ARTICLES
https://doi.org/10.1038/s41590-020-0791-5

nature immunology

Check for updates

Two subsets of stem-like CD8+ memory T cell progenitors with distinct fate commitments in humans

Giovanni Galletti[1,21], Gabriele De Simone[1,21], Emilia M. C. Mazza[1,21], Simone Puccio[1], Claudia Mezzanotte[2], Timothy M. Bi[3], Alexey N. Davydov[4], Maria Metsger[4], Eloise Scamardella[1], Giorgia Alvisi[1], Federica De Paoli[1], Veronica Zanon[1], Alice Scarpa[1], Barbara Camisa[2], Federico S. Colombo[5], Achille Anselmo[5], Clelia Peano[6,7], Sara Polletti[8], Domenico Mavilio[9,10], Luca Gattinoni[11,12,13], Shannon K. Boi[14], Benjamin A. Youngblood[14], Rhiannon E. Jones[15], Duncan M. Baird[15], Emma Gostick[16], Sian Llewellyn-Lacey[16], Kristin Ladell[16], David A. Price[16,17], Dmitriy M. Chudakov[18,19,20], Evan W. Newell[3], Monica Casucci[2] and Enrico Lugli[1,5]✉

T cell memory relies on the generation of antigen-specific progenitors with stem-like properties. However, the identity of these progenitors has remained unclear, precluding a full understanding of the differentiation trajectories that underpin the heterogeneity of antigen-experienced T cells. We used a systematic approach guided by single-cell RNA-sequencing data to map the organizational structure of the human CD8+ memory T cell pool under physiological conditions. We identified two previously unrecognized subsets of clonally, epigenetically, functionally, phenotypically and transcriptionally distinct stem-like CD8+ memory T cells. Progenitors lacking the inhibitory receptors programmed death-1 (PD-1) and T cell immunoreceptor with Ig and ITIM domains (TIGIT) were committed to a functional lineage, whereas progenitors expressing PD-1 and TIGIT were committed to a dysfunctional, exhausted-like lineage. Collectively, these data reveal the existence of parallel differentiation programs in the human CD8+ memory T cell pool, with potentially broad implications for the development of immunotherapies and vaccines.

Antigen recognition by CD8+ naive T cells initiates a program of clonal expansion and effector differentiation that leads to the clearance of infected or malignant cells and the subsequent formation of heterogeneous memory populations that confer durable immunity[1]. These memory populations are thought to be organized in a developmental hierarchy, according to which stem-cell memory T (T_{SCM}) cells self-renew and generate long-lived central memory T (T_{CM}) cells and short-lived effector memory T (T_{EM}) cells[2–6]. However, the mechanisms that underlie the enhanced multipotency of T_{SCM} cells relative to T_{CM} cells have not been clearly defined in molecular terms[7].

Memory T cell differentiation can become corrupted under conditions of persistent antigenic stimulation, as has been observed during chronic viral infections and progressive malignancies, which promote a state of T cell exhaustion that is characterized by an orderly loss of effector functions, impaired proliferation and the upregulation of inhibitory receptors[8]. This dynamic process occurs over a period of weeks after the initial priming event[9,10] and involves the genome-wide accumulation of epigenetic modifications[11,12]. Recent studies have shown that exhausted T (T_{EX}) cell populations are developmentally and functionally heterogeneous, incorporating stem-like progenitors that express T cell factor 1 (TCF1), which give rise to highly differentiated T_{EX} cells that are constitutively dysfunctional and lack TCF1 (refs. [13–16]). Importantly, the therapeutic benefits of immune checkpoint blockade in the context of chronic viral infections and various cancers are thought to operate via these TCF1+ progenitors, which appear to be susceptible to interventions that specifically target the inhibitory receptor programmed death-1 (PD-1)[13,15,17–20].

Current evidence therefore suggests that exhausted and functional memory T cells arise from separate populations of stem-like progenitors committed to distinct fates. However, the precise nature

[1]Laboratory of Translational Immunology, Humanitas Clinical and Research Center – IRCCS, Rozzano, Milan, Italy. [2]Innovative Immunotherapies Unit, Division of Immunology, Transplantation, and Infectious Diseases, IRCCS San Raffaele Scientific Institute, Milan, Italy. [3]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [4]Central European Institute of Technology, Brno, Czech Republic. [5]Humanitas Flow Cytometry Core, Humanitas Clinical and Research Center – IRCCS, Rozzano, Milan, Italy. [6]Institute of Genetic and Biomedical Research, UoS Milan, National Research Council, Rozzano, Milan, Italy. [7]Genomic Unit, Humanitas Clinical and Research Center – IRCCS, Rozzano, Milan, Italy. [8]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy. [9]Unit of Clinical and Experimental Immunology, Humanitas Clinical and Research Center – IRCCS, Rozzano, Milan, Italy. [10]Department of Medical Biotechnologies and Translational Medicine, University of Milan, Milan, Italy. [11]Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA. [12]Regensburg Center for Interventional Immunology, Regensburg, Germany. [13]University of Regensburg, Regensburg, Germany. [14]St. Jude Children's Research Hospital, Memphis, TN, USA. [15]Division of Cancer and Genetics, Cardiff University School of Medicine, Cardiff, UK. [16]Division of Infection and Immunity, Cardiff University School of Medicine, Cardiff, UK. [17]Systems Immunity Research Institute, Cardiff University School of Medicine, Cardiff, UK. [18]Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia. [19]Pirogov Russian National Research Medical University, Moscow, Russia. [20]Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia. [21]These authors contributed equally: Giovanni Galletti, Gabriele De Simone, Emilia M. C. Mazza. ✉e-mail: enrico.lugli@humanitasresearch.it

of these stem-like progenitors, which shape the adaptive immune response and influence the outcome of many globally relevant diseases, has remained obscure. In this study, we used a comprehensive and unbiased approach to map the origins of dysfunctional and functional human CD8$^+$ memory T cells. Our data identified two distinct subsets of CCR7$^+$ progenitors in healthy individuals, distinguished by the expression of PD-1 and TIGIT. Progenitors committed to the generation of dysfunctional, exhausted-like progeny expressed both of these inhibitory receptors, whereas progenitors committed to the generation of a more functional progeny lacked both of these inhibitory receptors. Differential inclusion of the transcriptionally distinct PD-1$^+$TIGIT$^+$ subset also explained most of the differences between T_{SCM} and T_{CM} cells, providing a clearer view of human CD8$^+$ memory T cell differentiation.

Results

Two subsets of stem-like CD8$^+$ memory T cell progenitors exist in humans. We initially used single-cell RNA sequencing (scRNA-seq; 10x Genomics Platform) to characterize the full spectrum of human CD8$^+$ memory T cells in peripheral blood (PB) samples obtained from healthy donors ($n=4$) (Supplementary Table 1). A total of 31,640 cells were isolated for this purpose via fluorescence-activated cell sorting (FACS) based on the expression of CD95 (Fig. 1a and Supplementary Fig. 1a), which identifies a vast majority of all memory T cells in humans[4]. Bioinformatic analysis of gene expression mapped in two dimensions via uniform manifold approximation and projection (UMAP)[21] identified 14 distinct clusters (denoted individually as C) (Fig. 1b). C1, C4, C5 and C7 were uniformly distant from the other cell populations and expressed high levels of *KLRB1*, which encodes CD161, and *IL7R*, which encodes the interleukin (IL)-7 receptor (IL-7R), also known as CD127. These clusters were therefore derived from mucosal associated invariant T (MAIT) cells[22]. An intermediate cluster, C9, which comprised less than 2% of all cells (Fig. 1c), overexpressed *TRDC* and *TRGC1*, which encode the constant regions of the T cell receptor (TCR) δ and γ chains, respectively, suggesting that γδ, rather than αβ, TCRs were expressed.

scRNA-seq further identified seven different clusters related to conventional memory T cells (Fig. 1b), the most abundant of which were C0, C2, C3, C6 and C10 (Fig. 1c). C2, C6 and C10 expressed genes associated with early differentiated memory T cells, including *CCR7*, *LEF1*, *SELL*, which encodes L-selectin (CD62L), and *FOXP1* (Fig. 1b,d and Supplementary Table 2). In contrast, the highly abundant C0 overexpressed multiple effector transcripts, including *GZMK* and *GZMM*, which encode serine proteases termed granzymes, *IFNG*, which encodes interferon (IFN)-γ, the chemokine (C-C motif) ligands *CCL4* and *CCL5*, and genes encoding human leukocyte antigen (HLA) class II molecules, consistent with the identification of T_{EM} cells (Fig. 1b,d and Supplementary Table 2). C3 displayed a gene-expression profile reminiscent of terminal effector T (T_{TE}) cells, featuring high levels of *GZMB*, *GNLY*, *NKG7*, *ZEB2* and *GZMA* (Fig. 1b,d and Supplementary Table 2). A lack of signature transcripts precluded the identification of C8 on the basis of current knowledge of the T cell differentiation pathway (Supplementary Table 2).

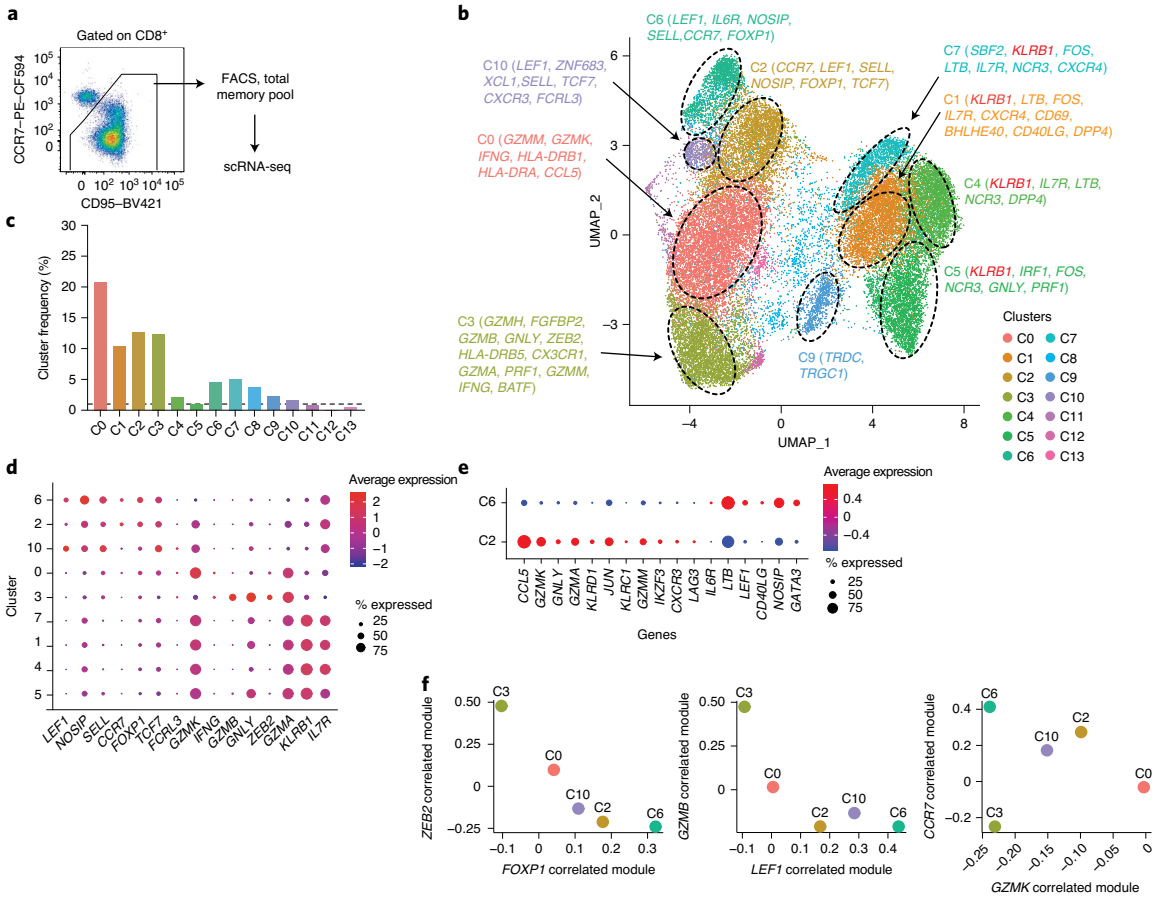
To explore the heterogeneity of the *CCR7*$^+$ memory T cell pool, we focused on C2 and C6, because C10 comprised only ~1% of all sorted cells (Fig. 1c). We identified 160 differentially expressed genes (DEGs) between C2 and C6 (Supplementary Table 3). C2 expressed higher levels of effector molecules, including *CCL5*, *GZMK*, *GNLY*, *GZMA*, *JUN*, *GZMM*, *HOPX*, *IKZF3*, *RUNX3* and *PRF1*, which encodes perforin, whereas C6 expressed higher levels of *IL6R*, *LTB*, *LEF1*, *NOSIP*, *GATA3* and *SELL* (Fig. 1e and Supplementary Table 3). We then used anchor genes selected from the most prominent DEGs to compute transcriptional modules associated with memory differentiation (correlated with *CCR7* and *LEF1*), quiescence (correlated with *FOXP1*), or cytotoxicity and terminal-effector differentiation (correlated with *GZMK* and *GZMB* and with *ZEB2*, respectively) among the five conventional memory T cell clusters (C0, C2, C3, C6 and C10). Using this approach, we found that C2 and C10 were similar, exhibiting intermediate memory and effector scores, whereas C6 was skewed toward a high memory score and C0 and C3 were skewed toward high effector scores (Fig. 1f). scRNA-seq analysis therefore identified four major subsets of conventional CD8$^+$ memory T cells, namely early differentiated *CCR7*$^+$*GZMK*$^-$ (C6) and *CCR7*$^+$*GZMK*$^+$ (C2), T_{EM}-like *CCR7*$^-$*GZMK*^{hi} (C0) and T_{TE}-like *GZMB*$^+$ (C3).

To confirm these findings at the protein level, we designed a high-dimensional flow-cytometry panel based on the cluster signature markers CCR7, LEF1, CD161, GZMB and GZMK (Fig. 1g). This panel was also equipped to detect memory and effector differentiation markers (CD27, CD28, CD45RO, CD127 and T-bet), activation markers (CD38 and HLA-DR), inhibitory receptors (PD-1 and TIGIT) and markers of tissue residency (CD69 and CD103) in CD8$^+$ T cells isolated from the PB and tissues (Supplementary Table 1). In line with the scRNA-seq data, UMAP analysis revealed that CD161^{hi} MAIT cells were largely distinct from other CD95$^+$ memory T cells (Fig. 1h). Among the effector subsets, CCR7$^-$ GZMK^{hi} cells expressed cytolytic molecules and generally lacked the memory markers LEF1, CD27, CD28 and CD127, whereas GZMB$^+$ cells also expressed granulysin (GNLY) and relatively high levels of T-bet (Fig. 1h). In contrast, the CCR7$^+$GMZK$^-$ subset expressed relatively high levels of LEF1, CD27, CD28 and CD127 and lacked effector molecules, activation markers and inhibitory receptors, whereas the CCR7$^+$GMZK$^+$ subset expressed intermediate levels of LEF1 alongside PD-1 and TIGIT, which were not detected in the scRNA-seq analysis, together with relatively high levels of CD27, CD28 and CD127 (Fig. 1h). These subsets displayed variable expression of CD45RO (Fig. 1h). As expected, CD69$^+$CD103$^+$ cells were detected only in tissues (Fig. 1h). A survey of different tissue sites revealed that CCR7$^+$GZMK$^-$PD-1$^-$TIGIT$^-$ and CCR7$^+$GZMK$^+$PD-1$^+$TIGIT$^+$ cells were relatively abundant in PB, lymph nodes and bone marrow, whereas CCR7$^-$GZMK^{hi} cells were ubiquitous, and GZMB$^+$ cells predominated in PB and lung tissue (Fig. 1i).

