The Journal of Immunology

RESEARCH ARTICLE | JANUARY 15 2022

Reinvestigation of Classic T Cell Subsets and Identification of Novel Cell Subpopulations by Single-Cell RNA Sequencing

Xuefei Wang; ... et. al

J Immunol (2022) 208 (2): 396–406.
https://doi.org/10.4049/jimmunol.2100581

Related Content

Identification of BST2 as a biomarker for alopecia areata in both mice and humans

J Immunol (May,2023)

Cutting Edge: Paradoxical Roles of BST2/Tetherin in Promoting Type I IFN Response and Viral Infection

J Immunol (March,2012)

Increased proliferation of epidermal gamma delta T cells and expression of the transmembrane protein, BST2, in Alopecia areata

J Immunol (May,2022)

Downloaded from http://journals.aai.org/jimmunol/article-pdf/208/2/396/1504672/ji2100581.pdf by guest on 30 September 2023

Reinvestigation of Classic T Cell Subsets and Identification of Novel Cell Subpopulations by Single-Cell RNA Sequencing

Xuefei Wang,[1] Xiangru Shen,[1] Shan Chen, Hongyi Liu, Ni Hong, Hanbing Zhong, Xi Chen, and Wenfei Jin

Classic T cell subsets are defined by a small set of cell surface markers, while single-cell RNA sequencing (scRNA-seq) clusters cells using genome-wide gene expression profiles. The relationship between scRNA-seq clustered populations (scCPops) and cell surface marker–defined classic T cell subsets remains unclear. In this article, we integrated six bead-enriched T cell subsets with 62,235 single-cell transcriptomes from human PBMCs and clustered them into nine scCPops. Bead-enriched $CD4^+$/$CD45RA^+$/$CD25^-$ naive T and $CD8^+$/$CD45RA^+$ naive T cells were mainly clustered into their scCPop counterparts, while cells from the other T cell subsets were assigned to multiple scCPops, including mucosal-associated invariant T cells and NKT cells. The multiple T cell subsets forming one scCPop exhibit similar expression patterns, but not vice versa, indicating scCPop is a more homogeneous cell population with similar cell states. Interestingly, we discovered and named IFN signaling–associated gene (ISAG) high T ($ISAG^{hi}$ T) cells, a T cell subpopulation that highly expressed ISAGs. We further enriched $ISAG^{hi}$ T cells from human PBMCs by FACS of BST2 for scRNA-seq analyses. The $ISAG^{hi}$ T cell cluster disappeared on *t*-distributed stochastic neighbor embedding plot after removing ISAGs, whereas the $ISAG^{hi}$ T cell cluster showed up by analysis of ISAGs alone, indicating ISAGs are the major contributor of the $ISAG^{hi}$ T cell cluster. $BST2^+$ and $BST2^-$ T cells showing different efficiencies of T cell activation indicate that a high level of ISAGs may contribute to quick immune responses. ***The Journal of Immunology*****, 2022, 208: 396–406.**

The T cells also known as T lymphocytes are an essential component of adaptive immunity for protecting us against various pathogens. In fact, T cells not only play a key role in elimination of invasive pathogens and cancer cells but also are important targets in the treatment of autoimmune diseases (1–5). The major T cells can be classified into CD4 T cells and CD8 T cells based on cell functions and cell surface markers. CD4 T cells, also known as Th cells, play a central role in adaptive immune response to pathogens, such as assisting B cells to produce Abs, recruiting granulocytes to infected sites, and producing cytokines and chemokines to orchestrate immune response (6). Many distinct CD4 T cell subsets have been identified since the identification of Th1 by Mosmann et al. (6, 7). $CD4^+CD25^+$ T cells, namely, regulatory T cells (Tregs), suppress self-reactive lymphocytes to maintain immunologic self-tolerance (8, 9). CD8 T cells are important for immune defense against intracellular pathogens (10, 11). In contrast, naive T cells egress from the thymus and are fairly quiescent, which could differentiate into distinct functional T cell subsets after they interact with cognate Ags, while memory T cells recognize previously encountered pathogens and respond to them quickly (3). These T cell subsets could be enriched or sorted out by FACS or magnetic beads because each of them is defined by a set of specific cell surface markers. Then population-average assays were routinely used to characterize the biochemical and molecular features of each T cell subset.

Single-cell RNA sequencing (scRNA-seq), profiling gene expression at single-cell resolution, has become an ideal approach for identifying cellular heterogeneity, searching new and/or rare cell types, pseudotime inference, and inferring regulators underlying lineage changes (5, 12–15). Most studies focused on searching cellular

Shenzhen Key Laboratory of Gene Regulation and Systems Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, China

[1]X.W. and X.S. contributed equally to this work.

ORCIDs: 0000-0003-0612-8745 (S.C.); 0000-0002-9404-3110 (N.H.); 0000-0001-9597-135X (H.Z.); 0000-0003-2648-3146 (X.C.); 0000-0002-0028-3739 (W.J.).

Received for publication June 16, 2021. Accepted for publication November 2, 2021.

This work was supported by the National Key Research and Development Program of China Grant 2018YFC1004500, National Natural Science Foundation of China Grants 81872330 and 31771603, Shenzhen Science and Technology Program Grants KQTD20180411143432337, JCYJ20170412153453623, and JCYJ20180504170158430, and Shenzhen Innovation Committee of Science and Technology Grant ZDSYS2020-0811144002008.

W.J. conceived and designed the project. S.C. did the experiment. X.W. and X.S. analyzed the data, with contributions from H.L. W.J., N.H., H.Z., and X.C. supervised the study. W.J. and X.W. wrote the manuscript. All authors have read and approved the manuscript.

The raw sequence data reported in this article have been submitted to the Genome Sequence Archive in Beijing Institute of Genomics Data Center (https://ngdc.cncb.ac.cn/gsa-human/browse/HRA000313) under accession number HRA000313.

Address correspondence and reprint requests to Prof. Wenfei Jin, School of Life Sciences, Southern University of Science and Technology, 1088 Xueyuan Road, Shenzhen 518055, China. E-mail address: jinwf@sustech.edu.cn

The online version of this article contains supplemental material.

Abbreviations used in this article: Cat., catalog; CD4 memory, $CD4^+$/$CD45RO^+$ memory T; CD4 naive, $CD4^+$/$CD45RA^+$/$CD25^-$ naive T; CD8 naive, $CD8^+$/$CD45RA^+$ naive T; CD4 Th, $CD4^+$ Th cell; CD4 Treg, $CD4^+$/$CD25^+$ regulatory T cell; DEG, differentially expressed gene; GO, Gene Ontology; ISAG, IFN signaling–associated gene; $ISAG^{hi}$ T, IFN signaling–associated gene high T; MAIT, mucosal-associated invariant T cell; PC, principal component; scCD4 T_{EMRA}, CD4 effector memory T cell re-expressing CD45RA scCPop; scCD8 T_{EMRA}/scCD8 T_{EFF}, CD8 effector memory T cell with CD45RA or CD8 effector cell; scCD4 T_M, CD4 memory T cell scCPop; scCD4 T_N, CD4 naive T cell scCPop; scCD8 T_N, CD8 naive T cell scCPop; scCPop, scRNA-seq clustered population; scMAIT, mucosal-associated invariant T cell scCPop; scRNA-seq, single-cell RNA sequencing; scT_M, scCPop #7; SUSTech, Southern University of Science and Technology; Treg, regulatory T cell; *t*-SNE, *t*-distributed stochastic neighbor embedding; UMI, unique molecular identifier.

This article is distributed under The American Association of Immunologists, Inc., Reuse Terms and Conditions for Author Choice articles.

Copyright © 2022 by The American Association of Immunologists, Inc. 0022-1767/22/$37.50

Downloaded from http://journals.aai.org/jimmunol/article-pdf/208/2/396/1504672/ji2100581.pdf by guest on 30 September 2023

heterogeneity and new cell types in well-defined T cell subsets because of the high statistical power of scRNA-seq to distinguish cellular heterogeneities (16–21). Only a few studies compared classic T cell subsets with scRNA-seq clustered populations (scCPops) (22); thus, our knowledge about the cellular relationship between classic T cell subsets and scCPops is limited. It remains unknown whether we could find the scCPop counterpart for each classic T cell subset, whether classic T cell subsets could represent well their scCPop counterparts, and whether more T cell subpopulations could be identified by reclustering multiple classic T cell subsets.

In this study, we integrated scRNA-seq data of six bead-enriched T cell subsets for investigating the cellular relationships between classic T cell subsets and scCPops. We found bead-enriched $CD4^+$/$CD45RA^+$/$CD25^-$ naive T cells (CD4 naive), $CD8^+$/$CD45RA^+$ naive T (CD8 naive) cells, and CD4 memory cells were mainly clustered into their scCPop counterparts, while cells from the other T cell subsets were assigned to multiple scCPops. The multiple T cell subsets that form a single scCPop exhibited similar expression patterns, but not vice versa, indicating that scCPops are more homogeneous cell populations. We unexpectedly identified a new T cell subpopulation that highly expresses IFN signaling–associated genes (ISAGs); we named it ISAG high T ($ISAG^{hi}$ T). Because *BST2* is located on the cell membrane and is highly expressed in $ISAG^{hi}$ T cells, it was selected to enrich $ISAG^{hi}$ T cells. $BST2^+$ T cells showed quick T cell activation upon anti-CD3/CD28 costimulation, indicating a high level of ISAGs may contribute to quick immune responses.

Materials and Methods

Isolation of PBMCs

This study was approved by Institutional Review Boards at Southern University of Science and Technology (SUSTech). All experiments were conducted following the protocols approved by Institutional Review Boards at SUSTech. Human peripheral blood samples were obtained from a healthy adult blood donor (donor B). PBMCs were isolated by density-gradient sedimentation using Lymphoprep (Axis-Shield, Oslo, Norway) following the manufacturer's instructions. In brief, a peripheral blood sample was diluted with an equal volume of PBS containing 2% FBS, then transferred into a Lymphoprep tube. After centrifugation at 800 × *g* for 20 min at room temperature, the middle white phase was transferred to a fresh tube. The enriched mononuclear cells were washed with PBS containing 2% FBS and centrifuged at 500 × *g* for 5 min twice. Cell count and viability were assessed by hemocytometer counting.

