Those observations suggested that those T cells were activated and expanded outside the tumor and circulated in the blood. Compared with healthy donors, cancer patients harbored more CD8⁺ T_{emra} cells in the blood (fig. S11C). In addition, in certain tumors, most of the pTRTs were T_{emra} cells (fig. S9A). Further, the most frequent cell state of pTRTs in the blood was also T_{emra} cell (fig. S9B), and this pattern applied to multiple cancer indications (fig. S9D).

Taken together, potentially tumor-reactive T cells emerged—T_{FH}/T_{H1} cells, TNFRSF9⁺ T_{reg} cells, CD8⁺ISG⁺ T cells, and four T_{ex} cell populations, representing a local antitumor immune response—whereas the expanded CD8⁺ T_{emra} cells might also harbor tumor-specific TCRs, which is consistent with the notion of a systemic immune response (21).

Common themes of CD8⁺ T_{ex} cell heterogeneity and dynamics

In the CD8⁺ compartment, the major potentially tumor-reactive metaclusters were the four CD8⁺ T_{ex} cell populations, all highly expressing multiple exhaustion markers, including TOX, TIGIT, CTLA4 (cytotoxic T lymphocyte-associated protein 4), and TNFRSF9 (TNF-receptor superfamily 9), but they differed in gene expression and pathway activities (Fig. 2A and fig. S12). A major population was terminal T_{ex} cells, which exhibited higher expression of the gene ENTPD1 (ectonucleoside triphosphate diphosphohydrolase 1), which is related to terminal differentiation (22, 23). The terminal T_{ex} cells also highly expressed IFNG (interferon- γ) and GZMB (granzyme B), implying its intrinsic antitumor effector potential, and certain genes with unknown roles in T cell exhaustion, including MYOIE (myosine IE) and MYO7A (Myosin VIIA). A relatively rare T_{ex} cell population was TCF7⁺ T_{ex} cells, which had a lower level of HAVCR2 (hepatitis A virus cellular receptor 2) and LAG3 (lymphocyte activating 3) but specifically expressed a high level of TCF7 (Fig. 2A). TCF7 has been considered to be the key regulator of stem-like T cells in tumors (24). Additionally, TCF7⁺ T_{ex} cells highly expressed CD200, GNG4 (G protein subunit γ 4), IGFBP4 (insulin-like growth factor binding protein 4), IGFL2 (insulin growth factor-like family member 2), and genes related to lymph node migration [such as CCR7 (C-C motif chemokine receptor 7) and SELL (selectin L)] (Fig. 2A).

Next, we combined gene expression and TCR data to dissect the trajectories of T cell exhaustion. First, on a global scale, the diffusion map (25) and RNA velocity (26) showed that CD8⁺ T cells could develop from naïve T cells to either T_{emra} or T_{ex} cells (fig. S13, A and B), which is consistent with previous reports (3, 9). Second, for the T_{ex} branch, combining

Monocle3 (27), RNA velocity, and graph inference based on uniform manifold approximation and projection (UMAP) (28) and STARTRAC pairwise transition indexes (pTrans) (14), we inferred two paths from naïve to T_{ex} cell: the first path (P1) going through GZMK⁺ T_{em} cells [naïve cells to IL7R⁺ memory T cells (T_m cells) to GZMK⁺ T cells to terminal T_{ex} cells], and the second (P2) going through ZNF683⁺ T_{rm} cells (naïve cells to IL7R⁺ T_m cells to ZNF683⁺CXCR6⁺ T_{rm} cells to terminal T_{ex} cells) (Fig. 2, B and C, and figs. S13, C to E, and S14). The state transition to T_{ex} cells from both paths could also be observed in individual tumors (fig. S15). Although certain tumors exhibited preferential usage of P1 or P2, the state transition of both paths was high in other tumors, implying that both T_{em} and T_{rm} cells were involved in the antitumor immunity in those tumors (fig. S15C). The terminal T_{ex} cells had moderate pTrans with ISG⁺ T cells, which in turn were highly connected with multiple metaclusters (fig. S16A). The ISG⁺ state was not an independent state, but a mixture of T_{em}, T_{rm}, and other cells (fig. S17). In addition, TCR clonotypes that contained multiple cell states—including ZNF683⁺CXCR6⁺ T_{rm} cells or GZMK⁺ T_{ex} cells, ISG⁺ T cells, and terminal T_{ex} cells—could be clearly identified in tumors (fig. S16, B and C). Because the ISG⁺ state represents an activation state possibly driven by TCR-triggered interferon- γ (IFN- γ) or induced by interferons directly (29), these observations suggested that nonexhausted T cells in P1 or P2 could become interferon-responsive, before entering “exhaustion,” which is reminiscent of observations in chronic virus infections in which the CD8⁺ T cell-intrinsic type I interferon signaling skewed the differentiation to a more terminal effector state (30) or exhaustion state (31).

It has been hypothesized that the progenitor cells that express CXCR5 or TCF7 give rise to terminal T_{ex} cells (23, 32, 33). We found that TCF7⁺ T_{ex} cells had a strong state transition connection with GZMK⁺ T_{ex} cells and even terminal T_{ex} cells (fig. S18). A fraction of GZMK⁺ early T_{em} cells also expressed high levels of CXCR5 or TCF7 (fig. S19A). Those CXCR5⁺ or TCF7⁺ cells from GZMK⁺ early T_{em} cells were located near the branch point to T_{emra} or T_{ex} cells (fig. S13, A and D), representing cells with the developmental potential to different fates. By contrast, the TCF7⁺ T_{ex} cells more likely represented committed T_{ex} cells with certain stemness. Compared with GZMK⁺ early T_{em} cells, those CXCR5⁺ or TCF7⁺ cells from the TCF7⁺ T_{ex} cell population had a higher frequency of cells expressing TOX and inhibitory receptors such as PDCDI, TIGIT, and CTLA4 (fig. S19B). TCF7⁺ T_{ex} cells did not have a strong state transition with GZMK⁺ early T_{em} cells but were highly connected with NME1⁺ (NME/NM23 nucleoside diphosphate kinase 1)