Collectively, these data identified CCR7$^+$GZMK$^-$ and CCR7$^+$GMZK$^+$ cells as distinct entities in the early differentiated CD8$^+$ memory T cell pool and further showed that these subsets could be distinguished by the expression of PD-1 and TIGIT.

Fig. 1 | Heterogeneity of the human CD8$^+$ memory T cell pool. a, Strategy for the isolation of CD8$^+$ memory T cells from PB via FACS. **b**, UMAP plot showing the distribution of 31,640 cells ($n=4$ donors). Cluster labels indicate selected DEGs. **c**, Histogram plot showing the median frequency of each cluster obtained in **b**. The dashed line is set at 1%. **d**, Balloon plot showing the average expression levels and expression frequencies of selected genes in each cluster obtained in **b**. **e**, Balloon plot showing the average expression levels and expression frequencies of selected genes in C2 versus C6. **f**, Bivariate plots depicting transcriptional module scores correlated with specific genes for selected clusters obtained in **b**. **g**, Flow-cytometric gating strategy for the identification of CD8$^+$ memory T cell subsets. Representative data are shown from PB. **h**, UMAP plot showing the expression of selected markers among CD8$^+$ memory T cells isolated from different tissues ($n=6$ donors per tissue with matched PB samples). Top left, overlays of the cell populations identified in **g**. **i**, Dot plot showing the tissue-specific frequencies of each subset identified in **g**. LN, lymph node; BM, bone marrow. Each dot represents 1 donor ($n=6$ per tissue with matched PB samples). Bars indicate mean ± s.e.m. Statistics were calculated for only the GZMK$^-$ and GZMK$^+$ populations. $^*P<0.05$, $^{**}P<0.01$ (two-tailed unpaired *t*-test for GZMK$^+$ in PB versus LN and lung versus LN, two-tailed Mann–Whitney *U* test for all other comparisons).

Exhausted-like CD8⁺ memory T cell progenitors express GZMK, PD-1 and TIGIT. Heterogeneity in the early differentiated memory T cell pool became apparent with the identification of multipotent T_{SCM} cells[4,23]. These cells exhibit a CCR7⁺CD45RO⁻CD95⁺ phenotype, in contrast to T_{CM} cells, which exhibit a CCR7⁺CD45RO⁺CD95⁺ phenotype. However, our scRNA-seq-guided flow-cytometric



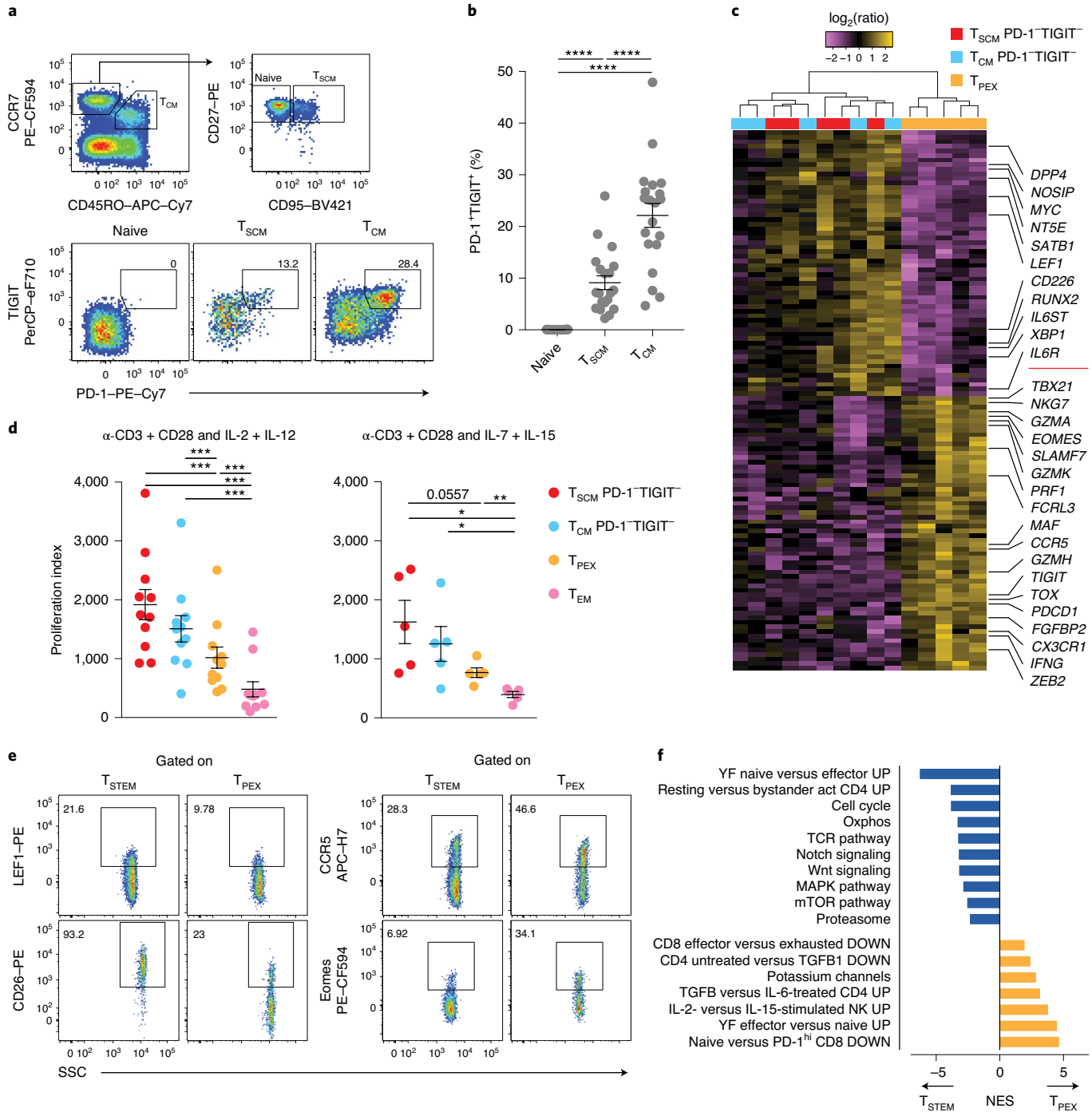


Fig. 2 | Identification of stem-like CD8⁺ memory T cell progenitors with differential expression of GZMK, PD-1 and TIGIT. a, Flow-cytometric gating strategy for the identification of PD-1⁺TIGIT⁺ cells in the CD8⁺ naive (CCR7⁺CD45RO⁻CD95⁻), T_{SCM} (CCR7⁺CD45RO⁻CD95⁺) and T_{CM} compartments (CCR7⁺CD45RO⁺CD95⁺). Numbers indicate percentages in the drawn gates. **b**, Dot plot summarizing the data obtained through the process outlined in **a**. Each dot represents 1 donor ($n = 20$ from 2 independent experiments). Bars indicate mean ± s.e.m. $^{****}P < 0.0001$ (1-way repeated-measures analysis of variance). **c**, Heatmap showing DEGs (adjusted P value < 0.01) for the indicated CD8⁺ memory T cell subsets ($n = 5$ donors). Labels highlight genes associated with memory or effector differentiation or exhaustion. Significance was evaluated using edgeR analysis with glmQLFTest and Benjamini–Hochberg correction. **d**, Proliferation of the indicated CD8⁺ memory T cell subsets in response to stimulation with anti-CD3 plus CD28 for 4 d in the presence of IL-2 and IL-12 ($n = 11$ donors from 6 independent experiments) or IL-7 and IL-15 ($n = 5$ donors from three independent experiments). Index calculations were based on the dilution of carboxyfluorescein succinimidyl ester (CFSE). Each dot represents one donor. Bars indicate mean ± s.e.m. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ (two-tailed Wilcoxon signed rank test for each population versus T_{EM} in the presence of IL-2 and IL-12, two-tailed paired t-test for all other comparisons). **e**, Representative flow-cytometric analysis of T_{STEM} (CCR7⁺PD-1⁻TIGIT⁻) and T_{PEX} cells (CCR7⁺PD-1⁺TIGIT⁺) showing the expression of markers selected from the DEGs identified in **c**. Numbers indicate percentages in the drawn gates. Similar data were obtained from other donors ($n = 5$ for LEF1 and CCR5, $n = 4$ for CD26 and Eomes). **f**, GSEA based on 1,000 permutations showing manually curated signatures that differed significantly (adjusted P value < 0.05) between T_{STEM} and T_{PEX} cells. UP and DOWN refer to genes upregulated or donwregulated, respectively, in the comparison. NES, normalized enrichment score.

analyses demonstrated that the CCR7$^+$GZMK$^-$PD-1$^-$TIGIT$^-$ and CCR7$^+$GZMK$^+$PD-1$^+$TIGIT$^+$ subsets could not be distinguished via the expression of CD45RO (Fig. 1h). To place these findings in context, we investigated the expression of PD-1 and TIGIT among classically defined T_{SCM} and T_{CM} cells. We found that 9.1 ± 1.3% of T_{SCM} cells and 22.1 ± 2.3% of T_{CM} cells (mean ± s.e.m.) expressed both PD-1 and TIGIT (Fig. 2a,b). Manual gating of the flow-cytometry data confirmed that PD-1 and TIGIT were preferentially expressed by CCR7$^+$GZMK$^+$ cells (Supplementary Fig. 1b).

On the basis of these results, we hypothesized that differential inclusion of the transcriptionally distinct CCR7$^+$GZMK$^+$PD-1$^+$TIGIT$^+$ subset could explain some of the previously reported differences between T_{SCM} and T_{CM} cells. To test this possibility, we analyzed the transcriptomes of T_{SCM} and T_{CM} cells after depletion of the CCR7$^+$PD-1$^+$TIGIT$^+$ (GZMK$^+$) population, hereafter termed T progenitor exhausted-like (T_{PEX}) (Supplementary Fig. 1a and Supplementary Fig. 2). In line with our hypothesis, T_{SCM} and T_{CM} cells depleted of T_{PEX} cells were very similar at the transcriptional level and could only be distinguished on the basis of 8 DEGs (adjusted P value < 0.01) (Supplementary Fig. 2 and Supplementary Table 4). One of these DEGs was *HNRNPLL*, which encodes heterogeneous nuclear ribonucleoprotein L-like, a master regulator of alternative splicing responsible for the expression of CD45RO[24], which is commonly used as a phenotypic marker to differentiate between T_{SCM} and T_{CM} cells[25]. In contrast, T_{PEX} cells were largely distinct, featuring lower expression levels of *SATB1*, which encodes a negative regulator of PD-1 expression[26], *MYC*, *DPP4*, which encodes CD26, *IL6ST*, *LEF1*, *IL6R* and *NT5E* and higher expression levels of transcription-factor (TF) genes recently associated with T cell exhaustion, including *TOX*[27–32], *EOMES*[33] and *MAF*[34], and other genes associated with effector differentiation and cytolytic activity, including *ZEB2*, *GZMK*, *GZMA*, *TBX21*, *PRF1*, *IFNG* and *NKG7* (Fig. 2c and Supplementary Table 4). As expected, *PDCD1*, which encodes PD-1, and *TIGIT* were also expressed at high levels, validating the integrity of cell isolation via FACS. Several other genes previously found to distinguish T_{SCM} from T_{CM} cells were identified among these DEGs[4] (Fig. 2c and Supplementary Table 4). In line with the transcriptional data, T_{PEX} cells stimulated with anti-CD3 plus CD28 and a combination of effector (IL-2 plus IL-12) or homeostatic cytokines (IL-7 plus IL-15) proliferated less vigorously than did PD-1$^-$TIGIT$^-$ T_{SCM} and T_{CM} cells under identical conditions (Fig. 2d). However, all three subsets proliferated similarly and remained phenotypically stable in response to IL-15, suggesting that they had equivalent self-renewal capabilities (Supplementary Fig. 3a,b). Accordingly, T_{SCM} and T_{CM} cells were better defined by the CCR7$^+$PD-1$^-$TIGIT$^-$ phenotype, hereafter termed stem-like T (T_{STEM}), whereas early differentiated memory cells with dysfunctional, exhausted-like traits were characterized by the CCR7$^+$PD-1$^+$TIGIT$^+$ phenotype (T_{PEX}). Of note, the gene expression profiles of T_{STEM} and T_{PEX} cells overlapped significantly with those of C6 and C2, respectively ($P < 0.01$ for each comparison using a hypergeometric test; data not shown), confirming the shared identity of subsets analyzed via scRNA-seq and flow cytometry (Fig. 1e).

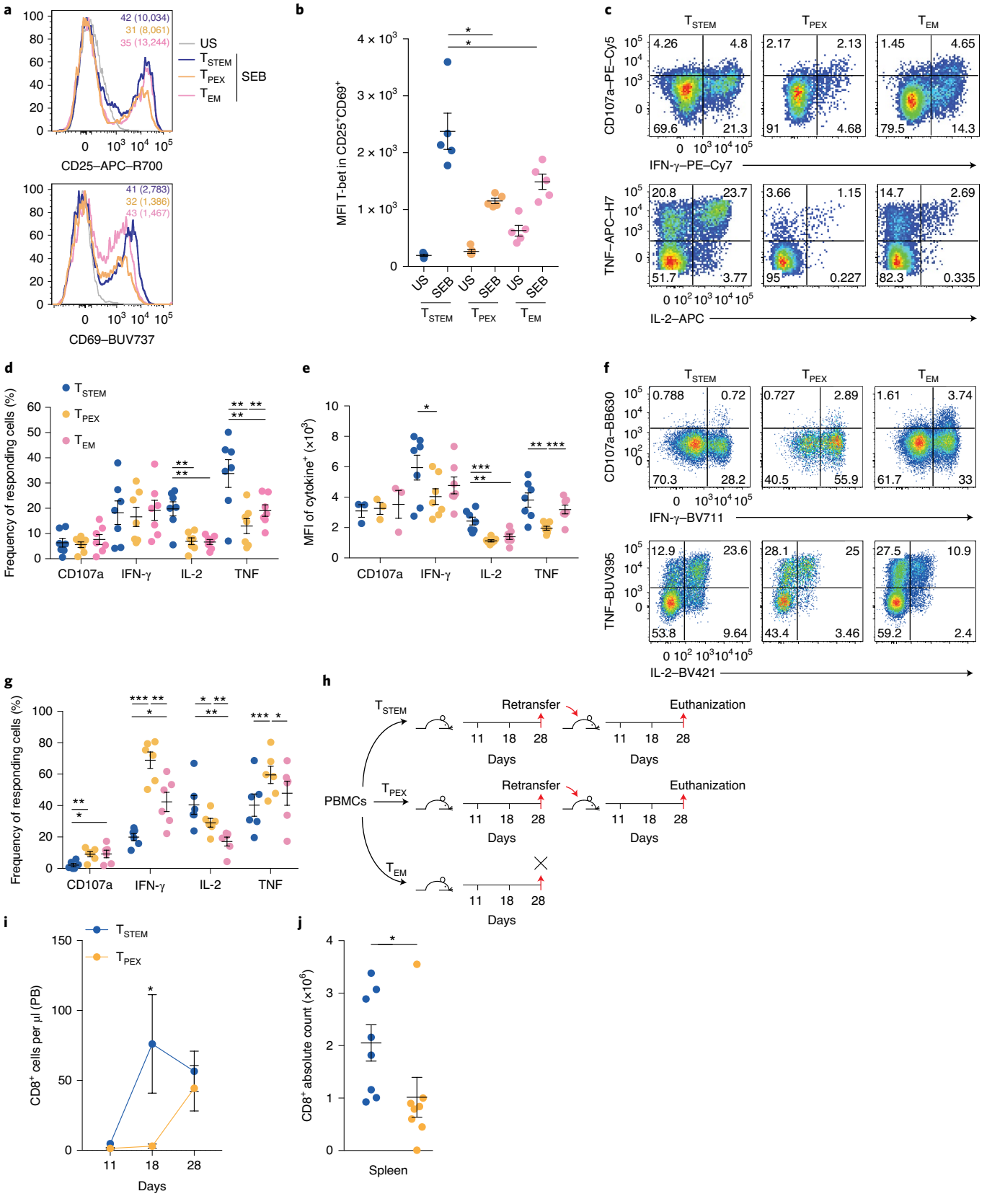
Flow-cytometric analyses demonstrated that T_{STEM} cells expressed CD26 and LEF1 more commonly, and CCR5 and Eomes less commonly, than did T_{PEX} cells (Fig. 2e). Gene-set enrichment analysis (GSEA) further revealed that T_{STEM} cells were characterized by transcripts associated with the naive state, quiescence, oxidative phosphorylation, the Wnt[35] and Notch signaling pathways[36] and proteasome activity[37], whereas T_{PEX} cells were characterized by transcripts associated with the transforming growth factor (TGF)-β signaling pathway[38], potassium regulation[39] and other mechanistic correlates of exhaustion, including the PD-1^{hi} state (Fig. 2f). Transcripts that were associated with the cell cycle and the TCR and mTOR signaling pathways, collectively suggesting a predisposition to antigen-driven proliferation and effector differentiation, were also upregulated in T_{STEM} versus T_{PEX} cells (Fig. 2f). Previous analyses have shown that progenitor exhausted CD8$^+$ T cells from tumors express stem-like genes along with *PDCD1*, *TIGIT* and *GZMK*[16], thereby suggesting a shared identity with T_{PEX} cells. Indeed, the transcriptional features of T_{PEX} cells aligned closely with those reported previously for progenitor exhausted-like (*CCR7*^{hi}*GZMK*^{hi}), but not memory-like (*CCR7*^{hi}*GZMK*^{lo}), CD8$^+$ T cells isolated from melanomas[18], whereas the opposite result was obtained in a parallel analysis of T_{STEM} cells ($P < 0.05$ for each comparison using a hypergeometric test) (Supplementary Table 4).