Isolation of T cells and T cell subpopulations

PBMCs were resuspended in FACS buffer (PBS containing 2% FBS) and incubated with Human TruStain FcX (Fc Receptor Blocking Solution; catalog [Cat.] #422302; BioLegend) for 15 min on ice. After anti-human CD14-allophycocyanin (61D3, Cat. #17-0149-42; eBioscience) and anti-human CD3-FITC (HIT3a, Cat. #11-0039-42; eBioscience) were added into PBMCs, they were incubated on ice for 30 min. These cells were washed twice with FACS buffer and counterstained with 1 μg/ml DAPI. These cells were sorted on FACS Aria-III (Becton Dickinson) with $CD14\text{-}CD3^+$ gating to obtain T cells. FlowJo (version 10) was used for FACS data analyses and figure plot. Besides, we simultaneously added anti-human CD317-PE (RS38E, Cat. #348405; BioLegend) during Ab incubation for sorting $BST2^+$ T or $BST2^{hi}$ T cells.

T cell activation

Following the manufacturer's instructions, FACS-sorted $BST2^+$ and $BST2^-$ T cells were stimulated with Dynabeads Human T-Activator CD3/CD28 for T Cell Expansion and Activation (Cat. #11161D; Thermo Fisher). In brief, 8×10^4 $BST2^+$ or $BST2^-$ T cells sorted by FACS were suspended with 100–200 μL medium, respectively. Prewashed Dynabeads were added at the bead/cell ratio of 1:1 and incubated in a humidified CO_2 incubator at 37°C. After 12 h, the bead-stimulated T cells were harvested by removing Dynabeads and then stained with anti-CD25 (Cat. #4302545; eBioscience) for FACS analyses of the efficiency of T cell activation.

scRNA-seq library preparation and sequencing

FACS-sorted $BST2^+$ T cells were directly processed for scRNA-seq with 10X Genomics 3′ kit v3 (10X Genomics, Pleasanton, CA) following the manufacturer's instructions. In short, $\sim 1.6 \times 10^4$ $BST2^+$ T cells were loaded into a single inlet of a 10X Genomics Chromium controller to generate the Gel Beads-in-Emulsions. Two independent biological replicates were prepared, with each replicate recovering ∼8000 cells. Gel Beads-in-Emulsions were used to complete the reverse transcription, cDNA amplification, and library preparation. scRNA-seq libraries were prepared using Chromium Single Cell 3′ Library Construction Kit v3 (P/N 1000078; 10X Genomics). Sequencing libraries were loaded at 2.4 pM on Illumina NovaSeq 6000 with 2 × 150 paired-end kits using the following read length: 28 bp Read1, 8 bp I7 Index, 8 bp I5 Index, and 91 bp Read2. About 40,000 reads per cell were generated for each sample.

scRNA-seq data of bead-enriched T cell subsets

We downloaded scRNA-seq data of six bead-enriched T cell subsets and PBMCs from the Web site of 10X Genomics (https://support.10xgenomics.com/single-cell-gene-expression/datasets). The six bead-enriched T cell subsets are $CD4^+$ Th cells (CD4 Th), CD4 naive, $CD4^+$/$CD25^+$ Tregs (CD4 Tregs), $CD4^+$/$CD45RO^+$ memory T cells (CD4 memory) cells, $CD8^+$ CTLs, and CD8 naive cells. FACS analyses showed the high purity of these T cell subsets: CD4 Th (99%), CD4 naive (98%), CD4 memory (98%), CD4 Treg (95%), CD8 naive (99%), and CD8 CTL (98%). All six bead-enriched T cell subsets and PBMCs in this part are from one healthy blood donor (donor A), and the basic information is available in Zheng et al. (23). These T cell subsets were processed by 10X Genomics Chromium with the aim of capturing 10,000 cells for each sample.

Data preprocessing and quality control

Cell ranger was used for generating the single-cell expression matrix. We filtered out the cells of bead-enriched T subsets meeting any of the following criteria: (1) <200 unique genes, (2) >1500 unique genes, (3) >50,00 unique molecular identifier (UMI) counts, and (4) >5% mitochondrial counts. However, the criterion for FACS-sorted $BST2^+$ T cells is (1) <200 unique genes, (2) >6000 unique genes, (3) >45,000 UMI counts, or (4) >20% mitochondrial counts. After quality control, we normalized the expression level of each gene to the total number of UMIs for each cell.

Dimension reduction and visualization of scRNA-seq data

Seurat (v3.0) (24) was used for data integration, data normalization, dimension reduction, cell clustering, and other basic scRNA-seq data analyses following our previous studies (5, 15). We calculated the cell-to-cell variation of each gene and selected the top 2000 genes exhibiting the highest variation for further analysis. To avoid highly expressed genes dominating in later analyses, we scaled the mean and variance of each gene across cells as 0 and 1, respectively. *t*-Distributed stochastic neighbor embedding (*t*-SNE) (15, 25, 26) in Seurat was used for data visualization. Cell embeddings of merged six bead-enriched T subsets cells, CD4 T cell subsets, and CD8 T cell subsets in the *t*-SNE plot were calculated based on 20 principal components (PCs), 0.6 resolution, and 35 perplexity. For

Downloaded from http://journals.aai.org/jimmunol/article-pdf/208/2/396/1504672/ji2100581.pdf by guest on 30 September 2023

fine analyses of CD4 T cell subsets, 1.0 resolution was used to obtain detail clusters. Although cell embeddings of FACS-sorted $BST2^{+}$ T cells were calculated based on 50 PCs, 2.0 resolution, and 30 perplexity, we displayed cluster-specific expression genes on the *t*-SNE plot, which provide nice visualization for distinguishing different cell clusters.

Gene Ontology enrichment analysis

To examine whether particular Gene Ontology (GO) terms were enriched in certain gene sets, we carried out GO enrichment analysis using DAVID (27). GO categories with adjusted $p < 0.05$ were considered as significant.

Relationships between bead-enriched T cell subsets and scCPops

The Sankey diagram is an intuitive visualization approach for depicting flows from one set of values to another. We plotted the Sankey diagram using Highcharts (https://www.highcharts.com/demo/sankey-diagram) to explore the relationships between bead-enriched T cell subsets and scCPops.

Single-cell trajectories inference

We implemented Slingshot (28) on scRNA-seq data to infer the trajectory of CD4 and CD8 T cells. The scCPops were used as input for Slingshot, and the top 20 PCs were used to infer the pseudotime. Cell lineages were visualized on the PC analysis three-dimensional diagram using plot3D in R.

Statistical analysis

The statistical tests and plots were conducted using R version 3.5.2 (December 12, 2018) (29).

Results

Single-cell mapping of bead-enriched T cell subsets

To analyze the relationship between classic T cell subsets and scCPops, we designed a workflow for single-cell analysis of six T cell subsets (Fig. 1A). The scRNA-seq data of six bead-enriched T cell subsets from a healthy donor (donor A), namely, CD4 Th cells, CD4 naive cells, CD4 memory cells, CD4 Tregs, CD8 CTLs, and

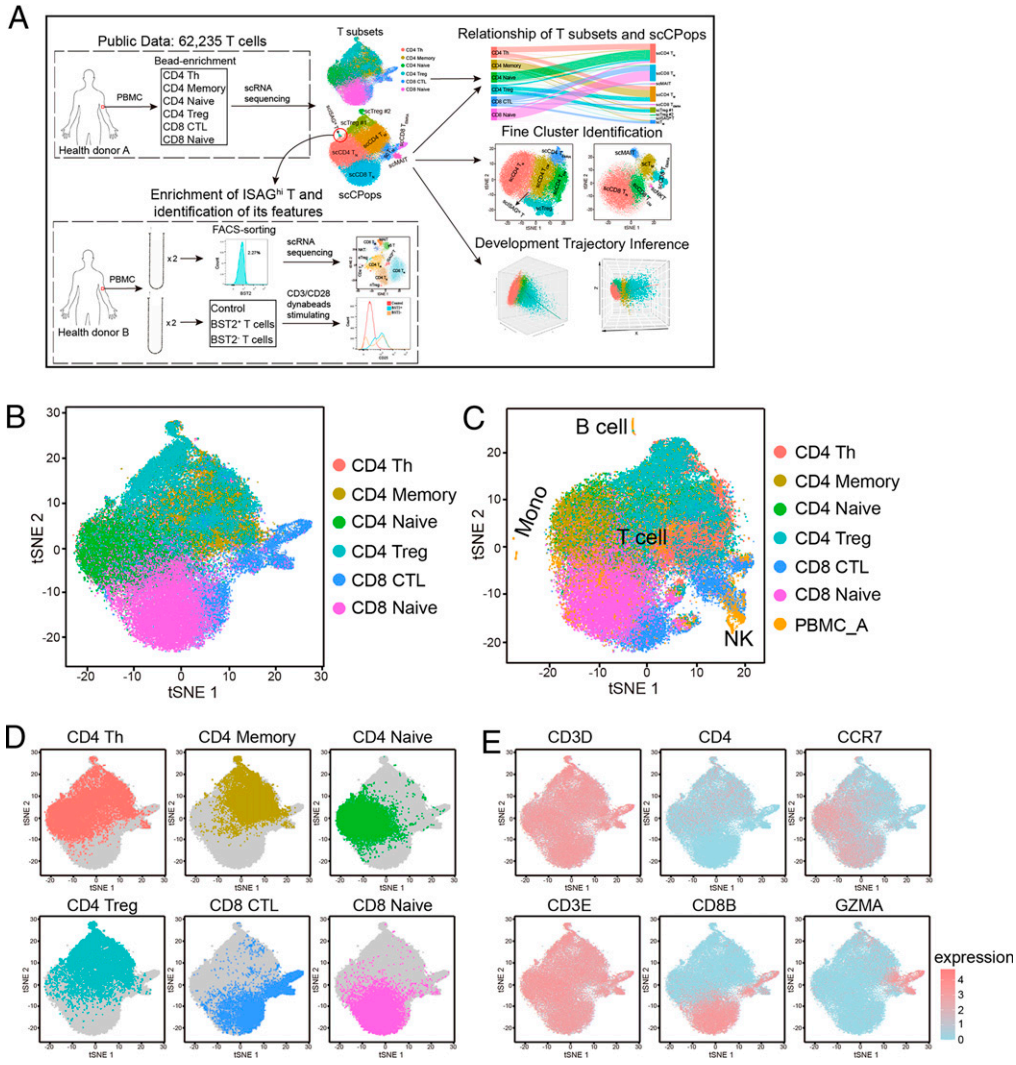

FIGURE 1. The six bead-enriched T cell subsets and their features. (**A**) The schematic of this study. (**B**) *t*-SNE projection of the six bead-enriched T cell subsets, colored by each T cell subset. (**C**) *t*-SNE projection of PBMCs and the six T cell subsets. The PBMCs and six T cell subsets are colored differently. (**D**) Distribution of T cell subset in a *t*-SNE plot of the six bead-enriched T cell subsets, highlighting one T cell subset in each panel. (**E**) *t*-SNE projection of the six T cell subsets, with each cell colored based on their normalized expression of *CD3D*, *CD3E*, *CD4*, *CD8B*, *CCR7*, and *GZMA*, respectively.