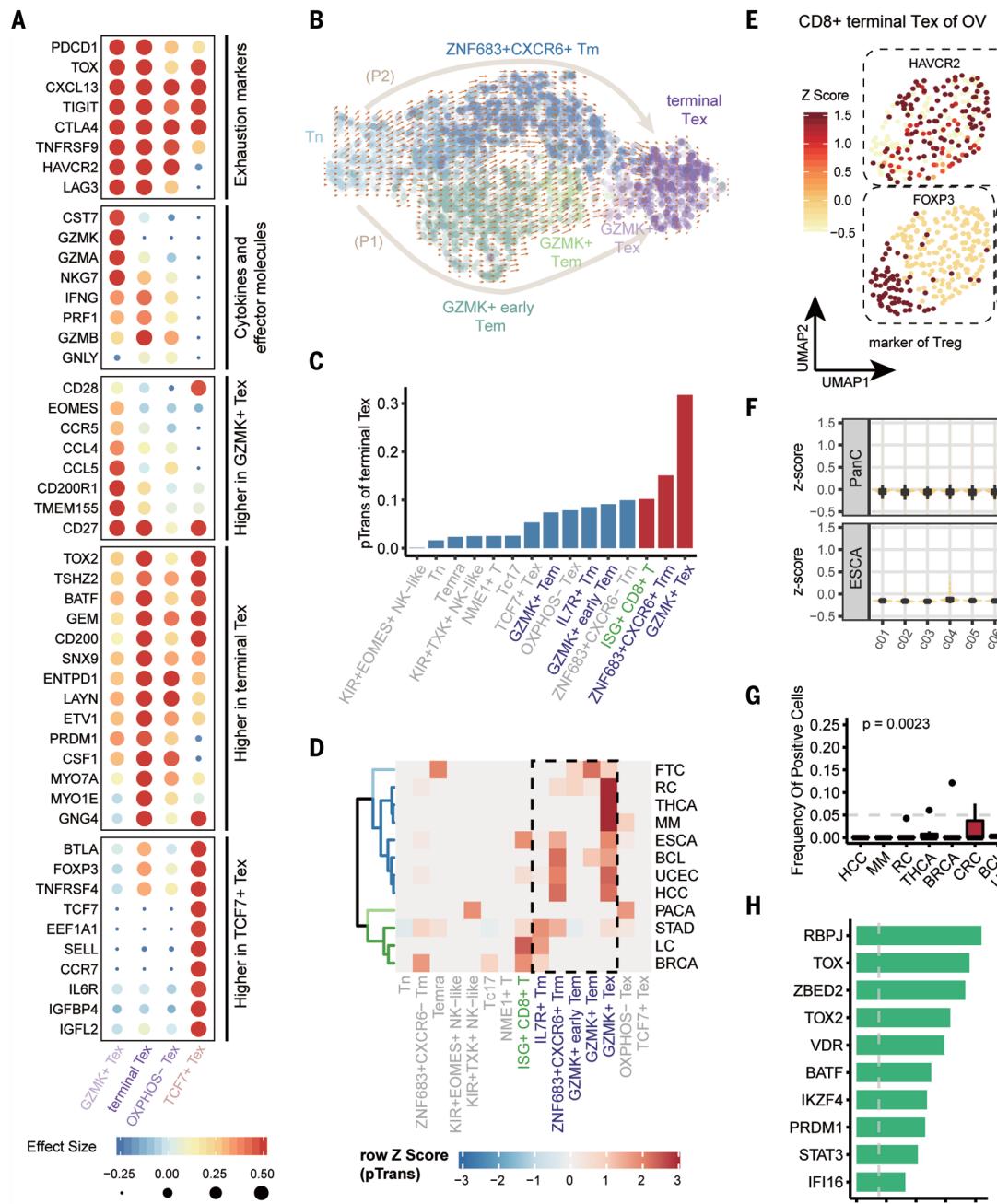


Fig. 2. Heterogeneity and dynamics of CD8⁺ exhausted T cells. (A) Dot plot showing the expression of signature genes of the four CD8⁺ exhausted T cells. Both color and size indicate the effect size. (B) RNA velocities overlaid on UMAP showing two major state transition paths from naïve to exhaustion. Arrows on a grid show the RNA velocity field, and dots are colored by metaclusters. (C) The pair-wise transition index (pTrans) of terminal T_{ex} cells. Three metaclusters, with which terminal T_{ex} cells are highly connected (pTrans > 0.1), are highlighted. (D) Heatmap showing the pTrans between terminal T_{ex} cells and other metaclusters, stratified by cancer types. Color represents the z-score scaled pTrans value by row. Metaclusters belonging to the two major paths to exhaustion are highlighted with a dashed line box. pTrans are whited out if the

value < 0.1, representing low transition potential. (E) UMAP visualization showing the expression of selected marker genes in terminal T_{ex} cells of ovarian cancer (OV). Color represents z-score-scaled gene expression values. (F) Violin and boxplot showing the expression of IL26 across metaclusters in pan-cancer (PanC) and esophageal cancer (ESCA). The expression pattern in pan-cancer serves as a “reference” comparing with the “ectopic” expression of genes IL26 in the terminal T_{ex} cells of ESCA. (G) Boxplot showing the frequency of IL26⁺ cells in terminal T_{ex} cells among different cancer types. The P value is calculated with the Kruskal-Wallis test. (H) Bar plot showing the effect size of the top 10 universal TFs of terminal T_{ex} cells. (I) Scatter plot showing the specificity scores of regulons of terminal T_{ex} cells. The top 10 regulons are highlighted.

T cells (figs. S18). Because of the scarcity of these two cell populations, it remains to be seen whether *TCF7⁺* T_{ex} cells were derived from *NMFI⁺* cells. Thus, we identified a T cell exhaustion path that cannot be defined with either *CXCR5* or *TCF7* alone. Taken together, we evidenced a more complex cellular pathway to T cell exhaustion.

Distinct paths to T cell exhaustion across different cancer types

To examine the prevalence of the major exhaustion paths among cancer types, we further stratified the TCR sharings by cancer types. In 11 out of the 12 cancer types with strong TCR sharings between terminal T_{ex} cells and other metaclusters, the terminal T_{ex} cells exhibited high pTrans values with *GZMK⁺* T_{ex} cells, *ZNF683⁺* T_{rm} cells, or other metaclusters in P1 or P2 (Fig. 2D), implicating P1 and P2 as the universal paths for T cell exhaustion. We also observed strong TCR sharings between terminal T_{ex} cells and metaclusters not in P1 or P2 in certain cancer types, including those with *KIR⁺TXK⁺* NK-like T cells in pancreatic cancer and those with Tc17 in breast cancer (Fig. 2D). These TCR sharing patterns suggested the presence of heterogeneous paths to T cell exhaustion in addition to the common paths through T_{em} and T_{rm} cells.

In addition to intercluster TCR sharing, the developmental connections between terminal T_{ex} cells and other cell populations could be confirmed by the expression of partial signature of nonexhaustion states in terminal T_{ex} cells. A subset of terminal T_{ex} cell of ovarian cancer expressed T_{reg} cell-dominant TF *FOXP3* (forkhead box P 3) (Fig. 2E and fig. S20A), and these CD8⁺*FOXP3⁺* T cells tended to share TCRs with CXCL13-expressing cells (fig. S20B), implying intracluster transition between CD8⁺*FOXP3⁺* T cells and CXCL13-expressing cells. Similarly, a subset of terminal T_{ex} cell of ovarian cancer expressed *KIR2DL3* (killer cell immunoglobulin like receptor, two Ig domains, and long cytoplasmic tail 3) and *TXK* (TXK tyrosine kinase), which are part of the signature genes of *KIR⁺TXK⁺* T cells (Fig. 2E and fig. S20A), and these cells shared TCRs with CXCL13-expressing cells (fig. S20B). In addition, the RNA velocity of *KIR2DL3⁺* cells pointed to CXCL13⁺ cells (fig. S20C), suggesting the transition from *KIR⁺TXK⁺* to exhaustion. Such state transitions might also occur in other cancer types because *FOXP3*-expressing cells, the *KIR⁺* T cell signature (fig. S20D), and similar TCR sharing patterns (fig. S20B) could be found in T_{ex} cells of multiple cancer types, although their frequencies varied among cancer types.

We further identified genes with cancer-type preference in terminal T_{ex} cells, including *IL26* (interleukin 26), *IL17A*, and *RORC* (RAR-related orphan receptor C). These genes are

primarily expressed in CD4⁺ T_{H17} cells or CD8⁺ Tc17 (18). We confirmed their expression in Tc17 (table S3), but in certain tumors of esophageal cancer, squamous cell carcinoma, and stomach adenocarcinoma, they were also expressed in a fraction of terminal T_{ex} cells (fig. S21A), especially for *IL26* (Fig. 2, F and G). Additionally, expanded clonotypes containing *IL26*, *IL17A*, or *RORC*-expressing T_{ex} cells tended to contain CXCL13-expressing T_{ex} cells (fig. S21B), implying the state transition connection between *IL26*, *IL17A*, and *RORC*-expressing cells and CXCL13-expressing cells. Furthermore, a fraction of the terminal T_{ex} cells expressed the semi-invariant α chain of MAIT, the signatures of exhaustion, and also a partial signature of Tc17 (fig. S22). Thus, there appeared to be a subset of T_{ex} cells with the capacity of secreting cytokines of type 17 response, which were likely derived from Tc17. The preferential expression of those genes—exemplified by *IL26*, which functions as an inflammatory mediator that induces the production of inflammatory cytokines in the mucosal tissues (34)—implied the multifunctional characteristic of T_{ex} cells in certain cancer types.

Universal and cancer type-specific transcriptional regulation of CD8⁺ T cell exhaustion

The identification of *TOX* as a critical TF of T cell exhaustion has sparked interest in finding additional regulators (4, 5). We systematically identified TFs associated with exhaustion. The signature genes of terminal T_{ex} cells encoding TFs and with significantly high expression [effect size > 0.15, false discovery rate (FDR) < 0.01 by meta-analysis] in >80% of cancer types were designated universal T_{ex} cell regulators. *TOX*, *TOX2*, *RBPJ* (recombination signal binding protein for immunoglobulin κ J region), *ZBED2* (zinc finger BED-type containing 1), *PRDM1* (PR domain zinc finger protein 1), *VDR* (vitamin D receptor), *IKZF4* (IKAROS family zinc finger 4), *BATF* (basic leucine zipper ATF-like transcription factor), *STAT3* (signal transducer and activator of transcription 3), and *IFI16* (interferon γ inducible protein 16) were ranked by effect size as the top 10 universal TFs (Fig. 2H, fig. S23A, and table S3). *TOX*, for example, showed statistical significance in all cancer types. Also, SCENIC analysis (35), which reconstructs regulons (TFs and their target genes), identified TFs that target *TOX*, including *NR5A2* (nuclear receptor subfamily 5 group A member 2), *ETV1* (ETS variant transcription factor 1), and *ARID5B* (AT-rich interaction domain 5B), which exhibited high regulon specificity in terminal T_{ex} cells (Fig. 2I and table S4) and showed statistical significance in >50% of cancer types (fig. S23A). Supporting such *TOX* regulation, a reanalysis of single-cell assay for transposase-accessible

chromatin with high-throughput sequencing (scATAC-seq) data of basal cell carcinoma (36) revealed that the high-accessibility peaks in the promoter or distinct intragenic enhancer of *TOX* in the terminal T_{ex} cell matched the motifs of these three TFs (fig. S23B). Although the exhaustion of T_{em} and T_{rm} cells shared multiple up-regulated TFs—such as *TOX*, *RBPJ*, and *ETV1*—pointing to these as common driving forces of T cell exhaustion, different exhaustion paths might also differ in the usage of TFs. For example, *BHLHE40* (basic helix-loop-helix family member E40) and *ZBTB32* (zinc finger and BTB domain containing 32) were featured in the late stage of P1, whereas *STAT1* and *IKZF3* were higher in the late stage of P2 (fig. S24). These observations revealed a finer regulation process of exhaustion.

