Resource

Immunity

Human epigenetic and transcriptional T cell differentiation atlas for identifying functional T cell-specific enhancers

Highlights

- RNA-seq and ATAC-seq atlas of 14 human T cell subsets from healthy donors
- The atlas provides a reference map for interpreting signatures of T cells from 3 diseases
- Prediction of *cis*-regulatory elements that control CD8 T cell subset gene expression
- Validation of functional enhancers for CXCR3 and GZMB using CRISPRi

Authors

Josephine R. Giles, Sasikanth Manne, Elizabeth Freilich, ..., Alexander C. Huang, Junwei Shi, E. John Wherry

Correspondence

wherry@pennmedicine.upenn.edu

In brief

Giles et al. generated an epigenomic T cell differentiation atlas from healthy human donor (HD) blood. This atlas was used to (1) identify transcriptional and epigenetic modules in human CD8 T cell differentiation, (2) interpret signatures of T cells from cancer and autoimmunity, and (3) predict and validate CD8 T cell enhancers.

Giles et al., 2022, Immunity *55*, 557–574
March 8, 2022 © 2022 Published by Elsevier Inc.
https://doi.org/10.1016/j.immuni.2022.02.004

CellPress

Immunity

CellPress

Resource

Human epigenetic and transcriptional T cell differentiation atlas for identifying functional T cell-specific enhancers

Josephine R. Giles,[1,2,3] Sasikanth Manne,[1,2] Elizabeth Freilich,[5] Derek A. Oldridge,[2,6,11] Amy E. Baxter,[1,2] Sangeeth George,[1,12] Zeyu Chen,[1,2,13] Hua Huang,[1,2,7] Lakshmi Chilukuri,[1,15] Mary Carberry,[4,8] Lydia Giles,[4,8] Nan-Ping P. Weng,[10] Regina M. Young,[11] Carl H. June,[3,6,11] Lynn M. Schuchter,[4,8] Ravi K. Amaravadi,[4,8] Xiaowei Xu,[6,8,11] Giorgos C. Karakousis,[8,9] Tara C. Mitchell,[4,8] Alexander C. Huang,[2,3,4,8] Junwei Shi,[5] and E. John Wherry[1,2,3,14,*]

[1]Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Philadelphia, PA, USA
[2]Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[3]Parker Institute for Cancer Immunotherapy, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[4]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[5]Department of Cancer Biology, University of Pennsylvania, Philadelphia, PA, USA
[6]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[7]Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, PA, USA
[8]Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[9]Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[10]Laboratory of Immunology, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA
[11]Center for Cellular Immunotherapies, Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[12]Present address: Bristol-Myers Squibb, Lawrenceville, NJ, USA
[13]Present address: Dana-Farber Cancer Institute, Boston, MA, USA
[14]Present address: Prostate Cancer Clinical Trials Consortium, New York, USA
[15]Lead contact
*Correspondence: wherry@pennmedicine.upenn.edu
https://doi.org/10.1016/j.immuni.2022.02.004

SUMMARY

The clinical benefit of T cell immunotherapies remains limited by incomplete understanding of T cell differentiation and dysfunction. We generated an epigenetic and transcriptional atlas of T cell differentiation from healthy humans that included exhausted CD8 T cells and applied this resource in three ways. First, we identified modules of gene expression and chromatin accessibility, revealing molecular coordination of differentiation after activation and between central memory and effector memory. Second, we applied this healthy molecular framework to three settings—a neoadjuvant anti-PD1 melanoma trial, a basal cell carcinoma scATAC-seq dataset, and autoimmune disease-associated SNPs—yielding insights into disease-specific biology. Third, we predicted genome-wide *cis*-regulatory elements and validated this approach for key effector genes using CRISPR interference, providing functional annotation and demonstrating the ability to identify targets for non-coding cellular engineering. These studies define epigenetic and transcriptional regulation of human T cells and illustrate the utility of interrogating disease in the context of a healthy T cell atlas.

