Genome Biology

RESEARCH

Open Access

Benchmarking single-cell tumor immune atlases and application for uncovering cell states related to immunotherapy response

Jing Yang[1,2], Yu Shyr[1,2*] and Qi Liu[1,2*]

*Correspondence: yu.shyr@vumc.org; qi.liu@vumc.org

[1] Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN 37203, USA
[2] Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN 37203, USA

Abstract

Background: The tumor immune microenvironment, containing a variety of immune cells with both tumor-promoting and anti-tumoral functions, plays a significant role in tumor immune surveillance and immunological evasion. Characterizing the landscape of the tumor immune microenvironment at the single-cell level is crucial for both cancer diagnosis and treatment strategy design. While current efforts to develop single-cell tumor immune atlases have laid a foundation for understanding the complexity and heterogeneity of the tumor immune microenvironment, existing atlases vary in their data sources, integration and annotation strategies, and the number and definition of cell types, posing challenges in selection.

Results: We systematically benchmarked five single-cell tumor immune atlases, comprising two pan-cancer and three cancer-specific ones. We first assessed their similarities and distinct characteristics of major immune cell subpopulations, including T, NK, B, macrophage, and dendritic cells. Next, we utilized each atlas as a reference to perform supervised annotation of six single-cell immuno-transcriptomics datasets, two with expert manual labels and four without. We evaluated annotation performance based on agreement with manual labels, mapping success, accuracy, clusterability, annotatability, and stability. Notably, supervised annotations consistently outperformed unsupervised clustering in identifying cell states related to immunotherapy response.

Conclusions: Our study provides insights into the characteristics and quality of existing atlases, demonstrating their utility in delivering harmonized annotations across datasets and uncovering crucial immune components associated with immunotherapy response. It also highlights key requirements and directions for the development of future atlases.

Keywords: Tumor immune microenvironment, Tumor immune atlas, Single-cell, Benchmarking, Immunotherapy

Background

The tumor microenvironment (TME), a highly structured and continuously evolving ecosystem surrounding cancer cells [1], comprises a variety of infiltrating and resident immune cells, stromal cells, secreted factors, and extracellular matrix [2, 3]. Notably, within this milieu, a heterogenous collection of immune cells with dual roles in tumor-promoting and anti-tumoral functions is a critical component, playing a significant role in tumor immune surveillance and immunological evasion [4, 5]. Therefore, characterizing the complexity and heterogeneity of the tumor immune microenvironment (TIME) at the single-cell level is imperative for both cancer diagnosis and treatment strategy design. The advancement of single-cell and spatial omics technologies has provided an unprecedented opportunity to profile and characterize the TIME [6, 7].

Given the profound variations in immune compositions and their functional states among patients, considerable efforts have been devoted to the development of single-cell tumor immune atlases by jointly analyzing extensive datasets from various studies and cancer types [8–14]. These tumor immune atlases, which are expected to lay the foundation of understanding the TIME heterogeneity, exhibit considerable diversity in terms of data sources, integration and annotation methodologies, as well as the number and definition of immune cell populations (Table 1). Some atlases focus on one specific cancer type, such as advanced biliary tract cancer (ABTC) [8], hepatocellular carcinoma (HCC) [9], and gastric cancer (STAD) [10]. Others are generated from multiple cancer types (pan-cancer), such as the Tumor Immune Cell Atlas (TICA) [11], Pan-Cancer Tumor T-cell Atlas (pan-T) [12], T cell reference map (TcellMap) [13], and pan-cancer blueprint of the heterogeneous tumor microenvironment (pan-blueprint) [14]. Moreover, these atlases employ different joint analysis strategies. For instance, TICA [11] and TcellMap [13] integrate data using reciprocal principal-component analysis, while pan-T

Table 1 Summary of existing single-cell tumor immune atlases

Atlas name	Sample size	Cell size	Populations				Cancer type	Reference
			T/NK cells	B cells	Macrophage	DC		
ABTC	16	45,851	11	4	1	1	1 (ABTC)	[8]
HCC	10	71,915	14	5	5	-	1 (HCC)	[9]
STAD	48	158,641	10	1	3	1	1 (STAD)	[10]
TICA	217	317,111	13	3	4	3	13*	[11]
Pan-T	316	397,810	14	-	-	-	21#	[12]
TcellMap$	171	308,048	32	-	-	-	16&	[13]
Pan-blueprint	36	233,591	10	4	6	4	4 (LC, CRC, OC, and BRCA)	[14]

BRCA breast cancer, *BCC* basal cell carcinoma, *SCC* squamous cell carcinoma, *EA* endometrial adenocarcinomas, *RCC* renal cell carcinomas, *ICC* intrahepatic cholangiocarcinoma, *HCC* hepatocellular carcinomas, *CRC* colorectal cancers, *PC* pancreatic adenocarcinomas, *OC* ovarian cancer, *NSCLC* non-small-cell lung cancer, *CM* cutaneous melanoma, *UM* uveal melanoma, *HNSCC* head and neck squamous cell carcinoma, *THCA* thyroid carcinoma, *STAD* stomach adenocarcinoma, *FTC* fallopian tube carcinoma, *UCEC* uterine corpus endometrial carcinoma, *NPC* nasopharyngeal carcinoma, *ESCA* esophageal cancer, *LC* lung cancer, *AML* acute myeloid leukemia, *BCL* B-cell lymphoma, *MM* multiple myeloma, *ABTC* advanced biliary tract cancer

$ Data unprovided

* 13 cell types contain BRCA, BCC, SCC, EA, RCC, ICC, HCC, CRC, PC, OC, NSCLC, CM, and UM

21 cell types include HNSCC, THCA, MELA, BCC, SCC, BRCA, STAD, PACA, OC, FTC, UCEC, NPC, ESCA, LC, HCC, ICC, RCC, CRC, AML, BCL and MM

& 16 cell types are FL, HCC, NSCLC, BCC, HNSC, PDAC, CRC, SKCM, BRCA, STAD, AML, DLBCL, GBM, LGG, OV and Brain

combines datasets using Harmony [15], and pan-blueprint [14] employs canonical correlation analysis for dataset integration. Notably, they also exhibit discrepancies in the number and definition of immune cell populations. For instance, TICA comprises 13 T/NK, three B, four macrophages, and three dendritic cell types, whereas pan-T includes only 14 T cell types, and ABTC contains 11 T/NK cell types but only a single type of macrophage or dendritic cell each (Table 1). Despite their anticipated utility in enabling automated and harmonized annotation of the TIME, the characteristics and quality of these atlases remain ambiguous. Substantial disparities among them make comparison and selection challenging.

In this study, we performed a comprehensive evaluation of five tumor immune atlases, consisting of two pan-cancer and three cancer-specific ones. We identified 18 T/NK, nine B cell, eight macrophage, and six dendritic cell populations with consistent marker genes, albeit named differently across various atlases. We also unveiled cell populations unique to or missing from each atlas. To assess the efficacy of these atlases as references for supervised annotation of the TIME, we compared atlas-based annotation with expert manual labels in two immuno-transcriptomics datasets. Besides, we designed five criteria to benchmark their performance in four datasets without manual annotations: mapping success; accuracy, measuring the similarity between supervised annotations and unsupervised clustering; clusterability, assessing distances within and between supervised cell populations; annotatability, evaluating the proportion of unsupervised clusters where majority of cells can be annotated to the same supervised cell population; and stability, examining the similarity of supervised cell populations across multiple query datasets. While these atlases exhibited complementary strengths, notably pan-blueprint performed well across all metrics. Importantly, supervised annotation surpassed unsupervised clustering in identifying cell states associated with immunotherapy response, specifically revealing consistent activation of $CD8^+$ exhausted T cells in responders compared to non-responders across multiple datasets. Our findings highlight both the similarities and differences among existing tumor immune atlases, emphasizing the importance of standardizing cell type nomenclatures, and to develop a comprehensive atlas. Furthermore, this study demonstrates the value of these atlases as references for supervised annotation, providing insights into the mechanisms underlying immunotherapy response.

Results

Overview of the benchmarking study

We conducted a comprehensive benchmarking of existing tumor immune atlases to reveal their similarities and differences, assess their utility for annotating the TIME, and evaluate their application in identifying cell states linked to immunotherapy response (Fig. 1). The study began with six publicly available atlases collected up to June 2022, comprising three pan-cancer atlases (TICA, pan-T, pan-blueprint) and three cancer-specific atlases (ABTC, HCC, STAD). Each atlas underwent uniform quality control (QC) and processing, including filtering of low-quality cells, normalization, and dimension reduction. The six atlases were then integrated through canonical correlation analysis (CCA) (Fig. 1). After integration, atlases were compared based on overlap of marker genes using the Jaccard index and expression correlations. Following the comparison,

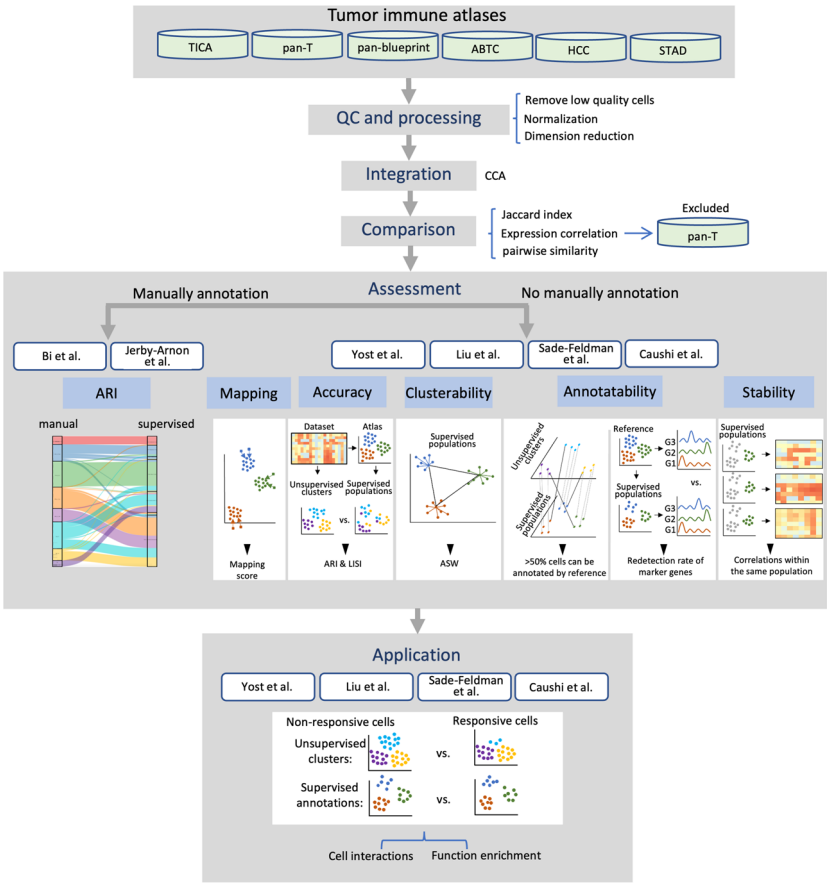

Fig. 1 Schematic overview of the benchwork study. The workflow includes quality control (QC) and processing of datasets, integration of tumor immune atlases, systematic comparison of their similarities and differences among atlases, assessment of their annotation performance, and application to identify functional immune cell states