Downloaded from http://journals.aai.org/jimmunol/article-pdf/208/2/396/1504672/ji2100581.pdf by guest on 30 September 2023

CD8 naive cells, were obtained from Zheng et al. (23) (Table I). Analyses of each T cell subset showed there was no obvious subpopulation in each T cell subset except CD8 CTLs (Supplemental Fig. 1A), which is essentially consistent with the original report that these T cell subsets are relatively pure (23). The 62,235 cells from the six bead-enriched T cell subsets were projected on the *t*-SNE plot and formed a single connected entity (Fig. 1B). The CD4 T cell subsets, namely, CD4 naive, CD4 memory, CD4 Treg, and CD4 Th cells, are mixed together on the *t*-SNE plot. There are five distinct cell clusters (T cells, NK cells, B cells, CD14 monocytes, and CD16 monocytes) in the scRNA-seq data of PBMCs from donor A (Supplemental Fig. 1B, 1C). The six bead-enriched T cell subsets are overlapped with the T cells from PBMCs on the *t*-SNE plot (Fig. 1C), indicating the cellular heterogeneity among T cells are quite small compared with that between T cells and other cell types. Highlighting the T cell subsets one by one showed that each T cell subset was concentrated on a specific location of the *t*-SNE plot (Fig. 1D).

We further calculated the cell–cell variation of each gene. The genes showing the lowest cell–cell variation are enriched in housekeeping genes and T lineage–associated genes (Supplemental Fig. 1D, 1E). T cell marker genes *CD3D* and *CD3E*, encoding the δ- and ε-chain of CD3 protein, respectively, are expressed in all T cell subsets (Fig. 1E, Supplemental Fig. 1D). In contrast, the genes showing the highest cell–cell variation are enriched in immune effector process, T cell activation, immune effector process, cytokine signaling, and cytotoxic-related genes, including *GNLY*, *CCL5*, *NKG7*, *GZMH*, *LYZ*, *GZMB*, and *GZMK* (Supplemental Fig. 1F, 1G). *CD4* and *CD8B* (encoding β-chain of CD8 protein) are highly expressed in the upper part and lower part of the *t*-SNE plot, respectively (Fig. 1E), consistent with the observation that CD4 and CD8 T cell subsets are in the upper part and lower part of the *t*-SNE plot, respectively. Expression of *CCR7* (naive T cell marker gene in scRNA-seq) and *GZMA* indicate the naive/naive-like T cells are on the left of the *t*-SNE plot and cells secreting cytotoxic protein are on the right of the *t*-SNE plot, respectively. Overall, expression of marker genes on the *t*-SNE plot indicates that clustering cells by scRNA-seq data could provide further biological insights.

scRNA-seq clustered populations

The 62,235 T cells from the six bead-enriched T cell subsets were clustered into nine clusters/scCPops by Seurat (24, 30) (Fig. 2A). The cell type of each scCPop was inferred based on scCPop-specific highly expressed genes (Fig. 2B, 2C, Supplemental Fig. 2A, 2B). The scCPops were named after their classic T cell counterpart with "sc" prefix. Among the nine T cell scCPops, five exhibit strong signals matching well-known classic T cell subsets: CD4 naive T cell (scCD4 T_N; *CD4*, *CCR7*, *LEF1*, *SELL*), CD8 naive T cell (scCD8 T_N; *CD8B*, *CCR7*, *LEF1*, *SELL*), CD4 memory T cell (scCD4 T_M; *CD4*, *TNFRSF4*, *TIMP1*, *USP10*), CD8 effector memory T cell with CD45RA or CD8 effector cells (scCD8 T_{EMRA}/scCD8 T_{EFF}; *CD8B*, *FGFBP2*, *GZMH*, *GZMB*), and mucosal-associated invariant T cell (scMAIT; *KLRB1*, *NCR3*, *ZBTB16*, *SLC4A10*, *RORC*). Two scCPops exhibit strong signals matching CD4 Tregs (*FOXP3*, *IL2RA*, *CCR10*), which were designated as CD4 Treg 1 (scTreg#1) and CD4 Treg 2 (scTreg#2), respectively. Compared with scTreg#1, the scTreg#2 population is significantly higher expressed in *MIR4435-1HG*, *LINC00152*, *KLRB1*, *LIMS1*, *NOSIP*, and *CD59* (Supplemental Fig. 2C), indicating scTreg#2 is in active proliferation (31, 32). The high expression of *TIMP1* and *TNFRSF4* in scCD4 T_M potentially explains the long life of memory T cells because both genes have antiapoptotic function (33, 34).

We identified 36 significantly high expressed genes in scCPop #5 by comparing scCPop #5 with the other cells. These most significant genes include *ISG15*, *IFI6*, *OAS1*, *MX1*, *MX2*, *IRF7*, *IFI44L*, *XAF1*, *STAT1*, *IFIT1*, *IFIT3*, and *BST2* (Fig. 2B, 2D), which does not match any predefined T cell subsets. GO analyses of the 36 genes showed the most significantly enriched GO terms were type I IFN signaling pathway (4.9×10^{-25}), defense response to virus (1.8×10^{-18}), negative regulation of viral genome replication (3.1×10^{-17}), response to virus (3.5×10^{-14}), response to IFN-β (1.1×10^{-8}), and IFN-γ–mediated signaling pathway (9.0×10^{-8}) (Fig. 2E). We named this scCPop as scISAG^{hi} T because all those significantly enriched GO terms belong to IFN signaling. The discovery of scISAG^{hi} T in bead-enriched T cell subsets further pinpointed the cellular heterogeneity in well-defined classic T cell subsets.

Different from the other scCPops either belonging to the CD4 T cell subset or the CD8 T cell subset, scCPop #7 contains 41.39% CD4 T cells and 58.61% CD8 T cells (Fig. 2F). This scCPop was named scT$_M$ because it highly expressed memory T cell–specific genes, such as *CXCR3*, *GZMK*, *TIMP1*, *LYAR*, *KLF6*, *DUSP1*, *JUN*, and *FOS* (Fig. 2B, 2G). The CD8 T cells in scT$_M$ highly expressed *NKG7*, *CCL5*, *CTSW*, and *CD7* (Fig. 2H), indicating those T cells are scCD8 T_{EM}, while CD4 T cells highly expressed *IL32*, *LTB*, *GZMA*, *S100A11*, *CXCR3*, and *GIMAP7* (Fig. 2H), indicating those cells are CD4 effector memory T cells re-expressing CD45RA (scCD4 T_{EMRA}) or CTLs (scCD4-CTLs). It is well known that CD4 and CD8 T cells diverged into two functional different lineages after double-positive T cells (4). The scCD4 T_{EMRA}/scCD4-CTL and scCD8 T_{EM} cells were clustered into one subpopulation (scT$_M$) potentially because both cell subsets express cytotoxic associated genes. The observation that CD4 and CD8 T cells could differentiate into similar phenotypes could be a model of functional convergence development.

Cellular relationships between classic T cell subsets and scCPops

The relationship between classic T cell subsets and scCPops was demonstrated by Sankey diagram (Fig. 3A). Bead-enriched CD4 naive and CD8 naive cells mainly clustered into their counterparts in scCPops, namely, scCD4 T_N and scCD8 T_N, respectively, indicating the consistency between classic T cell enrichment and scRNA-seq clustering on those cell populations. These bead-enriched T cells

Table I. The scRNA-seq data of six bead-enriched T cell subsets and FACS-sorted BST2^{hi} T cells after quality control

T Cell Subsets/Sample Names	Median Genes per Cell	Median UMI Counts per Cell	No. of Genes	No. of Cells
Donor A [Zheng et al., 2017 (23)]				
CD4 Th cells	546	1328	13,572	10,776
CD4 Tregs	547	1222	13,486	9,806
CD4 memory cells	547	1503	13,730	9,982
CD4 naive cells	493	1193	13,033	10,132
CD8 CTLs	572	1626	13,704	9,946
CD8 naive cells	502	1442	13,247	11,593
Donor B (this study)				
BST2^{hi} #1	1406	4935	18,585	6,352
BST2^{hi} #2	1373	5604	18,799	7,325

Downloaded from http://journals.aai.org/jimmunol/article-pdf/208/2/396/1504672/ji2100581.pdf by guest on 30 September 2023