INTRODUCTION

T cells have become a major target of immunotherapies including checkpoint blockade and engineered cellular therapies. However, the design of optimal T cell therapeutics is limited by an incomplete understanding of epigenetic and transcriptional mechanisms controlling human T cell differentiation and function. Although some T cell differentiation states are thought to be more relevant for therapeutics than others (Huang et al., 2017; Gattinoni et al., 2011), the ability to manipulate the differentiation trajectory of human T cells to a specific outcome remains limited. Defining the epigenetic and transcriptional landscape of T cell differentiation in healthy humans can inform the mechanisms of T cell dysfunction in diseases and improve T cell therapies.

Studies in mice have revealed key insights into transcriptional and epigenetic mechanisms underlying T cell differentiation. Models of acutely resolved infections have provided insights into naive T cell activation and differentiation into effector and/or long-lived memory T cells (Kaech and Cui, 2012).

Next-generation sequencing approaches have connected the phenotype and function to the underlying transcriptional programs and epigenetic changes of effector and memory T cells, as well as exhausted T cells found in chronic infections and cancer (Crompton et al., 2016; Pauken et al., 2016; Best et al., 2013; He et al., 2016; Scharer et al., 2013; Yu et al., 2017; Scott-Browne et al., 2016; Scharer et al., 2017; Sen et al., 2016). These studies in mice have provided a foundation for understanding the key steps in transcriptional and epigenetic control of T cell differentiation. Chromatin accessibility and transcriptional profiling also support the idea of discrete human T cell differentiation states and have identified potential roles for transcription factors (TFs) discovered in mouse models (Araki et al., 2009; Abdelsamed et al., 2017; Ucar et al., 2017; Qu et al., 2015). Epigenomic and transcriptional profiling of human antiviral CD8 T cells (Akondy et al., 2017; Sen et al., 2016), tumor-infiltrating T cells (Satpathy et al., 2019), or follicular helper CD4 T cells (Vella et al., 2019) have defined the patterns of chromatin accessibility associated with individual human T cell subtypes and/or specific disease settings. Nevertheless, the transcriptional circuits and epigenomic changes associated with distinct states of human T cell differentiation remain poorly defined, and a comprehensive epigenetic and transcriptional landscape map does not yet exist for human T cells. A framework built on canonical human T cell subsets would enable the identification of the underlying epigenomic mechanisms that control differentiation and function as well as provide a reference atlas for T cell states in diseases, including single-cell profiles. Such a foundation could aid in developing optimal T-cell-targeted therapeutics, including checkpoint blockade and chimeric antigen receptor (CAR) T cells, for the treatment of cancer and other diseases.

We generated RNA-sequencing (RNA-seq) and assay for transposase-accessible chromatin sequencing (ATAC-seq) data from 14 circulating T cell subsets from healthy human donors (HDs) and constructed a transcriptional and epigenetic human T cell atlas. We applied this atlas in three ways to link the subset phenotype and function to the underlying transcriptional and epigenetic regulation. First, we investigated the transcriptional and epigenetic programs in HD CD8 T cell subsets and identified the gene expression and chromatin accessibility modules associated with differentiation states and state transitions. Second, we applied this HD molecular framework to multiple disease datasets including a melanoma clinical trial, single-cell ATAC-seq (scATAC-seq) from basal cell carcinoma (BCC), and single-nucleotide polymorphisms (SNPs) associated with immune diseases from genome-wide association studies (GWASs). Third, we used machine learning to translate the HD transcriptional and epigenetic information into a genome-wide resource of predicted *cis*-regulatory elements for human T cell gene expression across differentiation states. Lastly, we experimentally validated the function of the predicted enhancers for *CXCR3* and *GZMB* in primary human CD8 T cells using CRISPR interference (CRISPRi), verifying this approach. By integrating transcriptomic and epigenetic information built on a set of canonical human T cell subsets, this human T cell atlas provides a foundation to dissect transcriptional and epigenetic mechanisms underlying human disease as well as a guide for improving T cell therapeutics, including genomic engineering of non-coding regions.