pan-T was excluded from further analysis due to its outlier status relative to others (Fig. 1). Next, we assessed the utility of the remaining five atlases for supervised annotations across six single-cell transcriptomics datasets, two with expert manual labels [16, 17] and four without but derived from patients undergoing immunotherapy [18–21] (Table 2). For datasets with manual labels, annotation performance was evaluated by agreement with manual labels using the adjusted rand index (ARI) [22]. For datasets lacking manual annotations, we established multiple criteria to evaluate the annotation quality. A robust and accurate annotation is expected to: 1) capture immune cell heterogeneity, encompassing subtle subpopulations; 2) demonstrate accuracy and robustness against technical and computational biases; 3) exhibit distinctive marker genes for each subpopulation; and 4) maintain consistency across multiple query datasets. These expectations are quantified using five criteria: 1) Mapping success: indicating the confidence in cell representation by each atlas; 2) Accuracy: measuring the similarity between supervised annotations and unsupervised clustering based on ARI and rescaled local inverse Simpson's index (LISI) [15, 23]; 3) Clusterability: assessing distances within and between cell populations from supervised annotation based on the average silhouette width (ASW) [24]; 4) Annotatability and marker gene re-detection rate: evaluating the

proportion of unsupervised clustering where > 50% of cells can be annotated to the same cell type by supervised annotations and the recovery of marker genes. 5) Stability: examining the similarity of cell populations assigned by supervised annotations across multiple query datasets (Methods) (Fig. 1). Finally, we applied atlases-based annotations to the four datasets derived from patients undergoing immunotherapy to identify cell states related to immunotherapy response and further explored relevant cell–cell interactions and pathways underlying the response (Fig. 1).

Similarities and differences across single-cell tumor immune atlases

To assess overall similarities among the six atlases, we first calculated expression-based distances and found that cell populations in pan-T did not cluster closely with those from the other atlases, indicating low expression similarity (Additional file 1: Fig. S1A). This discrepancy is likely attributable to pan-T's use of mini-cluster–level rather than single-cell–level expression profiles, which may amplify batch effects and technical biases. Additionally, several pan-T cell populations lacked canonical lineage markers, making it difficult to resolve even basic subsets such as $CD4^+$ and $CD8^+$ T cells. This limitation was further corroborated by poor performance when applying pan-T–based annotations to two datasets with expert manual labels. For example, $CD8^+$ T cells in the Bi et al. dataset were misclassified as $CD4^+$ cell populations by pan-T (Additional file 1: Fig. S1B), and over half of the cells from Jerby-Arnon et al. were mapped to CD8.c10. Trm.ZNF683, including a substantial portion of $CD4^+$ Treg cells (Additional file 1: Fig. S1C). Due to these inconsistencies, we excluded pan-T from subsequent analysis. We therefore focused on the remaining five atlases and conducted a detailed comparison involving 58 T/NK cell populations, 17 B cell populations, 19 macrophage populations, and nine dendritic cell populations.

We first assessed the similarities among the 58 T/NK cell populations by analyzing both the overlap and expression correlation of their marker genes (Fig. 2A and C). This analysis revealed 18 distinct subgroups, which were further categorized into four major classes: $CD4^+$/$CD8^+$ proliferative T cells, $CD4^+$ T cells, $CD8^+$ T cells, and NK cells. Notably, 15 of these subgroups encompassed cell populations from different atlases with consistent marker genes (Fig. 2B), despite being named differently across various atlases.

The group of $CD4^+$/$CD8^+$ proliferative T cells was characterized by elevated expression of cell cycle-related genes such as *STMN1*, *MKI67*, along with either *CD4* or *CD8* (Fig. 2B). $CD4^+$ proliferative T cells comprised three cell types from three different atlases: proliferative T cells from TICA and STAD, and 32_CD4_other from HCC. In contrast, $CD8^+$ proliferative T cells included only 51_CD8_other T cells from HCC (Fig. 2A and B). Notably, although pan-blueprint and ABTC atlases did not contain specific cell types labeled as proliferative T cells, *STMN1* and *MKI67* were highly expressed in their exhausted $CD8^+$ T cells (ABTC) and $CD8^+$ exhausted cytotoxic T cells (pan-blueprint), suggesting that proliferative T cells may have been grouped under exhausted $CD8^+$ T cells in those atlases.

The group of $CD4^+$ T cells included four subgroups: $CD4^+$ regulatory T cells, $CD4^+$ naive T cells, $CD4^+$ memory T cells, and $CD4^+$ exhausted effector T cells (Th2/Th17-like). The $CD4^+$ regulatory T cells, marked by high expression of *FOXP3* and *TNFRSF4*, encompassed five cell types, one from each atlas. The $CD4^+$ naive T cells was

Table 2 Summary of datasets used in the study

Dataset type	Dataset name	Sample size	Cell size	Responder	Non-responder	Pre-treatment	Post-treatment	Age	Gender	Cancer type	Reference
Manually labeled datasets	Bi et al.	4	8187	2	2	NA	NA	Y	Y	Renal cell carcinoma	[17]
	Jerby-Arnon et al.	32	6173	NA	NA	16	16	NA	NA	Melanoma	[16]
Datasets without manual annotation	Yost et al.	15	60,014	8	7	15	15	NA	NA	Basal or squamous cell carcinoma	[18]
	Liu et al.	47	150,574	14	2	8	8	NA	NA	NSCLC	[19]
	Sade-Feldman et al.	47	11,656	16	31	19	28	Y	Y	Melanoma	[20]
	Caushi et al.	15	115,953	6	9	0	15	Y	Y	NSCLC	[21]

NA, not available; *Y*, yes

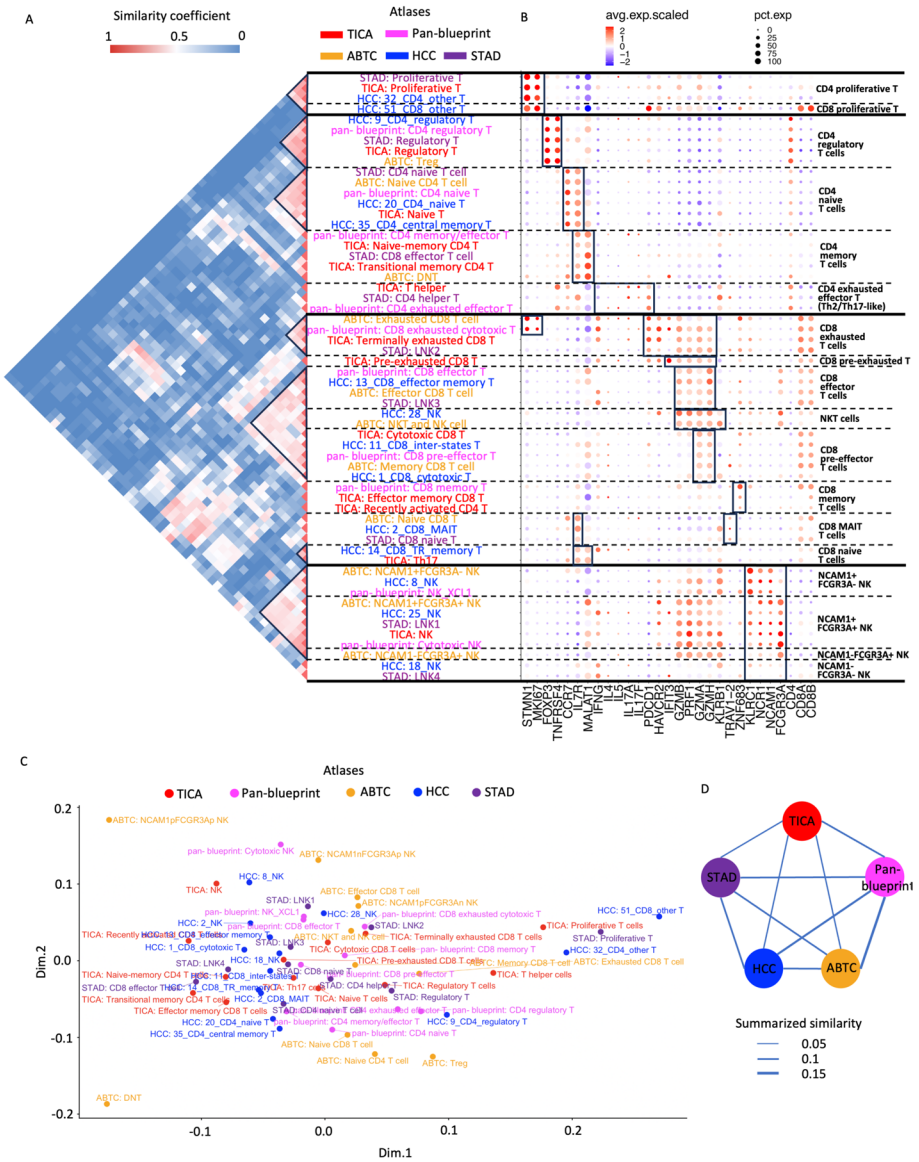

Fig. 2 Similarities and differences among the five atlases. **A** Heatmap showing the overlap of marker genes among 58 T/NK cell populations defined across the five atlases. **B** Dot plot illustrating the expression of marker genes in each cell population. Dot color indicates expression level, and dot size represents the percentage of cells expressing the gene. **C** Multidimensional scaling (MDS) plot visualizing the similarity between cell populations based on expression correlations of marker genes. **D** Network summarizing inter-atlas similarities. Edge width reflects the average pairwise correlation between cell populations across two atlases. DNT: double negative T cell ($CD4^{-}CD8^{-}$ T cell); 14_CD8_TR_memory: $CD8^{+}$ tissue-resident memory T cell; 11_CD8_inter-states T: $CD8^{+}$ intermediate state between the naive T and central memory T cells; 2_CD8_MAIT: $CD8^{+}$ mucosal-associated invariant T cells; LNK: lymphoid of NK-like cells

characterized by elevated expression of *CCR7* and *IL7R*, canonical markers of naïve T cells. This subgroup included six cell types, with five $CD4^{+}$ naive T cells from the five atlases and an additional 35_CD4_central memory T cells from HCC. Although 35_CD4_central memory T cells were not specifically annotated as $CD4^{+}$ naive cells, prior studies suggested that central memory cells were very similar to naive cells, circulating as the first subset in a linear model of $CD4^{+}$ T cell differentiation [25]. The $CD4^{+}$

memory T cells consisted of $CD4^+$ memory/effector T cells from pan-blueprint, naive-memory $CD4^+$ T cells and transitional memory $CD4^+$ T cells from TICA, $CD8^+$ effector T cells from STAD, and double-negative T cells (DNT) from ABTC, but did not include any cell type from HCC. This subgroup was marked by elevated expression of *IL7R*, a hallmark for memory T cells [26], and *MALAT1*, a regulator of effector cells and effector memory cells [27]. The inclusion of $CD8^+$ effector T cells from STAD in this group was due to their high expression of *MALAT1* and relatively low expression of *CD8*, suggesting potential mislabeling in STAD. The $CD4^+$ exhausted effector T cells (Th2/Th17-like) was a heterogeneous group, including T helper from TICA and STAD, and $CD4^+$ exhausted T cells from pan-blueprint. To characterize this group, we examined cytokine and exhaustion-related marker expression: IFNG is a marker of Th1 cells [28]; IL4 and IL5 are markers of Th2 cells [29]; IL17A and IL17F are canonical Th17 cytokines [30]; PDCD1 is a classic inhibitory receptor regulating T cell exhaustion [31]. We found that T helper of TICA expressed *IL5*, *IL17A* and *IL17F*, and *PDCD1*, but didn't express *IFNG*. $CD4^+$ helper T of STAD expressed *IL17A* and *IL17F*, and *PDCD1*, but didn't express *IFNG* or *IL4/IL5*. Pan-blueprint: $CD4^+$ exhausted effector T expressed *IL4*, *IL5*, and *PDCD1*. Therefore, we named this subgroup as $CD4^+$ exhausted effector T cells (Th2/Th17-like). Among the four $CD4^+$ T cell subgroups, $CD4^+$ regulatory T cells and $CD4^+$ naive T cells demonstrated the highest cross-atlas consistency, reflected in their strong similarity coefficients and highly correlated marker gene expression profiles.