Collectively, these data revealed that T_{SCM} and T_{CM} cells were largely homogenous after depletion of the T_{PEX} subset, indicating a need to refine current models of CD8$^+$ memory T cell differentiation.

T_{STEM} cells are functionally superior to T_{PEX} cells. To validate our transcriptional and phenotypic data, we compared the functional properties of FACS-purified T_{STEM} and T_{PEX} cells. In response to TCR-dependent stimulation with Staphylococcal enterotoxin B (SEB), T_{STEM} cells upregulated CD25 and CD69 to a greater extent than did T_{PEX} cells (Fig. 3a), and activated CD25$^+$CD69$^+$ T_{STEM} cells expressed higher levels of T-bet than did activated CD25$^+$CD69$^+$ T_{PEX} cells (Fig. 3b). Likewise, T_{STEM} cells produced cytokines (IL-2 and TNF) at higher frequencies and at higher levels on a per-cell basis (IFN-γ, IL-2 and TNF) than did T_{PEX} cells in response to stimulation with anti-CD3 plus CD28 (Fig. 3c–e). No clear

Fig. 3 | Functional properties of T_{STEM} and T_{PEX} cells. a, Representative flow-cytometric analysis of FACS-purified T_{STEM}, T_{PEX} and T_{EM} cells, showing the expression of CD25 and CD69 before (US, unstimulated) and after stimulation with SEB for 24 h. Similar data were obtained from other donors ($n = 4$). Numbers colored to match each subset indicate the percentage of marker$^+$ cells, with the corresponding median fluorescence intensity (MFI) in brackets. **b**, Dot plot showing the expression of T-bet among CD25$^+$CD69$^+$ T_{STEM}, T_{PEX} and T_{EM} cells before (US) and after stimulation, as in **a**. Data are shown in terms of MFI. Each dot represents one donor ($n = 5$ from two independent experiments). Bars indicate mean ± s.e.m. *$P < 0.05$ (two-tailed paired t-test). **c**, Representative flow-cytometric analysis showing the expression of CD107a, IFN-γ, IL-2 and TNF among T_{STEM}, T_{PEX} and T_{EM} cells stimulated with anti-CD3 plus CD28 for 12 h. Numbers indicate the percentages in the drawn gates. **d**, Dot plot summarizing the data obtained as in **c**. Each dot represents one donor ($n = 7$ from four independent experiments). Bars indicate mean ± s.e.m. **$P < 0.01$ (two-tailed paired t-test). **e**, Dot plot showing the function$^+$ populations identified in **d** in terms of MFI. Each dot represents one donor ($n = 3$ from 4 independent experiments for CD107a, $n = 7$ from 4 independent experiments for IFN-γ, IL-2 and TNF). Bars indicate mean ± s.e.m. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ (two-tailed paired t-test). **f**, Representative flow-cytometric analysis showing the expression of CD107a, IFN-γ, IL-2 and TNF among T_{STEM}, T_{PEX} and T_{EM} cells after stimulation of magnetically enriched CD8$^+$ T cells with PMA and ionomycin for 3 h. Numbers indicate percentages in the drawn gates. Subsets were gated as CCR7$^+$GZMK$^-$ (T_{STEM}), CCR7$^+$GZMK$^+$ (T_{PEX}) or CCR7$^-$CD45RO$^+$CD95$^+$ (T_{EM}). **g**, Dot plot summarizing the data obtained as in **f**. Each dot represents 1 donor ($n = 6$ from 3 independent experiments). Bars indicate mean ± s.e.m. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$ (two-tailed paired t-test). **h**, Schematic layout of the serial-transfer experiments. **i**, Line chart showing the absolute numbers of CD8$^+$ T cells in PB on days 11, 18 and 28 after transfer of T_{STEM} or T_{PEX} cells into secondary NSG recipients. Data were pooled from two independent experiments (total $n = 8$ mice). Bars indicate mean ± s.e.m. *$P = 0.0117$ (two-way ANOVA). PBMCs, PB mononuclear cells. **j**, Dot plot showing the absolute numbers of CD8$^+$ T cells in spleen on day 28 after transfer of T_{STEM} or T_{PEX} cells into secondary NSG recipients. Each dot represents 1 mouse ($n = 8$ from 2 independent experiments). Bars indicate mean ± s.e.m. *$P = 0.0148$ (two-tailed Mann–Whitney U test).

differences were observed between T_{STEM} and T_{PEX} cells with respect to degranulation, measured via the surface mobilization of CD107a (Fig. 3c–e). In response to TCR-independent stimulation with phorbol myristate acetate (PMA) and ionomycin, however, T_{PEX} cells produced IFN-γ and TNF and mobilized CD107a at much higher frequencies than did T_{STEM} cells, the functional superiority



of which was therefore limited to conditions that mimicked antigen-recognition events (Fig. 3f,g).

To determine the in vivo relevance of these observations, we performed serial adoptive-cell transfers (ACTs) in NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG) humanized mice (Fig. 3h). T_{EM} cells failed to repopulate these mice efficiently after the first ACT (data not shown), as has been reported previously[4]. Although the early memory subsets both expanded to similar numbers in primary hosts (data not shown), T_{STEM} cells proliferated more rapidly in PB (Fig. 3i) and repopulated the spleen more efficiently in secondary hosts compared with T_{PEX} cells (Fig. 3j). Of note, the suboptimal proliferative capabilities of T_{PEX} cells observed in vitro and in vivo were not associated with differences in telomere length relative to that of T_{STEM} cells, whereas T_{EM} cells harbored shorter telomeres ex vivo than did either T_{STEM} or T_{PEX} cells (Supplementary Fig. 3c).

Collectively, these data showed that T_{STEM} cells were functionally superior to T_{PEX} cells, both under homeostatic conditions and in response to stimulation via the TCR.

T_{PEX} cells are committed to a terminally dysfunctional state. Epigenetic regulation plays a key role in T cell fate decisions[40]. We therefore employed the assay for transposase-accessible chromatin using sequencing (ATAC-seq) to compare the open chromatin landscapes of T_{STEM} and T_{PEX} cells in terms of differentially accessible regions (DARs). Naive and T_{EM} cells were analyzed in parallel as lineage controls. Principal-component analysis (PCA) revealed that T_{STEM} and T_{PEX} cells were globally similar, although T_{STEM} cells mapped toward the naive subset, whereas T_{PEX} cells mapped toward the T_{EM} subset (Fig. 4a). However, we also identified a total of 13,414 DARs between T_{STEM} and T_{PEX} cells (Fig. 4b). Genes associated with T cell dysfunction (for example, *TOX*, *TOX2*, *TIGIT*, *PDCD1*, *NFATC2* and *MAF*), terminal differentiation (for example, *ZEB2* and *BATF*) and other immune-related processes previously identified at the messenger RNA level (for example, *EOMES* and *GZMA*) were more accessible in T_{PEX} versus T_{STEM} cells (Fig. 4b,c). In contrast, genes associated with T cell memory (for example, *LEF1*, *SELL*, *CCR7*, *BACH2* and *SATB1*) and effector functions (for example, *GZMB* and *RORA*[41]) were more accessible in T_{STEM} versus T_{PEX} cells (Fig. 4b,c). Computational analysis of these DARs further identified differentially accessible TF-binding motifs (TFBMs). Motifs linked to TFs associated with thymocytes and naive and early memory cells (RUNX2, RUNX1, LEF1 and FOXP1), effector differentiation (RORA)[41] and cytokine signaling (STAT5, STAT4 and STAT1) were enriched in T_{STEM} versus T_{PEX} cells, whereas the TBX21 (T-bet), EOMES and combined TBOX–SMAD motifs were enriched in T_{PEX} versus T_{STEM} cells (Fig. 4d).

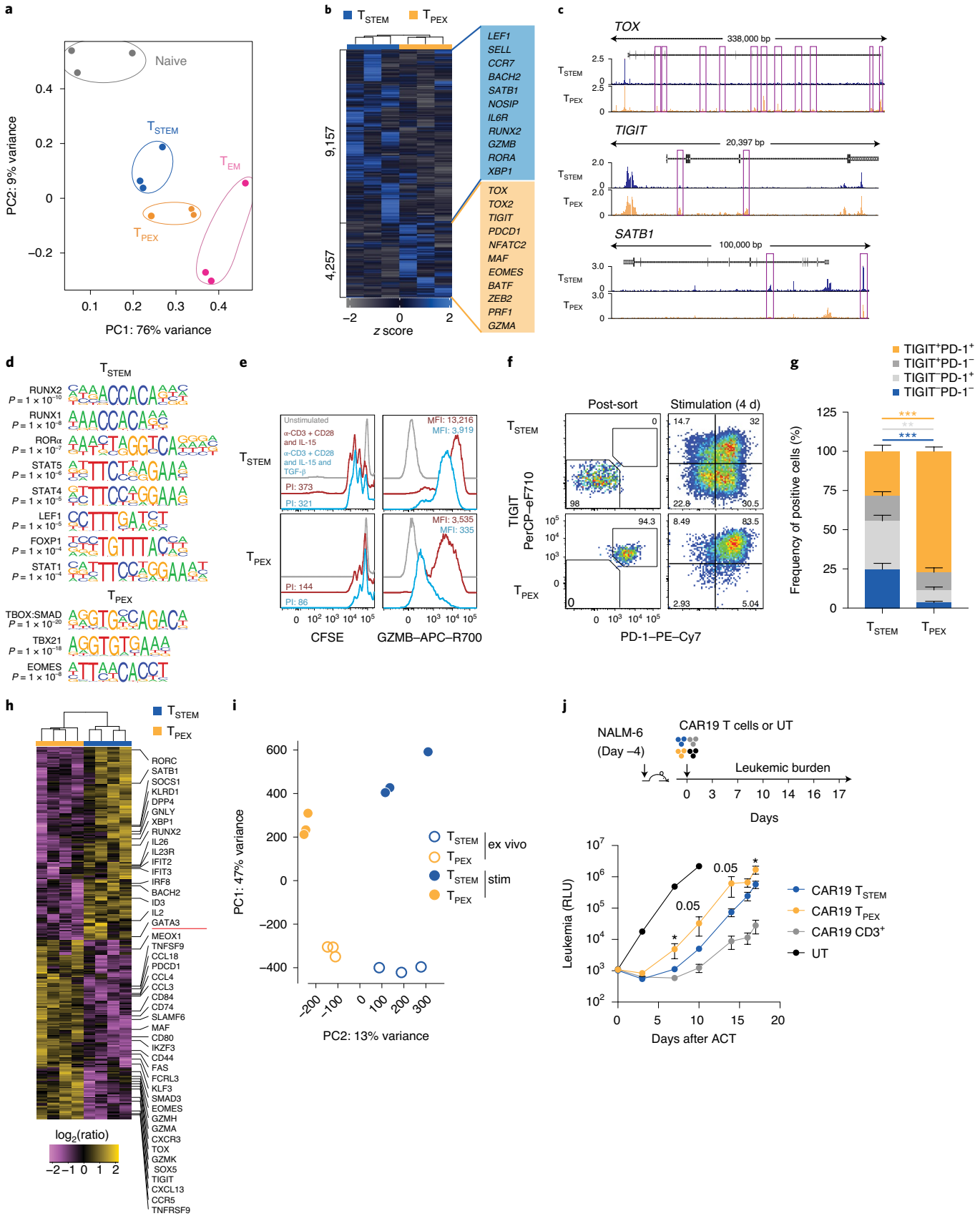
The chromatin-accessibility data suggested that T_{PEX} cells were predisposed to the generation of dysfunctional progeny and susceptible to the inhibitory effects of TGF-β signaling via SMADs. Accordingly, T_{PEX} cells proliferated to a lesser extent and produced less GZMB than did T_{STEM} cells in response to stimulation with anti-CD3 plus CD28 and IL-15 (Fig. 4e). The addition of TGF-β further inhibited these responses, especially the production of GZMB, in parallel cultures of T_{PEX} cells, whereas minimal effects were observed in parallel cultures of T_{STEM} cells (Fig. 4e). Importantly, most T_{PEX} cells retained a PD-1$^+$ TIGIT$^+$ phenotype after stimulation with anti-CD3 plus CD28 in the presence of the effector cytokines IL-2 and IL-12 (Fig. 4f,g) or the homeostatic cytokines IL-7 and IL-15 (Supplementary Fig. 3d). In contrast, T_{STEM} cells generated all possible combinations of phenotypes defined by PD-1 and TIGIT (Fig. 4f,g). We then used RNA-seq to profile the transcriptomes of T_{STEM} and T_{PEX} cells after stimulation with anti-CD3 plus CD28 in the presence of IL-2 and IL-12 (Fig. 4h). Activated T_{STEM} cells overexpressed the memory-related genes *BACH2*, *ID3*, *IL2* and *SATB1* alongside the effector-related genes *IRF8*, *RORC*, *GNLY*, *XBP1*, *IL26* and *IL23R*, whereas activated T_{PEX} cells overexpressed the dysfunction/exhaustion-related molecules *TOX*, *PDCD1*, *TIGIT*, *MAF* and *CXCL13* (ref. [42]), together with various chemokine genes, *IKZF3*, which encodes an inhibitor of IL-2 production, *SMAD3* and genes associated with cytolytic activity, including *GZMK*, *GZMH* and *GZMA* (Fig. 4h and Supplementary Table 5). Some of these genes were also differentially expressed between the corresponding subsets in ex vivo analyses (Fig. 2c). GSEA further demonstrated that activated T_{STEM} cells were preferentially enriched for gene sets associated with early differentiation and proliferation, whereas activated T_{PEX} cells were preferentially enriched for gene sets associated with the TGF-β and PD-1 signaling pathways and exhaustion in the tumor microenvironment[18] (Supplementary Fig. 4a). PCA of ATAC-seq data from paired ex vivo and activated samples revealed that stimulation profoundly altered the chromatin accessibility landscape in T_{STEM} and T_{PEX} cells (Fig. 4i). However, the major epigenetic differences between these subsets in the ex vivo state were maintained after stimulation (Fig. 4i), both at the level of specific genes (Supplementary Fig. 4b) and in terms of enrichment for particular TFBMs (Supplementary Fig. 4c).