RESULTS

Generation of HD T cell differentiation atlas

We sorted 14 human T cell subsets from the blood of HDs and generated an atlas of epigenetic and transcriptional data using ATAC-seq and RNA-seq. This atlas included samples from young and older HDs (10 donors aged 23–28 years, 14 donors aged 62–75 years); 4–21 ATAC-seq and 6–21 RNA-seq samples were sorted per T cell subset (Figure 1A; Table S1). A high-dimensional cytometry strategy was used to sort major T cell populations from the blood (Figures 1B–1D). Four subsets of CD4 T cells were purified: T regulatory T cells (Tregs), T follicular helper cells (Tfh), bulk naive cells, and bulk non-naive cells. For CD8 T cells, ten subsets were isolated. First, bulk naive and non-naive CD8 T cells were sorted using CD45RA and CD27. To capture heterogeneity within these bulk populations, we purified 8 additional populations. We distinguished a second stringently defined naive population (naive) from stem cell memory (SCM). Within the SCM population, we separated $CXCR3^+$ and $CXCR3^-$ subpopulations (SCM-$R3^+$ and SCM-$R3^-$). Memory and effector memory CD8 T cell populations from humans can be defined using CD45RA and CD27 or CCR7 (Hamann et al., 1997; Sallusto et al., 1999). We used all 3 surface proteins to define central memory (CM) and effector memory RA (EMRA) and fractionated the effector memory population into effector memory 1 (EM1, $CD27^+$) and effector memory 2 (EM2, $CD27^-$). We also purified a putative exhausted T cell (Tex) population based on the co-expression of PD1 and CD39. $PD1^+CD39^+$ CD8 T cells in chronic viral infection and cancer are Tex (Gupta et al., 2015; Canale et al., 2018; Bengsch et al., 2018), but such cells have not been characterized in HDs. These 14 purified human T cell populations were then subjected to RNA-seq and ATAC-seq.

We first examined global relationships between human T cell subsets based on gene expression or chromatin accessibility. Principal component analysis (PCA) of ATAC-seq and RNA-seq revealed samples from different subjects clustered according to a sorted subset label (Figures 1E and 1F). T cell subsets formed a gradient across PC1 with naive CD4 and CD8 T cells located at one extreme, memory subsets (SCM and CM) in the middle, and EMRA CD8 T cells at the opposite end. $PD1^+CD39^+$ CD8 T cells were located between memory (SCM-$R3^+$, SCM-$R3^-$, CM) and effector memory samples (EM1, EM2, and EMRA). The samples were then clustered using distal accessible chromatin regions (ACRs) from ATAC-seq (> 2kb from the nearest transcriptional start site [TSS]) (Figure 1G), proximal ACRs from ATAC-seq (< 2kb from the nearest TSS) (Figure 1H), or RNA-seq (Figure 1I). Similar to other cell types (Corces et al., 2016; Yoshida et al., 2019), distal ACRs resulted in clearest distinctions between subsets. To quantify this difference, we used information quality ratio (IQR), a normalized entropy metric, to confirm that distal ACRs contained more information about T cell subset identity than proximal ACRs or transcriptomic data (Figure 1J).

We also investigated the effect of age (Figures S1A and S1B). As expected (Goronzy and Weyand, 2019), there were fewer naive and more non-naive T cells in older HDs (Figure S1C). However, comparing purified T cell subsets directly revealed few differentially expressed genes (DEGs) in young