The group of $CD8^+$ T cells consisted of eight subgroups: $CD8^+$ exhausted, $CD8^+$ pre-exhausted, $CD8^+$ effector, NKT, $CD8^+$ pre-effector, $CD8^+$ memory, $CD8^+$ MAIT, and $CD8^+$ naive T cells (Fig. 2A and B). Pre-exhausted $CD8^+$ T cells were distinguished by the high expression of both *IFIT3* and *HAVCR2* (*HAVCR2*$^+$*IFIT3*$^+$), whereas terminally exhausted $CD8^+$ T cells expressed only *HAVCR2* but not *IFIT3* (*HAVCR2*$^+$*IFIT3*$^-$) [11]. The $CD8^+$ exhausted T cells comprised exhausted $CD8^+$ T cells from ABTC, $CD8^+$ exhausted cytotoxic T cells from pan-blueprint, terminally exhausted $CD8^+$ T cells from TICA, and LNK2 from STAD. Notably, LNK2 from STAD, displaying similar expression patterns as the cell type in TICA, should represent terminally exhausted $CD8^+$ T cells due to the absence of *IFIT3* expression. On the other hand, the exhausted $CD8^+$ T cells from ABTC and $CD8^+$ exhausted cytotoxic T cells from pan-blueprint seem to encompass a mixture of pre- and terminally exhausted $CD8^+$ T cells, as indicated by their high expression of *IFIT3* and moderate expression of *HAVCR2*. The $CD8^+$ effector T cells included effector $CD8^+$ T cells from pan-blueprint, 13_CD8_effctor memory T cells from HCC, effector $CD8^+$ T cells from ABTC, and LNK3 from STAD. This subgroup was characterized by high expression of Granzymes B and perforins (*GZMB* and *PRF1*), key components of the apoptosis pathway used by cytotoxic lymphocytes and natural killers to eliminate targets [32] and known markers for the activated effector state of $CD8^+$ effector T cells. LNK3, originally annotated as NK cells in the atlases, but showed high expression of *CD8A* and *CD8B,* and clusters with $CD8^+$ effector T cell signatures. We therefore classified it within the $CD8^+$ effector subgroup. The NKT cell subgroup comprises 28_NK from HCC and NKT and NK cells from ABTC. Although these populations expressed markers of $CD8^+$ effector T cells, their *CD8A* and *CD8B* expression were markedly lower than that of classical $CD8^+$ effector T cells. Additionally, both populations expressed *KLRB1* (*CD161*), a cell surface receptor found on NKT cells [33].

The subgroup of $CD8^+$ pre-effector T cells consisted of cytotoxic $CD8^+$ T cells from TICA, 11_CD8_inter-states and 1_CD8_cytotoxic T cells from HCC, $CD8^+$ pre-effector T cells from pan-blueprint, and memory $CD8^+$ T cells from ABTC. This subgroup was distinguished by high expression of *GZMA* and *GZMH*, coupled with low expression of *GZMB* and *PRF1*. The common marker genes and strong correlations among cytotoxic-related $CD8^+$ T cell types indicated a potential state transition path: pre-effector, effector, and pre-/terminal-exhausted cells. Notably, pan-blueprint and ABTC included all cell types covering the full differentiating path. In contrast, TICA lacked $CD8^+$ effector T cells, HCC missed exhausted $CD8^+$ T cells, and STAD lacked pre-effector $CD8^+$ T cells. The subgroup of $CD8^+$ memory T cells included $CD8^+$ memory T cells from pan-blueprint, as well as effector memory $CD8^+$ and recently activated $CD4^+$ T cells from TICA. Despite being annotated as a type of $CD4^+$ T cell, the recently activated $CD4^+$ T cells from TICA expressed *CD8B,* suggesting potential mislabeling. The marker gene of this group was *ZNF683* (also known as *HOBIT*), a key transcription factor expressed in tissue-resident memory T cells [34]. The subgroup of $CD8^+$ mucosal-associated invariant T cells (MAIT) included naïve $CD8^+$ T cells from ABTC and STAD and 2_CD8_MAIT from HCC, characterized by high expression of *TRAV1-2* and *IL7R* but low *CCR7*, consistent with a $CD8^+$ MAIT phenotype that recognizes microbial riboflavin metabolites through the semi-invariant TRAV1-2 TCRα chain [35, 36]. The subgroup of CD8 + naive T cells, characterized by high expression of *CCR7* and *IL7R*, included 14_CD8_TR_memory T cells from HCC and Th17 cells from TICA. Th17 cells are typically considered a subset of $CD4^+$ T cells, however, Th17 from TICA expressed *CD8A* instead of *CD4*, and showed minimal expression of *IL17A* or *IL17F,* suggesting misclassification in the TICA. Notably, $CD8^+$ naive T cells displayed an expression pattern similar to that of $CD4^+$ naive T cells.

The group of NK cells comprised four subgroups: $NCAM1^+FCGR3A^-$, $NCAM1^+FCGR3A^+$, $NCAM1^-FCGR3A^+$, and $NCAM1^-FCGR3A^-$ NK cells, distinguished by the expression of *KLRC1*, *NCR1*, *NCAM1 (CD56)* and *FCGR3A (CD16)* (Fig. 2A and B). KLRC1 and NCR1 act as activating receptors on NK cells and were expressed in all NK populations except LNK4 from STAD. NCAM1 is associated with the immunoregulatory and cytokine-producing capacity [37], while FCGR3A mediates target cell killing through the secretion of perforin and granzymes [38]. All atlases included $NCAM1^+FCGR3A^+$ NK cells, whereas only pan-blueprint, ABTC, and HCC contained $NCAM1^+FCGR3A^-$ NK cells. Neither TICA nor pan-blueprint defined any cell types representing $NCAM1^-$ NK cells. $NCAM^-FCGR3A^+$ NK cells were unique to ABTC.

Similarly, we identified nine subgroups of B cells (Additional file 1: Fig. S2A), including plasma B cells, proliferative B cells, memory IgM B cells, mature-naïve B cells, $CD27^+$/$IGHM^-$ B cells, and four subgroups with mixed phenotypes. The plasma B cell subgroup included only 34_B from HCC and was characterized by high *PRDM1* expression, a transcription factor essential for plasma cell differentiation [39]. The proliferative B cell subgroup comprised transitional B cells from ABTC, proliferative B cells from TICA, and 50_B from HCC, all expressing canonical proliferation markers *STMN1* and *MKI67*. The memory IgM B cell subgroup, marked by *CD27*, *CCR7*, and *GPR183*, included C2_$CD27^+$/$IGHM^+$ from pan-blueprint, 40_B from HCC, and resting memory B cells from

ABTC, with *CD27* being a classical memory marker for B cells [40]. The mature-naïve B cell subgroup, characterized by *IGHD* expression and lack of *CD27* [41], included B cells from TICA, C1_CD27$^-$/IGHD$^+$ from pan-blueprint, 37_B from HCC, and naïve unswitched B cells from ABTC. The C3_CD27$^+$/IGHM$^-$ cluster from pan-blueprint may represent a memory B cell population, as it expressed *CD27*, *IGHA1*, and *GPR183*. The remaining four subgroups exhibited expression of marker genes from multiple cell types, suggesting mixed phenotypes.

Next, we reclassified 19 macrophage cell populations from the five atlases into eight subgroups, including M1-like, early M1-like, ISG$^+$ M1-like, proliferative, M2-like, tumor-associated macrophages (TAMs), and two subgroups with mixed phenotypes (Additional file 1: Fig. S2B). M1-like (pro-inflammatory) macrophages were marked by high expression of pro-inflammatory cytokines such as *IL1B* and *IL6* [42], and represented a shared cell type across four atlases (16_Myeloid from HCC, proinflammatory TAMs from TICA, MYM1 from STAD, and C3_CCR2 from pan-blueprint). The early M1-like subgroup, derived from 26_Myeloid in HCC, exhibited high expression of *S100A8* and *S100A9* instead of *IL1B*/*IL6*; these genes are involved in early inflammatory responses [43]. The ISG$^+$ M1-like subgroup included only MYM2 from STAD and was defined by high expression of interferon-stimulated genes (ISGs), such as *ISG20* and *IFITM2*, which regulate macrophage activation and polarization through induction of pro-inflammatory genes [44]. The proliferative macrophage subgroup, characterized by cell cycle-related genes such as *MKI67* and *TOP2A*, included proliferative macrophages from TICA. The M2-like (anti-inflammatory) macrophage subgroup included C9_LYVE1 and C4_CCL2 from pan-blueprint, and was marked by high expression of *CD163*, involved in anti-inflammatory responses [45], and *MRC1* (*CD206*), a phagocytic receptor that supports anti-inflammatory function [46]. The TAM subgroup included SPP1$^+$ TAMs from TICA, C6_MMP9 and C7_CX3CR1 from pan-blueprint, and 23_Myeloid from HCC. These cells were characterized by expression of *SPP1*, *MMP9*, and *MMP14*, well-known markers of TAMs [44]. The remaining two subgroups consisted of (1) a mixed TAM/M2-like macrophage population, expressing CD163, *MRC1*, *MMP9*, *MMP14*, and *SPP1*, and (2) a heterogeneous macrophage population from ABTC, exhibiting markers of M1-like, M2-like, and TAMs.

Finally, we regrouped nine dendritic cells into six subtypes, including plasmacytoid DCs (pDCs), conventional DC type 2 (cDC2), conventional DC type 1 (cDC1), migratory DCs (mDCs), and two mixed clusters (Additional file 1: Fig. S2C). The pDC subgroup, characterized by *IRF7* and *BST2* (*CD317*) expression, included C4_LILRA4 from pan-blueprint and pDC from TICA. The cDC2 subgroup, marked by high expression of *ITGAM* and *CLEC10A*, comprised mDC from TICA and C2_CLEC10A from pan-blueprint. The cDC1 subgroup, defined by high expression of *XCR1* and *CLEC9A*, was represented by C1_CLEC9A from pan-blueprint. The mDC subgroup was defined by high *CCR7* expression and included C3_CCR7 from pan-blueprint. A mixed cDC1/mDC subgroup from TICA co-expressed *XCR1*, *CLEC9A*, and *CCR7*, suggesting overlapping characteristics. The remaining mixed subgroup, which included MYD from STAD and DC from ABTC, likely represented aggregated dendritic cell populations without clear subtype distinction.