To assess the in vivo relevance of these findings, we employed a stringent ACT protocol in which T_{STEM} and T_{PEX} cells were redirected using a chimeric antigen receptor (CAR) targeting CD19 and were transferred in the absence of autologous CD4$^+$ T cells or exogenous cytokines into NSG mice previously injected with the acute lymphoblastic leukemia cell line NALM6. In line with the in vitro data, T_{STEM} cells displayed enhanced control of leukemic burden compared with that of T_{PEX} cells at multiple time points after ACT (Fig. 4j).

Collectively, these data indicated that T_{STEM} cells were relatively resistant to exhaustion, facilitating their more-efficient control of tumor growth in vivo than that of T_{PEX} cells, which were hardwired to a dysfunctional signature.

Fig. 4 | Fate commitments of T_{STEM} and T_{PEX} cells. a, PCA plot showing the top 1,000 hypervariable peaks obtained from ex vivo ATAC-seq analysis (adjusted *P* value < 0.01) of T_{STEM}, T_{PEX} and T_{EM} cells. Each dot represents 1 donor ($n=3$). **b**, Heatmap showing DARs. Labels highlight accessible genes associated with memory or effector differentiation or exhaustion. **c**, Representative genomic regions showing the ATAC-seq profiles of *TOX*, *TIGIT* and *SATB1* in T_{STEM} and T_{PEX} cells. DARs are highlighted in purple. **d**, TFBMs enriched among the DARs shown in **b**. Enrichment was assessed using a one-sided hypergeometric test in HOMER with correction for false-discovery rate. **e**, Representative overlay histograms showing CFSE dilution (left) and GZMB-expression profiles (right) for T_{STEM} and T_{PEX} cells stimulated with anti-CD3 plus CD28 and IL-15 for 3 d in the absence or presence of TGF-β. Unstimulated controls are shown for comparison. Similar data were obtained from other donors in the absence ($n=8$ from 4 independent experiments) or presence of TGF-β ($n=6$ from 4 independent experiments). PI, proliferation index. **f**, Representative flow-cytometric analysis of T_{STEM} and T_{PEX} cells showing the expression of PD-1 and TIGIT after stimulation with anti-CD3 plus CD28 for 4 d in the presence of IL-2 and IL-12. Numbers indicate percentages in the drawn gates. **g**, Bar graph summarizing the data obtained as in **f** ($n=5$ donors from 3 independent experiments). Bars indicate mean ± s.e.m. $^{**}P<0.01$, $^{***}P<0.001$ (two-tailed Mann–Whitney *U* test). **h**, Heatmap showing selected DEGs (adjusted *P* value < 0.05) for T_{STEM} and T_{PEX} cells stimulated as in **f** ($n=4$ donors). Significance was evaluated using edgeR analysis with glmQLFTest and Benjamini–Hochberg correction. **i**, PCA plot as in **a** comparing T_{STEM} and T_{PEX} cells before and after stimulation as in **f**. **j**, Top, schematic layout of the adoptive-transfer experiment. Bottom, time-series plot showing the growth of NALM6 cells in NSG mice ($n=5$ mice per group) adoptively transferred with T_{STEM} or T_{PEX} cells expressing a CAR specific for CD19 (CAR19). RLU, relative light unit; UT, untransduced CD3$^+$ cells. Follow-up was stopped when RLU values $\geq 1\times10^6$ were observed in more than 75% of mice in 1 of the treated groups. Bars indicate mean ± s.e.m. $^{*}P<0.05$ (two-tailed unpaired *t*-test for CAR19 T_{STEM} versus CAR19 T_{PEX} on day 17, two-tailed Mann–Whitney *U* test for all other comparisons).

T_{PEX} cells are abundant in persistent infections and clonally distinct from T_{STEM} cells. $CD8^+$ T cell dysfunction and exhaustion develop in response to persistent antigenic stimulation via cognate TCRs[43]. We therefore reasoned that acute viral infections would preferentially generate antigen-specific T_{STEM} cells, whereas chronic viral infections would preferentially generate antigen-specific



T_{PEX} cells. A single round of yellow-fever virus (YFV) vaccination is known to induce long-lived memory cells with an early differentiated T_{SCM}-like CCR7$^+$CD45RA$^+$ (or CD45RO$^-$) CD95$^+$ phenotype[44,45]. To determine the composition of these T_{SCM}-like populations in terms of CCR7$^+$PD-1$^-$TIGIT$^-$ T_{STEM} and CCR7$^+$PD-1$^+$TIGIT$^+$ T_{PEX} cells, we compared publicly available gene expression data from vaccinated individuals[45] with our RNA-seq data (Fig. 2c). In line with our hypothesis, we found that YFV-specific CCR7$^+$CD45RA$^+$CD95$^+$ cells analyzed years after vaccination were transcriptionally related to T_{STEM}, but not T_{PEX}, cells (Fig. 5a).

To extend these findings, we used peptide–HLA class I tetramers in conjunction with mass cytometry by time of flight (CyTOF) to investigate the phenotypic characteristics of CCR7$^+$CD8$^+$ memory T cells specific for acute (influenza virus or rotavirus) or chronic viruses (cytomegalovirus (CMV) or Epstein–Barr virus (EBV)) in healthy donors ($n=3$) and HIV$^+$ people ($n=2$). Signature markers of T_{PEX} cells, namely PD-1, TIGIT, GZMK, GZMA and CCR5, were expressed at higher levels among EBV-specific and, to a lesser extent, CMV-specific CD8$^+$ T cells than among influenza-virus-specific and rotavirus-specific CD8$^+$ T cells (Fig. 5b,c and Supplementary Fig. 5a,b). Chronic virus-specific CD8$^+$ T cells also overexpressed 2B4. Moreover, high-avidity CMV-specific CD8$^+$ T cell populations, selectively identified using a double-point-mutated peptide-HLA class I tetramer[46], incorporated T_{PEX} cells at frequencies equivalent to those detected among the corresponding total CMV-specific CD8$^+$ T cell populations, suggesting that persistent antigenic drive rather than signal strength determined the acquisition and maintenance of dysfunctional, exhausted-like traits[43] (Supplementary Fig. 5c).

In further experiments, we used a high-throughput approach (TCR-seq) to profile the clonotypic repertoires of T_{STEM} and T_{PEX} cells. As expected, these early-differentiated subsets exhibited similarly diverse repertoires, measured via the normalized Shannon–Weiner index, whereas the corresponding T_{EM} subsets exhibited comparatively less-diverse repertoires (Fig. 5d). An additional estimator based on abundance, the Chao1 index, which accounts for the distribution of infrequent clonotypes, revealed similar trends and further identified greater levels of diversity among the T_{STEM} subsets compared than among the T_{PEX} subsets, potentially reflecting distinct broadness of specificities (Supplementary Fig. 5d). Although the stem-like subsets both shared clonotypes with the corresponding T_{EM} subsets, minimal repertoire overlap was detected between T_{STEM} and T_{PEX} cells, quantified in terms of the absolute numbers (Fig. 5e) or normalized counts of shared clonotypes (Fig. 5f), suggesting distinct spectra of recognized antigens. Comparable data were obtained using two additional metrics: F2, which accounts for the size of each clonotype (Supplementary Fig. 5e), and R, which estimates correlations of clonotype frequencies (Supplementary Fig. 5f).

Collectively, these data revealed that T_{STEM} and T_{PEX} cells were clonally distinct and committed to parallel differentiation programs, the relative prevalence of which was determined by the dynamics of antigen exposure within any given specificity (Fig. 5g).

Discussion

In this study, we used an unbiased approach guided by scRNA-seq to capture the extensive heterogeneity that exists in the human CD8$^+$ memory T cell pool under physiological conditions. We identified two previously unrecognized subsets of stem-like CD8$^+$ memory T cells, neither of which corresponded with previous descriptions of early-differentiated progenitors on the basis of the expression of CCR7, CD45RA/RO and CD95. These subsets were defined by core transcriptional signatures that could be distilled phenotypically into simple profiles, namely CCR7$^+$PD-1$^-$TIGIT$^-$ (T_{STEM} cells) and CCR7$^+$PD-1$^+$TIGIT$^+$ (T_{PEX} cells). Moreover, the distinct gene-expression profiles of T_{SCM} and T_{CM} cells were mostly attributable to the differential inclusion of T_{PEX} cells, indicating a need for refined models to understand the process of human CD8$^+$ memory T cell differentiation.

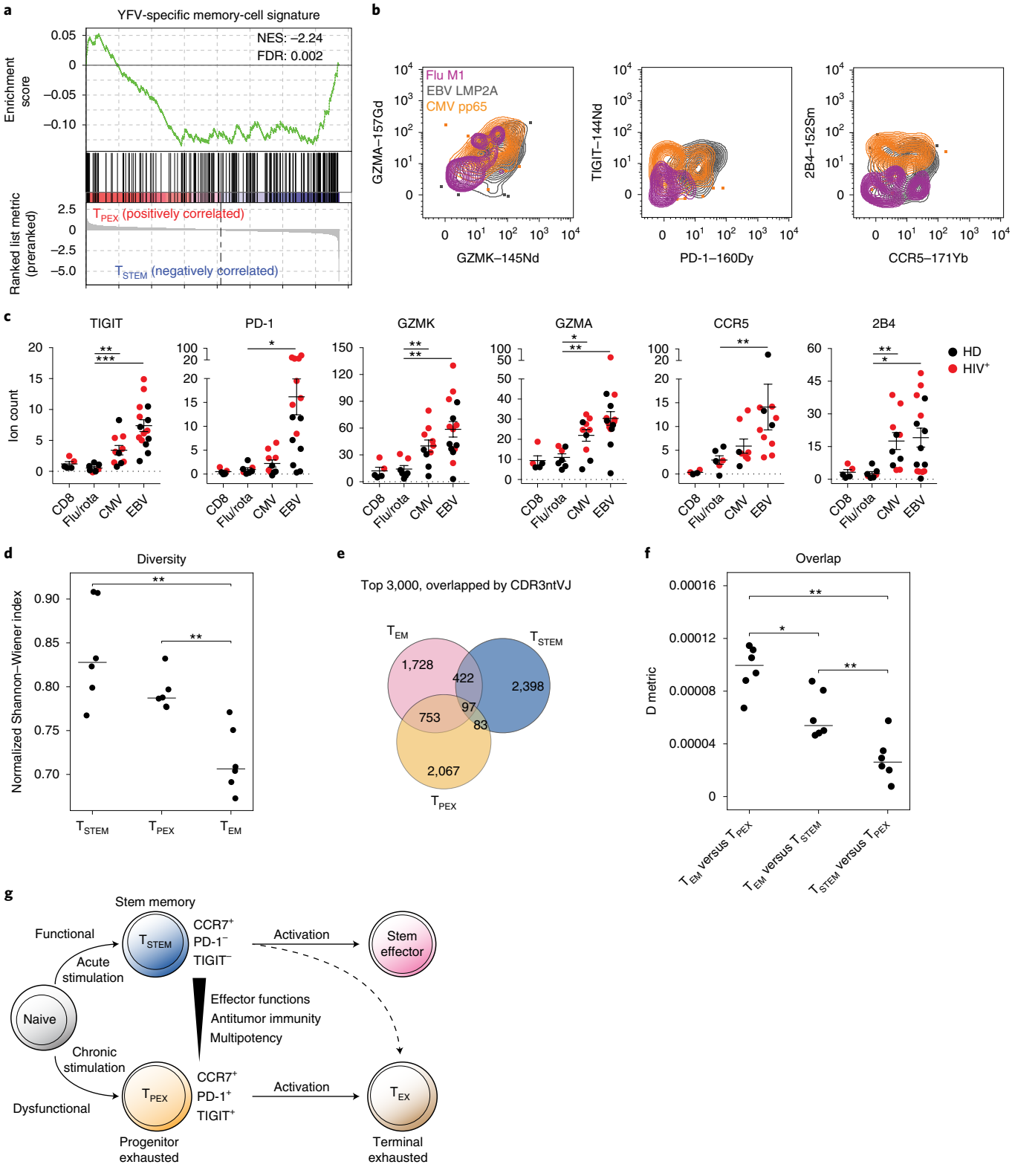
T_{STEM} cells proliferated vigorously in response to activation and generated a diverse array of memory and effector progeny, collectively enabling functionally superior immunity in vivo. Of note, the ex vivo frequencies of T_{STEM} cells were sufficient to overcome current limitations associated with the relative paucity of T_{SCM} cells, potentially facilitating immunotherapies that rely on specificity redirection by providing an alternative source of progenitors with self-renewal capabilities and a propensity for effector differentiation. In contrast, T_{PEX} cells were committed to the generation of progeny with reduced functionality and proliferated less efficiently in response to activation, at least via the TCR. Importantly, T_{STEM} and T_{PEX} cells were also clonally, epigenetically and transcriptionally distinct, suggesting a branching point in the early memory compartment associated with the initial antigen recognition event(s). This interpretation was supported by the observation that persistent antigenic stimulation was preferentially associated with the development of T_{PEX} cells. On the basis of these findings, we propose a revised model of T cell differentiation, according to which T_{PEX} cells become hardwired to a dysfunctional-like signature after immune activation and effector differentiation, compatible with the generation of a parallel lineage[47] defined by genome-wide epigenetic modifications[48,49], whereas T_{STEM} cells remain multipotent and relatively resistant to exhaustion, resulting in enhanced functionality and protective immunity in vivo. This model is likely to need further refinement to accommodate a degree of plasticity within the T_{STEM} subset, given that initially functional CD8$^+$ memory T cells can become exhausted as a result of continuous exposure to high-dose antigen in mice chronically infected with lymphocytic choriomeningitis virus (LCMV)[50].

The acquisition of dysfunctional traits associated with exhaustion was not accompanied by a substantial loss of memory-like features in the T_{PEX} subset. Instead, these characteristics were found to coexist in individual cells, potentially indicating functional adaptation to persistent antigenic stimulation[3,16]. Such adaptations may be

Fig. 5 | Antigen specificity and clonal identity of T_{STEM} and T_{PEX} cells. a, GSEA of the YFV-specific CD8$^+$ memory T cell signature[45] in T_{STEM} versus T_{PEX} cells. **b**, Representative CyTOF analysis showing the expression of GZMK, GZMA, PD-1, TIGIT, CCR5 and 2B4 among CCR7$^+$ virus-specific CD8$^+$ T cell populations from healthy ($n=3$) and HIV$^+$ donors ($n=2$). **c**, Dot plots summarizing the data obtained as in **b**. Epitopes derived from influenza virus ($n=6$) and rotavirus ($n=1$) were pooled for simplicity. Comparative data are shown for the corresponding total CD8$^+$ T cell populations. Each dot represents 1 specificity in 1 donor ($n=3$ healthy donors, $n=2$ HIV$^+$ donors). Bars indicate mean ± s.e.m. $^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$ (two-tailed Mann–Whitney *U* test for CCR5 comparisons, two-tailed unpaired *t*-test for all other marker comparisons). HD, healthy donor. **d**, Dot plot showing the normalized Shannon–Wiener diversity index for TCR-β repertoires obtained from the T_{STEM}, T_{PEX} and T_{EM} subsets. Each dot represents 1 donor ($n=6$). Bars indicate median values. $^{**}P<0.01$ (two-tailed paired *t*-test with Bonferroni correction). **e**, Venn diagram showing the numbers of shared and unique clonotypes among T_{STEM}, T_{PEX} and T_{EM} cells from a representative donor. Shared clonotypes defined by identical CDR3 nt sequence and VJ segments, CDR3ntVJ. Similar data were obtained from other donors ($n=5$). Analysis was restricted to the top 3,000 clonotypes. **f**, Dot plot summarizing the pairwise comparisons among subsets illustrated in **e**. Relative overlap diversity, D metric in VDJtools. Each dot represents 1 donor ($n=6$). Bars indicate median values. $^*P<0.05$, $^{**}P<0.01$ (two-tailed paired *t*-test with Bonferroni correction). **g**, Proposed model showing the origins and differentiation trajectories of T_{STEM} and T_{PEX} cells.

necessary in this context to maintain a diverse repertoire and simultaneously minimize the risk of immunopathology[28,30,47]. Further studies will nonetheless be required to address these issues in patients with progressive malignancies or uncontrolled viral infections. In contrast, activated T_{STEM} cells coexpressed stem-like and effector genes, consistent with the notion of a functionally sustainable hybrid state[3]. Accordingly, we propose that T_{STEM} cells represent a naturally occurring lineage with optimal features for the induction of potent long-term immunity.