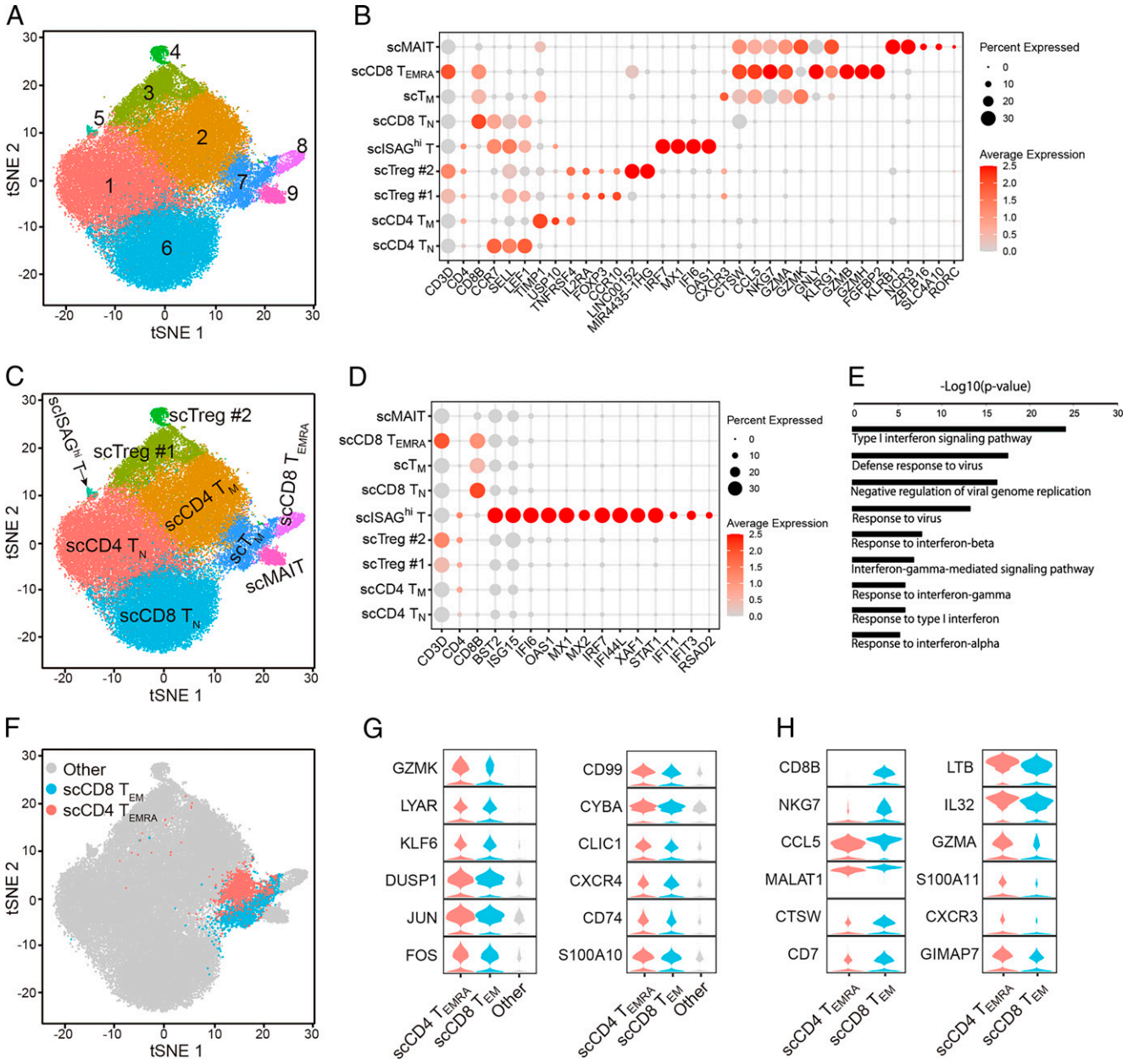

FIGURE 2. The nine scCPops and their features. (**A**) *t*-SNE projection of the 62,235 T cells, where each cell was grouped into one of the nine clusters distinguished by colors. Numbers in the panel indicate cluster identifier. (**B**) Normalized expression level and expression percentage of the cell-type-specific genes in the nine scCPops. Color represents normalized expression level, and dot size represents expression percentage. (**C**) *t*-SNE projection of the 62,235 T cells, colored by the nine scCPops, with inferred cell types on the panel. (**D**) Dot plot of scISAG^{hi}-specific expressed genes. Color represents normalized expression level, and dot size represents expression percentage. (**E**) GO enrichment analysis of scISAG^{hi}-specific expressed genes. (**F**) The number of CD4 T cells and CD8 T cells almost equally contributed to scT$_M$. (**G**) scT$_M$-specific expressed genes in their CD4 T subset and CD8 T subset. (**H**) DEGs between the CD4 T subset and the CD8 T subset of scT$_M$.

solely clustering into their scCPop counterparts also indicated those T cell subsets are relatively homogeneous. In contrast, CD4 Tregs and CD4 memory cells were assigned to multiple scCPops, indicating high heterogeneity in the two T cell subsets. For instance, Tregs were mainly assigned to five different scCPops, namely, scCD4 T$_N$, scCD4 T$_M$, scT$_M$, scTreg#1, and scTreg#2. As expected, CD4 Th cells and CD8 CTLs contributed to all CD4 scCPops and all CD8 scCPops, respectively. Although there are differences between scCPops and their classic counterparts, classic T subsets and their scCPop counterparts showed the highest similarity among all T cell populations (Supplemental Fig. 2E). Furthermore, the correlations between naive T subsets and their counterparts are higher than that of the other T cell subsets. Differentially expressed gene (DEG) analyses showed the number of DEGs between naive T cells and their counterparts was much less than that of other T cell subsets (Supplemental Fig. 2D). These results are consistent with the Sankey diagram that naive T cell subsets are much more similar to their counterparts than other T cell subsets.

Cell composition analyses of bead-enriched T cell subsets based on scCPops directly demonstrated the cellular heterogeneity in classic T cell subsets. The CD4 naive subset consists of 91.21% scCD4 T$_N$, 5.56% scCD8 T$_N$, 1.74% scCD4 T$_M$, and 0.63% scTreg#1 (Fig. 3B). The CD8 naive subset consists of 85.23% scCD8 T$_N$ and

Downloaded from http://journals.aai.org/jimmunol/article-pdf/208/2/396/1504672/ji2100581.pdf by guest on 30 September 2023

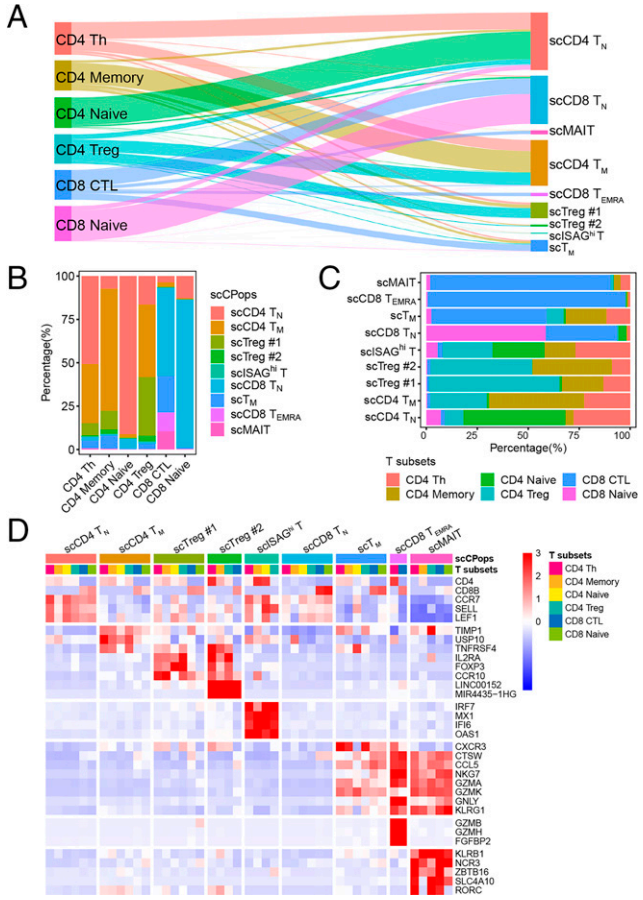

FIGURE 3. Relationships between bead-enriched T cell subsets and scCPops. (**A**) Cellular relationships between the six bead-enriched T cell subsets and the nine scCPops by Sankey diagram. Each node represents a cell population, and the width of line between two nodes represents number of shared cells. (**B**) Cell compositions of bead-enriched T cell subsets by bar chart. (**C**) Bar plot displays cell compositions of scCPops based on bead-enriched T cell subsets. (**D**) Gene expression heatmap of cell subpopulations generated by combinations of scCPops and bead-enriched T cell subsets.

12.79% scCD4 T_N. The CD4 memory subset consists of 70.35% scCD4 T_M, 10.79% scTreg#1, 7.47% scT_M, and 7.35% scCD4 T_N (Fig. 3B). The CD4 Treg subset consists of 41.72% scCD4 T_M, 33.86% scTreg#1, 16.57% scCD4 T_N, 3.26% scTreg#2, and 3.07% scT_M (Fig. 3B). The CD4 Th cell subset consists of 50.75% scCD4 T_N, 34.11% scCD4 T_M, 6.98% scT_{Reg}#1, 4.32% scT_M, and 2.14% scCD8 T_N (Fig. 3B). CD8 CTLs consist of 50.96% scCD8 T_N, 21.03% scT_M, 10.82% scCD8 T_{EMRA}, 10.49% scMAIT, 3.66% scCD4 T_N, and 2.31% scCD4 T_M, which should be the most heterogeneous among the six bead-enriched T cell subsets.

We further calculated the cell composition of scCPops based on classic T cell subsets. Both scMAIT and scCD8 T_{EMRA} were almost solely derived from CD8 CTLs (Fig. 3A, 3C). However, the cell compositions of the other seven scCPops were quite complex, with each scCPop deriving from several bead-enriched T cell subsets (Fig. 3C). In particular, all six bead-enriched T cell subsets contributed to scISAG^{hi} T, arranging from 2.54 to 26.92% (Fig. 3C). These results indicated scISAG^{hi} T was a rare cell population accounting for 0.26% of the analyzed cells, with CD4 T cell subsets containing a high fraction of scISAG^{hi} T. scISAG^{hi} T accounts for ~0.4% of CD4 naive, CD4 Th, and CD4 Tregs and 0.24% of CD4 memory cells, while it accounts for only 0.04% of CD8 CTLs and 0.09% of CD8 naive cells (Supplemental Fig. 2F).

The T cells were separated into 54 cell subpopulations by combinations of nine scCPops and six bead-enriched T cell subsets. Finally, 45 cell subpopulations were left after we removed the subpopulations with the number of cells <10. Cell subpopulations deriving from different bead-enriched T cell subsets while belonging to same scCPop always showed similar expression patterns (Fig. 3D). In contrast, cell subpopulations deriving from different scCPops while belonging to same bead-enriched T cell subset usually showed quite different expression patterns (Fig. 3D, Supplemental Fig. 2G). These results are consistent with our assumption that scCPops better represent the cell states than that of bead-enriched T cell subsets.

Fine analyses of CD4 T cell subsets

The CD4 T cell subsets (Supplemental Fig. 3A) were further analyzed to better understand the subpopulations and lineages of CD4 T cells. The 40,696 cells from the CD4 Th, CD4 naive, CD4 memory, and CD4 Treg populations formed a single connected entity on the *t*-SNE plot and different CD4 T cell subsets mixed together (Fig. 4A). We clustered these T cells into six scCPops (Fig. 4B, 4C), among which four scCPops have been identified as mentioned earlier, namely, scCD4 T_N (*CCR7*, *SELL*, *LEF1*), scISAG^{hi} T (*OAS1*, *IFI6*, *MX1*, *BST2*), scCD4 T_{EMRA} (*GZMA*, *CCL5*, *GZMK*, *CST7*, *LYAR*), and scTreg (*FOXP3*, *IL2RA*, *IL10RA*, *CD59*). The original scCD4 T_M cells were separated into central memory T cells (scCD4 T_{CM}) (*CCR7*, *SELL*, *TCF7*) and effector memory T cells (scT_{EM}s) (*CCR7*, *TIMP1*, *LGALS1*, *USP10*), respectively (Fig. 4B, 4C). The Sankey diagram showed that the cellular relationships between the four bead-enriched CD4 T cell subsets and the six scCPops were consistent with the earlier-mentioned analyses of all T cells. In brief, bead-enriched CD4 naive cells almost solely clustered into the counterpart scCD4 T_N. In contrast, CD4 Th, CD4 memory, and CD4 Tregs contributed to all six scCPops (Fig. 4D).