We also identified key transcription regulators not previously known to be associated with T cell exhaustion, including *SOX4* and *FOXP3* (Fig. 2I and table S3). *SOX4*, in particular, had a high regulon specificity score and showed statistical significance in two-thirds of all cancer types (fig. S23A). As a downstream target of transforming growth factor- β (TGF- β), *SOX4* has been reported to play an important role in CXCL13-producing T_H cells (37) and to up-regulate the expression of exhaustion marker *ENTPD1* (CD39) in T_{reg} cells (38). Thus, we inferred that *SOX4* exerted similar functions in T_{ex} cells, although this should be further verified. *FOXP3*, also with a high regulon specificity score, showed statistical significance in only 47% of all cancer types (fig. S23A). Because *FOXP3* is important for T_{reg} cell functions, the connection between CD8⁺*FOXP3⁺* T cells (CD8⁺ T_{reg} cells) and exhaustion deserves further investigation. The frequencies of cells expressing *SOX4* or *FOXP3* varied significantly among cancer types ($P < 0.01$, Kruskal-Wallis tests) (figs. S20D and S23C), reflecting the differential impact of distinct TMEs on the phenotypes of T_{ex} cells.

Properties of potentially tumor-reactive T cells in the CD4⁺ compartment

In the CD4⁺ compartment, the major potentially tumor-reactive metaclusters were *IFNG⁺* T_{FH}/T_{H1} and *TNFRSF9⁺* T_{reg} cells. The global diffusion map and RNA velocity analyses revealed that CD4⁺ T cells could develop from naive T cells to T_{emra} cells, T_{FH}/T_{H1} cells, or *TNFRSF9⁺* T_{reg} cells separately (Fig. 3A). To gain a finer resolution of their developmental trajectories, we performed trajectory inference on each direction separately. The two T_{FH} cell-related metaclusters showed a gradual transition process from the classical *IL21⁺* T_{FH} cell to *IFNG⁺* T_{FH}/T_{H1} cells (Fig. 3, A and B, and fig. S25, A and B). Along this transition process, the type I response-related cytokines and cytotoxic effector molecules—including *IFNG*, *GZMB*, and *PRF1*—significantly increased (FDR < 0.01,

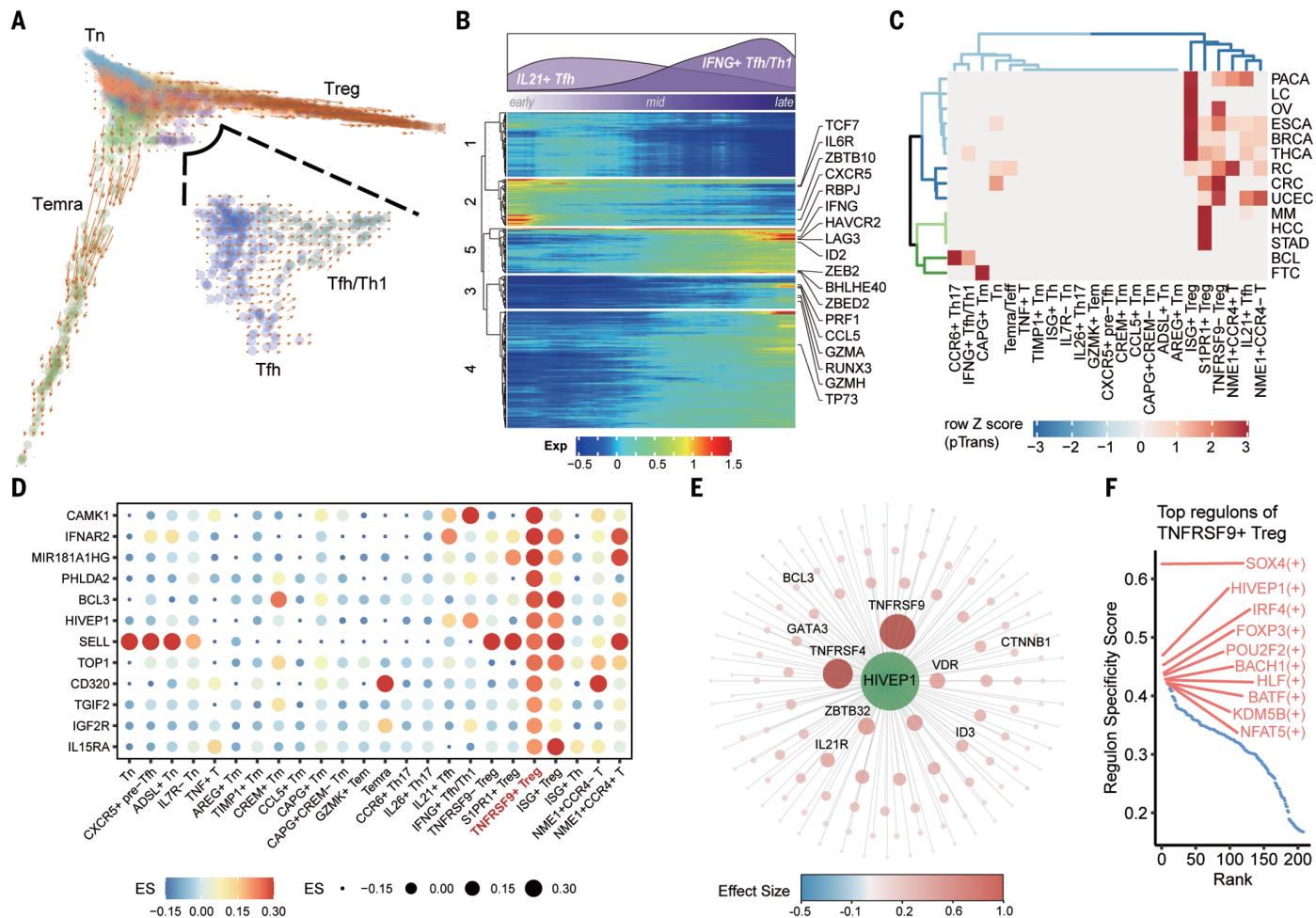


Fig. 3. Properties of potentially tumor-reactive CD4⁺ T cells. (A) Diffusion map of CD4⁺ T cells. Arrows on a grid show the RNA velocity field. (Inset) A similar diffusion map to zoom in the two T_{FH} cell metaclusters. (B) Heatmap showing genes with significant expression (absolute coefficient > 0.5 and FDR < 0.01, generalized additive model) changes along with the pseudotime. Color represents the z-score-scaled expression. The density plot of the distribution of the two T_{FH} cell metaclusters along the pseudotime is shown on top of the heatmap. (C) Heatmap showing the pTrans between

TNFRSF9⁺ T_{reg} cells and other metaclusters, stratified by cancer types. Color represents the z-score-scaled pTrans value; pTrans are whited out if the value < 0.01. (D) Dot plot showing expression of representative signature genes of TNFRSF9⁺ T_{reg} cells. Both color and size indicate the effect size (ES). (E) The transcriptional regulatory network showing the target genes of TF HIVEP1. Color represents the effect size. (F) Scatter plot showing the specificity scores of regulons of TNFRSF9⁺ T_{reg} cells. The top 10 regulons are highlighted.

generalized additive model) (Fig. 3B). Also, the TF *RUNX3* (runt-related transcription factor 3) exhibited elevated expression at a point at which a high density of T_{FH}/T_{H1} cells emerged (fig. S25C), which is consistent with a previous report that *RUNX3* regulated the cytotoxic phenotype in CD4⁺ cytotoxic T cells (39). TF *TP73* (tumor protein P73) appeared at an earlier point of pseudotime (fig. S25C) and was identified as the regulator of a regulon with high specificity in T_{FH}/T_{H1} cells (fig. S25D). These observations suggested other key players in acquiring and maintaining the phenotype of T_{FH}/T_{H1} cells.