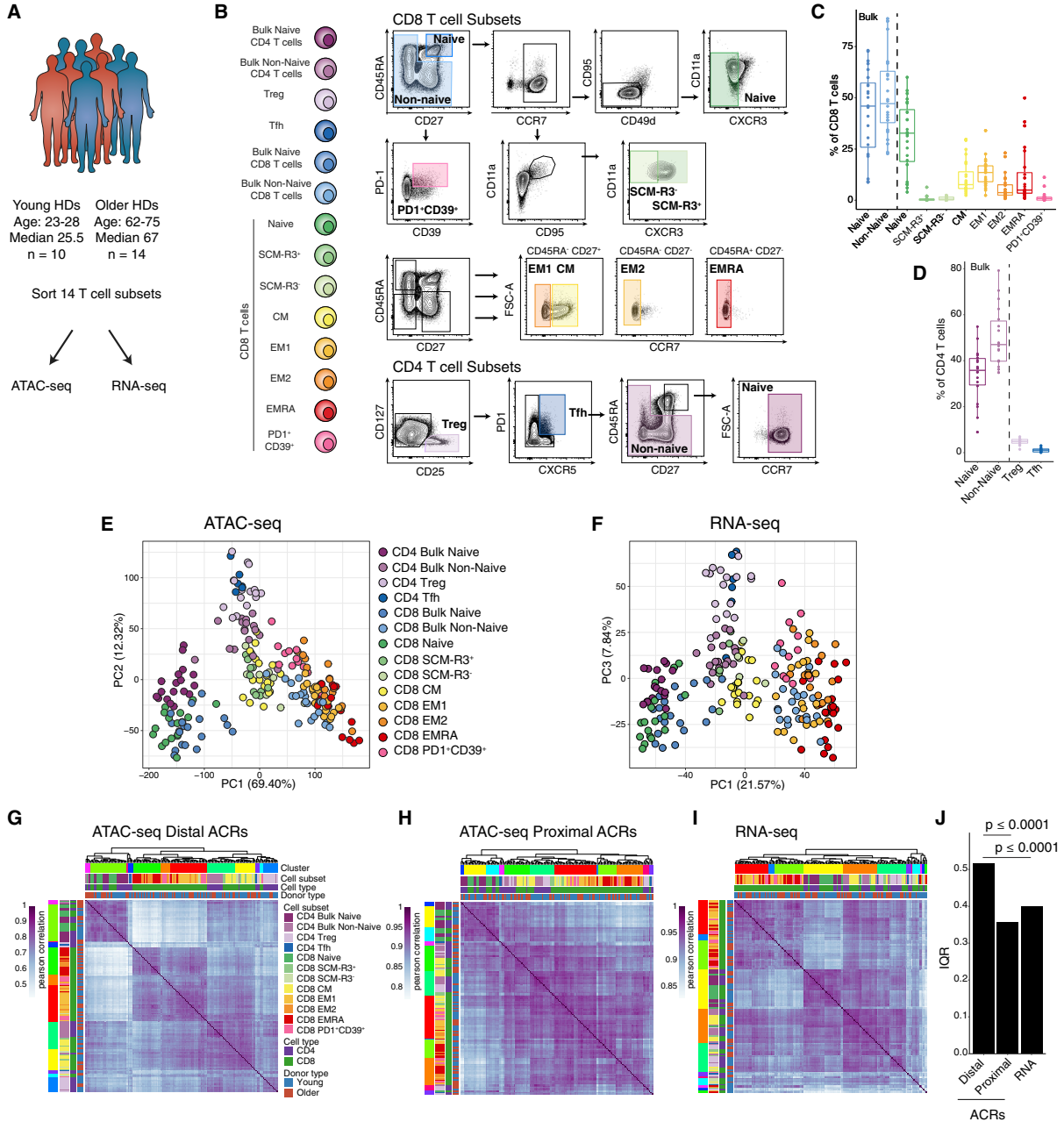


Figure 1. Human T cell transcriptional and epigenetic landscape

(A) Experimental schematic.

(B) Sorting strategy. Cells were gated as live singlets, then CD8$^+$, or CD4$^+$.

(C and D) Enumeration of cells gated in (B) for CD8 T cells (C) and CD4 T cells (D).

(E and F) PCA of ATAC-seq (E) and RNA-seq (F).

(G–I) Sample-to-sample Pearson correlation and hierarchical clustering using (G) ATAC-seq distal ACRs (≥ 2kb from the nearest TSS), (H) ATAC-seq proximal ACRs (≤ 2kb from the nearest TSS), or (I) RNA-seq.

(J) Information quality ratio (IQR) comparing sorted subset label with cluster label. p values are determined by calculating null distribution. See also Figure S1.