In summary, we have identified two subsets of human stem-like CD8$^+$ memory T cell progenitors with distinct fate commitments and lineage relationships. Although further work is required to characterize the molecular mechanisms that underlie the early dichotomy between T_{STEM} and T_{PEX} cells, we anticipate that such



efforts will reveal new targets for therapeutic interventions designed to inhibit or reverse the process of exhaustion, with obvious implications for the treatment of persistent infections and various cancers.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-020-0791-5.

Received: 21 April 2020; Accepted: 14 August 2020;
Published online: 12 October 2020

References

1. Kaech, S. M., Wherry, E. J. & Ahmed, R. Effector and memory T-cell differentiation: implications for vaccine development. *Nat. Rev. Immunol.* **2**, 251–262 (2002).
2. Gattinoni, L., Speiser, D. E., Lichterfeld, M. & Bonini, C. T memory stem cells in health and disease. *Nat. Med.* **23**, 18–27 (2017).
3. Lugli, E., Galletti, G., Boi, S. K. & Youngblood, B. A. Stem, effector, and hybrid states of memory CD8$^+$ T cells. *Trends Immunol.* **41**, 17–28 (2020).
4. Gattinoni, L. et al. A human memory T cell subset with stem cell-like properties. *Nat. Med.* **17**, 1290–1297 (2011).
5. Biasco, L. et al. In vivo tracking of T cells in humans unveils decade-long survival and activity of genetically modified T memory stem cells. *Sci. Transl. Med.* **7**, 273ra13 (2015).
6. Mahnke, Y. D., Brodie, T. M., Sallusto, F., Roederer, M. & Lugli, E. The who's who of T-cell differentiation: human memory T-cell subsets. *Eur. J. Immunol.* **43**, 2797–2809 (2013).
7. Graef, P. et al. Serial transfer of single-cell-derived immunocompetence reveals stemness of CD8$^+$ central memory T cells. *Immunity* **41**, 116–126 (2014).
8. Wherry, E. J. T cell exhaustion. *Nat. Immunol.* **12**, 492–499 (2011).
9. Angelosanto, J. M., Blackburn, S. D., Crawford, A. & Wherry, E. J. Progressive loss of memory T cell potential and commitment to exhaustion during chronic viral infection. *J. Virol.* **86**, 8161–8170 (2012).
10. Schietinger, A. et al. Tumor-specific T cell dysfunction is a dynamic antigen-driven differentiation program initiated early during tumorigenesis. *Immunity* **45**, 389–401 (2016).
11. Philip, M. et al. Chromatin states define tumour-specific T cell dysfunction and reprogramming. *Nature* **545**, 452–456 (2017).
12. Sen, D. R. et al. The epigenetic landscape of T cell exhaustion. *Science* **354**, 1165–1169 (2016).
13. Im, S. J. et al. Defining CD8$^+$ T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537**, 417–421 (2016).
14. Leong, Y. A. et al. CXCR5$^+$ follicular cytotoxic T cells control viral infection in B cell follicles. *Nat. Immunol.* **17**, 1187–1196 (2016).
15. Utzschneider, D. T. et al. T cell factor 1-expressing memory-like CD8$^+$ T cells sustain the immune response to chronic viral infections. *Immunity* **45**, 415–427 (2016).
16. Brummelman, J. et al. High-dimensional single cell analysis identifies stem-like cytotoxic CD8$^+$ T cells infiltrating human tumors. *J. Exp. Med.* **215**, 2520–2535 (2018).
17. He, R. et al. Follicular CXCR5-expressing CD8$^+$ T cells curtail chronic viral infection. *Nature* **537**, 412–428 (2016).
18. Sade-Feldman, M. et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell* **175**, 998–1013.e20 (2018).
19. Siddiqui, I. et al. Intratumoral Tcf1$^+$PD-1$^+$CD8$^+$ T cells with stem-like properties promote tumor control in response to vaccination and checkpoint blockade immunotherapy. *Immunity* **50**, 195–211.e10 (2019).
20. Miller, B. C. et al. Subsets of exhausted CD8$^+$ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019).
21. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2018).
22. Dusseaux, M. et al. Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161^{hi} IL-17-secreting T cells. *Blood* **117**, 1250–1259 (2011).
23. Lugli, E. et al. Superior T memory stem cell persistence supports long-lived T cell memory. *J. Clin. Invest.* **123**, 594–599 (2013).
24. Oberdoerffer, S. et al. Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPLL. *Science* **321**, 686–691 (2008).
25. Lugli, E. et al. Identification, isolation and in vitro expansion of human and nonhuman primate T stem cell memory cells. *Nat. Protoc.* **8**, 33–42 (2013).
26. Stephen, T. L. et al. SATB1 expression governs epigenetic repression of PD-1 in tumor-reactive T cells. *Immunity* **46**, 51–64 (2017).
27. Yao, C. et al. Single-cell RNA-seq reveals TOX as a key regulator of CD8$^+$ T cell persistence in chronic infection. *Nat. Immunol.* **20**, 890–901 (2019).
28. Alfei, F. et al. TOX reinforces the phenotype and longevity of exhausted T cells in chronic viral infection. *Nature* **571**, 265–269 (2019).
29. Scott, A. C. et al. TOX is a critical regulator of tumour-specific T cell differentiation. *Nature* **571**, 270–274 (2019).
30. Khan, O. et al. TOX transcriptionally and epigenetically programs CD8$^+$ T cell exhaustion. *Nature* **571**, 211–218 (2019).
31. Wang, X. et al. TOX promotes the exhaustion of antitumor CD8$^+$ T cells by preventing PD1 degradation in hepatocellular carcinoma. *J. Hepatol.* **71**, 731–741 (2019).
32. Seo, H. et al. TOX and TOX2 transcription factors cooperate with NR4A transcription factors to impose CD8$^+$ T cell exhaustion. *Proc. Natl Acad. Sci. USA* **116**, 12410–12415 (2019).
33. Li, J., He, Y., Hao, J., Ni, L. & Dong, C. High levels of Eomes promote exhaustion of anti-tumor CD8$^+$ T cells. *Front. Immunol.* **9**, 2981 (2018).
34. Giordano, M. et al. Molecular profiling of CD8 T cells in autochthonous melanoma identifies Maf as driver of exhaustion. *EMBO J.* **34**, 2042–2058 (2015).
35. Gattinoni, L. et al. Wnt signaling arrests effector T cell differentiation and generates CD8$^+$ memory stem cells. *Nat. Med.* **15**, 808–813 (2009).
36. Kondo, T. et al. Notch-mediated conversion of activated T cells into stem cell memory-like T cells for adoptive immunotherapy. *Nat. Commun.* **8**, 15338 (2017).
37. Widjaja, C. E. et al. Proteasome activity regulates CD8$^+$ T lymphocyte metabolism and fate specification. *J. Clin. Invest.* **127**, 3609–3623 (2017).
38. Yang, Z. Z. et al. TGF-β upregulates CD70 expression and induces exhaustion of effector memory T cells in B-cell non-Hodgkin's lymphoma. *Leukemia* **28**, 1872–1884 (2014).
39. Vodnala, S. K. et al. T cell stemness and dysfunction in tumors are triggered by a common mechanism. *Science* **363**, eaau0135 (2019).
40. Henning, A. N., Roychoudhuri, R. & Restifo, N. P. Epigenetic control of CD8$^+$ T cell differentiation. *Nat. Rev. Immunol.* **18**, 340–356 (2018).
41. Doering, T. A. et al. Network analysis reveals centrally connected genes and pathways involved in CD8$^+$ T cell exhaustion versus memory. *Immunity* **37**, 1130–1144 (2012).
42. Thommen, D. S. et al. A transcriptionally and functionally distinct PD-1$^+$ CD8$^+$ T cell pool with predictive potential in non-small-cell lung cancer treated with PD-1 blockade. *Nat. Med.* **24**, 994–1004 (2018).
43. Utzschneider, D. T. et al. High antigen levels induce an exhausted phenotype in a chronic infection without impairing T cell expansion and survival. *J. Exp. Med.* **213**, 1819–1834 (2016).
44. Fuertes Marraco, S. A. et al. Long-lasting stem cell–like memory CD8$^+$ T cells with a naive-like profile upon yellow fever vaccination. *Sci. Transl. Med.* **7**, 282ra48 (2015).
45. Akondy, R. S. et al. Origin and differentiation of human memory CD8 T cells after vaccination. *Nature* **552**, 362–367 (2017).
46. Price, D. A. et al. Avidity for antigen shapes clonal dominance in CD8$^+$ T cell populations specific for persistent DNA viruses. *J. Exp. Med.* **202**, 1349–1361 (2005).
47. Blank, C. U. et al. Defining 'T cell exhaustion'. *Nat. Rev. Immunol.* **19**, 665–674 (2019).
48. Pauken, K. E. et al. Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science* **354**, 1160–1165 (2016).
49. Ghoneim, H. E. et al. De novo epigenetic programs inhibit PD-1 blockade-mediated T cell rejuvenation. *Cell* **170**, 142–157.e19 (2017).
50. West, E. E. et al. Tight regulation of memory CD8$^+$ T cells limits their effectiveness during sustained high viral load. *Immunity* **35**, 285–298 (2011).

Methods

Study approval. The use of human samples was approved by the Humanitas Clinical and Research Center Institutional Review Board under the following protocols: buffy coats from healthy donors (28/01/2016), LNs and PB from people with head and neck cancer (700/2010), adjacent cancer-free lung tissue and PB from people with non-small-cell lung cancer (1501) and BM and PB from healthy donors (1397). Healthy and HIV$^+$ donors from the Fred Hutchinson Cancer Research Center were obtained via the HIV Vaccine Trials Network (HVTN). All donors provided written informed consent in accordance with the Declaration of Helsinki. Mice were housed and bred in a specific-pathogen-free animal facility and treated in accordance with the European Union Guideline on Animal Experiments. Mouse protocols were approved by the Italian Ministry of Health, the Humanitas Institutional Animal Care and Use Committee (256/2015-PR) and the San Raffaele Institutional Animal Care and Use Committee (646).

Cells. PBMCs were isolated from buffy coats via density-gradient separation and were either used fresh or cryopreserved in FBS supplemented with 10% DMSO. Tissue samples were processed as has been described previously[16,51,52]. Total CD8$^+$ T cells were enriched via negative magnetic separation using an EasySep Human CD8$^+$ T Cell Isolation Kit (Stem Cell Technologies) or a MojoSort Human CD8$^+$ T Cell Isolation Kit (BioLegend). Total CD3$^+$ T cells were enriched via negative magnetic separation using a MojoSort Human CD3$^+$ T Cell Isolation Kit (BioLegend). CD8-depleted PBMCs were obtained via negative magnetic separation using CD8 MicroBeads (Miltenyi Biotec). The human NALM-6 cell line (DSMZ) was tested for *Mycoplasma* (Eurofins Genomics) and transduced with a lentiviral vector encoding secreted luciferase (Lucia$^+$NGFR$^+$ NALM-6)[53].

Flow cytometry and cell sorting. High-dimensional flow cytometry was performed as has been described previously[54]. Dead cells were excluded from all analyses, using Zombie Aqua (BioLegend). Fluorochrome-conjugated monoclonal antibodies were purchased from commercial vendors (Supplementary Table 6). All reagents were titrated prior to use to determine optimal concentrations. Cells were fixed/permeabilized for intracellular analyses using a Cytofix/Cytoperm Kit (BD Biosciences). TFs and intranuclear molecules were measured in conjunction with a FoxP3 Transcription Factor Staining Buffer Set (eBioscience) or a Transcription Factor Buffer Set (BD Biosciences). Cell proliferation was determined by measuring the progressive dilution of CFSE (Thermo Fisher Scientific). The proliferation index (PI) was calculated as follows: (MFI of the non-proliferating fraction / MFI of the proliferating fraction) × (percent of cells with diluted CFSE)[16]. Samples were acquired using a FACSymphony A5 or an LSR Fortessa equipped with FACSDiva software version 8.0.1 (all from BD Biosciences). Electronic compensation was performed using single-stained controls prepared with antibody-capture beads (BD Biosciences)[54]. T cell subsets were sorted to purity using a FACSAria III (BD Biosciences) as shown in Supplementary Fig. 1a.

Flow-cytometry standard (FCS) 3.0 files were imported into FlowJo software version 9 (FlowJo). A conventional gating strategy was used to remove aggregates and dead cells, and 5,000 CD95$^+$ bulk CD8$^+$ memory T cells per sample (Supplementary Table 1) were exported into FlowJo software version 10 (FlowJo). Data were then biexponentially transformed and exported for further analysis in Python version 3.7.3 using a custom-written script incorporating PhenoGraph retrieved from the scikit-learn package (https://github.com/luglilab/Cytophenograph). Tissue samples were labeled with a unique computational barcode for further identification, converted into comma separated (CSV) files, and concatenated in a single matrix using the merge function in pandas (https://pandas.pydata.org/). The *K* value, indicating the number of nearest neighbors identified in the first iteration of the algorithm, was set to 1,000. UMAP was retrieved from Python. Data were visualized using FlowJo version 10 (FlowJo).

Peptide-HLA class I tetramers. Biotinylated wild-type and D227K/T228A (KA) HLA-A*0201 complexes refolded with CMV pp65$_{495–503}$ NLVPMVATV (NV9) were multimerized with streptavidin–PE (Sigma-Aldrich) as described previously[52]. Cells were stained with each tetramer at a concentration of 5 µg ml^{-1} for 15 min at 37 °C.

CyTOF. Cryopreserved PBMCs from three healthy donors and two HIV$^+$ donors were obtained from the HVTN. Purified monoclonal antibodies were purchased from commercial vendors (Supplementary Table 7) and labeled according to the Maxpar Antibody Labeling Kit Protocol (Fluidigm). Streptavidin was produced and labeled as described previously[55]. Myc-tagged peptide-HLA class I monomers were synthesized and biotinylated as described previously (Supplementary Table 7)[56]. Peptide-HLA class I tetramers were generated via the addition of heavy-metal-labeled streptavidins in a triple-coding scheme and were used to stain cells in a cocktail format as has been described previously[55]. All reagents were titrated prior to use to determine optimal concentrations. Antibody staining, live/dead discrimination and DNA staining were performed as has been described previously[57].