For T_{reg} cells, a trajectory from TNFRSF9⁻ T_{reg} cell to TNFRSF9⁺ T_{reg} cell emerged (Fig. 3A and fig. S26, A to D), indicating a gradual transition from the resting state (TNFRSF9⁻)

to activated state (TNFRSF9⁺). The ISG⁺ T_{reg} cells were located at the center of the trajectory, suggesting that a fraction of T_{reg} cells responded to type I interferons in TME during activation, which is consistent with a recent report suggesting high ISG as a feature of intermediate state during CD4⁺ T cell activation (29). Such a T_{reg} cell developmental trajectory was common across cancer types (Fig. 3C). At the pan-cancer level, we did not observe obvious induction of T_{reg} cells from non-T_{reg} cell conventional T_H cells, but TNFRSF9⁺ T_{reg} cells exhibited certain state transition potentials with non-T_{reg} cells in a few cancer types (Fig. 3C). For example, the TNFRSF9⁺ T_{reg} cells were mainly connected with CCR6⁺ T_{H17} and T_{FH}/T_{H1} cells in B cell lymphoma, but instead had a connection with T_{FH} cells in the uterine corpus endometrial

carcinoma and pancreatic cancer. Thus, various conventional CD4⁺ T cell populations had conversion relationships with T_{reg} cells, but such conversion patterns were diverse and varied among cancer types.

We identified multiple TNFRSF9⁺ T_{reg} cell signature genes that have not been previously found (fig. S27), including those encoding membrane proteins with kinase activities [*CAMK1* (calcium/calmodulin dependent protein kinase I) and *IGF2R* (insulin like growth factor 2 receptor)], cytokine receptor [*IL15RA* (interleukin 15 receptor α)], known drug targets [*IFNAR2* (interferon α and β receptor subunit 2) and *TOP1* (DNA topoisomerase I)], and TFs [*TGIF2* (TGFB-induced factor 2 protein) and *HIVEP1* (HIVEP zinc finger 1)] (Fig. 3D). *HIVEP1* was inferred as a key regulator of 143

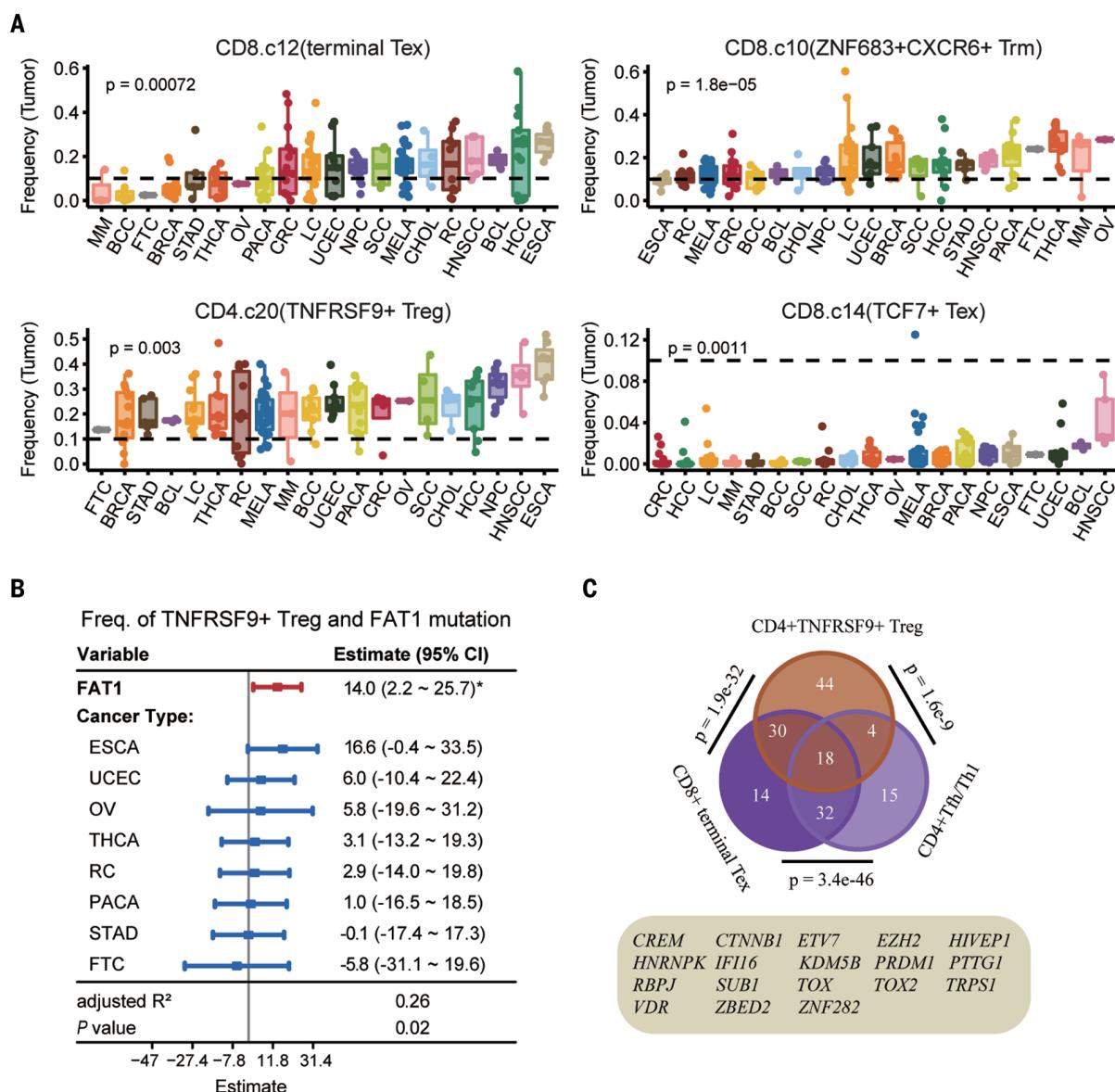


Fig. 4. TME shaping the landscape of tumor-infiltrating T cells. (A) Box plots showing the frequencies of metaclusters in tumors across cancer types. Only four metaclusters with significant differences (ANOVA, $P < 0.05$) among cancer types are shown (fig. S30). (B) Forest plot showing the association between *FAT1* mutation and the frequency of *TNFRSF9*⁺ T_{reg} cells in the tumor. The estimated coefficients and their 95% confidence intervals, the goodness of fitting (adjusted R^2), and the significance of the model are reported. * $P < 0.05$. (C) Venn diagram illustrating the overlap of signature genes encoding TFs of three metaclusters. The P values are calculated with hypergeometric tests, and the 18 signature genes shared by all the three metaclusters are highlighted.

target genes, most of which were also signature genes of *TNFRSF9*⁺ T_{reg} cells, including *TNFRSF4*, *TNFRSF9*, *ID3* (inhibitor of DNA binding 3), *IL2IR*, and *VDR* (Fig. 3, E and F, and table S4). *HIVEP1* accelerated its expression in a late stage of the T_{reg} cell trajectory, which is distinct from the pattern of other known T_{reg} cell TFs such as *FOXP3* and *BATF* (fig. S26E) that were “turned on” at an early stage of the T_{reg} cell trajectory.

TME shaping the landscape of tumor-infiltrating T cells

We examined extrinsic factors associated with distinct T cell compositions in the tumor.

Although we did not find a tight association between the frequency of each metacluster with age, gender (fig. S28A), or the clinical stages (fig. S28B), we detected a strong association between the frequency of *IL2IR*⁺ T_{FH} cells and the body mass index (BMI) [proportion of variance explained (PVE) >20%] (fig. S28C). Additionally, certain metaclusters clearly exhibited tissue specificity, supporting the effect by host tissues, particularly the liver-enriched T_{CI7} cells (figs. S28D and S29).

Cancer types exert an extensive impact on the frequencies of T cell populations because we observed distinct T cell distribution patterns across cancer types (Fig. 4A and figs. S28E

and S30). The median frequencies of terminal T_{ex} cells, for example, ranged from highly abundant (26.64% of $CD8^{+}$ T cells) in esophageal carcinoma to barely detectable in multiple myeloma (0.15%) [analysis of variance (ANOVA) $P = 0.00072$, PVE 13.1%]. Similarly, although the median frequencies of *TNFRSF9*⁺ T_{reg} cells were high in all cancer types (>10%), the variability was still high across cancer types (ANOVA $P = 0.003$, PVE 13.2%), with esophageal carcinoma as well as head and neck cancer exhibiting approximately twofold higher frequencies than that of breast and stomach cancers (Fig. 4A). Additionally, although the *TCF7*⁺ T_{ex} cell was a rare population, with

the head and neck squamous cell carcinoma showing the highest median value (2.63%), a subset of melanoma samples could reach a frequency of 12.5%. Further, although the CD8⁺ T_{emra} cell was underrepresented in the tumor, a few tumor samples of lung cancer and melanoma and a large fraction of renal tumors showed much higher frequencies of such cells (fig. S30). Thus, therapeutic strategies targeting the above cell subtypes should consider the variability across cancer types.