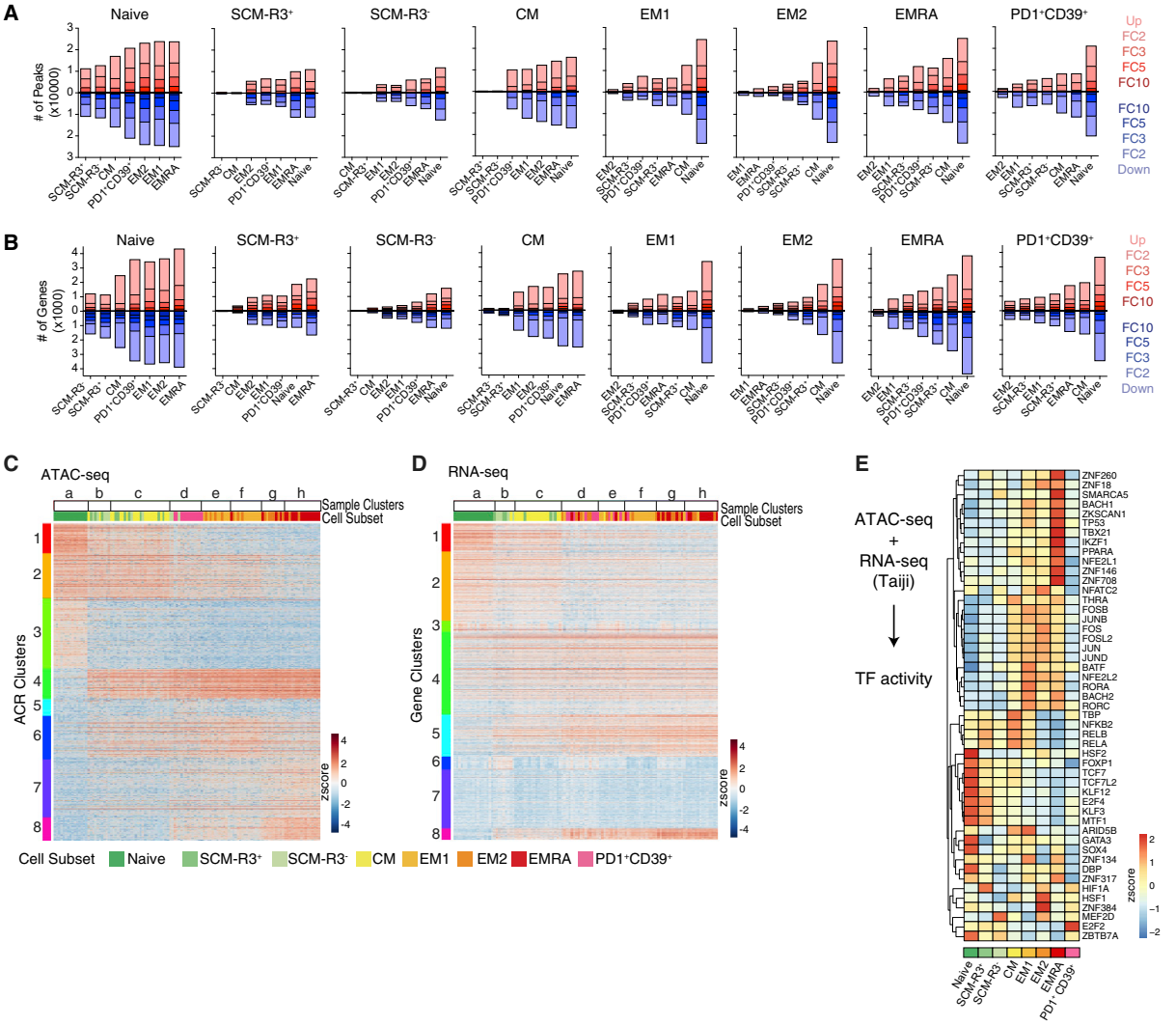


Figure 2. Global analyses reveal overlapping epigenetic regulation and gene expression in CD8 T cell subsets

(A and B) Pairwise comparison of CD8 subsets: ACRs (A) or genes (B) increased and decreased, colored by fold change as indicated.

(C and D) Sample (column) and ACR (row) clustering of ATAC-seq (C) and RNA-seq (D), with column scaling.