We further clustered the CD4 T cells into 10 scCPops for better understanding of its fine substructure (Fig. 4E). The original scTregs were separated into central memory Tregs (*FOXP3*, *CCR7*, *LEF1*, *CTLA4^{low}*) and effector memory Tregs (*FOXP3*, *IL2RA*, *LGALS1*, *ANXA2*, *CTLA4*) (Supplemental Fig. 3B). The original scCD4 T_N cells were separated into scCD4 T_N#1 and scCD4 T_N#2, with most CD4 naive-specific genes, such as *CCR7*, *SELL*, *LEF1*, and *TCF7*, expressing higher in scCD4 T_N#1 (Supplemental Fig. 3B). The original scCD4 T_{EM} cells were divided into scCD4 T_{EM}#1 and scCD4 T_{EM}#2 (Fig. 4E), with *KLRB1* and *PTGER2* expressing higher in scCD4 T_{EM}#1, potentially indicating scCD4 T_{EM}#1 is an intermediate state between scCD4 T_{EM}#2 and scCD4 T_{EMRA}. The Sankey diagram showed that CD4 Th, CD4 memory, and CD4 Tregs contributed to almost all 10 fine scCPops, indicating the high cell heterogeneity of bead-enriched T cells subsets (Supplemental Fig. 3C).

Application of Slingshot (28) on these CD4 T cells identified four pseudotime lineages: (1) scCD4 T_N → scCD4 T_{CM} → scCD4 T_{EM} → scCD4 T_{EMRA}; (2) scCD4 T_N → scCD4 T_{CM} → scCD4 T_{EM}; (3) scCD4 T_N → scCD4 T_{CM} → scCD4 Treg; and (4) scCD4 T_N → scISAG^{hi} (Fig. 4F). The CD4 T cell differentiation was a continuous process in which naive cells first progressed into central memory–like cells and then developed into different functional CD4 T cells. The first inferred lineage was consistent with recent reports on CD4 T cell differentiation (20, 35, 36), which was the most important lineage because it included most of the CD4 T cells.

Fine analyses of CTLs

We integrated three T cell populations, namely, bead-enriched CD8 CTLs, bead-enriched CD8 naive cells, and scCD4 T_{EMRA}/scCD4-CTLs in scT_M (Supplemental Fig. 3D), for detecting fine substructure in these T cell subsets and providing insight into the

Downloaded from http://journals.aai.org/jimmunol/article-pdf/208/2/396/1504672/ji2100581.pdf by guest on 30 September 2023

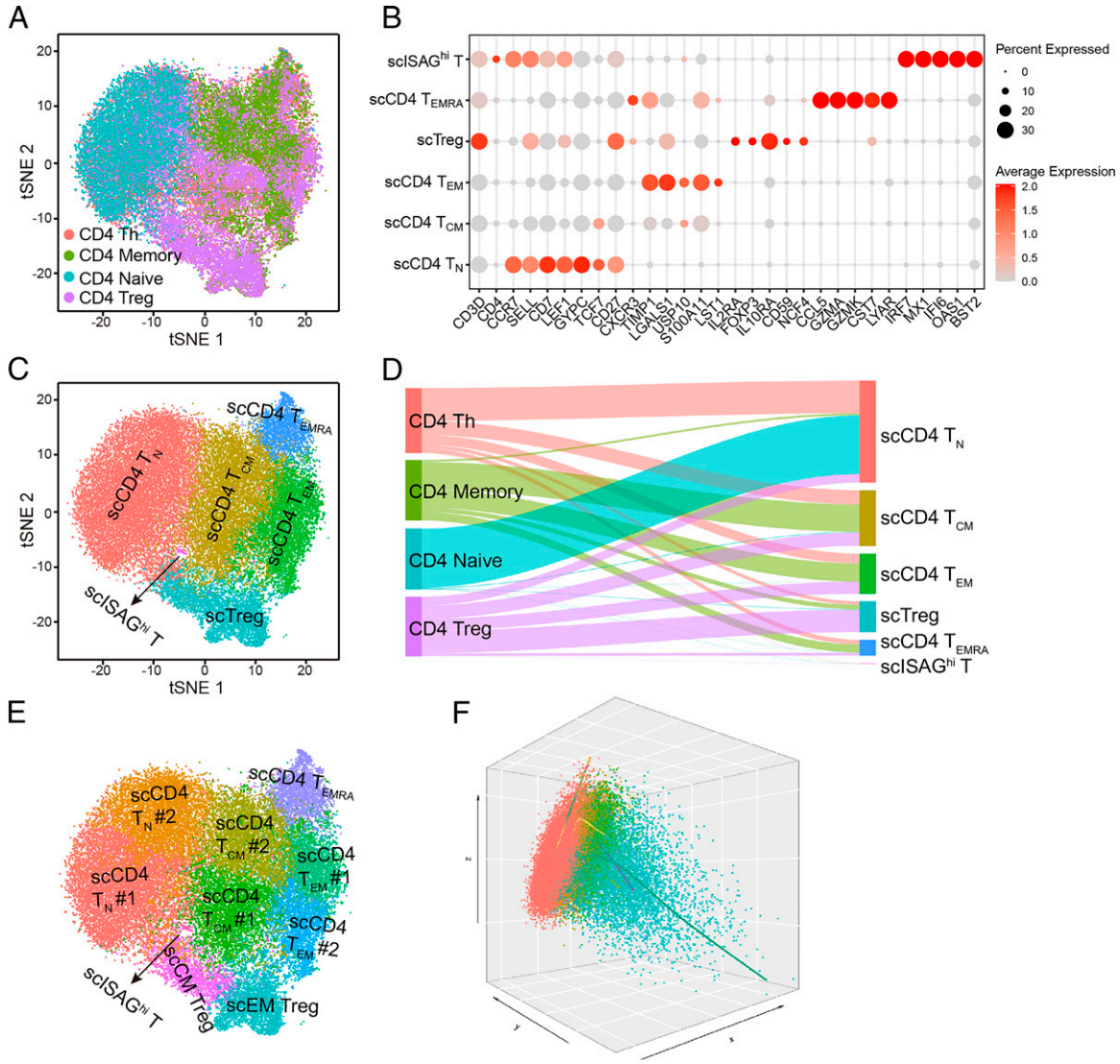

FIGURE 4. Reclassification of CD4 T cells and single-cell trajectory. (**A**) *t*-SNE projection of CD4 T cells, with each cell colored according to bead-enriched T cell subsets that are distinguished by colors. (**B**) Normalized expression level and expression percentage of the cell-type-specific genes in six CD4 T cell scCPops. (**C**) *t*-SNE projection of CD4 T cells, where each cell is grouped into one of the six scCPops, with inferred cell types on the panel. (**D**) Cellular relationships between the four CD4 T cell subsets and the six scCPops by Sankey diagram. Each node represents a cell population, and the width of line between two nodes represents number of shared cells. (**E**) The original 6 scCPops were further separated into 10 scCPops by setting a high resolution. (**F**) Single-cell trajectories of CD4 T cells inferred by Slingshot. The color of each scCPop is the same as that in (C).

relationship between CD4 CTLs and CD8 CTLs. The three T cell populations formed one major group, which mainly consisted of CD8 naive cells and several small groups in which scCD4 T_{EMRA} overlapped with some CD8 CTLs on the *t*-SNE plot (Fig. 5A). We further clustered these T cells into six scCPops (Fig. 5B, 5C), namely, scCD8 T_N (*CCR7*, *SELL*, *LEF*, *CD69*^{low}), scCD8 T_{CM} (*CCR7*, *SELL*, *LEF1*, *CD69*), scT_M (*CXCR3*, *TRADD*, *GZMA*, *GZMK*, *CCL5*), scCD8 T_{EMRA} (*FGFBP2*, *GZMH*, *GZMB*, *GNLY*), scMAIT (*KLRB1*, *ZBTB16*, *NCR3*, *SLC4A10*, *RORC*), and NK cells (scNKT; *NCR3*, *KLRC2*, *ZNF683*, *XCL1*) (Fig. 5B, 5C). The expression pattern and cell composition of scT_M is the same as the aforementioned scT_M in Fig. 2. The expression of commonly used T cell markers on the *t*-SNE plot showed obvious scCPop-specific features (Fig. 5D). A volcano plot showed that both MAIT and NKT expressed many cell-type-specific genes (Supplemental Fig. 3E, 3F), further indicating that MAITs and NKTs are outliers in CD8 CTLs.

Bead-enriched CD8 naive cells were mainly clustered into its counterpart scCD8 T_N, with a small fraction clustered into scCD8 T_{CM} (Fig. 5E). scCD4 T_{EMRA}, together with a fraction of cells from CD8 CTL, formed scT_M, consistent with our earlier results. The cells from CD8 CTLs contributed to each of the six scCPops (Fig. 5E), indicating multiple T cell subsets in commonly used CD8 CTLs. Especially, the existence of an unignorable amount of MAITs and NKTs in CD8 CTLs indicated that the composition of classic CD8 CTL is complex.

We removed scMAIT, scCD4 T_{EMRA}, and scNKT cells for inferring CD8 T cell differentiation lineage because the three cell populations do not belong to well-defined CD8 T cells. By conducting Slingshot (28), we inferred the pseudotime trajectory of CD8 T cell differentiation: scCD8 $T_N \rightarrow$ scCD8 $T_{CM} \rightarrow$ scCD8 $T_{EM} \rightarrow$ scCD8 T_{EMRA} (Fig. 5F), which was similar to the CD4 main lineage. Based on inferred lineage and cell projection on the top PCs, the differentiation from CD8 naive T to scCD8 T_{EMRA} cells was a continuous process (Fig. 5F).