The tumor mutation burden (TMB) has been associated with the efficacy of ICB (40). By partitioning tumors into TMB-high and TMB-low groups (fig. S31A), we found that only the frequency of CD4⁺ T_{FH}/T_{H1} cells showed a strong correlation with TMB (FDR < 0.001, PVE > 26.3%) (figs. S28F and S32A). This association could be validated in pan-cancer and multiple individual cancer types using the bulk TCGA (The Cancer Genome Atlas) data (fig. S32B). We also identified a positive association between *FAT1* (fatty acid translocase 1) mutations and TNFRSF9⁺ T_{reg} cells (Fig. 4B and figs. S31B and S32C). Thus, the T cell composition in the tumor could be affected not only by the number of potential neoantigens, reflected by TMB but also by specific somatic mutations of cancer cells.

To reveal the overall pattern of T cell compositions across cancer types, we inspected the frequency correlations among metaclusters in the tumor and identified several highly correlated modules of metaclusters (fig. S33A). One module consisted of the three ISG⁺ metaclusters, whereas the four CD8⁺ T_{ex} cell populations, T_{FH}/T_{H1} cells, and TNFRSF9⁺ T_{reg} cells formed another module. Metaclusters with similar gene signatures but from different compartments (CD8⁺ Tc, CD4⁺ T_H, and CD4⁺ T_{reg} cells) tended to cluster together (for example, CD8⁺ Tc17 and CD4⁺ T_{H17} cells). The potentially tumor-reactive metaclusters were negatively correlated with certain metaclusters, which could be explained by several mechanisms, including the dynamic state transitions between metaclusters. For example, CD8⁺ terminal T_{ex} metacluster was negatively correlated with CD8⁺ ZNF683⁺ T_{rm} and CD8⁺ GZMK⁺ T_{em} metaclusters (fig. S33B) and showed an aforementioned state transition relationship. The positively correlated metaclusters usually had significant overlap between signature genes ($P < 0.01$, hypergeometric tests) (Fig. 4C and fig. S34), suggesting that the same regulators induced similar transcriptional programs in different T cell populations.

Using those overlapped signature genes and the NicheNet algorithm (41) to find shared ligands, we found that *RORC* and other type-17 response-related genes were potentially induced in both T_{H17} and Tc17 by ligands such as *IL23A* and *IL15* (fig. S35A). We identified shared ligands for the potentially tumor-reactive metaclusters. For the ISG⁺ metaclusters,

the type I interferon-encoding genes were expectedly ranked as the top ligands by the NicheNet analysis (fig. S35B). For terminal T_{ex}, T_{FH}/T_{H1}, and TNFRSF9⁺ T_{reg} cells, a total of 325 shared signature genes were identified, 18 of which were TFs (for example, *TOX*, *TOX2*, *VDR*, *ZBED2*, *ETV7*, *ZNF282*, and *HIVEP1*) (Fig. 4C and fig. S34). Thus, similar transcriptional machineries were likely used by the three different cell populations to respond to TME stimuli. Further, both terminal T_{ex} and T_{FH}/T_{H1} cells produced CXCL13 and type 1 response cytokines, and *TGFB1* was inferred as one of the top potential ligands inducing their shared signature genes (fig. S35C). Meanwhile, TNFRSF9⁺ T_{reg} cells highly express TGF-β-induced TFs, including *SOX4* and *TGIF2* (table S3). Moreover, *IFNB1* was inferred as one of the top potential ligands inducing the shared signature genes between TNFRSF9⁺ T_{reg} cells and terminal T_{ex} or T_{FH}/T_{H1} cells (fig. S35, D and E). These observations suggested that TGF-β and interferons may affect the transcriptional program and abundance of the potentially tumor-reactive T cells.

Immune types of pan-cancer defined by the composition of T cells

Next, using the frequencies of these correlated metaclusters, we found that tumor samples could be clustered into eight groups (C1 to C8) (Fig. 5A). The C1 and C2 harbored high frequencies of terminal T_{ex} cells, and C1 also had the highest frequency of TNFRSF9⁺ T_{reg} cells. Tumors of C3 to C8 harbored a low frequency of terminal T_{ex} cells and high frequency of CD8⁺ ZNF683⁺ CXCR6⁺ T_{rm} cells and could be further divided into groups dominated by naïve T cell (C7), enriched naïve T cell (C8), enriched T_{emra} cell (C6), enriched Tc17 or T_{H17} cell (C4), and with a low frequency of TNFRSF9⁺ T_{reg} cell (C5), respectively. On the basis of the linear model analysis, we found that our grouping could explain more variabilities of the T cell composition of the tumor than other factors (figs. S28 and S36). Although each immune type included mixed cancer types, certain cancer types exhibited clear preferences (fig. S37). For example, nearly half of esophageal and nasopharyngeal carcinoma tumors were of C1. By contrast, thyroid carcinoma and uterine corpus endometrial carcinoma were enriched in C3, suggesting that a large proportion of these two cancer types with high T cell suppression might still benefit from immunotherapy because of the presence of ISG⁺ activating T cells and the low abundance of terminal T_{ex} cells. More than half of melanomas were of C2, which has high T_{ex} cell but lower TNFRSF9⁺ T_{reg} cell frequency, which is consistent with their tendency to respond to ICB. Both basal cell carcinoma and hepatocellular carcinoma were enriched in C4, indicating that their tendencies are inflammatory

through IL17-producing T cells. These T cell-based immune types provide a reference to understand the overall tumor-infiltrating T cell properties, which may help guide the development of newer therapies and patient stratification instead of the conventional cancer type metrics.

Such immune type classification may have clinical implications. Using immune type signatures to stratify the TCGA cancer patients, we found that the T_{ex}^{lo}T_{rm}^{hi} (C3 to C8) tumors had better overall survival than that of T_{ex}^{hi}T_{rm}^{lo} tumors (C1 and C2) across cancer types or in multiple individual cancer types, including lung adenocarcinoma, hepatocellular carcinoma, and renal papillary cell carcinoma (Fig. 5B and fig. S38A). Because T cells are the direct target for many immunotherapies, the T cell-based immune types could logically be associated with the treatment efficacies. Reanalysis of published data of PD-1 antibody treatment for melanoma (42) indicated that responsive tumors had a lower frequency of terminal T_{ex} cells and a higher frequency of naïve T cells (Fig. 5C and fig. S38B). The T_{ex} cell connection was reproduced in another dataset (43), showing that the responder group was enriched with more T_{ex}^{lo}T_{rm}^{hi} tumors than that of the nonresponders (Fisher's exact test, $P = 0.025$) (fig. S38C). Pretreatment tumors in responders also had a higher frequency of Tc17 (Fig. 5C), implicating an important role of Tc17 in ICB treatment. Further investigation is needed to reveal how this finding is tied to the notion that Tc17 could also go into exhaustion in the tumor.

Discussion

We systematically characterized the T cells from various human cancers, investigating different aspects from gene expression signature and heterogeneity to state transitions and regulations. Multiple tumor-enriched metaclusters—including T_{ex}, T_{FH}/T_{H1}, and TNFRSF9⁺ T_{reg} cells—deserve particular attention because of their potential tumor reactivities. Our analyses revealed diverse paths to T cell exhaustion and the cancer type preference of those paths (fig. S39). Such landscape depiction deepens our understanding of cancer immunity and will facilitate therapeutic development.