(E) TF score per cell subset calculated by integrating ATAC-seq and RNA-seq.

versus older subjects (Figure S1D). Bulk naive cells had the greatest number of DEGs, but few changes were observed comparing the more stringently defined naive subset in young versus older subjects. In the bulk naive comparison, activation-related genes, such as *TBX21*, *PRF1*, and *GZMA*, were increased in older HDs (Figure S1E). These transcriptional differences likely reflect more antigen-experienced cells in this less rigorously defined population. Although larger numbers of samples could reveal subtle age-related changes in each T cell subset, these data suggest that major age-related changes in human T cell biology reflect different proportions of circulating T cell subpopulations rather than large age-dependent transcriptional differences within specific T cell subsets.

Defining transcriptional and epigenetic regulatory landscape of human CD8 T cell differentiation

We first applied this HD epigenomic atlas to investigate the molecular programs of human CD8 T differentiation states. We identified differentially accessible peaks (DAPs) and DEGs in all pairwise comparisons (Figures 2A and 2B; Tables S2 and S3). Naive CD8 T cells had an equivalent number of DEGs and DAPs that were increased or decreased, suggesting that naive T cells are not simply quiescent but actively maintain the naive state through specific transcriptional and epigenetic programs. There were few transcriptional or epigenetic differences between SCM-R3$^+$, SCM-R3$^-$, and CM, indicating that these subsets were relatively similar in HDs. There were also few individual differences between EM1 and EM2, although EM1 had