Moreover, we summarized the similarities among the five atlases across the 58 T and NK cell populations (Fig. 2D). B cells, macrophages, and dendritic cells were excluded from this analysis because ABTC contained only one macrophage and DC population, and STAD included only one B cell and one DC type (Table 1). We found that both pan-cancer and cancer-specific atlases exhibited similar correlation patterns, suggesting a generally consistent heterogeneity of T and NK cell populations across different cancer types. In other words, an atlas developed from one or multiple cancer types may be applicable to unseen cancer types.

In conclusion, comparing similarities and differences across atlases allows us to identify shared, missing, and potentially misannotated cell types in each resource, while underscoring the importance of standardizing cell type nomenclature. This analysis enhances our understanding of the heterogeneity and quality of existing atlases and supports their use as references for supervised annotation of the TIME.

Performance evaluation of atlas-guided supervised annotation of the TIME

We evaluated the performance of each atlas when used as a reference for supervised annotation of the TIME across six single-cell immuno-transcriptomics datasets. Two of these datasets, Bi et al. and Jerby-Arnon et al. included manual cell type labels [16, 17] (Fig. 3). Annotation accuracy was assessed by measuring agreement with the manual labels using ARI [22]. In the Bi et al. dataset (Fig. 3A), pan-blueprint achieved the highest agreement (ARI = 0.40), followed by ABTC (ARI = 0.31) and STAD (ARI = 0.30). In contrast, TICA (ARI = 0.13) and HCC (ARI = 0.17) showed the lowest concordance. In the Jerby-Arnon et al. dataset (Fig. 3B), TICA yielded the best overall match (ARI = 0.32), followed closely by ABTC and pan-blueprint (ARI = 0.30 each). Annotation misalignment frequently arose from missing subtypes in the reference atlases. For example, the absence of memory and exhausted T cell subtypes often led to cross-mapping errors. Proliferative T cells were commonly misclassified as exhausted or cytotoxic $CD8^+$ T cells in atlases such as pan-blueprint and ABTC, both of which lacked a dedicated proliferative T cell cluster. Similarly, NK and NKT cells were sometimes misannotated as cytotoxic $CD8^+$ T cells, particularly in TICA, where key NK subsets (e.g., $NCAM1^-$ or $FCGR3A^-$) were absent. γδ T cells were not represented in any of the five atlases, resulting in their misannotation in the Jerby-Arnon et al. dataset (Fig. 3B).

For the other four single-cell datasets lacking manual annotations, including Yost et al., Liu et al., Sade-Feldman et al., and Caushi et al. [18–21], we established five criteria: mapping success, accuracy, clusterability, annotatability, and stability, to evaluate the performance of the five atlases utilized as references for supervised annotations (Fig. 1 and Additional file 1: Fig. S3). Mapping success, designed by Seurat [47], assesses whether query cells are represented accurately in the atlases (Methods). Consequently, STAD achieved higher mapping success than others across all four datasets. Conversely, HCC exhibited the lowest mapping success (Fig. 4A). Accuracy was devised to assess the agreement between supervised annotations and unsupervised clustering. We utilized two metrics for this evaluation: ARI [22] and rescaled LISI [15, 23] (Methods). Higher ARI/LISI scores indicate a higher agreement between supervised annotations and unsupervised clustering. We experimented with different resolutions for unsupervised clustering to mitigate the impact of the number of clusters on the ARI/LISI score (Fig. 4A).

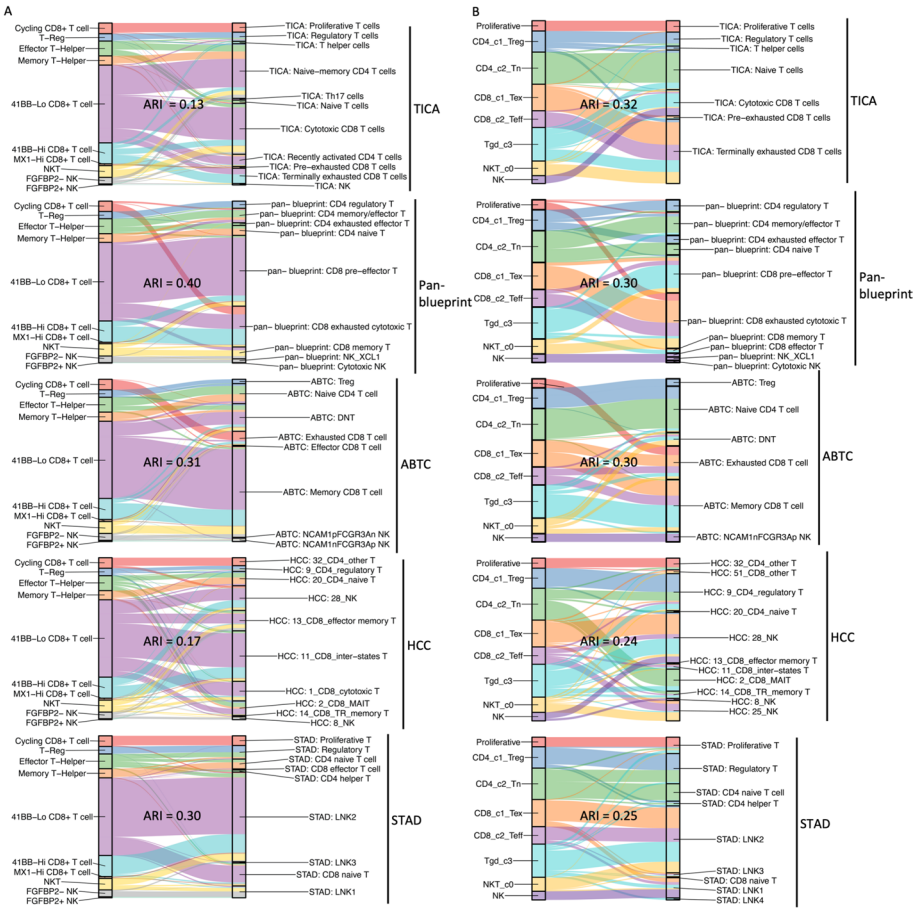

Fig. 3 Performance of supervised annotations on the Bi et al. and Jerby-Arnon et al. datasets using each atlas as a reference. **A** Sankey diagrams showing the concordance between manual labels (left) and supervised annotations (right) for the Bi et al. dataset. ARI is shown for each atlas. **B** Sankey diagrams for the Jerby-Arnon et al. dataset. Ribbon width is proportional to the number of cells

We observed that pan-blueprint consistently outperformed others in ARI regardless of clustering resolution, while STAD exhibited slightly higher LISI scores than others across multiple resolutions. Clusterability was employed to evaluate the quality of populations grouped by supervised annotations, where robust grouping should exhibit high expression cohesion within the same population and clear separation from other populations. We utilized the ASW [24] to assess the grouping quality (Methods). A higher ASW value indicates closer distances within a population and farther distances between populations, indicating better performance of atlases. We observed that HCC attained the highest ASW score, closely followed by TICA (Fig. 4A). In contrast, ABTC exhibited the lowest ASW score. Annotatability of each atlas was initially assessed by calculating the proportion of populations from supervised annotations in which more than 50% of cells originated from the same unsupervised cluster. We experimented with various resolutions for unsupervised clustering to mitigate the impact of the number of clusters. ABTC achieved the highest score, suggesting high confidence in cell annotation (Fig. 4A). Annotatability was further evaluated by the redetection of marker genes. Ideally, supervised annotations should be capable of reproducing consistent markers across

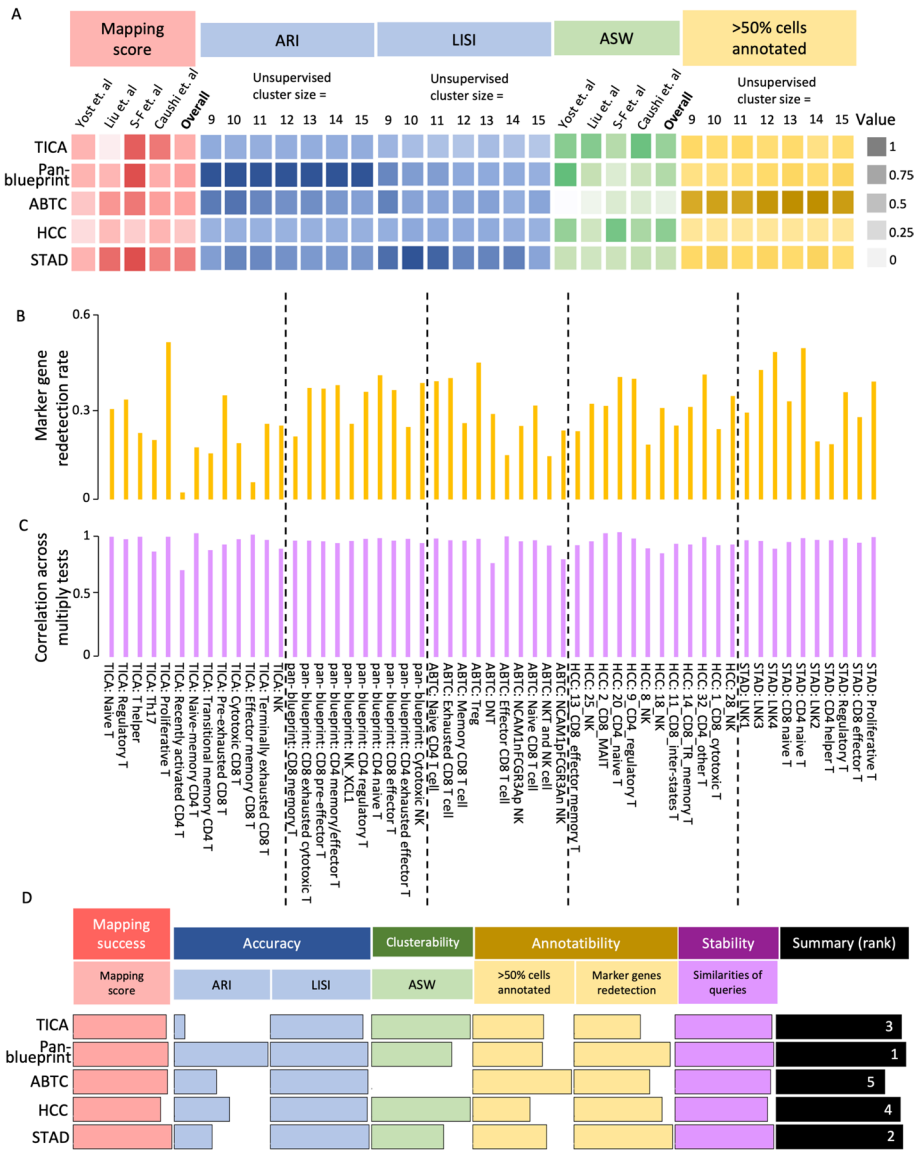

Fig. 4 Performance of supervised annotations on the four datasets based on five evaluation criteria defined by seven metrics: mapping success, accuracy, clusterability, annotatability and stability. **A** Heat-map plots for mapping scores across the four datasets, ARI and LISI scores across multiple cluster sizes, ASW scores across query datasets and proportions of supervised annotations whose > 50% cells were projected from the same unsupervised clusters across multiple cluster sizes. **B** Results of marker gene redetection rate for each population. **C** Correlations of supervised annotations across multiple queries. **D** The summary of the evaluation through summing the five criteria, mapping success, accuracy, clusterability, annotatability, and stability