The T cell states and infiltration in tumors are affected by multifaceted elements, such as tumor-intrinsic and metabolic factors (8). In our data, the TMB shows a positive association with T_{FH}/T_{H1} cells, whereas the BMI exhibits a positive association with T_{FH} cells. Because both TMB (40) and BMI (44) have been previously linked to ICB responses, our findings highlight the importance of T_{FH}-related cell populations in the antitumor response. Additionally, specific mutations could affect T cell compositions in the TME. *FAT1* mutations are positively correlated with TNFRSF9⁺ T_{reg} cell

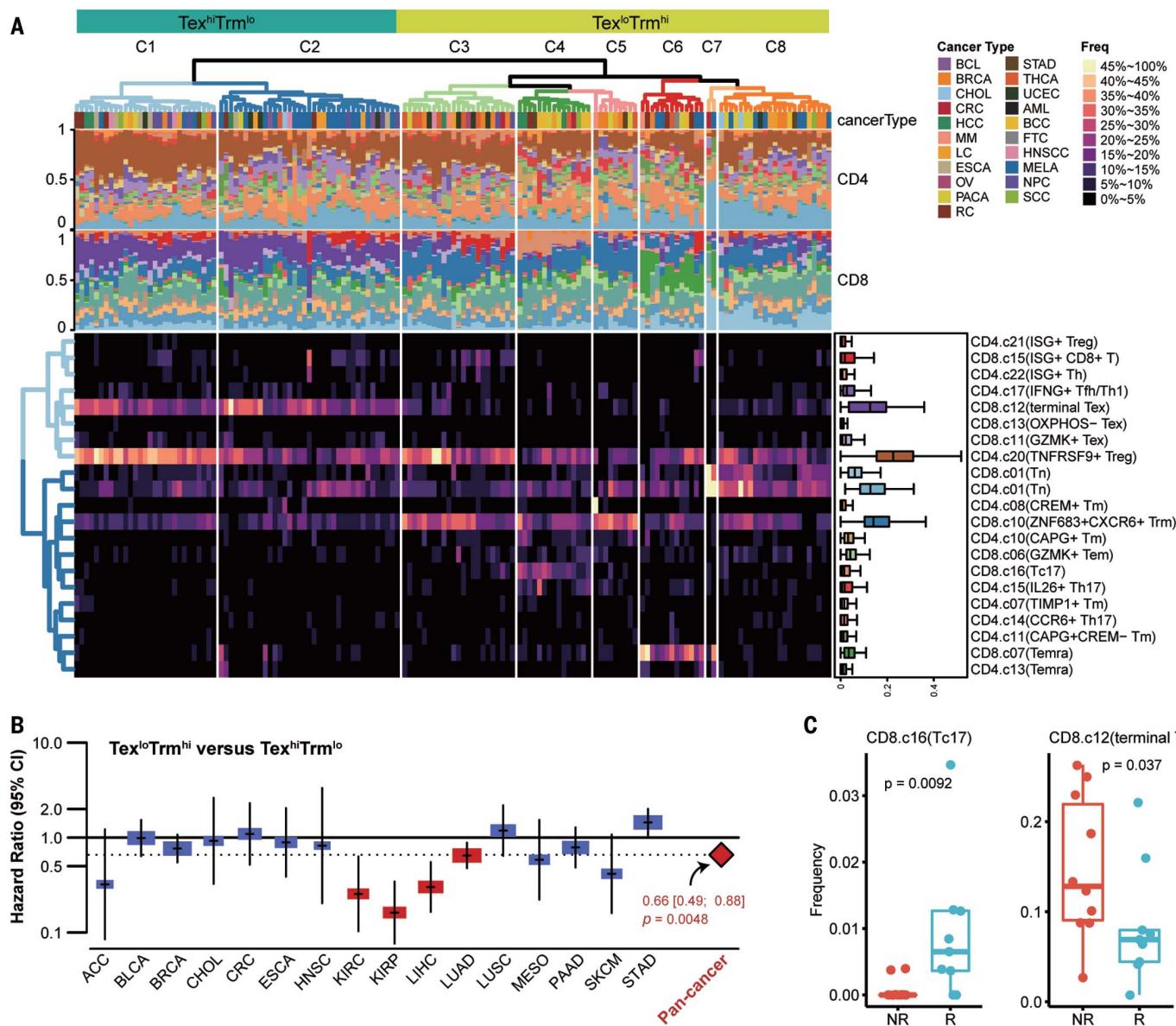


Fig. 5. Immune types of pan-cancer defined by the T cell composition.

(A) Heatmap reporting the frequencies of metaclusters in the tumor. Only metaclusters that have a correlation of >0.35 with at least one other metacluster are shown. The rows for metaclusters, the columns for tumor samples. For columns, stacked bar plots illustrate the proportion of metaclusters; for rows, a box plot illustrates the distribution of the metacluster proportion across samples. Hierarchical clustering based on Euclidean distance is applied. (B) Forest plot

reporting the effect of the two major T cell-based immune types on overall survival. The hazard ratios are calculated by using Cox regression models with the age, gender, and stage corrected. The black solid line for hazard ratio 1 (meaning no effect). Red for FDR < 0.05. (C) Boxplots comparing the frequencies of metaclusters in melanoma between nonresponders (NR) and responders (R) with antibody-to-PD-1 treatment. Published data from (42) is used. The P values by Wilcoxon tests are shown.

frequencies. This association might be mediated by the *FAT1*-Hippo-YAP1 (yes-associated protein 1) pathway dysfunction in cancer cells (45–47). The effect of tumor mutations on infiltrating immune cells is emerging, and mechanisms of such connection will likely be actively pursued in the future. Furthermore, our study demonstrates notable differences of T cell compositions in distinct TMEs of various cancer types and suggests an immune-typing scheme that leverages the overall tumor-

infiltrating T cell properties. With further tuning of the single cell-based immune-typing that could faithfully recapitulate the complex tumor-infiltrating T cell properties, we will be better informed when developing future immunotherapies that can be personalized to achieve maximal clinical benefit.

Methods summary

The scRNA-seq data of T cells were collected from both newly sequenced and previously

published datasets. For newly generated data, the cancer patients of origin were enrolled, pathologically diagnosed, and surgically biopsied at PKU Cancer Hospital and Institute, with approval of their Research and Ethical Committee. Written informed consents were obtained. The tumors and adjacent noncancer tissues were digested on the basis of gentleMACS and the related kit (Miltenyi, USA). CD45⁺ living cells were sorted by means of a BD FACSaria III sorter (BD Biosciences, USA)

from single-cell suspensions. The libraries of single-cell transcriptome and single-cell TCR were prepared by means of a 10x Chromium Single-cell 5' and VDJ library construction kit, then sequenced by means of a Hiseq X Ten sequencer (Illumina, USA).

We applied Cell Ranger (version 3.0) for gene expression quantification, TCR sequence assembly, and cell identification. Scrublet was used to remove potential doublets. Seurat v3 was used to identify T and NK cells. The CD3⁺CD8⁺CD4⁻ and CD3⁺CD4⁺CD8⁻ T cells were isolated according to computational gating and processed separately in downstream clustering and signature gene analysis.

To integrate heterogeneous data from different sources, a three-step procedure was applied. We first performed per-cell size-factor normalization and per-gene z-score scaling across cells for each dataset. Then, cells within each dataset were partitioned into small groups (miniclusters) to reduce noise. Subsequently, a batch effect correction algorithm, Harmony, was applied to further improve the integration. On the basis of the Harmony result, Seurat was applied to identify clusters, termed metaclusters. We used limma to identify differentially expressed genes among metaclusters. After estimating the moderated effect size of each dataset, the combined effect size was calculated by weighted averaging of the effect sizes. The Gene Set Enrichment Analysis (GSEA) (version 4.0.3) was performed to evaluate the pathway activities of metaclusters.

To characterize the metaclusters, using TCRs as markers, we applied STARTRAC to quantify the magnitude of T cell clonal expansion, migration potential, and state transition potential. A proliferation index, indicating the ongoing proliferation activity of a metacluster, was defined as the frequency of proliferating cells in a metacluster. The OR was used to characterize the tissue distribution of metaclusters.

To model the T cell state transition among metaclusters, we used multiple methodologies, including diffusion map, UMAP, monocle3, and RNA velocity. Specific clonotypes spanning different cell states with high likelihood ratios were also identified, providing direct and intuitive evidence for cell state transitions. We used SCENIC to construct the TF regulatory network. The NicheNet was applied to identify the potential ligands that induced the expression of genes of interest.

The bulk tumor and peripheral blood of patients were subjected to whole-exome sequencing for somatic mutation calling. TMB was calculated and tumors were divided as TMB-high and -low groups by using a cutoff of 10. Patient-matched tumors were also used for RNA-seq, and gene expression quantification was performed following the UCSC Xena Tool RNAseq pipeline.