Cell culture and stimulation conditions. Cells were cultured in RPMI 1640 medium supplemented with 10% FBS, 1% penicillin–streptomycin and 2 mM L-glutamine. To induce cytokine production, magnetically enriched CD8$^+$ T cells were stimulated for 3 h with PMA (10 ng ml^{-1}) and ionomycin (500 ng ml^{-1}) (both from Sigma-Aldrich) in the presence of anti-CD107a (clone H4A3, BD Biosciences) and the protein-transport inhibitors GolgiPlug (brefeldin A, 1 µl ml^{-1}, BD Biosciences) and GolgiStop (monensin, 0.67 µl ml^{-1}, BD Biosciences). Subsets were identified among bulk CD8$^+$ memory T cells by gating as specified in the legend for Fig. 3f. Alternatively, FACS-purified CD8$^+$ T cell subsets were stimulated for 12 h with anti-CD3 plus CD28 DynaBeads (bead-to-cell ratio 1:2, Thermo Fisher Scientific). To evaluate the expression of activation markers and T-bet, FACS-purified CD8$^+$ T cell subsets were mixed with CD8-depleted autologous PBMCs (cell-to-cell ratio 1:4) and stimulated for 24 h with SEB (1 µg ml^{-1}, Sigma-Aldrich). To evaluate differentiation and proliferation, FACS-purified CD8$^+$ T cell subsets were stimulated for 3 or 4 d with anti-CD3 plus CD28 DynaBeads (bead-to-cell ratio 1:2, Thermo Fisher Scientific) in combination with various cocktails of human cytokines, including TGF-β, IL-2, IL-7, IL-12 and IL-15 (each at 10 ng ml^{-1}, Peprotech). To evaluate self-renewal capacity, FACS-purified CD8$^+$ T cell subsets were stimulated for 10 d with IL-15 (25 ng ml^{-1}, Peprotech). Unstimulated samples were used as controls in all assays.

scRNA-seq. FACS-purified CD95$^+$CD8$^+$ T cells were resuspended in 1 ml of PBS without Ca^{2+} and Mg^{2+} containing 0.04% BSA, washed twice by centrifugation at 450 rcf for 7 min, resuspended in 100 µl of the same medium and counted using a Countess II Automatic Cell Counter (Thermo Fisher Scientific). Approximately 20,000 cells from each sample were then loaded into one channel of Single Cell Chip A using a Chromium Single Cell 3′ v2 Reagent Kit (10x Genomics). After capture and lysis, complementary DNA was synthesized and amplified over 14 cycles according to the manufacturer's protocol (10x Genomics). Libraries were prepared from 50 ng amplified cDNA. Sequencing was performed using a NovaSeq 6000 System (Illumina). An average sequencing depth of at least 50,000 reads per cell was obtained for each sample.

scRNA-seq data analysis. Sample demultiplexing, barcode processing and unique molecular identifier (UMI) counting were performed using Cell Ranger version 2.1.1 (10x Genomics). Briefly, raw base call files were demultiplexed in FASTQ format using the 'cellranger mkfastq' pipeline, and the 'cellranger count' pipeline was run with '–transcriptome=refdata-cellranger-GRCh38-1.2.0' for each sample. Outputs from 'cellranger count' were concatenated in a single matrix. Libraries were then normalized to an identical sequencing depth using the 'cellranger aggr' pipeline. Pooled data were imported into R version 3.5.1 using Seurat version 3.0.1 (ref. [58]). Genes detected in fewer than three cells or cells containing fewer than 200 features were excluded from the analysis. Cells with unique feature counts less than 200 or greater than 3,500 were also filtered out, along with cells containing mitochondrial counts above 10%. The resulting dataset was normalized using a global scaling method converted by a scale factor (10,000) and log-transformed using the 'ScaleData' function in Seurat version 3.0.1. Data were then subjected to cluster analysis using standard package procedures and the 'FindClusters' function in Seurat version 3.0.1. Parameters were set to the first 20 principal components and a resolution of 0.6. DEGs for each cluster were identified using the Wilcoxon rank-sum test with default correction for multiple comparisons in Seurat version 3.0.1.

Anchor gene analysis of memory cell clusters from scRNA-seq data. Six different gene modules were computed from the scRNA-seq dataset using selected genes as anchors. Transcriptional scores were built by calculating the mean expression profiles of the top 100 genes most correlated with the anchors (Pearson's correlation).

Bulk RNA-seq. RNA was extracted from 50,000 FACS-purified CD8$^+$ T cells per subset using a Direct-Zol RNA Microprep Kit (Zymo Research) and stored at −80 °C. Quality control was performed using a High Sensitivity RNA ScreenTape Assay with a 4200 TapeStation System (Agilent). Libraries for mRNA sequencing were prepared from 5 ng total RNA using the SMART-Seq v4 Ultra Low Input RNA Kit (Clontech-Takara). Full-length cDNAs were processed using a Nextera XT DNA Library Preparation Kit (Illumina). Quality control was performed using a High Sensitivity DNA ScreenTape Assay with a 4200 TapeStation System (Agilent). Libraries were then multiplexed in an equimolar pool and sequenced using a NextSeq 500/550 Platform (Illumina). An average of 11 million single-end 75-base-pair (bp) reads were generated per sample. Libraries for total RNA sequencing were prepared from 1 ng total RNA using a SMART-Seq Stranded Kit (Clontech-Takara). Quality control and sequencing were performed as described for the mRNA libraries, generating an average of 103 million paired-end 75-bp reads per sample.

Bulk RNA-seq data analysis. Raw sequence data were quality-controlled using FastQC version 0.11.8 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Single-end reads were aligned to the human genome (GENCODE Human Release 29, reference genome sequence GRCh38/hg38) using STAR version 2.5.1b (ref. [59]). Alignments were performed using default parameters. Reads associated with annotated genes were counted using the STAR aligner option '-quantMode geneCounts'. Differential gene expression was assessed using the edgeR package

version 3.20.9. Benjamini–Hochberg correction was applied to estimate the false discovery rate (FDR). Paired-end reads were processed similarly after removing adapter sequences and poor-quality bases with Trimmomatic version 0.36.

Overrepresentation analysis. GSEA was applied to the entire list of genes in the RNA-seq expression matrix. Genes were ranked based on $\log_2$(fold change) calculated using the edgeR package version 3.20.9. GSEA was performed in preranked mode using a 'classic' enrichment statistic. Gene sets of interest were retrieved from collections C2 and C7 in the Molecular Signatures Database version 6.2 and integrated with those corresponding to exhausted T cell clusters G6 and G9 in Sade-Feldman et al.[18] or with those obtained via a reanalysis of the dataset in Akondy et al.[45].

Enrichment analysis. Normalized scRNA-seq counts were downloaded from the Gene Expression Omnibus (GSE120575). Analysis was restricted to cells belonging to clusters G5 or G10 as defined in Sade-Feldman et al.[18]. DEGs with an adjusted $P < 0.01$ in the pairwise comparisons of G5 versus G10 were identified using the 'FindAllMarkers' function in Seurat version 3.0.1. Hypergeometric tests were used to compare the G5 or G10 signatures with the combined T_{SCM}/T_{CM} or T_{PEX} signatures in 'phyper' R.

Microarray data analysis. Normalized data matrices from Akondy et al.[45] were downloaded from GEO (GSE26347). To identify the signatures 'YF naive versus effector UP' and 'YF effector versus naive UP', the expression profiles of effector CD8$^+$ T cells isolated 14 d after vaccination with YF-17D (GSM837587, GSM837588, GSM837589 and GSM837590) were compared with those of naive CD8$^+$ T cells (GSM837584, GSM837585 and GSM837586) using the limma algorithm in R version 3.34.9 (ref. [60]). The gene set of interest arbitrarily included the top 200 genes with the highest log fold change among DEGs with an adjusted $P < 0.05$. All samples hybridized on the Human Genome U133 Plus 2.0 Array (the second set of samples from the GSE26347 dataset) were concatenated with those hybridized on the HT Human Genome U133A Array (the third set of samples from the GSE26347 dataset). Probe sets were matched on both chips. Batch effects were eliminated by adjusting gene-expression values in the combined data matrix with the empirical Bayes method ComBat in SVA version 3.26.0 (ref. [61]). ComBat was applied with default parameters, except for the adjustment variables, which were imputed as a vector of platform type labels. To identify genes comprising the 'YFV-specific memory-cell signature' shown in Fig. 5a, the expression profiles of YF-17D virus NS4B-214 epitope-specific human CD8$^+$ memory T cells (GSM837594, GSM837595, GSM837596, GSM837597, GSM837598 and GSM837599) were compared with those of YFV-tetramer$^+$ effector CD8$^+$ T cells (GSM837587, GSM837588, GSM837589 and GSM837590) using the limma algorithm, arbitrarily selecting the top 200 genes with the highest log fold change among DEGs with an adjusted P value < 0.05.

ATAC-seq. Libraries were prepared using a protocol adapted from Buenrostro et al.[62]. Briefly, 50,000 FACS-purified CD8$^+$ T cells per subset were washed in PBS without Ca^{2+} and Mg^{2+} and resuspended in 50 µl lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 0.1% IPEGAL CA-630). Nuclei were pelleted by centrifugation for 10 min at 500*g* and resuspended in a final reaction volume of 50 µl comprising 1 µl of Tn5 transposase (made in-house), 10 µl of 5× transposase buffer (50 mM Tris-HCl pH 8.4, 25 mM $MgCl_2$) and 39 µl of ultrapure water (Milli-Q). The reaction was incubated with mixing at 300 r.p.m. for 30 min at 37 °C, supplemented with 10 µl clean-up buffer (900 mM NaCl, 30 mM EDTA), 5 µl of 20% SDS, 0.7 µl of ultrapure water (Milli-Q) and 4.3 µl of proteinase K (18.6 µg µl^{-1}, Thermo Fisher Scientific), and incubated for a further 30 min at 40 °C. Tagmented DNA was isolated using 2× SPRI Beads (Beckman Coulter) and amplified via PCR. Fragments smaller than 600 bp were isolated via negative size selection using 0.65× SPRI Beads (Beckman Coulter) and purified using 1.8× SPRI Beads (Beckman Coulter). Quality control was performed using a 4200 TapeStation System (Agilent) in conjunction with a Qubit 2.0 Fluorometer (Thermo Fisher Scientific). Libraries were then multiplexed in an equimolar pool and sequenced using a NextSeq 500/550 Platform (Illumina). At least 20 million single-end 75-bp reads were generated per sample.

ATAC-seq data analysis. Read quality was assessed using FastQC version 0.11.8 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Adapters and poor-quality bases were trimmed using Cutadapt version 1.16 (ref. [63]). Samples were aligned to the human reference genome GRCh38 using default parameters in BWA–MEM version 0.7.17. Mitochondrial reads were removed using SAMtools version 1.9 (ref. [64]). PCR duplicates were removed using the 'MarkDuplicates' function in Picard Tools version 2.19 (http://broadinstitute.github.io/picard/). Open chromatin was detected using MACS2 version 2.1.2 (ref. [65]) with an FDR < 0.01. The number of reads in each peak was determined using featureCounts version 1.6.4 (ref. [66]). Differentially accessible peaks were identified using an FDR cut-off below 0.05 after normalization in DESeq2 version 1.20 (Bioconductor). Peaks were annotated using the 'annotatePeaks.pl' function and scanned for motifs using the 'findMotifsGenome.pl' function in HOMER version 4.9.1.

Single-telomere length analysis. DNA was extracted from 6,000 FACS-purified CD8$^+$ T cells per subset using a QIAamp DNA Micro Kit (Qiagen). Single-telomere length analysis was carried out at the XpYp telomere as described previously[67]. Briefly, genomic DNA was eluted in 35 µl of Tris (10 mM) containing 0.75 µl of the Telorette-2 linker (10 µM). Multiple PCRs were then performed for each test DNA. Each reaction was set up in a final volume of 10 µl containing 1 µl of DNA and 0.5 µM of the telomere-adjacent and teltail primers in 75 mM Tris-HCl pH 8.8, 20 mM $(NH_4)_2SO_4$, 0.01% Tween-20 and 1.5 mM $MgCl_2$, with 0.5 U of a 10:1 mixture of Taq (Thermo Fisher Scientific) and Pwo polymerase (Roche Molecular Biochemicals). The reactions were processed in a Tetrad2 Thermal Cycler (Bio-Rad). DNA fragments were resolved via 0.5% Tris-acetate-EDTA agarose gel electrophoresis and identified via Southern hybridization with a random-primed anti-^{32}P-labeled (PerkinElmer) TTAGGG repeat probe, together with probes specific for molecular-weight markers at 1 kb (Stratagene) and 2.5 kb (Bio-Rad). Hybridized fragments were detected using a Typhoon FLA 9500 Phosphorimager (GE Healthcare). The molecular weights of the DNA fragments were calculated using a Phoretix 1D Quantifier (Nonlinear Dynamics).

TCR-seq. Total RNA was extracted separately from duplicate vials of 150,000 FACS-purified CD8$^+$ T cells per subset using an RNeasy Mini Kit (Qiagen). UMI-labeled 5′ RACE TCRβ sequencing libraries were prepared using a Human TCR Profiling Kit (MiLaboratory). Libraries were prepared in parallel using the same number of PCR cycles and sequenced using a NextSeq 500/550 High-Output Kit with a NextSeq 500 Platform (Illumina). Approximately 3×10^7 reads were obtained in total and assembled into 2.2×10^6 UMI-labeled cDNA molecules (up to 10^5 per library). UMI extraction and consensus assembly were performed using MIGEC software version 1.2.9 (ref. [68]) with a threshold of at least three reads per UMI. In-frame CDR3β repertoires were extracted using MiXCR software version 3.0.3 (ref. [69]). Each library contained from 3,000 to 18,000 functional CDR3β clonotypes. Diversity metrics were calculated using VDJtools software version 1.2.1 (ref. [70]) after normalization to 42,000 randomly selected UMIs per sample. D, R and F2 metrics were calculated for the top 3,000 clones from each pair of samples using VDJtools software version 1.2.1 (ref. [70]).

T cell transduction and culture conditions. FACS-purified CD8$^+$ T cell subsets were stimulated with a MACS-GMP T Cell TransAct Cocktail (Miltenyi Biotec). Stimulated cells were transduced the following day with a bidirectional lentiviral vector encoding a CD19-specific CAR with a CD28 costimulus in sense and the LNGFR marker gene in antisense and then cultured for 13 d in TexMACS Medium (Miltenyi Biotec) supplemented with 3% FBS, 1% penicillin/streptomycin, IL-7 (25 U ml^{-1}, Miltenyi Biotec),and IL-15 (50 U ml^{-1}, Miltenyi Biotec). Magnetically purified CD3$^+$ cells were processed similarly for control purposes.

Mouse studies. T_{STEM} and T_{PEX} cells were isolated from the PB of healthy donors on the basis of differential expression of PD-1. Eight-week-old female NSG mice (Charles River) were infused retroorbitally with FACS-purified T_{STEM}, T_{PEX} or T_{EM} cells (1×10^6 cells per mouse) and autologous CD8$^-$ PBMCs (6×10^6 cells per mouse), and were euthanized on day 28. To maximize recovery, spleen and lung cells were mixed from the same experimental group, normalized in terms of the CD4:CD8 ratio, and injected as above into secondary NSG recipients (1×10^6 CD8$^+$ T cells per mouse). Spleens were collected on day 28 and processed to single-cell suspensions. Absolute numbers of T cells in blood were determined using CountBright Absolute Counting Beads (Thermo Fisher Scientific). The frequencies of human CD4$^+$ and CD8$^+$ T cell subsets were determined by flow cytometry. For tumor experiments, eight-week-old female NSG mice were injected intravenously with 0.5×10^6 Lucia$^+$NGFR$^+$ NALM-6 cells. After 4 d, mice were further injected with 3×10^6 CAR19-redirected T_{STEM}, T_{PEX} or total CD3$^+$ T cells. Untransduced CD3$^+$ T cells were used as controls. Tumor progression was monitored weekly via bioluminescence detection using QUANTI-Luc (InvivoGen) and expressed as relative light units (RLUs).

Statistics. Statistical analyses were performed using Prism version 7.0c (GraphPad) or R software version 3.4.4. Significance was assigned at $P < 0.05$, unless stated otherwise. Specific tests are indicated in the relevant figure legends.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Publicly available scRNA-seq data were retrieved from the Gene Expression Omnibus via accession code GSE120575. Microarray data from YFV-17D-specific CD8$^+$ T cells were retrieved from the Gene Expression Omnibus via accession code GSE26347. Gene sets of interest were retrieved from the Molecular Signatures Database (http://www.broadinstitute.org/gsea/msigdb/index.jsp). The ATAC-seq data reported in this paper are available on request. The bulk RNA-seq and scRNA-seq data reported in this paper have been deposited in the Gene Expression Omnibus under accession code GSE147398. The TCR-seq data reported in this paper have been deposited at the European Bioinformatic Institute under accession

code E-MTAB-8892. All other data that support the findings of this study are available on request from the corresponding author.