different queries. Therefore, we introduced a metric called marker gene redetection rate (Methods) to assess the overlap of markers between different supervised annotations (Fig. 4B). Overall, we found that STAD excelled in reproducing markers, closely followed by pan-blueprint (Fig. 4B). Meanwhile, we observed that $CD4^+$ regulatory T cells were the most or second-most conserved populations in ABTC and HCC. Naive $CD4^+$ T cells exhibited the highest marker redetection rate in pan-blueprint and STAD and were the second-highest in HCC. TICA obtained the highest marker gene redetection

rate for proliferative T cells (Fig. 4B). Across the five atlases, $CD4^+/CD8^+$ proliferative T cells obtained the highest marker redetection rate, closely followed by $CD4^+$ regulatory T cells and $CD4^+$ naive T cells (Additional file 1: Fig. S4), consistent with previous findings that these three cell types were the most conserved across atlases (Fig. 2A). In contrast, $CD8^+$ memory T cells exhibited the lowest marker redetection rate. Stability was employed to assess whether similar cells across multiple datasets were annotated as the same population by supervised annotations. A high stability indicates robustness of the atlas against technical and computational biases. We observed that most cell populations from each atlas exhibited high stability. Among them, naive-memory $CD4^+$ T cells from TICA and $CD4^+$ naive T cells from pan-blueprint achieved the highest scores (Fig. 4C). Overall, STAD and pan-blueprint outperformed others in stability measurement.

Furthermore, we conducted a comparative analysis of supervised versus unsupervised annotation methods, focusing on their performance in identifying both abundant and rare cell populations. We applied pan-blueprint–guided supervised annotation and unsupervised clustering to the Yost et al. dataset [18]. To ensure a fair comparison, the resolution for unsupervised clustering was adjusted to yield the same number of cell subtypes (nine) as identified by the supervised method (Additional file 1: Fig. S5A). The supervised approach successfully identified three rare populations—$CD8^+$ memory T cells, $CD8^+$ effector T cells, and NK_XCL1 cells—each supported by expression of canonical marker genes (Additional file 1: Fig. S5B). In contrast, the unsupervised method detected only NK_XCL1 cells and failed to resolve the other two rare populations. Moreover, for the six abundant cell types, supervised annotation produced higher expression levels of canonical lineage markers compared to the unsupervised method, except for $CD4^+$ exhausted T cells (Additional file 1: Fig. S5B). These findings demonstrate that supervised approaches more accurately resolve both rare and abundant cell populations.

To summarize, we observed substantial variability in performance across different metrics, indicating that each atlas exhibited complementary strengths (Fig. 4D). Among them, pan-blueprint consistently performed well across all evaluation metrics in all six datasets, including those with and without manual annotations. Additionally, supervised annotation outperformed unsupervised clustering in accurately identifying both abundant and rare cell populations.

Application of supervised annotation in discovering cell states related to immunotherapy response

The transition of T cell states, particularly from a resting to an activated state, is a complex process involving proliferation, effector function, and differentiation [48]. The presence and functional state of T cells have been identified as independent positive prognostic markers of immunotherapy response in various malignancies [49, 50]. Single-cell tumor immune atlases capture the full spectrum of T cell subtypes and activation states, providing a foundation to investigate how specific T cell states correlate with immunotherapy response.

To evaluate the ability of the five atlases in identifying immune cell states associated with immunotherapy response, we first quantified activation scores for each cell population defined by supervised annotation. We then assessed their correlation with

immunotherapy outcome, adjusting for potential biological confounders such as age and gender, when available (Methods). Notably, TICA consistently revealed significantly higher activation scores in responders compared to non-responders within terminally exhausted CD8$^+$ T cells across all four datasets (Fig. 5A). Pan-blueprint and ABTC also identified significantly higher activation scores in CD8$^+$ exhausted T cells in responders across all datasets except Liu et al., where the scores were either significantly lower or unchanged. STAD detected significantly higher activation scores in the LNK2 population in three datasets; based on our previous findings (Figs. 2 and 5A), this population should be reclassified as CD8$^+$ exhausted T cells. In contrast, HCC lacked the CD8$^+$ exhausted T cell population and therefore failed to identify consistent signals across datasets. Furthermore, we investigated whether the activation of CD8$^+$ exhausted T cells was pre-existing or induced by treatment. Among the four datasets, the Caushi et al. dataset included only post-treatment samples, while the remaining three contained both pre- and post-treatment samples. To address this, we reanalyzed the data separately for pre- and post-treatment conditions. We found that CD8$^+$ exhausted T cell activity remained significantly associated with immunotherapy response in all three post-treatment datasets when using annotations guided by pan-blueprint, ABTC, and STAD (Yost et al., Sade-Feldman et al., and Caushi et al.), but only in one pre-treatment dataset (Yost et al.) (Additional file 1: Fig. S6). These findings suggest that, although CD8$^+$ exhausted T cells typically express inhibitory checkpoint molecules such as *PD-1* and *CTLA-4* [51], they can regain effector function in response to immunotherapy. This is consistent with previous studies showing that immune checkpoint blockade can reactivate exhausted T cells [52, 53]. Overall, our results support the interpretation that the observed activation reflects therapy-induced reactivation rather than a baseline predictive marker. Additionally, no significant association was observed between other immune cell subpopulations (e.g., B, macrophage, and dendritic cells) with immunotherapy response (Additional file 1: Fig. S7).

In comparison, we evaluated the efficacy of unsupervised clustering in identifying response-related cell states. We first determined the optimal clustering resolution by maximizing the ASW, which best captured the inherent data structure (Methods). The resulting clusters were then annotated using canonical marker genes. Finally, we assessed the association between the activity of each cluster and immunotherapy response across the four datasets (Fig. 5B). Although a cluster corresponding to CD8$^+$ exhausted T cells was present in all four datasets, a significant association between its activation scores and response was observed only in the Caushi et al. dataset. None of the cell types demonstrated consistent and reproducible correlations with immunotherapy response across more than one dataset (Fig. 5B).

In summary, supervised annotations enabled more reliable identification of CD8$^+$ exhausted T cells across datasets, consistently linking their activity to positive immunotherapy outcomes. This comparison highlights the limitations of unsupervised clustering in resolving functionally relevant, treatment-associated T cell states, and underscores the value of expert-guided or atlas-based annotations in cross-dataset immune profiling. Moreover, supervised annotations provide harmonized, reproducible, and interpretable labels that are easily comparable across datasets. In contrast, unsupervised clustering requires careful tuning of computational

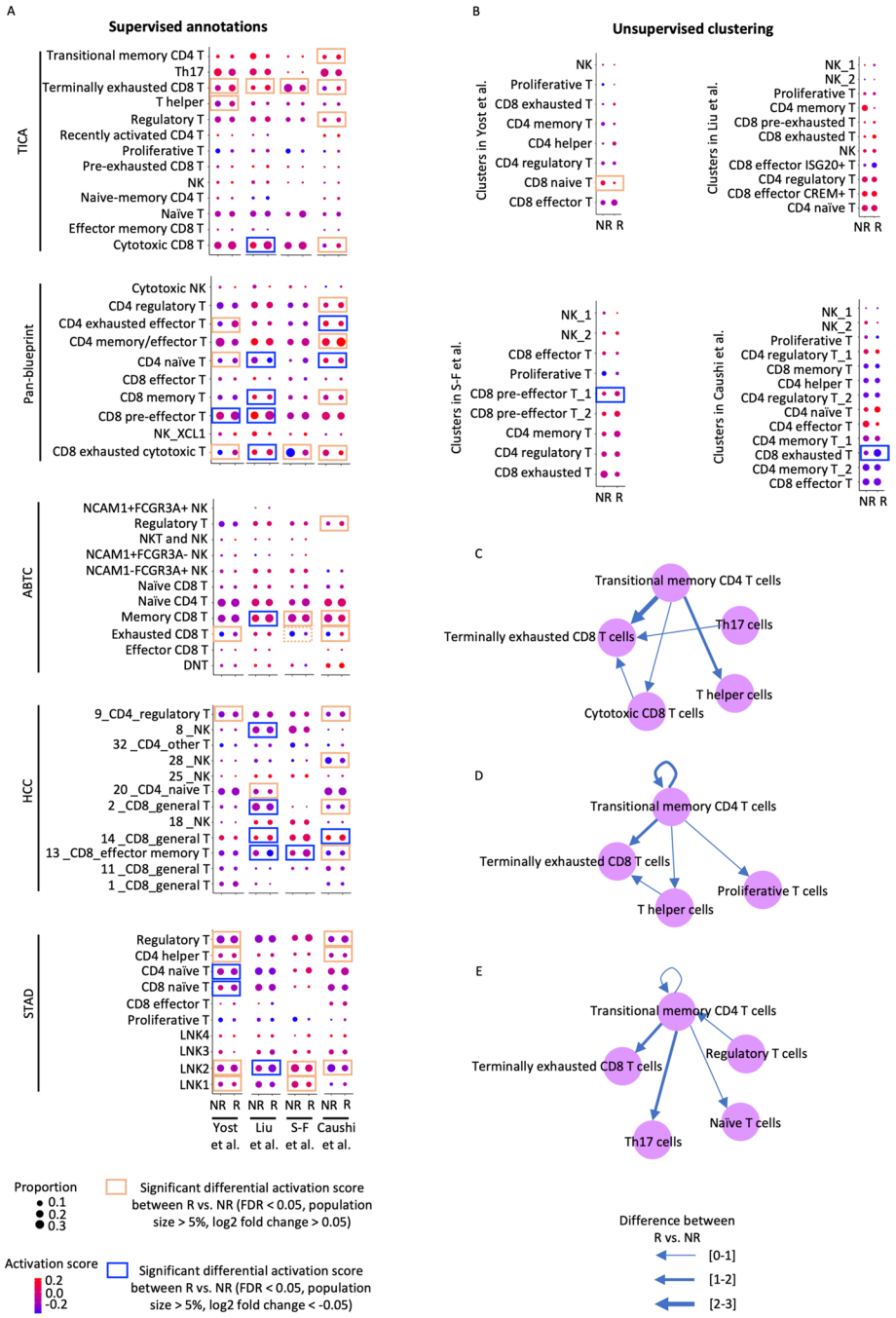

Fig. 5 Performance of supervised annotations versus unsupervised clustering in identifying cell states associated with immunotherapy response. **A** Differential analysis of activation scores across cell populations defined by supervised annotations using each atlas as a reference between responders and non-responders. Dot color indicates the activation score, and dot size represents the proportion of the cell population. The dashed orange rectangle denotes comparisons with FDR = 0.06. **B** Differential analysis of activation scores across unsupervised clusters using the optimal resolution for each dataset. **C**–**E** Top five differential interactions of the VEGFB–FLT1 ligand–receptor pairs between responders and non-responders in the Yost et al. (C), Liu et al. (D), and Caushi et al. (E) datasets. R: responder; NR: non-responder

parameters and specialized expertise, often producing results that are less reproducible, harder to interpret, and difficult to compare across studies.