REFERENCES AND NOTES

- L. Chen, D. B. Flies, Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat. Rev. Immunol.* **13**, 227–242 (2013). doi: [10.1038/nri3405](https://doi.org/10.1038/nri3405); pmid: [23470321](https://pubmed.ncbi.nlm.nih.gov/23470321/)
- H. Li *et al.*, Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma. *Cell* **181**, 747 (2020). doi: [10.1016/j.cell.2020.04.017](https://doi.org/10.1016/j.cell.2020.04.017); pmid: [32359441](https://pubmed.ncbi.nlm.nih.gov/32359441/)
- X. Guo *et al.*, Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24**, 978–985 (2018). doi: [10.1038/s41591-018-0045-3](https://doi.org/10.1038/s41591-018-0045-3); pmid: [29942094](https://pubmed.ncbi.nlm.nih.gov/29942094/)
- A. C. Scott *et al.*, TOX is a critical regulator of tumour-specific T cell differentiation. *Nature* **571**, 270–274 (2019). doi: [10.1038/s41586-019-1324-y](https://doi.org/10.1038/s41586-019-1324-y); pmid: [31207604](https://pubmed.ncbi.nlm.nih.gov/31207604/)
- O. Khan *et al.*, TOX transcriptionally and epigenetically programs CD8⁺ T cell exhaustion. *Nature* **571**, 211–218 (2019). doi: [10.1038/s41586-019-1325-x](https://doi.org/10.1038/s41586-019-1325-x); pmid: [31207603](https://pubmed.ncbi.nlm.nih.gov/31207603/)
- B. C. Miller *et al.*, Subsets of exhausted CD8⁺ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019). doi: [10.1038/s41590-019-0312-6](https://doi.org/10.1038/s41590-019-0312-6); pmid: [30778252](https://pubmed.ncbi.nlm.nih.gov/30778252/)
- D. R. Sen *et al.*, The epigenetic landscape of T cell exhaustion. *Science* **354**, 1165–1169 (2016). doi: [10.1126/science.aae0491](https://doi.org/10.1126/science.aae0491); pmid: [27789799](https://pubmed.ncbi.nlm.nih.gov/27789799/)
- D. S. Thommen, T. N. Schumacher, T cell dysfunction in cancer. *Cancer Cell* **33**, 547–562 (2018). doi: [10.1016/j.ccr.2018.03.012](https://doi.org/10.1016/j.ccr.2018.03.012); pmid: [29634943](https://pubmed.ncbi.nlm.nih.gov/29634943/)
- L. Zhang *et al.*, Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**, 268–272 (2018). doi: [10.1038/s41586-018-0694-x](https://doi.org/10.1038/s41586-018-0694-x); pmid: [30479382](https://pubmed.ncbi.nlm.nih.gov/30479382/)
- O. Zavidij *et al.*, Single-cell RNA sequencing reveals compromised immune microenvironment in precursor stages of multiple myeloma. *Nat. Cancer* **1**, 493–506 (2020). doi: [10.1038/s43018-020-0053-3](https://doi.org/10.1038/s43018-020-0053-3); pmid: [33409501](https://pubmed.ncbi.nlm.nih.gov/33409501/)
- T. D. Wu *et al.*, Peripheral T cell expansion predicts tumour infiltration and clinical response. *Nature* **579**, 274–278 (2020). doi: [10.1038/s41586-020-2056-8](https://doi.org/10.1038/s41586-020-2056-8); pmid: [32103181](https://pubmed.ncbi.nlm.nih.gov/32103181/)
- J. Qian *et al.*, A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res.* **30**, 745–762 (2020). doi: [10.1038/s41422-020-0355-0](https://doi.org/10.1038/s41422-020-0355-0); pmid: [32561858](https://pubmed.ncbi.nlm.nih.gov/32561858/)
- I. Korsunsky *et al.*, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019). doi: [10.1038/s41592-019-0619-0](https://doi.org/10.1038/s41592-019-0619-0); pmid: [31740819](https://pubmed.ncbi.nlm.nih.gov/31740819/)
- Materials and methods are available as supplementary materials.
- G. X. Y. Zheng *et al.*, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017). doi: [10.1038/ncomms14049](https://doi.org/10.1038/ncomms14049); pmid: [28091601](https://pubmed.ncbi.nlm.nih.gov/28091601/)
- S. Picelli *et al.*, Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014). doi: [10.1038/nprot.2014.006](https://doi.org/10.1038/nprot.2014.006); pmid: [24385147](https://pubmed.ncbi.nlm.nih.gov/24385147/)
- M. Dusseaux *et al.*, Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161hi IL-17-secreting T cells. *Blood* **117**, 1250–1259 (2011). doi: [10.1182/blood-2010-08-303339](https://doi.org/10.1182/blood-2010-08-303339); pmid: [21084709](https://pubmed.ncbi.nlm.nih.gov/21084709/)
- M. St Paul, P. S. Ohashi, The roles of CD8⁺ T cell subsets in antitumor immunity. *Trends Cell Biol.* **30**, 695–704 (2020). doi: [10.1016/j.tcb.2020.06.003](https://doi.org/10.1016/j.tcb.2020.06.003); pmid: [32624246](https://pubmed.ncbi.nlm.nih.gov/32624246/)
- C. E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948). doi: [10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x)
- A. M. van der Leur, D. S. Thommen, T. N. Schumacher, CD8⁺ T cell states in human cancer: Insights from single-cell analysis. *Nat. Rev. Cancer* **20**, 218–232 (2020). doi: [10.1038/s41568-019-0235-4](https://doi.org/10.1038/s41568-019-0235-4); pmid: [32024970](https://pubmed.ncbi.nlm.nih.gov/32024970/)
- M. H. Spitzer *et al.*, Systemic immunity is required for effective cancer immunotherapy. *Cell* **168**, 487–502.e15 (2017). doi: [10.1016/j.cell.2016.12.022](https://doi.org/10.1016/j.cell.2016.12.022); pmid: [28111070](https://pubmed.ncbi.nlm.nih.gov/28111070/)
- P. K. Gupta *et al.*, CD39 expression identifies terminally exhausted CD8⁺ T cells. *PLOS Pathog.* **11**, e1005177 (2015). doi: [10.1371/journal.ppat.1005177](https://doi.org/10.1371/journal.ppat.1005177); pmid: [26485519](https://pubmed.ncbi.nlm.nih.gov/26485519/)
- Z. Chen *et al.*, TCF-1-centered transcriptional network drives an effector versus exhausted CD8 T cell-fate decision. *Immunity* **51**, 840–855.e5 (2019). doi: [10.1016/j.immuni.2019.09.013](https://doi.org/10.1016/j.immuni.2019.09.013); pmid: [31606264](https://pubmed.ncbi.nlm.nih.gov/31606264/)
- I. Siddiqui *et al.*, Intratumoral Tcf1⁺PD-1⁺CD8⁺ T cells with stem-like properties promote tumor control in response to vaccination and checkpoint blockade immunotherapy. *Immunity* **50**, 195–211.e10 (2019). doi: [10.1016/j.immuni.2018.12.021](https://doi.org/10.1016/j.immuni.2018.12.021); pmid: [30635237](https://pubmed.ncbi.nlm.nih.gov/30635237/)
- L. Haghverdi, F. Buettner, F. J. Theis, Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015). doi: [10.1093/bioinformatics/btv325](https://doi.org/10.1093/bioinformatics/btv325); pmid: [26002886](https://pubmed.ncbi.nlm.nih.gov/26002886/)
- V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020). doi: [10.1038/s41587-020-0591-3](https://doi.org/10.1038/s41587-020-0591-3); pmid: [32747759](https://pubmed.ncbi.nlm.nih.gov/32747759/)
- J. Cao *et al.*, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019). doi: [10.1038/s41586-019-0969-x](https://doi.org/10.1038/s41586-019-0969-x); pmid: [30787437](https://pubmed.ncbi.nlm.nih.gov/30787437/)
- E. Becht *et al.*, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2018). doi: [10.1038/nbt.4314](https://doi.org/10.1038/nbt.4314); pmid: [30531897](https://pubmed.ncbi.nlm.nih.gov/30531897/)
- P. A. Szabo *et al.*, Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, 4706 (2019). doi: [10.1038/s41467-019-12464-3](https://doi.org/10.1038/s41467-019-12464-3); pmid: [31624246](https://pubmed.ncbi.nlm.nih.gov/31624246/)
- E. Stelekati *et al.*, Bystander chronic infection negatively impacts development of CD8⁺ T cell memory. *Immunity* **40**, 801–813 (2014). doi: [10.1016/j.jimmuni.2014.04.010](https://doi.org/10.1016/j.jimmuni.2014.04.010); pmid: [24837104](https://pubmed.ncbi.nlm.nih.gov/24837104/)
- T. Wu *et al.*, The TCF1-Bcl6 axis counteracts type I interferon to repress exhaustion and maintain T cell sterility. *Sci. Immunol.* **1**, eaai8593 (2016). doi: [10.1126/sciimmunol.aai8593](https://doi.org/10.1126/sciimmunol.aai8593); pmid: [28018990](https://pubmed.ncbi.nlm.nih.gov/28018990/)
- S. J. Im *et al.*, Defining CD8⁺ T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537**, 417–421 (2016). doi: [10.1038/nature19330](https://doi.org/10.1038/nature19330); pmid: [27501248](https://pubmed.ncbi.nlm.nih.gov/27501248/)
- D. T. Utschneider *et al.*, T cell factor 1-expressing memory-like CD8(+) T cells sustain the immune response to chronic viral infections. *Immunity* **45**, 415–427 (2016). doi: [10.1016/j.jimmuni.2016.07.021](https://doi.org/10.1016/j.jimmuni.2016.07.021); pmid: [27533016](https://pubmed.ncbi.nlm.nih.gov/27533016/)
- V. Larocette *et al.*, IL-26, a cytokine with roles in extracellular DNA-induced inflammation and microbial defense. *Front. Immunol.* **10**, 204 (2019). doi: [10.3389/fimmu.2019.00204](https://doi.org/10.3389/fimmu.2019.00204); pmid: [30809226](https://pubmed.ncbi.nlm.nih.gov/30809226/)
- S. Albar *et al.*, SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017). doi: [10.1038/nmeth.4463](https://doi.org/10.1038/nmeth.4463); pmid: [28991892](https://pubmed.ncbi.nlm.nih.gov/28991892/)
- A. T. Satpathy *et al.*, Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019). doi: [10.1038/s41587-019-0206-z](https://doi.org/10.1038/s41587-019-0206-z); pmid: [31375813](https://pubmed.ncbi.nlm.nih.gov/31375813/)
- H. Yoshitomi *et al.*, Human Sox4 facilitates the development of CXCL13-producing helper T cells in inflammatory environments. *Nat. Commun.* **9**, 3762 (2018). doi: [10.1038/s41467-018-06180-p](https://doi.org/10.1038/s41467-018-06180-p); pmid: [30232328](https://pubmed.ncbi.nlm.nih.gov/30232328/)
- M. C. Gerner *et al.*, The TGF-β/SOX4 axis and ROS-driven autophagy co-mediate CD39 expression in regulatory T-cells. *FASEB J.* **34**, 8367–8384 (2020). doi: [10.1096/fj.201902664](https://doi.org/10.1096/fj.201902664); pmid: [32319705](https://pubmed.ncbi.nlm.nih.gov/32319705/)
- H. Cheroutre, M. M. Husain, CD4 CTL: Living up to the challenge. *Semin. Immunol.* **25**, 273–281 (2013). doi: [10.1016/j.smim.2013.10.022](https://doi.org/10.1016/j.smim.2013.10.022); pmid: [24246226](https://pubmed.ncbi.nlm.nih.gov/24246226/)
- J. J. Havel, D. Chowell, T. A. Chan, The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* **19**, 133–150 (2019). doi: [10.1038/s41568-019-0116-x](https://doi.org/10.1038/s41568-019-0116-x); pmid: [30755690](https://pubmed.ncbi.nlm.nih.gov/30755690/)
- R. Browaeys, W. Saenels, Y. Saeyns, NicheNet: Modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020). doi: [10.1038/s41592-019-0667-5](https://doi.org/10.1038/s41592-019-0667-5); pmid: [31819264](https://pubmed.ncbi.nlm.nih.gov/31819264/)
- M. Sade-Feldman *et al.*, Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell* **176**, 404 (2019). doi: [10.1016/j.cell.2018.12.034](https://doi.org/10.1016/j.cell.2018.12.034); pmid: [30633907](https://pubmed.ncbi.nlm.nih.gov/30633907/)
- W. Hugo *et al.*, Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* **168**, 542 (2017). doi: [10.1016/j.cell.2017.01.010](https://doi.org/10.1016/j.cell.2017.01.010); pmid: [28129544](https://pubmed.ncbi.nlm.nih.gov/28129544/)
- Z. Wang *et al.*, Paradoxical effects of obesity on T cell function during tumor progression and PD-1 checkpoint blockade. *Nat. Med.* **25**, 141–151 (2019). doi: [10.1038/s41591-018-0221-5](https://doi.org/10.1038/s41591-018-0221-5); pmid: [30420753](https://pubmed.ncbi.nlm.nih.gov/30420753/)
- D. Martin *et al.*, Assembly and activation of the Hippo signalingome by FAT1 tumor suppressor. *Nat. Commun.* **9**, 2372 (2018). doi: [10.1038/s41467-018-04590-1](https://doi.org/10.1038/s41467-018-04590-1); pmid: [29985391](https://pubmed.ncbi.nlm.nih.gov/29985391/)
- M. Shibata, K. Ham, M. O. Hoque, A time for YAP1: Tumorigenesis, immunosuppression and targeted therapy. *Int. J. Cancer* **143**, 2133–2144 (2018). doi: [10.1002/ijc.31561](https://doi.org/10.1002/ijc.31561); pmid: [29696628](https://pubmed.ncbi.nlm.nih.gov/29696628/)
- L. M. Francisco *et al.*, PD-L1 regulates the development, maintenance, and function of induced regulatory T cells. *J. Exp. Med.* **206**, 3015–3029 (2009). doi: [10.1084/jem.20090847](https://doi.org/10.1084/jem.20090847); pmid: [20008522](https://pubmed.ncbi.nlm.nih.gov/20008522/)
- L. Zheng, S. Qin, Codes for the paper “Pan-cancer single-cell landscape of tumor-infiltrating T cells.” Zenodo (2021). doi: [10.5281/zenodo.5461803](https://doi.org/10.5281/zenodo.5461803)