Downloaded from http://journals.aai.org/jimmunol/article-pdf/208/2/396/1504672/ji2100581.pdf by guest on 30 September 2023

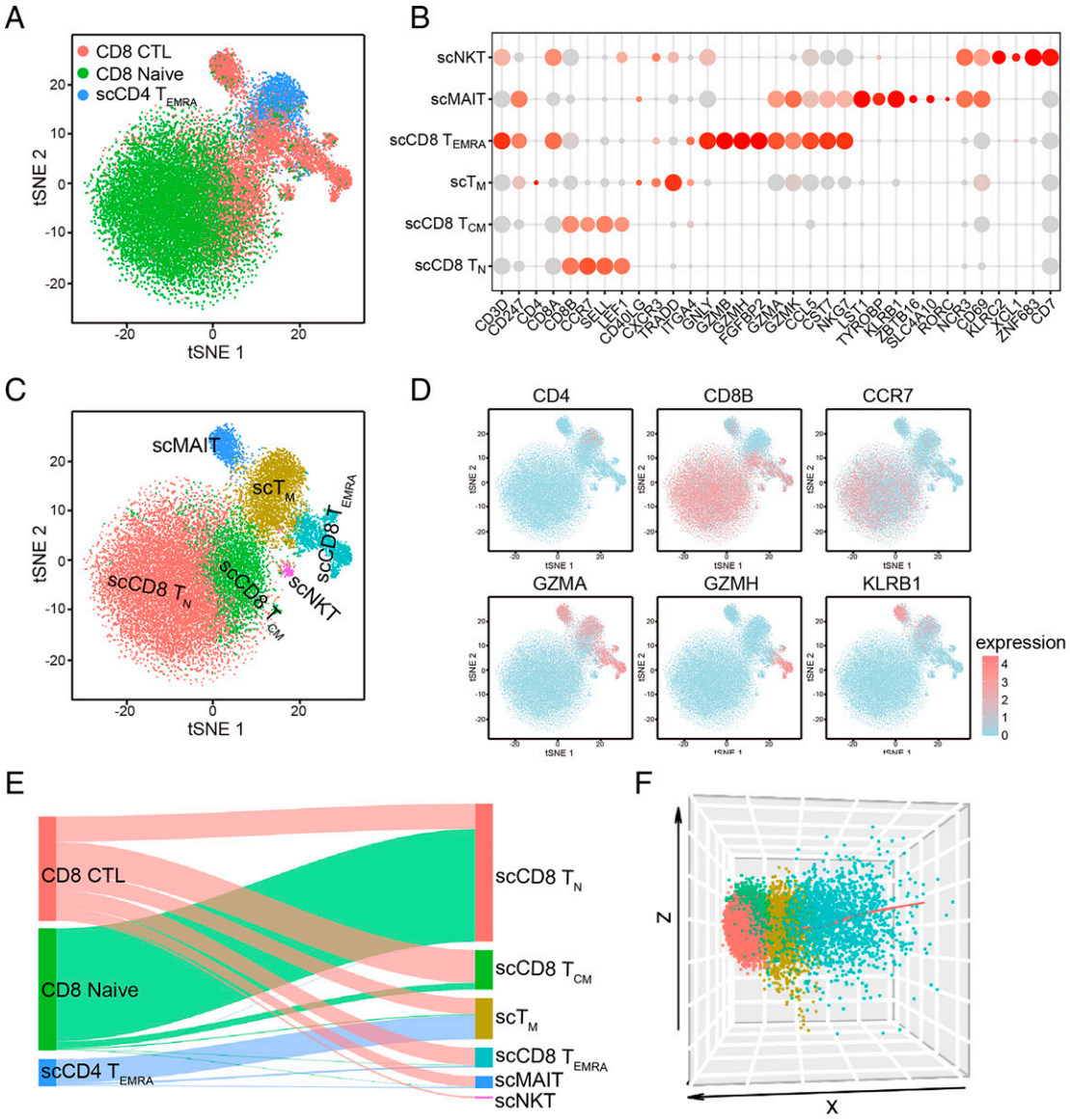

FIGURE 5. Reclassification of potential CTLs based on scRNA-seq and single-cell trajectory. (**A**) *t*-SNE projection of CD8 naive cells, CD8 CTLs, and scCD4 T_{EMRA} cells, with each cell colored according to cell population. (**B**) Normalized expression level and expression percentage of the cell-type-specific genes in the six scCPops. (**C**) *t*-SNE projection of CTLs, where each cell is grouped into one of the six scCPops, with inferred cell types indicated in the panel. (**D**) *t*-SNE projection of CTLs, with each cell colored based on its normalized expression of *CD4*, *CD8B*, *CCR7*, *GZMA*, *GZMH*, and *KLRB1*, respectively. (**E**) Relationships between the initial T cell subsets and the six scCPops by Sankey diagram. Each node represents a cell population, and the width of line between two nodes represents the number of shared cells. (**F**) Single-cell trajectory of CD8 CTL inferred by Slingshot. The color of each scCPops is the same as that in (C).

Enrichment of ISAG^{hi} T cells and identification of their features

Although *BST2* showed weak specificity for only scISAG^{hi} T cells (Fig. 2D), it is the only cell surface gene with a well-established Ab among ISAGs. We enriched scISAG^{hi} T by FACS of BST2^{hi} cells from PBMCs and made two independent scRNA-seq libraries, BST2^{hi} T#1 and BST2^{hi} T#2 (Fig. 6A, Table I, Supplemental Fig. 4A). A total of 13,677 FACS-sorted BST2^{hi} T cells were clustered into ISAG^{hi} T, CD4 T_N, CD8 T_N, CD4 T_M, CD8 T_M, proliferated CD4 T, MAIT, γδ T, NKT, effector Treg, and naive Tregs (Fig. 6B, Supplemental Fig. 4B). Although the enrichment of ISAG^{hi} T cells has significantly increased, ISAG^{hi} T accounts for only a moderate fraction of the BST2^{hi} cell because *BST2* is expressed in many T cell subsets. The expression level of *BST2* in ISAG^{hi} T is significantly higher than that of other cells based on the scRNA-seq data (Fig. 6C). We further separated the non-ISAG^{hi} T cells into BST2^{+} and BST2^{-} T cells based on whether *BST2* was detected in scRNA-seq data (Fig. 6D). The expression profile of ISAG^{hi} T cells is quite different from that of BST2^{+} and BST2^{-} T cells, while the expression profiles of BST2^{+} and BST2^{-} T cells are more similar (Fig. 6E, 6F). ISAG^{hi} T cell has 89 genes with significantly higher expression, including *IF144L*, *MX1*, *IF1616*, *XAF1*, *ISG15*, *OAS3*, *IFI44*, *IRF7*, *MX2*, *IFITM1*, *OASL* and *IFI35*, than *BST2*$^{+}$ T cells (Fig. 6G). These ISAG^{hi} T cell–specific genes were significantly enriched in IFN signaling, defense response to virus, and response to IFN-γ and antiviral mechanism by IFN-stimulated genes (Supplemental Fig. 4C). In summary, high expression of ISAGs is the most significant feature of ISAG^{hi} T cells.

ISAGs shape ISAG^{hi} T cells and contribute to quick T cell activation

We reconducted dimension reduction and clustering analysis on BST2^{hi} T cells using gene set–excluded ISAGs; the scCPops did

Downloaded from http://journals.aai.org/jimmunol/article-pdf/208/2/396/1504672/ji2100581.pdf by guest on 30 September 2023