Given the pivotal role of $CD8^+$ exhausted T cells in immunotherapy response, we further investigated their interactions with other immune subsets. We identified dysregulated ligand-receptor interactions between responders and non-responders using scLR [54]. Interestingly, the VEGFB–FLT1 interaction between various $CD4^+$/$CD8^+$ T cell subsets and $CD8^+$ exhausted cytotoxic T cells was consistently upregulated in responders compared to non-responders across all three datasets with available raw count data, including the Yost et al., Liu et al., and Caushi et al. datasets (Fig. 5C–E) (FDR < 0.05). *FLT1* (*VEGFR1*), a receptor of VEGFB, is expressed on $CD8^+$ exhausted T cells, while *VEGFB* is secreted by both $CD4^+$ and $CD8^+$ T cells [55]. VEGFB–FLT1 signaling may influence $CD8^+$ exhausted T cells by promoting both pro-inflammatory and regulatory effects. For example, VEGF has been shown to promote T cell activation and migration to inflammatory sites [56]. Engagement of FLT1 can enhance T-cell activation and promote chemotaxis to inflamed tissues [56], suggesting that VEGFB signaling may re-invigorate $CD8^+$ exhausted T cells and facilitate their recruitment into tumors during immunotherapy. Additionally, we performed Gene Set Enrichment Analysis (GSEA) to investigate pathways that may contribute to the reactivation of $CD8^+$ exhausted T cells in responders across the four datasets. TNF-α signaling emerged as significantly enriched in all datasets (Additional file 1: Fig. S8; FDR < 0.05). Specifically, this pathway was significantly upregulated in responders with both non-melanoma skin cancers (Yost et al., FDR = 0.001) and melanoma (Sade-Feldman et al., FDR = 0.016) (Table 2; Additional file 1: Fig. S8A and C), while it was downregulated in non-small cell lung cancer (NSCLC) responders (Liu et al., and Caushi et al. with FDR = 0.024 and FDR = 0, respectively) (Additional file 1: Fig. S8B and D). These results suggest that TNF-α signaling may play a context-dependent role in modulating $CD8^+$ T cell responses during immunotherapy.

Discussion

The development of single-cell tumor immune atlases, derived from extensive analysis of large-scale datasets spanning various studies and cancer types, aims to lay the foundation for characterizing the heterogeneity and complexity of the TIME. However, these atlases vary considerably, making selection challenging. In this study, we explored the similarities and differences among existing tumor immune atlases. We identified 18 T/NK cell components, nine subgroups of B cells, eight macrophage populations, and six subtypes of dendritic cells across pan-cancer and cancer-specific atlases, despite variations in their nomenclature (Fig. 2A and B, and Additional file 1: Fig. S2). Notably, $CD4^+$/$CD8^+$ proliferative cells, $CD4^+$ regulatory T cells, $CD4^+$ naive T cells, $CD8^+$ exhausted T cells, $CD8^+$ effector cells, and subtypes of NK cells demonstrated the most consistent gene signatures across atlases. Conversely, $CD4^+$ and $CD8^+$ memory cells exhibited less conservation, likely due to subtype definitions, different resolution levels, and limited functional exploration in original studies. Various types of memory cells, such as central memory, resident memory, and effector memory cells, result in difficulties in marker gene discovery. For instance, TICA includes transitional memory

$CD4^+$ T cells, naive-memory $CD4^+$ T cells, and effector memory $CD8^+$ T cells. Pan-blueprint and ABTC have only one type of $CD8^+$ memory T cells. HCC comprises three types of memory cells, 35_CD4_central memory T cells, 13_CD8_effector memory T cells, and 14_CD8_TR_memory T cells. STAD has no memory cells at all (Fig. 2A and B). Memory cells dissected at different resolution levels and non-generalized expression profiles greatly limit the definition of subpopulations. Meanwhile, $CD8^+$ effector T cells and a subset of NK cells exhibited similarity and elevated expression of cytotoxic-related genes, reflecting their shared role as cytotoxic effector cells of the immune system [57]. Distinguishing between them can be challenging due to their similar functions and expression profiles, underscoring the importance of comprehensive characterization for accurate classification. Our study provides important insights into the characteristics and annotation quality of each atlas, underscoring the need for standardized nomenclature of immune cell populations. We also identify missing or underrepresented cell types in each atlas, highlighting the importance of refining and expanding the definition of fine-grained immune subpopulations. For example, proliferative T cells are absent from the pan-blueprint atlas; $NCAM1^+$ $FCGR3A^-$ NK cells are missing from both TICA and pan-blueprint; $CD8^+$ exhausted T cells are not represented in the HCC atlas; and memory T cell subtypes are entirely absent from STAD. Since supervised annotations are heavily dependent on the cell types represented in the reference atlas, a clear understanding of these gaps is crucial. Such knowledge can help researchers make informed choices when selecting atlases for downstream analyses.

Furthermore, we evaluated the effectiveness of each atlas as a reference for supervised annotation of the TIME across multiple immune-transcriptomic datasets. While the pan-blueprint atlas performed well across all evaluation metrics, each atlas demonstrated complementary strengths. These results suggest that integrating multiple atlases may enhance annotation comprehensiveness and accuracy, offering a more robust foundation for downstream analysis and interpretation of the TIME.

Finally, we demonstrated the power of supervised annotations in identifying functional cell states associated with immunotherapy response. Specifically, $CD8^+$ exhausted T cells exhibited significantly higher activation in responders versus non-responders, consistent with prior studies showing that immune checkpoint blockade reactivates exhausted T cells [52, 53]. Interestingly, we observed no significant difference in the volume of $CD8^+$ exhausted T cells between responders and non-responders, suggesting that a pre-existing pool of tumor-recognizing exhausted T cells may be a prerequisite for effective immunotherapy. Compared to unsupervised clustering, supervised annotations are more interpretable, reproducible, and less prone to computational biases. However, their performance is constrained by the cell types present in the reference atlas, limiting their ability to detect novel or rare cell states. Hybrid approaches may offer a promising solution by combining the strengths of both strategies—for instance, applying supervised annotation to confidently assign known cell types, followed by unsupervised clustering to refine or discover novel subpopulations among unmapped cells.

While our benchmark provides a systematic comparison of tumor immune atlases and demonstrates their utility for annotating single-cell datasets and revealing functional cell states, several limitations remain. First, despite the use of a uniform

preprocessing pipeline and batch correction to minimize technical variation across atlases, observed performance differences may still reflect a mix of biological heterogeneity and residual technical artifacts inherent to each dataset. Fully disentangling these factors would require a more controlled benchmarking framework with matched technical conditions. Second, certain cell states or populations not represented in current atlases, such as γδ T cells or innate lymphoid cells (ILCs), leading to incomplete characterization when using supervised annotation alone. Third, while our study focused on six well-curated atlases, we did not establish an automated framework to integrate newly emerging atlases or harmonize cell type nomenclature across references. As the field rapidly evolves, developing such a framework will be essential for maintaining up-to-date, interoperable annotations and enabling broader cross-study comparisons.

To enable further improvement beyond current resources, future efforts should prioritize the development of a more comprehensive and dynamic reference atlas. This includes publishing standardized marker gene schemas and harmonized cell type labels to ensure consistency and reproducibility across studies. In addition, creating automated frameworks to integrate newly developed atlases and standardize cell annotations across references would support regular updates and facilitate more effective cross-study comparisons. Moreover, integrating other single-cell omics modalities—such as proteomics, scATAC-seq, and spatial transcriptomics—into multimodal atlases will further enhance the resolution and biological interpretability of immune cell states beyond what transcriptomic data alone can achieve.

Conclusions

We systematically benchmark tumor immune atlases to evaluate their effectiveness in annotating single-cell immuno-transcriptomic datasets, uncover key features of the tumor immune microenvironment relevant to immunotherapy response, and identify directions for the development of future reference atlases.

Methods

Collection of single-cell tumor immune atlases

We collected all publicly available tumor immune atlases released on or before 30 June 2022 that met the following criteria: (1) unrestricted access to gene-cell matrices and cell annotations; (2) human single-cell RNA sequencing or equivalent multimodal transcriptomic technology; (3) inclusion of ≥ 10 patient samples to capture inter-patient variability; and (4) immune-focused annotation. Ten candidates were initially identified: TICA [11], pan-T [12], pan-blueprint [14], ABTC [8], HCC [9], STAD [10], TISCH [58], ProjectTILs [59], the PDAC mass cytometry atlas [60], and HCC_iCCA [61]. Four were excluded for not meeting these criteria: ProjectTILs is a mouse atlas; TISCH and HCC_iCCA lacked accessible data or annotations; and the PDAC atlas is proteomics-based rather than transcriptomics-based. As a result, we benchmarked the remaining six atlases: TICA, pan-T, pan-blueprint, ABTC, HCC, and STAD, comprising three pan-cancer (TICA, pan-T, pan-blueprint) and three cancer-specific atlases (ABTC, HCC, STAD).

Cell annotation tables (including gene-cell matrix and cell types) were obtained from the original publications. T, NK, B, macrophage, dendritic cell clusters were selected based on original cell type annotation. We used Seurat (v.4.1.1) [47] with default parameters to process each atlas to ensure uniform interpretation. For TICA and pan-T cell, Seurat objects were provided in original articles. We sort out T, NK, B, macrophages and dendritic cells separately, and then applied the *ScaleData* and *RunPCA* (npcs = 30) functions to the integrated matrix with default parameters. We then used *RunUMAP* (dims = 1:30) function with return.model = T to store the UMAP model so that query datasets could be interpreted in the context of the existing reference. Regarding to pan-blueprint, ABTC, HCC and STAD, we normalized expression matrix of each atlas by the *NormalizeData* (normalization.method = "LogNormalize") function with default parameters since only raw data were provided. Then *FindVariableFeatures* (selection.method = "vst", nfeatures = 2000) function was used to detect highly variable genes (HVGs). Then within each atlas, the *FindIntegrationAnchors* (scale = TRUE, reduction = "cca") and *IntegrateData* (dim = 1:30) functions were then applied to choose genes and integrate multiple samples. Finally, *ScaleData*, *RunPCA* and *RunUMAP* with *return.model = T* were used for each atlas. All analyses were run in R 4.4.1. R packages used in this study were posted in conda/R environment of GitHub (https://github.com/JingYangSciBio/Tumor-immune-atlases).