Code availability

Scripts used to analyze the ATAC-seq data are available at https://github.com/luglilab/SP018_CD8_Galletti_et_al. All other codes are available on request.

References

51. Lugli, E. et al. IL-15 delays suppression and fails to promote immune reconstitution in virally suppressed chronically SIV-infected macaques. *Blood* **118**, 2520–2529 (2011).
52. Roberto, A. et al. Role of naive-derived T memory stem cells in T-cell reconstitution following allogeneic transplantation. *Blood* **125**, 2855–2864 (2015).
53. Falcone, L. & Casucci, M. Exploiting secreted luciferases to monitor tumor progression in vivo. *Methods Mol. Biol.* **1393**, 105–111 (2016).
54. Brummelman, J. et al. Development, application and computational analysis of high-dimensional fluorescent antibody panels for single-cell flow cytometry. *Nat. Protoc.* **14**, 1946–1969 (2019).
55. Newell, E. W. et al. Combinatorial tetramer staining and mass cytometry analysis facilitate T-cell epitope mapping and characterization. *Nat. Biotechnol.* **31**, 623–629 (2013).
56. Bakker, A. H. et al. Conditional MHC class I ligands and peptide exchange technology for the human MHC gene products HLA-A1, -A3, -A11, and -B7. *Proc. Natl Acad. Sci. USA* **105**, 3825–3830 (2008).
57. Simoni, Y. et al. Bystander CD8$^+$ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* **557**, 575–579 (2018).
58. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
59. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
60. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
61. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
62. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
63. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
64. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
65. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
66. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
67. Capper, R. et al. The nature of telomere fusion and a definition of the critical telomere length in human cells. *Genes Dev.* **21**, 2495–2508 (2007).
68. Shugay, M. et al. Towards error-free profiling of immune repertoires. *Nat. Methods* **11**, 653–655 (2014).
69. Bolotin, D. A. et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
70. Shugay, M. et al. VDJtools: unifying post-analysis of T cell receptor repertoires. *PLoS Comput. Biol.* **11**, e1004503 (2015).

Acknowledgements

The authors thank G. Natoli (European Institute of Oncology, Milan) for assistance with the ATAC-seq protocol, R. Roychoudhuri (University of Cambridge) and M. Iannacone (San Raffaele Scientific Institute, Milan) for critical discussions, and G. Cugini and G. Colombo (Humanitas Clinical and Research Center, Milan) for the provision of lymph node samples. This work was funded by the European Research Council (ERC-2014-STG PERSYST no. 640511 to E.L.) and by the Associazione Italiana per la Ricerca sul Cancro (AIRC IG 20676 to E.L.). Additional support was provided by the Associazione Italiana per la Ricerca sul Cancro (AIRC IG 21567 to D.M.), the Italian Ministry of Health (Bando Ricerca Finalizzata PE-2016-02363915 to D.M.), the Intramural Research Fund of the Humanitas Clinical and Research Center (5 ×1000 2019 Program to D.M.) and Cancer Research UK (C17199/A18246/A29202 to D.M.B.). G.G., G.D.S., S.P. and E.S. were supported by Fellowships from the Fondazione Italiana per la Ricerca sul Cancro-Associazione Italiana per la Ricerca sul Cancro (FIRC-AIRC). A.N.D. and M.M. were supported by the Ministry of Education, Youth, and Sports of the Czech Republic (CEITEC 2020 LQ1601). D.A.P. was supported by a Wellcome Trust Senior Investigator Award (100326/Z/12/Z). D.M.C. was supported by the Ministry of Health of the Russian Federation (0908300057056). The purchase of a FACSSymphony A5 was defrayed in part by a grant from the Italian Ministry of Health (Agreement 82/2015).

Author contributions

G.G. and E.L. conceived the study; G.G., G.D.S., E.M.C.M., S.P., C.M., T.M.B., A.N.D., M.M., E.S., G.A., F.D.P., V.Z., A.S., B.C., F.S.C., A.A., C.P., S.P., L.G., R.E.J., D.M.B., E.G., S.L.-L. and K.L. performed experiments; G.G., G.D.S., E.M.C.M., S.P., T.M.B., A.N.D., D.M.B., D.M.C., E.W.N., M.C. and E.L. analyzed data; D.M., S.K.B., B.A.Y. and D.A.P. provided critical expertise and resources; E.L. supervised the study; G.G., D.A.P. and E.L. wrote the manuscript. All authors contributed intellectually and approved the manuscript.

Competing interests

The Laboratory of Translational Immunology receives reagents in kind as part of a collaborative research agreement with BD Biosciences (Italy). L.G. and E.L. are inventors on a patent describing methods for the generation and isolation of T_{SCM} cells. E.L. has a consulting agreement with Achilles Therapeutics. L.G. has consulting agreements with Lyell Immunopharma and Advaxis Immunotherapies. E.W.N. is a cofounder and advisor for ImmunoScape. The other authors have no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41590-020-0791-5.

Correspondence and requests for materials should be addressed to E.L.

Peer review information Peer reviewer reports are available. Zoltan Fehervari was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Enrico Lugli

Last updated by author(s): Aug 3, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Flow cytometry data were collected using a FACSymphony A5 or an LSR Fortessa with FACSDiva Software Version 8.0.1 (BD Pharmingen). Mass cytometry data were collected using a Helios (Fluidigm). Sequencing data were collected using a NovaSeq 6000 or NextSeq 500/550 (Illumina). For single telomere length analysis hybridized fragments were detected using a Typhoon FLA 9500 Phosphorimager (GE Healthcare) and the molecular weights of the DNA fragments were calculated using a Phoretix 1D Quantifier (Nonlinear Dynamics).
Data analysis	FACS: Flow Cytometry Standard (FCS) 3.0 files were analyzed with FlowJo 9 and 10 (FlowJo LLC). In dedicated experiments samples were further analyzed with Python version 3.7.3 using a custom-written script incorporating PhenoGraph retrieved from the scikit-learn package (https://github.com/luglilab/Cytophenograph). UMAP was obtained by UMAP Python package and visualized in FlowJo 10. scRNA-Seq: Sample demultiplexing, barcode processing, and UMI counting were performed using Cell Ranger version 2.1.1 (10X Genomics). Pooled data were imported into R version 3.5.1 using Seurat version 3.0.1. RNA-Seq: Raw sequence data were quality-controlled using FastQC version 0.11.8 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Single-end reads were aligned to the human genome (GENCODE Human Release 29; Reference genome sequence: GRCh38/hg38) using STAR version 2.5.1b. Alignments were performed using default parameters. Reads associated with annotated genes were counted using the STAR aligner option "-quantMode geneCounts". Differential gene expression was assessed using the edgeR package version 3.20.9. Benjamini-Hochberg correction was applied to estimate the false discovery rate (FDR). Paired-end reads were processed similarly after removing adapter sequences and poor-quality bases with Trimmomatic version 0.36. Public data analyses: To identify the signatures "YF_naive vs. effector UP" and "YF_effector vs. naive UP", the expression profiles of effector CD8+ T cells isolated 14 d after vaccination with YF-17D were compared with those of naive CD8+ T cells using the limma algorithm in the same R package version 3.34.9. Public data obtained from Akondy et al., Nature, 2017. Batch effects were eliminated by adjusting gene expression values in the combined data matrix with the empirical Bayes method ComBat coded in the SVA package version 3.26.0. For GSEA, gene sets of interest were retrieved from collections C2 and C7 in the Molecular Signatures Database v6.2. To perform hypergeometric tests between G5 or G10 signatures and TSCM / TCM or TPEX subsets: differentially expressed genes (adjusted

p-value < 0.01) in the pairwise comparison of G5 versus G10 T cell clusters were determined by the "FindAllMarkers" function coded in the Seurat R package (version 3.0.1). Hypergeometric tests were run with the "phyper" R function. Public data obtained from Sade-Feldman et al., Cell, 2018.

ATAC-Seq: Read quality was assessed using FastQC version 0.11.8 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Adaptors and poor-quality bases were trimmed using Cutadapt version 1.16. Samples were aligned to the human reference genome GRCh38 using default parameters in BWA-MEM version 0.7.17. Mitochondrial reads were removed using SAMtools version 1.9. PCR duplicates were removed using the "MarkDuplicates" function in Picard Tools version 2.19 (http://broadinstitute.github.io/picard/). Open chromatin was detected using MACS2 version 2.1.2 with an FDR < 0.01. The number of reads in each peak was determined using featureCounts version 1.6.4. Differentially accessible peaks were identified using an FDR cut-off below 0.05 after normalization in DESeq2 version 1.20 (Bioconductor). Peaks were annotated using the "annotatePeaks.pl" function and scanned for motifs using the "findMotifsGenome.pl" function in HOMER version 4.9.1.

TCR-Seq: UMI extraction and consensus assembly were performed using MIGEC software version 1.2.9 with a threshold of at least three reads per UMI. In-frame CDR3β repertoires were extracted using MiXCR software version 3.0.3. Each library contained from 3,000 to 18,000 functional CDR3β clonotypes. Diversity metrics were calculated using VDJtools software version 1.2.1 after normalization to 42,000 randomly selected UMIs per sample. D, R, and F2 metrics were calculated for the top 3,000 clones from each pair of samples using VDJtools software version 1.2.1.

Other: Microsoft Excel version 15.37, GraphPad Prism version 7.0c and R software version 3.4.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Publicly available data were retrieved from the Gene Expression Omnibus: GSE120575 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120575; processed data for single cell RNA sequencing) and GSE26347 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26347; microarray data of YFV-17D specific CD8 T cells in humans). Gene sets of interest were retrieved from the Molecular Signatures Database (http://www.broadinstitute.org/gsea/msigdb/index.jsp). The ATAC-seq data reported in this paper are available on request. The bulk RNA-seq and scRNA-seq data reported in this paper have been deposited in the Gene Expression Omnibus under accession code GSE147398. The TCR-seq data reported in this paper have been deposited at the European Bioinformatic Institute under accession code E-MTAB-8892. All other data that support the findings of this study are available from the corresponding author upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was chosen taking into account the means of the target values between the different experimental groups, the standard error and the statistical analyses used. Additionally, the selection of sample size was based on previous studies conducted by this laboratory which allow for statistically valid comparisons.
Data exclusions	Eight-week-old female JAX NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (NSG, #005557, Charles River) mice were infused by retroorbital injection with TSTEM, TPEX or TEM cells (1x10^6 per mouse), freshly sorted from two healthy donors' PBMCs (Fig. 3h-j). An experiment from a third donor led to a poor recovery of cells following transfer in primary recipients. Recovered cells did not pass the pre-established criteria of inclusion which implies to retransfer 1x10^6 cells into secondary recipient mice. Thus, the experiment was excluded. No other data were excluded.
Replication	All data were reliably reproduced in at least two independent experiments with the exception of scRNA-seq (once from four donors), high-dimensional flow cytometry (once from 6 donors per tissue with matched peripheral blood), bulk RNA-seq (ex vivo, once from 5 donors; activated cells, once from 4 donors), ATAC-seq (ex vivo, once from 3 donors; activated cells, once from 3 donors), TCR-seq (once from 6 donors) and STELA assay (once from 6 donors). The experiment evaluating in vivo CAR19-redirected T cell subsets was also performed once.
Randomization	For in vivo tumor experiments, mice were randomized prior to adoptive cell transfer. For all the remaining experiments, donors (healthy and patients) were randomly selected in order to avoid potential biases (see Supplementary Table 1 for more details). No manual randomization was performed on these samples.
Blinding	All the data were collected and analyzed in a non-blind fashion because did not involve subjective measurements.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

FLOW CYTOMETRY
Anti-human CD45 - PE-Cy7 - 1:666 - HI30 - BioLegend - Cat # 304016 - RRID:AB_314404
Anti-human CD8 - BUV 805 - 1:166 - SK1 - BD Biosciences - Cat # 564912 - RRID:AB_2744465
Anti-human CD8 - BV786 - 1:322 - RPA-T8 - BD Biosciences - Cat # 563823 - RRID:AB_2687487
Anti-human CD4 - FITC - 1:80 - M-T477 - BD Biosciences - Cat # 556615 - RRID:AB_396487
Anti-human CCR7 - BV711 - 1:40 - G043H7 - BioLegend - Cat # 353228 - RRID:AB_2563865
Anti-human CD45RA - BV480 - 1:166 - HI100 - BD Biosciences - Cat # 566114 - RRID:AB_2739516
Anti-human CD45RO - APC-H7 - 1:80 - UCHL1 - BD Biosciences - Cat # 561137 - RRID:AB_10562194
Anti-human CCR7 - BB660 - 1:100 - 150503 - BD Biosciences - Cat # 625454 - N/A
Anti-human CD95 - BV421 - 1:166 - DX2 - BioLegend - Cat # 305624 - RRID:AB_2561830
Anti-human CD95 - BUV563 - 1:40 - DX2 - BD Biosciences - Cat # 624284 - N/A
Anti-human CD95 - APC - 1:333 - DX2 - BD Biosciences - Cat # 558814 - RRID:AB_398659
Anti-human CD45RO - BUV395 - 1:80 - UCHL1 - BD Biosciences - Cat # 562491 - N/A
Anti-human CD3 - BV650 - 1:40 - OKT3 - BioLegend - Cat # 317324 - RRID:AB_2563352
Anti-human CD4 - BV570 - 1:80 - RPA-T4 - BioLegend - Cat # 300534 - RRID:AB_2563791
Anti-human CD4 - BUV615 - 1:3,333 - SK3 - BD Biosciences - Cat # 624297 - N/A
Anti-human CD3 - BUV496 - 1:40 - UCHT1 - BD Biosciences - Cat # 564809 - RRID:AB_2744388
Anti-human CD27 - BV570 - 1:40 - O323 - BioLegend - Cat # 302825 - RRID:AB_11149686
Anti-human TNF-α - APC-Cy7 - 1:80 - mAB11 - BioLegend - Cat # 502944 - RRID:AB_2562870
Anti-human IFN-γ - PE-Cy7 - 1:666 - B27 - BioLegend - Cat # 506518 - RRID:AB_2123321
Anti-human IL-2 - APC - 1:322 - MQ1-17H12 - BD Biosciences - Cat # 554567 - RRID:AB_398571
Anti-human CD45 - Pacific Blue - 1:166 - HI30 - BioLegend - Cat # 304022 - RRID:AB_493655
Anti-human CD3 - PE-Cy5 - 1:20 - HIT-3A - BD Biosciences - Cat # 555341 - RRID:AB_395747
Anti-human CCR7 - PE-CF594 - 1:40 - 150503 - BD Biosciences - Cat # 562381 - RRID:AB_11153301
Anti-human CD27 - PE - 1:10 - M-T271 - BD Biosciences - Cat # 560985 - RRID:AB_10563213
Anti-human CD28 - BV785 - 1:166 - CD28.2 - BioLegend - Cat # 302950 - RRID:AB_2632607
Anti-human CD25 - APC-R700 - 1:80 - 2A3 - BD Biosciences - Cat # 565106 - RRID:AB_2744339
Anti-human CD69 - BUV737 - 1:322 - FN50 - BD Biosciences - Cat # 564439 - RRID:AB_2722502
Anti-human CD107a - PE-Cy5 - 1:80 - H4A3 - BD Biosciences - Cat # 555802 - RRID:AB_396136
Anti-human CD107a - BB630 - 1:2,500 - H4A3 - BD Biosciences - Cat # 624294 - N/A
Anti-human CD127 - PE-Cy5 - 1:40 - eBioRDR5 - eBioscience - Cat # 15-1278-42 - RRID:AB_2043801
Anti-human PD-1 - BV480 - 1:27 - EH12.1 - BD Biosciences - Cat # 566112 - RRID:AB_2739514
Anti-human PD-1 - PE-Cy7 - 1:166 - EH12.2H7 - BioLegend - Cat # 329918 - RRID:AB_2159324
Anti-human TIGIT - PerCP-eFluor 710 - 1:40 - MBSA43 - eBioscience - Cat # 46-9500-42 - RRID:AB_10853679
Anti-human TIGIT - FITC - 1:40 - MBSA43 - eBioscience - Cat # 11-9500-41 - RRID:AB_2572529
Anti-human TIGIT - BV421 - 1:80 - A15153G - BioLegend - Cat # 372710 - RRID:AB_2632925
Anti-human HLA-DR - BUV661 - 1:166 - G46-6 - BD Biosciences - Cat # 565073 - RRID:AB_2722500
Anti-human CD38 - BV711 - 1:166 - HIT2 - BioLegend - Cat # 303528 - RRID:AB_2563811
Anti-human CD103 - BV421 - 1:166 - Ber-ACT8 - BioLegend - Cat # 350213 - RRID:AB_2563513
Anti-human CD161 - BV605 - 1:20 - HP-3G10 - BioLegend - Cat # 339916 - RRID:AB_2563607
Anti-human CD14 - BV510 - 1:20 - M5E2 - BioLegend - Cat # 301842 - RRID:AB_2561946
Anti-human Granulysin - Alexa Fluor 488 - 1:20 - RB1 - BD Biosciences - Cat # 558254 - N/A
Anti-human EOMES - PE-eFluor 610 - 1:80 - WD1928 - eBioscience - Cat # 61-4877-41 - RRID:AB_2574615
Anti-human Granzyme B - APC-R700 - 1:80 - GB11 - BD Biosciences - Cat # 561016 - RRID:AB_2033973
Anti-human Granzyme K - PE - 1:166 - GM6C3 - Santa Cruz - Cat # sc-56125 PE - RRID:AB_2263772
Anti-human Granzyme K - Alexa Fluor 647 - 1:322 - GM6C3 - Santa Cruz - Cat # sc-56125 AF647 - RRID:AB_2263772
Anti-human IRF4 - Alexa Fluor 488 - 1:322 - IRF4.3E4 - BioLegend - Cat # 646406 - RRID:AB_256326
Anti-human IRF8 - APC - 1:666 - V3GYWCH - eBioscience - Cat # 17-9852-80 - RRID:AB_2573317
Anti-human T-bet - PE-Cy7 - 1:666 - 4-B10 - eBioscience - Cat # 25-5825-82 - RRID:AB_11042699
Anti-human LEF1 - PE - 1:111 - C12A5 - Cell Signaling - Cat # 14440 - N/A
Anti-mouse CD45.1 - PE-Cy7 - 1:100 - A20 - BD Biosciences - Cat # 560578 - RRID:AB_1727488
Anti-mouse CD45 - PerCP-Cy5.5 - 1:1,250 - 30-F11 - BioLegend - Cat # 103132 - RRID:AB_893340