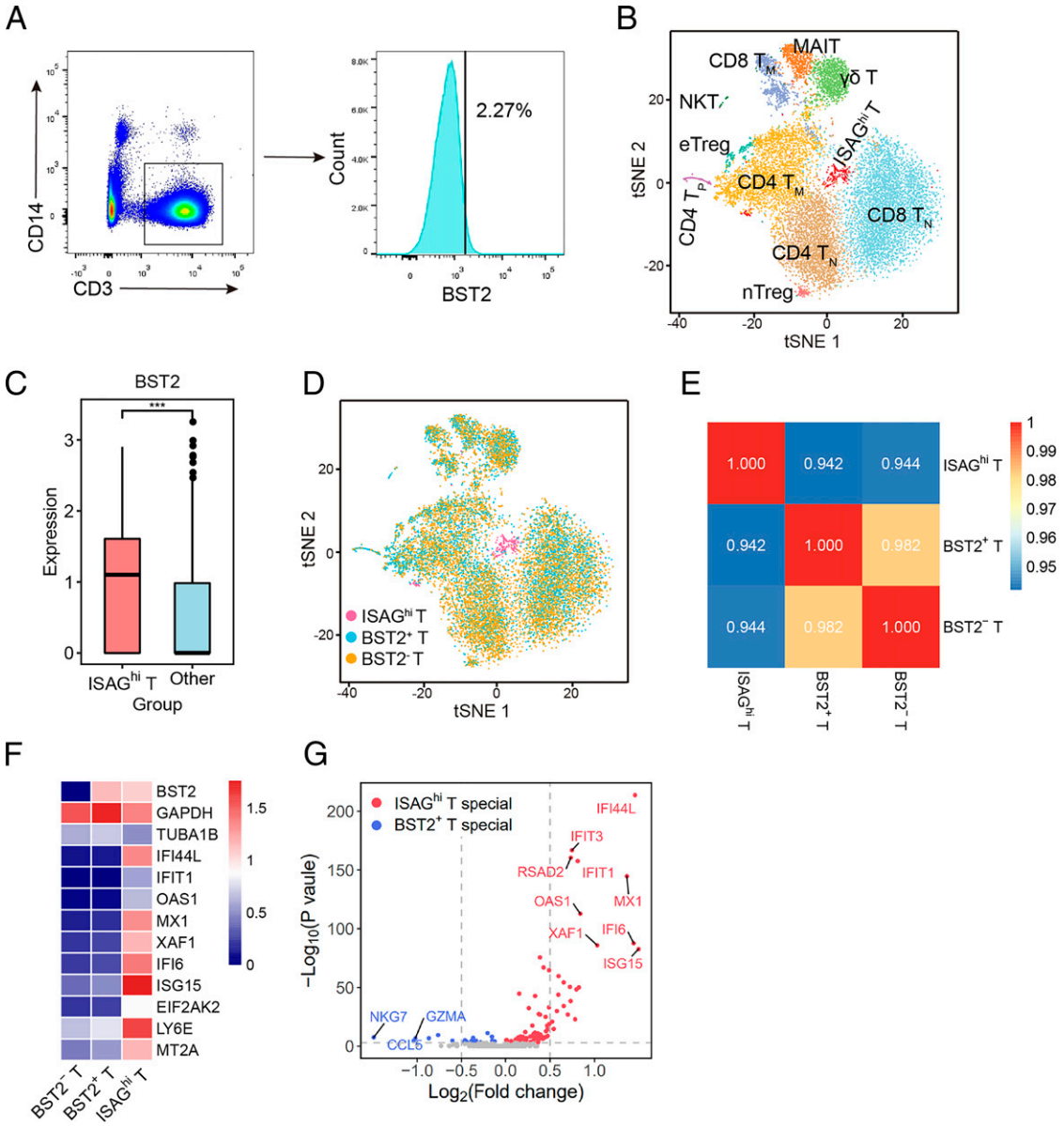

FIGURE 6. Enrichment of ISAG^{hi} T cells by FACS and features of ISAG^{hi} T cells. (**A**) Sorting of BST2^{hi} T cells from PBMCs by FACS. (**B**) *t*-SNE projection of 13,644 sorted BST2^{hi} T cells, with each cell colored according to cell subpopulation. (**C**) Box plot of *BST2* expression of ISAG^{hi} T cells and other cells in BST2^{hi} T cells. Asterisks represent a significant difference (***$p < 0.001$). (**D**) *t*-SNE projection of BST2^{hi} T cells, with each cell colored according to ISAG^{hi} T, BST2^{+} T, and BST2^{-} T cells. BST2^{+} and BST2^{-} T cells are defined by expression of BST2 in scRNA-seq data. (**E**) Spearman correlation coefficient between ISAG^{hi} T, BST2^{+} T, and BST2^{-} T cells. Color represents correlation coefficient between two groups. (**F**) Heatmap of DEGs between ISAG^{hi} T, BST2^{+} T, and BST2^{-} T cells. Color represents average expression of genes in each group. (**G**) Volcano plot for DEGs between ISAG^{hi} T and BST2^{+} T cells ($p < 0.01$). Red points represent ISAG^{hi} T cell–specific high expressed genes, while blue points represent BST2^{+} T cell–specific high expressed genes. CD4 Tp, proliferated CD4 T; eTreg, effector Treg; nTreg, naive Treg.

not change, except the ISAG^{hi} T cluster disappeared (Fig. 7A). On the *t*-SNE plot based on gene set without ISAGs, ISAG^{hi} T cells dispersed into other scCPops, in particular concentrated on CD8 T_N (39.41%), CD4 T_N (28.39%), and CD4 T_M (28.39%) cells, although only a few ISAG^{hi} T cells dispersed into innate-like T cells, such as MAIT, γδ T, and NKT cells (Fig. 7B, Supplemental Fig. 4D). These results are expected because the expression pattern of ISAG^{hi} T cells is very similar to CD4 T_N, CD8 T_N, and CD4 T_M cells (Supplemental Fig. 4B). After normalizing the cell number, the percentage of ISAG^{hi} T cells in CD8 T_N, CD4 T_N, and CD4 T_M subsets is similar (Supplemental Fig. 4E). These results are consistent with our earlier-mentioned observations that bead-enriched CD4 T cell subsets have a much higher fraction of cells that were clustered into ISAG^{hi} T cells (Supplemental Fig. 2F). Interestingly, the ISAG^{hi} T cell cluster appeared when using only ISAGs to do the same analysis (Fig. 7C), further indicating that ISAGs are a major contributor for the formation of ISAG^{hi} T cell cluster.

Constitutive expression of ISAGs plays pivotal roles in host responses to malignant cells and pathogens, such as virus (37–39). We used FACS to sort the T cells into BST2^{+} and BST2^{-} T cells because *BST2* level indicates activation level of ISAGs in some way. The sorted BST2^{+} and BST2^{-} T cells were activated by Dynabeads human T activator CD3/CD28 for 12 h. FACS analysis of CD25 showed almost all BST2^{+} T cells become activated upon anti-CD3/CD28 costimulation, whereas only a fraction of BST2^{-} T cells become activated (Fig. 7D, 7E, Supplemental Fig. 4F, 4G).

Downloaded from http://journals.aai.org/jimmunol/article-pdf/208/2/396/1504672/ji2100581.pdf by guest on 30 September 2023

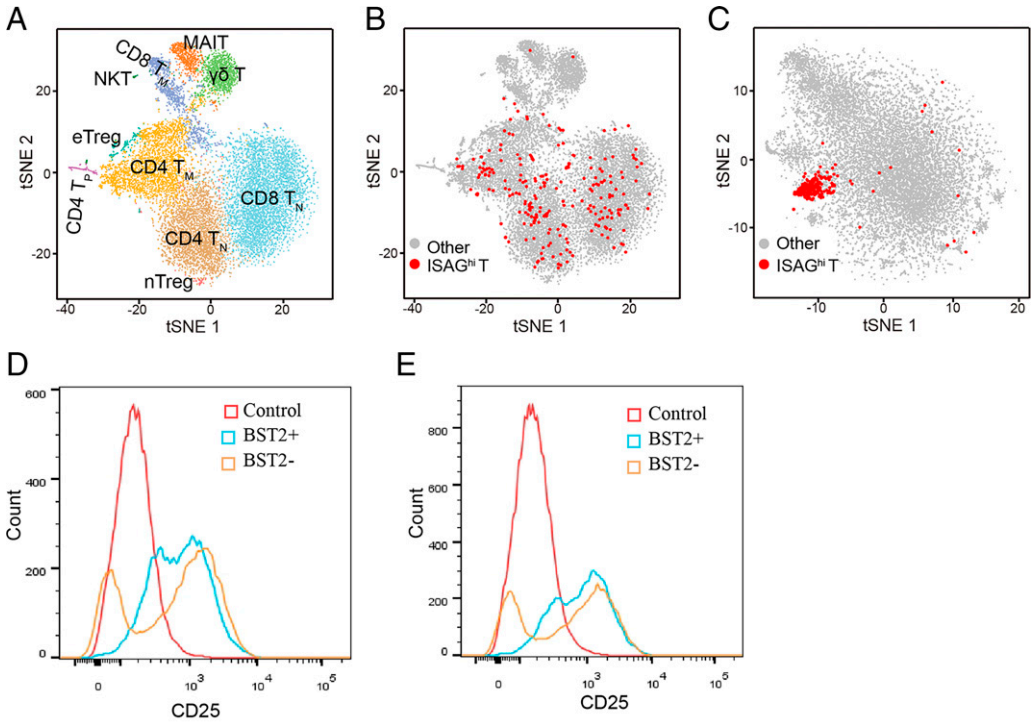

FIGURE 7. ISAGs are the major contributor of ISAG^{hi} T cell cluster, and a high level of ISAGs may contribute to quick T cell activation. (**A**) *t*-SNE projection of BST2^{hi} T cells using gene sets that removed ISAGs, in which ISAG^{hi} T cell cluster disappears. (**B**) *t*-SNE projection of BST2^{hi} T cells using gene sets that removed ISAGs, with red points highlighting the original identified ISAG^{hi} T cells. (**C**) *t*-SNE projection of BST2^{hi} T cells using ISAGs, with red points highlighting original identified ISAG^{hi} T cells. (**D** and **E**) FACS analysis of CD25 showed almost all BST2^{+} T cells become activated upon anti-CD3/CD28 beads for 12 h, while only a fraction of BST2^{-} T cells become activated at the same time (two experiment repeats).

These results may indicate that almost all BST2^{+} T cells have been well prepared for T cell activation, whereas only a fraction of BST2^{-} T cells have been prepared for quick T cell activation.

Discussion

It is assumed that each T cell subset has specific functions, and several T cell subsets may coordinate with each other to complete a specific function (40, 41). The recent advance of scRNA-seq provided an alternative approach to identify the T cell subpopulations and cell heterogeneity. In this study, we reclustered the six bead-enriched T cell subsets using scRNA-seq data. Among the six T cell subsets, naive T cells and memory T cells account for the majority of the T cells in this study (Figs. 2B, 2C, 3A), and the signals from memory T cells and naive T cells were strong. Therefore, we mainly focused on memory T cells and naive T cells, while ignoring other T cell subsets, such as Th1, Th2, and Th17.

We inferred the cell type of each scCPop and its counterpart in bead-enriched T cell subsets using scCPop-specific genes. The naive T subsets were mainly clustered into their scCPop counterparts, while other classic T subsets were dispersed into multiple scCPops. Correlation analyses and DEG analyses showed that the similarities between naive T cell subsets and their scCPop counterparts are much greater than that of other T cell subsets, indicating the naive T cell subsets are relatively pure compared with other bead-enriched T cell subsets. Because the expression of population-specific genes in scCPops are stronger and more homogeneous than that of their bead-enriched T cell counterpart, the statistical parameters of scCPops should be good representatives of features of the T cell population. Furthermore, we increased clustering resolution and identified very specific T cell subpopulations, such as scCD4 T_{CM}, scCD4 T_{EM}, scCD4 T_{EMRA}, central memory Tregs, effector memory Tregs, scCD8 T_{CM}, scCD8 T_{EM}, and scCD8 T_{EMRA}. These results showed that scRNA-seq clustering is essentially consistent with the classic FACS/bead-enriched approach. In addition, scRNA-seq clustering identified rare cell subpopulations, such as ISAG^{hi} T, and inferred very specific subpopulations, indicating its advantage compared with the classic approach. Further analyses revealed that ISAG^{hi} T showed specific features and may contribute to the quick T cell activation.

Although CD4 and CD8 T cells essentially are distinguishable based on cell surface markers, the boundaries between CD4 and CD8 T cells were blurred based on scRNA-seq data (Fig. 1B), indicating the overall transcriptional profiles between CD4 and CD8 T cells were not completely different. In particular, we identified the scT_M cluster that contains both CD4 and CD8 T cells. Furthermore, we inferred the trajectories of CD4 and CD8 T cells, which showed that CD4 naive T cells and CD8 naive T cells differentiated into CD4 T_{EMRA} and CD8 T_{EMRA}, respectively. Reinvestigation of T cell subsets by scRNA-seq is of great significance for understanding T cell subsets and helps us to further understand the mechanism of the human adaptive immune response, which also has important implications for the ongoing immunotherapy.

Acknowledgments

We thank Dr. Keji Zhao for critical reading of the manuscript. We thank members of the Flow Cytometry facility in SUSTech for the cell sorting. We thank members of the Center for Computational Science and Engineering in SUSTech for computational support.