Quality control, data filtering and supervised annotation of query scRNA-seq datasets

Raw scRNA-seq data were obtained from public repositories or original publications. Cells with unique feature counts over 5,000 or less than 200 were removed. We further filtered cells that have > 20% mitochondrial counts. Normalization was performed using Seurat, dividing the unique molecular identifier (UMI) counts of each gene by total UMI count of each cell. Then cell clusters were selected based on the identification using Seurat functions with default parameters, including *FindVariableFeatures*, *ScaleData*, *RunPCA*, and *RunUMAP*. Then cells were filtered using over-representation analysis for cell type annotation. *FindVariableFeatures* and *ScaleData* were used again to calculate and scale highly variable genes. Regarding supervised annotation, *FindTransferAnchors* function with default parameters was used to find out "anchor" cells between reference atlases and query datasets, that represent similar profiling. Then *MapQuery* function was employed to project queries onto the reference UMAP structure.

Similarities between reference atlases

Marker genes of atlas populations were identified via differential expression by comparing one population to all other populations. *FindAllMarkers* function of Seurat R package was used within each atlas. Differentially expressed genes (DEGs) were filtered based on the following criteria: expressed in at least 25% of population cells and FDR value < 0.05. Then top80 DEGs were kept as marker genes.

The Jaccard similarity index was used for comparing overlap markers of two populations. It's a measure of similarity for the two sets of data, with a range from 0 to 100%. The higher the percentage, the more similar the two populations.

$$J(P_i, P_j) = |P_i \cap P_j| / |P_i \cup P_j|$$

where P_i and P_j are the maker genes of population i and j. Pearson correlation coefficient was employed to estimate the correlation between any two populations. Then the matrix of classical multidimensional scaling (MDS) was generated based on the correlations. To remove potential batch-effect across atlases, we used *removeBatchEffect* function in 'limm' R package along with a vector that preserved different atlases.

Mapping success

This metric was designed by *MappingScore* function of Seurat to estimate if query cells were well represented in the reference data. Query cells were projected to a reference-defined space and then projected them back onto the query dataset. By comparing the neighborhoods before and after projection, cells whose local neighborhoods were most altered by the transformation would be identified and assigned low mapping scores. Cells with consistent neighbors would get high mapping scores (see https://satijalab.org/seurat/reference/mappingscore).

Adjusted rand index (ARI), local inverse Simpson's index (LISI) and average silhouette width (ASW)

The ARI measures the percentage of matches between two label lists. In our study, unsupervised clustering and supervised annotations were compared using *adjustedRandIndex* function of the mclust R package [62]. The LISI can be used to assess goodness of cell type integration [63]. The Simpson index describes the probability that two cell type labels are taken at random, and the inverse represents the effective number of cell types in the neighborhood. LISI scores range from 1 to B (where B denotes the number of cell types), indicating perfect separation and perfect mixing, respectively. We rescaled LISI to the range 1 to 0 by min–max scaling.

$$LISI_{norm} = \frac{\max(LISI) - LISI}{\max(LISI) - \min(LISI)}$$

$LISI_{norm}$ of 1 corresponds to perfect separation which means cells from a supervised annotation projected to an unsupervised cluster without any mixing of other supervised annotations. A value closer to 0 denotes that the supervised annotations contain mixed unsupervised clusters. The code to compute LISI is available at https://github.com/immunogenomics/LISI.

ASW uses the distance difference between cells within clusters and between clusters to measure the distribution of cells [24]. The resulting score ranges from −1 to 1, where a high score denotes that the cell fits well in the current supervised annotations. The average score of all cells is used to measure overall cell type purity.

Proportions of supervised annotations whose > 50% cells were projected from the same unsupervised clusters and re-detection rate of marker genes

If > 50% cells of a supervised population were from a same unsupervised cluster, this supervised population would be a well-built population since it captured original expression features of query datasets. The proportions of supervised populations with > 50%

cells came from the same unsupervised clusters was defined as one metric of annotatability. Regarding re-detection rate of marker gene between reference atlases and query datasets, the Jaccard similarity index was used for measuring the similarities of two marker gene lists which were identified from supervised populations and unsupervised clusters.

Activation scores of cell states

Gene sets NK_T_CELL_ACTIVATION, NK_T_CELL_DIFFERENTIATION, NK_T_CELL_PROLIFERATION, and T_CELL_MEDIATED_CYTOTOXICITY were downloaded from MSigDB [64]. These gene sets represent canonical markers associated with T cell activation, differentiation, proliferation, and cytotoxic function. To quantify pathway activity at the single-cell level, we applied the *CellCycleScoring* function from the Seurat package, which computes a score for each cell based on its expression of the relevant marker genes. Differences in pathway activity between responders and non-responders were assessed using the Wilcoxon rank-sum test. For datasets with available age and sex metadata, the Wilcoxon test was adjusted accordingly. All p-values were corrected for multiple testing across cell populations using the Benjamini–Hochberg method.

Unsupervised clustering

For each of the four immunotherapy datasets lacking manual labels we generated clusters across resolutions 0.1–2.0 (yielding 4–25 clusters). We then computed the ASW for every resolution and selected the resolution with the highest ASW (resolution = 0.10 in Yost et al.; 0.20 in Liu et al.; 0.35 in Sade-Feldman et al.; 0.31 in Caushi et al.). This analysis identified the partition most supported by the dataset's internal structure, showing eight clusters in Yost et al., 11 clusters in Liu et al., nine clusters in Sade-Feldman et al., and 13 clusters in Caushi et al. We then calculated the DEGs of each cluster by using *FindAllMarkers* function of Seurat R package. Each unsupervised cluster was annotated by comparing DEGs to canonical marker genes, such as *CD4*, *CD8A/B*, *PDCD1*, *HAVCR2*, *IL7R*, *GZMB*, *TRAV1-2*, etc.

Ligand-receptor interaction analysis

To explore inter-cellular communication, we applied scLR [54] to identify differentially regulated ligand–receptor interactions between responders and non-responders across the three immunotherapy datasets (Yost et al., Liu et al., and Caushi et al.). Sade-Feldman et al. was excluded due to lack of raw counts. Cell types were defined using the TICA-based supervised annotation, providing a consistent reference frame. For each dataset we quantified all possible ligand–receptor pairs between annotated cell populations and computed the log-fold change in interaction strength between responders and non-responders.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-026-04055-5.

Additional file 1.

Acknowledgements
We thank authors of the TICA, Pan-blueprint, ABTC, HCC, STAD and Pan-T atlases for promptly providing source files and clarifying metadata in response to our e-mail requests. We also appreciate the insights and feedback offered by Dr. Bryan Helm, who reviewed an early version of the manuscript and suggested several improvements to the analysis workflow and presentation.

Peer review information
Andrew Cosgrove was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions
J. Y. designed the study, acquired the data, analyzed the data, and wrote the manuscript. Y. S. acquired the funding and reviewed the manuscript. Q. L. supervised the analysis, revised the manuscript and acquired the funding. All authors have read and approved the final manuscript.

Funding
This work is supported by National Cancer Institute grants (U2C CA233291, P01CA229123 and U54 CA274367), National Institutes of Health (P01 AI139449), and Cancer Center Support Grant (P30CA068485).

Data availability
All pre-processed single-cell immune atlases can be found at Cancer-Immu (https://bioinfo.vanderbilt.edu/database/Cancer-Immu/) or on Zenodo at https://doi.org/10.5281/zenodo.15867143 [65]. All R scripts used in the manuscript and a virtual Conda environment are available on GitHub at https://github.com/JingYangSciBio/Tumor-immune-atlases [66] under an MIT licence or on Zenodo at https://doi.org/10.5281/zenodo.18989719 [67] under a Creative Commons Attribution 4.0 International License.
The six single-cell transcriptomics datasets analyzed in this study are publicly available through their original publications. Specifically, the Bi et al. dataset was downloaded from Single Cell Portal [68]. The Jerby-Arnon et al. dataset was obtained from the Gene Expression Omnibus (GEO) under accession number GSE115978 [69]. The datasets from Yost et al., Liu et al., Sade-Feldman et al., and Caushi et al. are available under GEO accession numbers GSE123814 [70], GSE179994 [71], GSE120575 [72], and GSE176021 [73], respectively.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 24 May 2024 Accepted: 20 March 2026
Published online: 31 March 2026

References

1. de Visser KE, Joyce JA. The evolving tumor microenvironment: from cancer initiation to metastatic outgrowth. Cancer Cell. 2023;41:374–403.
2. Bozyk A, Wojas-Krawczyk K, Krawczyk P, Milanowski J. Tumor microenvironment-a short review of cellular and interaction diversity. Biology (Basel). 2022;11:929.
3. Arner EN, Rathmell JC. Metabolic programming and immune suppression in the tumor microenvironment. Cancer Cell. 2023;41:421–33.
4. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 2018;24:541–50.
5. Wang M, Zhao J, Zhang L, Wei F, Lian Y, Wu Y, et al. Role of tumor microenvironment in tumorigenesis. J Cancer. 2017;8:761–73.
6. Ren X, Zhang L, Zhang Y, Li Z, Siemers N, Zhang Z. Insights gained from single-cell analysis of immune cells in the tumor microenvironment. Annu Rev Immunol. 2021;39:583–609.
7. Elhanani O, Ben-Uri R, Keren L. Spatial profiling technologies illuminate the tumor microenvironment. Cancer Cell. 2023;41:404–20.
8. Shi X, Li Z, Yao R, Cheng Q, Li W, Wu R, et al. Single-cell atlas of diverse immune populations in the advanced biliary tract cancer microenvironment. NPJ Precis Oncol. 2022;6:58.
9. Lu Y, Yang A, Quan C, Pan Y, Zhang H, Li Y, et al. A single-cell atlas of the multicellular ecosystem of primary and metastatic hepatocellular carcinoma. Nat Commun. 2022;13:4594.
10. Kumar V, Ramnarayanan K, Sundar R, Padmanabhan N, Srivastava S, Koiwa M, et al. Single-cell atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer. Cancer Discov. 2022;12:670–91.
11. Nieto P, Elosua-Bayes M, Trincado JL, Marchese D, Massoni-Badosa R, Salvany M, et al. A single-cell tumor immune atlas for precision oncology. Genome Res. 2021;31:1913–26.