ACKNOWLEDGMENTS

We thank Y. Yang, B. Zhang, Y. Gao, Y. Li, M. Gao, K. Wang, D. Lin, L. Zhou, and L. Jiang for clinical sample collection and F. Wang and X. Zhang for assistance with fluorescence-activated cell sorting (FACS). We thank the Computing Platform of the Center for Life Science and the National Center for Protein Sciences Beijing (Peking University). **Funding:** This project was supported by the Beijing Advanced Innovation Centre for Genomics at Peking University, the Beijing Municipal Science and Technology Commission (Z201100005320014), the National Natural Science Foundation of China (81988101, 91959000, 91942307, and 31991171), the Beijing Natural Science Foundation (7182075), Capital's Funds for Health Improvement and Research (2020-4-40916), and the Beijing Natural Science Foundation (7204327). W.S. was supported in part by the Postdoctoral Fellowship of

Peking-Tsinghua Center for Life Sciences. **Author**

contributions: Conceptualization: Z.Z., J.J., X.H., Z.B., and L.Z. Resources: A.W., X.W., J.Z., B.X., N.W., N.Z., H.Z., H.O., K.C., Z.B., and J.J. Methodology: Z.Z., X.H., L.Z., S.Q., W.S., and X.R. Investigation: R.G. and L.W. Formal analysis: L.Z. and S.Q. Writing, original draft: L.Z., S.Q., W.S., X.H., and Z.Z.

Competing interests: Z.Z. is a founder of Analytical BioSciences and is a consultant for InnoCare Pharma and ArsenalBio. **Data and materials availability:** The data presented in this manuscript are tabulated in the supplementary materials. Sequencing data are available at Genome Sequence Archive (accession no. PRJCA001702), and processed gene expression data are deposited in Gene Expression Omnibus (accession no. GSE156728) and can be accessed through an online data browser

(http://cancer-pku.cn:3838/PanC_T). All codes used for analysis are available from the Zenodo repository (48).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abe6474

Materials and Methods

Figs. S1 to S39

References (49–65)

MDAR Reproducibility Checklist

Tables S1 to S4

[View/request a protocol for this paper from Bio-protocol.](#)

4 September 2020; resubmitted 17 May 2021

Accepted 26 October 2021

10.1126/science.abe6474

Pan-cancer single-cell landscape of tumor-infiltrating T cells

Liangtao Zheng Shishang Qin Wen Si Anqiang Wang Baocai Xing Ranran Gao Xianwen Ren Li Wang Xiaojiang Wu Ji Zhang Nan Wu Ning Zhang Hong Zheng Hanqiang Ouyang Keyuan Chen Zhaode Bu Xueda Hu Jiafu Ji Zemin Zhang

Science, 374 (6574), abe6474. • DOI: 10.1126/science.abe6474

An atlas of cancer-associated T cells

The tumor microenvironment contains many different kinds of immune cells, the composition, function, and roles of which are unclear. Using single-cell RNA sequencing of T cells in 21 cancer types from more than 300 patients, Zheng *et al.* identified differences in transcript composition that could be used to catalog different T cell types (see the Perspective by van der Leun and Schumacher). These annotations identified the different roles of specific types of CD4 and CD8 T cells among the different tumor types. Some of these clusters revealed evidence for two developmental paths for T cells, one of which shows a trajectory toward the “exhausted” T cell state, knowledge of which may be useful in developing future cancer immunotherapies. —LMZ

View the article online

<https://www.science.org/doi/10.1126/science.abe6474>

Permissions

<https://www.science.org/help/reprints-and-permissions>