Disclosures

The authors have no financial conflicts of interest.

References

1. Zhang, L., X. Yu, L. Zheng, Y. Zhang, Y. Li, Q. Fang, R. Gao, B. Kang, Q. Zhang, J. Y. Huang, et al. 2018. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* 564: 268–272.
2. Farh, K. K., A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shoresh, H. Whitton, R. J. Ryan, A. A. Shishkin, et al. 2015. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518: 337–343.
3. Farber, D. L., N. A. Yudanin, and N. P. Restifo. 2014. Human memory T cells: generation, compartmentalization and homeostasis. *Nat. Rev. Immunol.* 14: 24–35.

Downloaded from http://journals.aai.org/jimmunol/article-pdf/208/2/396/1504672/ji2100581.pdf by guest on 30 September 2023

4. Hu, G., K. Cui, D. Fang, S. Hirose, X. Wang, D. Wangsa, W. Jin, T. Ried, P. Liu, J. Zhu, et al. 2018. Transformation of accessible chromatin and 3D nucleome underlies lineage commitment of early T cells. *Immunity* 48: 227–242.e8.
5. Qin, P., Y. Pang, W. Hou, R. Fu, Y. Zhang, X. Wang, G. Meng, Q. Liu, X. Zhu, N. Hong, et al. 2021. Integrated decoding hematopoiesis and leukemogenesis using single-cell sequencing and its medical implication. *Cell Discov.* 7: 2.
6. Zhu, J., and W. E. Paul. 2008. CD4 T cells: fates, functions, and faults. *Blood* 112: 1557–1569.
7. Mosmann, T. R., H. Cherwinski, M. W. Bond, M. A. Giedlin, and R. L. Coffman. 1986. Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *J. Immunol.* 136: 2348–2357.
8. Sakaguchi, S., N. Sakaguchi, M. Asano, M. Itoh, and M. Toda. 1995. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J. Immunol.* 155: 1151–1164.
9. Fontenot, J. D., M. A. Gavin, and A. Y. Rudensky. 2003. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat. Immunol.* 4: 330–336.
10. Suni, M. A., S. A. Ghanekar, D. W. Houck, H. T. Maecker, S. B. Wormsley, L. J. Picker, R. B. Moss, and V. C. Maino. 2001. CD4(+)CD8(dim) T lymphocytes exhibit enhanced cytokine expression, proliferation and cytotoxic activity in response to HCMV and HIV-1 antigens. *Eur. J. Immunol.* 31: 2512–2520.
11. Intlekofer, A. M., N. Takemoto, E. J. Wherry, S. A. Longworth, J. T. Northrup, V. R. Palanivel, A. C. Mullen, C. R. Gasink, S. M. Kaech, J. D. Miller, et al. 2005. Effector and memory CD8+ T cell fate coupled by T-bet and eomesodermin. [Published erratum appears in 2006 *Nat. Immunol.* 7: 113.] *Nat. Immunol.* 6: 1236–1244.
12. Wang, W., G. Ren, N. Hong, and W. Jin. 2019. Exploring the changing landscape of cell-to-cell variation after CTCF knockdown via single cell RNA-seq. *BMC Genomics* 20: 1015.
13. Macosko, E. Z., A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, et al. 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161: 1202–1214.
14. Paul, F., Y. Arkin, A. Giladi, D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, D. Winter, D. Lara-Astiaso, M. Gury, A. Weiner, et al. 2015. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. [Published erratum appears in 2016 *Cell.* 164: 325.] *Cell* 163: 1663–1677.
15. Zhou, B., and W. Jin. 2020. Visualization of single cell RNA-seq data using t-SNE in R. *Methods Mol. Biol.* 2117: 159–167.
16. Yu, Y., J. C. Tsang, C. Wang, S. Clare, J. Wang, X. Chen, C. Brandt, L. Kane, L. S. Campos, L. Lu, et al. 2016. Single-cell RNA-seq identifies a PD-1^{hi} ILC progenitor and defines its development pathway. *Nature* 539: 102–106.
17. Proserpio, V., A. Piccolo, L. Haim-Vilmovsky, G. Kar, T. Lönnberg, V. Svensson, J. Pramanik, K. N. Natarajan, W. Zhai, X. Zhang, et al. 2016. Single-cell analysis of CD4+ T-cell differentiation reveals three major cell states and progressive acceleration of proliferation. [Published erratum appears in 2016 *Genome Biol.* 17: 133.] *Genome Biol.* 17: 103.
18. Waickman, A. T., K. Victor, T. Li, K. Hatch, W. Rutvisuttinunt, C. Medin, B. Gabriel, R. G. Jarman, H. Friberg, and J. R. Currier. 2019. Dissecting the heterogeneity of DENV vaccine-elicited cellular immunity using single-cell RNA sequencing and metabolic profiling. *Nat. Commun.* 10: 3666.
19. Zemmour, D., R. Zilionis, E. Kiner, A. M. Klein, D. Mathis, and C. Benoist. 2018. Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. [Published erratum appears in 2018 *Nat. Immunol.* 19: 645.] *Nat. Immunol.* 19: 291–301.
20. Patil, V. S., A. Madrigal, B. J. Schmiedel, J. Clarke, P. O'Rourke, A. D. de Silva, E. Harris, B. Peters, G. Seumois, D. Weiskopf, et al. 2018. Precursors of human $CD4^+$ cytotoxic T lymphocytes identified by single-cell transcriptome analysis. *Sci. Immunol.* 3: eaan8664.
21. Chattopadhyay, P. K., T. M. Gierahn, M. Roederer, and J. C. Love. 2014. Single-cell technologies for monitoring immune systems. *Nat. Immunol.* 15: 128–135.
22. Hao, Y., S. Hao, E. Andersen-Nissen, W. M. Mauck III, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, et al. 2021. Integrated analysis of multimodal single-cell data. *Cell* 184: 3573–3587.e29.
23. Zheng, G. X., J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8: 14049.
24. Butler, A., P. Hoffman, P. Smibert, E. Papalexi, and R. Satija. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36: 411–420.
25. van der Maaten, L., and G. Hinton. 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9: 2579–2605.
26. Mahfouz, A., M. van de Giessen, L. van der Maaten, S. Huisman, M. Reinders, M. J. Hawrylycz, and B. P. Lelieveldt. 2015. Visualizing the spatial gene expression organization in the brain through non-linear similarity embeddings. *Methods* 73: 79–89.
27. Huang, D.W., B. T. Sherman, and R. A. Lempicki. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44–57.
28. Street, K., D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, and S. Dudoit. 2018. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 19: 477.
29. R Core Team. 2018. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.
30. Stuart, T., A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, 3rd, Y. Hao, M. Stoeckius, P. Smibert, and R. Satija. 2019. Comprehensive integration of single-cell data. *Cell* 177: 1888–1902.e21.
31. Van Grembergen, O., M. Bizet, E. J. de Bony, E. Calonne, P. Putmans, S. Brohée, C. Olsen, M. Guo, G. Bontempi, C. Sotiriou, et al. 2016. Portraying breast cancers with long noncoding RNAs. *Sci. Adv.* 2: e1600220.
32. Zhou, J., X. Zhi, L. Wang, W. Wang, Z. Li, J. Tang, J. Wang, Q. Zhang, and Z. Xu. 2015. Linc00152 promotes proliferation in gastric cancer through the EGFR-dependent pathway. [Published erratum appears in 2016 *J. Exp. Clin. Cancer Res.* 35: 30.] *J. Exp. Clin. Cancer Res.* 34: 135.
33. Hayakawa, T., K. Yamashita, K. Tanzawa, E. Uchijima, and K. Iwata. 1992. Growth-promoting activity of tissue inhibitor of metalloproteinases-1 (TIMP-1) for a wide range of cells. A possible new growth factor in serum. *FEBS Lett.* 298: 29–32.
34. Rogers, P. R., J. Song, I. Gramaglia, N. Killeen, and M. Croft. 2001. OX40 promotes Bcl-xL and Bcl-2 expression and is essential for long-term survival of CD4 T cells. *Immunity* 15: 445–455.
35. Hashimoto, K., T. Kouno, T. Ikawa, N. Hayatsu, Y. Miyajima, H. Yabukami, T. Terooatea, T. Sasaki, T. Suzuki, M. Valentine, et al. 2019. Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. *Proc. Natl. Acad. Sci. USA* 116: 24242–24251.
36. Cano-Gamez, E., B. Soskic, T. I. Roumeliotis, E. So, D. J. Smyth, M. Baldrighi, D. Willé, N. Nakic, J. Esparza-Gordillo, C. G. C. Larminie, et al. 2020. Single-cell transcriptomics identifies an effectorness gradient shaping the response of $CD4^+$ T cells to cytokines. *Nat. Commun.* 11: 1801.
37. Urata, S., E. Kenyon, D. Nayak, B. Cubitt, Y. Kurosaki, J. Yasuda, J. C. de la Torre, and D. B. McGavern. 2018. BST-2 controls T cell proliferation and exhaustion by shaping the early distribution of a persistent viral infection. *PLoS Pathog.* 14: e1007172.
38. Oriol-Tordera, B., M. Berdasco, A. Llano, B. Mothe, C. Gálvez, J. Martinez-Picado, J. Carrillo, J. Blanco, C. Duran-Castells, C. Ganoza, et al. 2020. Methylation regulation of Antiviral host factors, Interferon Stimulated Genes (ISGs) and T-cell responses associated with natural HIV control. *PLoS Pathog.* 16: e1008678.
39. Kisand, K., M. Link, A. S. Wolff, A. Meager, L. Tserel, T. Org, A. Murumägi, R. Uibo, N. Willcox, K. Trebusak Podkrajsek, et al. 2008. Interferon autoantibodies associated with AIRE deficiency decrease the expression of IFN-stimulated genes. *Blood* 112: 2657–2666.
40. Jameson, S. C., and D. Masopust. 2018. Understanding subset diversity in T cell memory. *Immunity* 48: 214–226.
41. Hoyer, S., S. Prommersberger, I. A. Pfeiffer, B. Schuler-Thurner, G. Schuler, J. Dörrie, and N. Schaft. 2014. Concurrent interaction of DCs with CD4(+) and CD8(+) T cells improves secondary CTL expansion: It takes three to tango. *Eur. J. Immunol.* 44: 3543–3559.

Downloaded from http://journals.aai.org/jimmunol/article-pdf/208/2/396/1504672/ji2100581.pdf by guest on 30 September 2023