12. Zheng L, Qin S, Si W, Wang A, Xing B, Gao R, et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. Science. 2021;374:abe6474.
13. Chu Y, Dai E, Li Y, Han G, Pei G, Ingram DR, Thakkar K, Qin JJ, Dang M, Le X, et al. Pan-cancer T cell atlas links a cellular stress response state to immunotherapy resistance. Nat Med. 2023.
14. Qian J, Olbrecht S, Boeckx B, Vos H, Laoui D, Etlioglu E, et al. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. Cell Res. 2020;30:745–62.
15. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16:1289–96.
16. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC, et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. Cell. 2018;175(984–997):e924.
17. Bi K, He MX, Bakouny Z, Kanodia A, Napolitano S, Wu J, et al. Tumor and immune reprogramming during immunotherapy in advanced renal cell carcinoma. Cancer Cell. 2021;39:649-661 e645.
18. Yost KE, Satpathy AT, Wells DK, Qi Y, Wang C, Kageyama R, et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. Nat Med. 2019;25:1251–9.
19. Liu B, Hu X, Feng K, Gao R, Xue Z, Zhang S, et al. Temporal single-cell tracing reveals clonal revival and expansion of precursor exhausted T cells during anti-PD-1 therapy in lung cancer. Nat Cancer. 2022;3:108–21.
20. Sade-Feldman M, Yizhak K, Bjorgaard SL, Ray JP, de Boer CG, Jenkins RW, et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. Cell. 2018;175:998-1013 e1020.
21. Caushi JX, Zhang J, Ji Z, Vaghasia A, Zhang B, Hsiue EH, et al. Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. Nature. 2021;596:126–32.
22. Hubert L, Arabie P. Comparing partitions. J Classif. 1985;2:193–218.
23. Buttner M, Miao Z, Wolf FA, Teichmann SA, Theis FJ. A test metric for assessing single-cell RNA-seq batch correction. Nat Methods. 2019;16:43–9.
24. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53–65.
25. Durek P, Nordstrom K, Gasparoni G, Salhab A, Kressler C, de Almeida M, et al. Epigenomic profiling of human CD4(+) T cells supports a linear differentiation model and highlights molecular regulators of memory development. Immunity. 2016;45:1148–61.
26. Kallies A. Distinct regulation of effector and memory T-cell differentiation. Immunol Cell Biol. 2008;86:325–32.
27. Kanbar JN, Ma S, Kim ES, Kurd NS, Tsai MS, Tysl T, Widjaja CE, Limary AE, Yee B, He Z, et al. The long noncoding RNA Malat1 regulates CD8+ T cell differentiation by mediating epigenetic repression. J Exp Med. 2022;219:e20211756.
28. Cao W, Chen Y, Alkan S, Subramaniam A, Long F, Liu H, et al. Human T helper (Th) cell lineage commitment is not directly linked to the secretion of IFN-gamma or IL-4: characterization of Th cells isolated by FACS based on IFN-gamma and IL-4 secretion. Eur J Immunol. 2005;35:2709–17.
29. Chu DK, Mohammed-Ali Z, Jimenez-Saiz R, Walker TD, Goncharova S, Llop-Guevara A, et al. T helper cell IL-4 drives intestinal Th2 priming to oral peanut antigen, under the control of OX40L and independent of innate-like lymphocytes. Mucosal Immunol. 2014;7:1395–404.
30. Salkowska A, Karas K, Karwaciak I, Walczak-Drzewiecka A, Krawczyk M, Sobalska-Kwapis M, et al. Identification of novel molecular markers of human Th17 cells. Cells. 2020;9:1611.
31. Lee J, Ahn E, Kissick HT, Ahmed R. Reinvigorating exhausted T cells by blockade of the PD-1 pathway. Front Immunopathol Dis Ther. 2015;6:7–17.
32. Cao X, Cai SF, Fehniger TA, Song J, Collins LI, Piwnica-Worms DR, et al. Granzyme B and perforin are important for regulatory T cell-mediated suppression of tumor clearance. Immunity. 2007;27:635–46.
33. Maas-Bauer K, Kohler N, Stell AV, Zwick M, Acharya S, Rensing-Ehl A, et al. Single-cell transcriptomics reveal different maturation stages and sublineage commitment of human thymic invariant natural killer T cells. J Leukoc Biol. 2024;115:401–9.
34. Mackay LK, Minnich M, Kragten NA, Liao Y, Nota B, Seillet C, et al. Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. Science. 2016;352:459–63.
35. Wong EB, Gold MC, Meermeier EW, Xulu BZ, Khuzwayo S, Sullivan ZA, et al. TRAV1-2(+) CD8(+) T-cells including oligoconal expansions of MAIT cells are enriched in the airways in human tuberculosis. Commun Biol. 2019;2:203.
36. Nel I, Bertrand L, Toubal A, Lehuen A. MAIT cells, guardians of skin and mucosa? Mucosal Immunol. 2021;14:803–14.
37. Cooper MA, Fehniger TA, Turner SC, Chen KS, Ghaheri BA, Ghayur T, et al. Human natural killer cells: a unique innate immunoregulatory role for the CD56(bright) subset. Blood. 2001;97:3146–51.
38. Jacobs R, Hintzen G, Kemper A, Beul K, Kempf S, Behrens G, et al. CD56bright cells differ in their KIR repertoire and cytotoxic features from CD56dim NK cells. Eur J Immunol. 2001;31:3121–7.
39. Cattoretti G, Angelin-Duclos C, Shaknovich R, Zhou H, Wang D, Alobeid B. PRDM1/Blimp-1 is expressed in human B-lymphocytes committed to the plasma cell lineage. J Pathol. 2005;206:76–86.
40. Grimsholm O. CD27 on human memory B cells-more than just a surface marker. Clin Exp Immunol. 2023;213:164–72.
41. Sanz I, Wei C, Jenks SA, Cashman KS, Tipton C, Woodruff MC, et al. Challenges and opportunities for consistent classification of human B cell and plasma cell populations. Front Immunol. 2019;10:2458.
42. Chen S, Saeed A, Liu Q, Jiang Q, Xu H, Xiao GG, et al. Macrophages in immunoregulation and therapeutics. Signal Transduct Target Ther. 2023;8:207.
43. Gong C, Ma J, Deng Y, Liu Q, Zhan Z, Gan H, et al. S100A9(-/-) alleviates LPS-induced acute lung injury by regulating M1 macrophage polarization and inhibiting pyroptosis via the TLR4/MyD88/NFkappaB signaling axis. Biomed Pharmacother. 2024;172:116233.
44. Orecchioni M, Ghosheh Y, Pramod AB, Ley K. Macrophage polarization: different gene signatures in M1(LPS+) vs. classically and M2(LPS-) vs. alternatively activated macrophages. Front Immunol. 2019;10:1084.
45. Chen L, Mei W, Song J, Chen K, Ni W, Wang L, et al. CD163 protein inhibits lipopolysaccharide-induced macrophage transformation from M2 to M1 involved in disruption of the TWEAK-Fn14 interaction. Heliyon. 2024;10:e23223.

46. Nasim S, Bichsel C, Dayneka S, Mannix R, Holm A, Vivero M, et al. MRC1 and LYVE1 expressing macrophages in vascular beds of GNAQ p.R183Q driven capillary malformations in Sturge Weber syndrome. Acta Neuropathol Commun. 2024;12:47.
47. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(3573–3587):e3529.
48. Mak TW, Saunders ME. 14 - T Cell Activation. In: Mak TW, Saunders ME, editors. The immune response. Burlington: Academic Press; 2006. p. 373–401.
49. van der Leun AM, Thommen DS, Schumacher TN. CD8(+) t cell states in human cancer: insights from single-cell analysis. Nat Rev Cancer. 2020;20:218–32.
50. Fridman WH, Pages F, Sautes-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer. 2012;12:298–306.
51. Ghoneim HE, Zamora AE, Thomas PG, Youngblood BA. Cell-intrinsic barriers of t cell-based immunotherapy. Trends Mol Med. 2016;22:1000–11.
52. Grosser R, Cherkassky L, Chintala N, Adusumilli PS. Combination immunotherapy with CAR t cells and checkpoint blockade for the treatment of solid tumors. Cancer Cell. 2019;36:471–82.
53. Cherkassky L, Morello A, Villena-Vargas J, Feng Y, Dimitrov DS, Jones DR, et al. Human CAR t cells with cell-intrinsic PD-1 checkpoint blockade resist tumor-mediated inhibition. J Clin Invest. 2016;126:3130–44.
54. Liu Q, Hsu CY, Li J, Shyr Y. Dysregulated ligand-receptor interactions from single-cell transcriptomics. Bioinformatics. 2022;38:3216–21.
55. He J, Chen Y, Ding H, Zhou JA, Xing Z, Yang X, et al. Autocrine VEGF-B signaling maintains lipid synthesis and mitochondrial fitness to support T cell immune responses. J Clin Invest. 2024;134:e176586.
56. Shin JY, Yoon IH, Kim JS, Kim B, Park CG. Vascular endothelial growth factor-induced chemotaxis and IL-10 from T cells. Cell Immunol. 2009;256:72–8.
57. Rosenberg J, Huang J. CD8(+) t cells and NK cells: parallel and complementary soldiers of immunotherapy. Curr Opin Chem Eng. 2018;19:9–20.
58. Sun D, Wang J, Han Y, Dong X, Ge J, Zheng R, et al. TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. Nucleic Acids Res. 2021;49:D1420–30.
59. Andreatta M, Corria-Osorio J, Muller S, Cubas R, Coukos G, Carmona SJ. Interpretation of T cell states from single-cell transcriptomics data using reference atlases. Nat Commun. 2021;12:2965.
60. Wang H, Chen L, Qi L, Jiang N, Zhang Z, Guo H, et al. A single-cell atlas of tumor-infiltrating immune cells in pancreatic ductal adenocarcinoma. Mol Cell Proteomics. 2022;21:100258.
61. Ma L, Wang L, Khatib SA, Chang CW, Heinrich S, Dominguez DA, et al. Single-cell atlas of tumor cell evolution in response to therapy in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. J Hepatol. 2021;75:1397–408.
62. Scrucca L, Fop M, Murphy TB, Raftery AE. Mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J. 2016;8:289–317.
63. Tran HTN, Ang KS, Chevrier M, Zhang X, Lee NYS, Goh M, et al. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. Genome Biol. 2020;21:12.
64. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.
65. Yang J. Tumor-immune-atlases. Zenodo. 2025. https://doi.org/10.5281/zenodo.15867143.
66. Yang J. Tumor-immune-atlases. Datasets, Github. 2025. https://github.com/JingYangSciBio/Tumor-immune-atlases.
67. Yang J. Tumor-immune-atlases. Zenodo. 2025. https://doi.org/10.5281/zenodo.18989719.
68. Bi K, He MX, Bakouny Z, Kanodia A, Napolitano S, Wu J, Grimaldi G, Braun DA, Cuoco MS, Mayorga A, et al. Tumor and immune reprogramming during immunotherapy in advanced renal cell carcinoma. Datasets, Single Cell Portal. 2021. https://singlecell.broadinstitute.org/single_cell/study/SCP1288/tumor-and-immune-reprogramming-during-immunotherapy-in-advanced-renal-cell-carcinoma#study-summary.
69. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC, Leeson R, Kanodia A, Mei S, Lin J, et al. Single-cell RNA-seq of melanoma ecosystems reveals sources of T cells exclusion linked to immunotherapy clinical outcomes. Datasets, Gene Expression Omnibus. 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115978.
70. Yost KE, Satpathy AT, Wells DK, Qi Y, et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. Datasets, Gene Expression Omnibus. 2019. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123814.
71. Liu B, Hu X, Feng K, Gao R, et al. Temporal single cell tracing reveals clonal revival and expansion of precursor exhausted T cells during anti-PD-1 therapy in lung cancer. Datasets, Gene Expression Omnibus. 2021. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE179994.
72. Sade-Feldman M, Yizhak K, Bjorgaard SL, Ray JP, et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. Datasets. 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120575.
73. Caushi JX, Zhang J, Ji Z, Vaghasia A, et al. Distinct transcriptional programs characterize neoantigen-specific T cells in lung cancers treated with neoadjuvant PD-1 blockade [single-cell RNA-seq]. Datasets. 2021. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE176021.

Publisher's Note