Tetramer CMV-pp65 (wt) - PE - 1:50 - Sequence NLVPMVATV
Tetramer CMV-pp65 (D227K/T228A [KA]) - PE - 1:50 - Sequence NLVPMVATV

MASS CYTOMETRY
Anti-human CD45 - 89 - 1:200 - Hi30 - Fluidigm - Cat # 3089003B
Anti-human CD14 - 112/114 - 1:200 - TüK4 - Invitrogen - Cat # Q10064
Anti-human CD57 - 115 - 1:400 - HCD57 - BioLegend - Cat # 322325 (discontinued)
Anti-human Granzyme B - 141 - 1:300 - 2C5/F5 - BD Biosciences - Cat # 550558
Anti-human HLA-DR - 142 - 1:100 - L243 - BioLegend - Cat # 307651
Anti-human ITB7 - 143 - 1:200 - FIB504 - BioLegend - Cat # 321202
Anti-human TIGIT - 144 - 1:50 - MAB7898 - R&D Systems - Cat # MAB7898
Anti-human Granzyme K - 145 - 1:50 - GM6C3 - Life Technologies - Cat # MA1-17755
Anti-human CD8a - 146 - 1:400 - SK1 - BioLegend - Cat # 344727
Anti-human CD4 - 147 - 1:400 - SK3 - BioLegend - Cat # 344625
Anti-human CD45RO - 148 - 1:200 - UCHL1 - BioLegend - Cat # 304239
Anti-human CD161 - 149 - 1:100 - HP-3G10 - BioLegend - Cat # 339919
Anti-human KLRG1 - 150 - 1:200 - 13F2F12 - eBioscience - Cat # 16-9488-85
Anti-human CD27 - 151 - 1:200 - LG.7F9 - eBioscience - Cat # 5012495
Anti-human 2B4 - 152 - 1:100 - C1.7 - BioLegend - Cat # 329502
Anti-human CD103 - 153 - 1:400 - B-Ly7 - eBioscience - Cat # 5012794
Anti-human TCRgd-PE - 154 - 1:100 - 5A6.E9 - Invitrogen - Cat # MHGD04
Anti-PE - 1:100 - PE001 - BioLegend - Cat # 408105
Anti-human CD95 - 155 - 1:100 - DX2 - BioLegend - Cat # 305631
Anti-human CD3 - 156 - 1:200 - UCHT1 - BioLegend - Cat # 300443
Anti-human Granzyme A - 157 - 1:200 - CB9 - BioLegend - Cat # 507202
Anti-human CD56 - 158 - 1:200 - NCAM16.2 - BD Biosciences - Cat # 559043
Streptavidin - 159 - in house
Anti-human PD-1 - 160 - 1:50 - eBioJ105 - eBioscience - Cat # 14-2799-80
Streptavidin - 161 - in house
Anti-human CD19 - 162 - 1:200 - HIB19 - BioLegend - Cat # 302247
Streptavidin - 163 - in house
Streptavidin - 164 - in house
Anti-human CXCR5 - 165 - 1:100 - RF8B2 - BD Biosciences - Cat # 552032
Anti-human CTLA4 - 166 - 1:200 - BNI3 - BD Biosciences - Cat # 555851
Streptavidin - 167 - in house
Anti-human CCR7 - 168 - 1:100 - MAB197-100 - R&D Systems - Cat # MAB197-100
Anti-human CD45RA - 169 - 1:100 - HI100 - BioLegend - Cat # 304143
Streptavidin - 170 - in house
Anti-human CCR5 - 171 - 1:100 - NP-6G4 - Fluidigm - Cat # 3171017A
Anti-human CD39 - 172 - 1:100 - A1 - BioLegend - Cat # 328221
Streptavidin - 173 - in house
Anti-human CD127 - 174 - 1:100 - A019D5 - BioLegend - Cat # 351337
Anti-human Perforin - 175 - 1:200 - B-D48 - Abcam - Cat # ab47225
Anti-human CD38 - 176 - 1:100 - HIT2 - BioLegend - Cat # 303535
DNA - 191/193 - 1:2000 - Fluidigm - Cat # 201192B
Cisplatin - 194 - 1:4000 - Sigma-Aldrich - Cat # 479306-1G
Anti-human CD16 - 209 - 1:200 - 3G8 - Fluidigm - Cat # 3209002B
Tetramer CMV-IE1 - Sequence VLEETSVML
Tetramer CMV-pp65 - Sequence NLVPMVATV
Tetramer CMV-pp65-2 - Sequence QMWQARLTV
Tetramer EBV-BRLF1 - Sequence YVLDHLIVV
Tetramer EBV-BLMF1 - Sequence GLCTLVAML
Tetramer EBV-LMP1-1 - Sequence YLLEMLWRL
Tetramer EBV-LMP1-2 - Sequence YLQQNWWTL
Tetramer EBV-LMP2A - Sequence CLGGLLTMV
Tetramer Flu-M1 - Sequence GILGFVFTL
Tetramer Flu-PB1 - Sequence NMLSTVLGV
Tetramer Rota-VP6 - Sequence TLLANVTAV

Validation

All the reagents for flow cytometry and mass cytometry used in this manuscript went through stringent validation steps as previously described (Simoni et al., Nature, 2018; Mazza et al., Cytometry A., 2018; Brummelman et al., Nat. Protoc., 2019). Antibodies were all titrated to determine the optimal concentration. Flow cytometry antibodies were serially diluted as follows: 1:20, 1:40, 1:80, 1:160, 1:320, 1:640 and 1:1,280. Mass cytometry antibodies were serially diluted as follows: 1:50, 1:100, 1:200, and 1:400. Titration stainings included additional markers in order to validate the expression patterns on known cell subsets (e.g. T cells). Optimal concentration was defined by comparing the expression with other previously validated clones of a given antibody or with other previously validated lots of the same antibody clone. Any further information on the validation performed by the manufacturer can be retrieved from the manufacturers' websites.

Flow cytometry:
Fluorochrome-conjugated monoclonal antibodies were purchased from commercial vendors. High-dimensional flow cytometry was performed as previously described (Brummelman et al., Nat. Protoc., 2019). Biotinylated wildtype and D227K/T228A (KA) HLA-A*0201 complexes refolded with CMV pp65495-503 NLVPMVATV (NV9) were multimerized with streptavidin-PE (Sigma-Aldrich) as described previously (Roberto et al., Blood, 2015). Cells were stained with each tetramer at a concentration of 5 µg/ml for 15 min at 37°C.

Cytometry by Time of Flight (CyTOF):
Purified monoclonal antibodies were purchased from commercial vendors (Supplementary Table 7) and labeled according to the Maxpar Antibody Labeling Kit Protocol (Fluidigm). Streptavidin was produced and labeled as described previously (Ramachandiran et al., J. Immunol. Methods, 2007; Newell et al., Nat. Biotechnol., 2013). Myc-tagged peptide-HLA class I monomers were synthesized and biotinylated as described previously (Toebes et al., Nat. Med., 2006; Bakker et al., Proc. Natl. Acad. Sci. U. S. A., 2008). Peptide-HLA class I tetramers were generated via the addition of heavy metal-labeled streptavidins in a triple coding scheme and used to stain cells in a cocktail format as described previously (Ramachandiran et al., J. Immunol. Methods, 2007; Newell et al., Nat. Biotechnol., 2013). Antibody staining, live/dead discrimination, and DNA staining were performed as described previously (Simoni et al., Nature, 2018).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The human NALM-6 cell line was purchased from DSMZ, Germany.
Authentication	NALM-6 cell line was validated by the vendor, and by the authors' assessment of cell morphology and of CD19 expression by FACS.
Mycoplasma contamination	NALM-6 cell line was confirmed mycoplasma negative (Mycoplasmacheck, Eurofins Genomics).
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Eight-week-old female JAX NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (NSG, #005557, Charles River) mice were used. Mice were housed and bred in a specific pathogen-free animal facility, treated in accordance with the European Union guidelines.
Wild animals	No wild animals were used in this study.
Field-collected samples	No field-collected samples were employed in this study.
Ethics oversight	All experiments using mice were conducted upon the approval of the institutional IACUCs (Humanitas Clinical and Research Center and IRCCS San Raffaele Scientific Institute) and the Italian Ministry of Health.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Samples from healthy donors, HIV infected individuals, patients with head and neck cancer and non-small cell lung cancer patients were utilized in this study. A total of 124 individuals (95 males, 28 females and 1 unknown gender) were included in the study. Age range: 22-77 years old (3 unknown age). Two individuals were HIV infected patients. Twelve individuals were cancer patients: six with head and neck cancer and six with non-small cell lung cancer. All cancer patients were treatment-naive. Please, refer to Supplementary Table 1 for further details.
Recruitment	Healthy donors were obtained from Humanitas Clinical and Research Center and Fred Hutchinson Cancer Research Center. HIV infected individuals were obtained from Fred Hutchinson Cancer Research Center. Patients with head and neck cancer and non-small cell lung cancer patients were obtained from Humanitas Clinical and Research Center. We randomly selected samples from healthy donors or HIV infected individuals or treatment-free cancer patients that underwent surgical resection of the adjacent tumor-free lung tissue or lymph node. All the samples were anonymized. No other self-selection biases are present.
Ethics oversight	Humanitas Clinical and Research Center IRB and Fred Hutchinson Cancer Research Center HTVN approved sample and data collection.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	In all assays, cells were stained for 15 min at room temperature with Zombie Aqua fixable viability dye (BioLegend) to exclude dead cells. Fluorochrome-conjugated monoclonal antibodies were purchased from BD Biosciences, BioLegend, eBioscience, Santa Cruz Biotechnology and Cell Signaling and titrated to determine optimal concentrations. Chemokine receptor expression was measured by incubating cells at 37°C for 20 min. Surface markers were evaluated by incubating cells at RT for 20 min. The Cytofix/Cytoperm kit (BD Biosciences) was used to detect intracellular cytokine expression on sorted T cell subsets. Measurement of transcription factors and intranuclear molecules was performed with FoxP3 Transcription Buffer Set (Invitrogen) or Transcription Factors Buffer Kit (BD) according to manufacturers' instructions. Cell proliferation was determined by the analysis of 5-(and 6)-carboxyfluorescein diacetate succinimidyl ester (CFSE) dilution (final concentration: 2 uM; used according to the manufacturer's protocol; Life Technologies). The proliferation index (PI) was calculated as: MFI non-proliferating fraction / MFI proliferating fraction × % cells with diluted CFSE, as previously reported.
Instrument	T cell subsets were FACS-sorted to purity by using a FACSAria III (BD Biosciences). All samples were acquired on FACSymphony A5 (equipped to detect 30 parameters) or LSR Fortessa flow cytometers.
Software	Flow cytometry and cell sorting: Flow cytometry data were collected with FACSDIVA software and analyzed and compensated with FlowJo 9 and 10 (FlowJo LLC) by using single-stained controls prepared with antibody-capture beads (BD Biosciences). High-dimensional flow cytometry data analysis: Flow Cytometry Standard (FCS) 3.0 files were imported into FlowJo software version 9, analyzed by standard gating to remove aggregates and dead cells, and CD95+ bulk memory CD8+ T cells were identified. 5,000 CD95+ T cells per sample were subsequently imported in FlowJo version 10, biexponentially transformed and exported for further analysis in Python (version 3.7.3) by a custom-made script that makes use of PhenoGraph (originally retrieved from the scikit-learn package; full script available at https://github.com/luglilab/Cytophenograph). Lymph nodes, bone marrow and adjacent cancer-free lung tissues were labeled with a unique computational barcode for further identification and converted in comma separated (CSV) files and concatenated in a single matrix by using the merge function of pandas package. K value, indicating the number of nearest neighbors identified in the first iteration of the algorithm, was set equal to 1000 for clustering. UMAP was obtained by UMAP Python package and visualized in FlowJo 10.
Cell population abundance	Post-sort fractions were considered pure when the subset of interest was >93% of the sample. Purity was evaluated by FACS acquisition immediately after sort.
Gating strategy	Flow cytometric gating strategy for the isolation of CD8+ naive, TSTEM, TSCM PD-1– TIGIT–, TCM PD-1– TIGIT–, TPEX, and TEM cells is provided in Supplementary Figure 1. Briefly, lymphocytes were first selected on the basis of physical parameters and doublets and dead cells were excluded. Then, CD8+ naive and memory T cells were selected on the basis of CD95 expression. Early memory compartment was further identified as CD27+CCR7+ while TEM were defined as CD27+CCR7- cells. TSTEM and TPEX were then distinguished as PD1-TIGIT- and PD1+TIGIT+, respectively. In dedicated experiments, TSTEM were further subdivided into TSCM PD-1– TIGIT– and TCM PD-1– TIGIT– according to preferential expression of CD45RA and CD45RO, respectively.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.