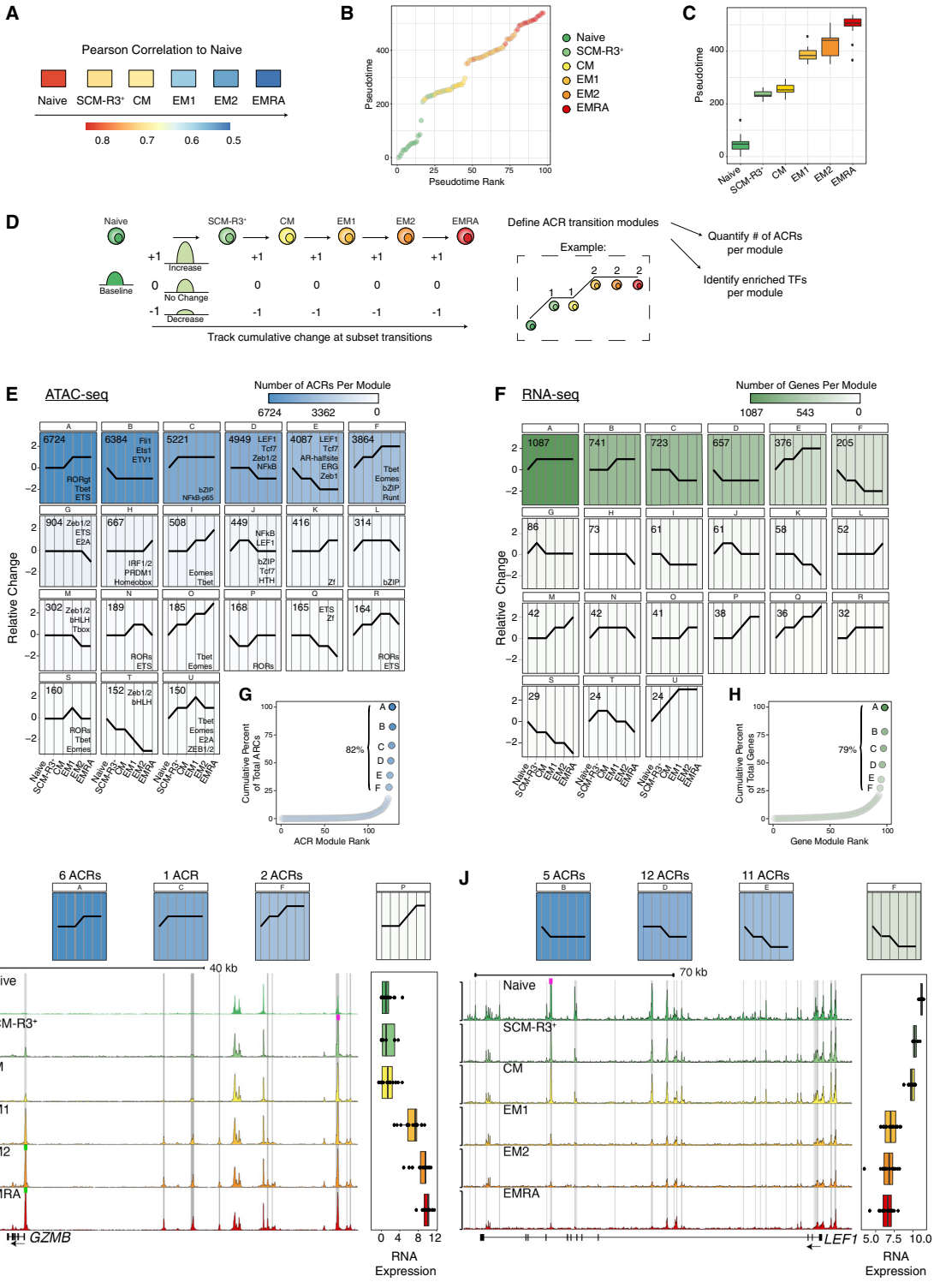


Figure 3. Trajectory analysis indicates two major inflection points in the molecular control of human CD8 T cell differentiation

(A) Median Pearson correlation of each subset to naive.

(B and C) Pseudotime value calculated by Slingshot of each sample (B) and summarized by subset (C).

(legend continued on next page)

higher expression of genes such as *CD28* and *TCF7*, whereas EM2 upregulated *ZEB2* and *GZMB*. Many of the differences between EM2 and EMRA were killer cell family genes (*KLRC2*, *KLRF1*, *KIR3DL1*, etc.), indicating that EMRA T cells express genes for cytotoxic pathways associated with NK cells. Similar genes have been reported in tumor-infiltrating T cells in glioblastoma (Mathewson et al., 2021), but these data suggested that the expression of these genes is not disease-specific but reflects effector programs used in HD CD8 T cells. $PD1^{+}CD39^{+}$ cells were most similar to EM2 but had higher expression of genes involved in progenitor biology (*LEF1*), cell cycle (*TOP2A* and *CDC7A*), and exhaustion (*TOX* and *TOX2*), consistent with Tex (McLane et al., 2019).

Next, we used an unbiased global approach to determine the overall structure of chromatin accessibility data for these purified human CD8 T cell populations. We used bi-clustering to simultaneously cluster ACRs (rows) and samples (columns). This method revealed the local patterns of chromatin accessibility across sample groups (Figure 2C). Most sample clusters (*a–h*; horizontal axis) contained samples from the same T cell subset (IQR analysis, $p < 0.00005$). Most ACR clusters (1–8; vertical axis) had a gradient of chromatin accessibility across sample clusters (*a–h*) and exhibited two primary patterns (Figure 2C). One pattern contained ACR clusters (1–3) that were most accessible in sample cluster *a* (containing naive cells), decreased in accessibility in sample clusters *b–g* (containing memory, exhaustion, and effector memory CD8 T cells, respectively), and were least accessible in sample cluster *h* (containing EMRA cells). The second major pattern, which included ACR clusters 4–8, was the opposite. This unbiased clustering revealed unique combinations of ACRs which could distinguish even closely related CD8 T cell subsets, indicating that subset-specific cellular and functional properties were driven by different combinatorial sets of regulatory chromatin regions.

We then applied the same bi-clustering method to our RNA-seq data. Consistent with the IQR analysis in Figure 1, sample clusters generated with transcriptional information contained a greater diversity of sorted subsets (Figure 2D) compared with the ATAC-seq data (Figure 2C) (RNA IQR = 0.410, ATAC IQR = 0.526, $p < 0.0001$). As with the ATAC-seq analysis, we found that most gene clusters (1–8) were expressed by multiple sample clusters (*a–h*). Thus, applying the patterns from this epigenomic atlas for different subtypes of CD8 T cells should allow the identification of associated biological features in T cell populations even when the surface phenotype or transcriptional signatures of cell type identity are less clear.

We next integrated gene expression and chromatin accessibility information to infer TF activity in each CD8 T cell subset (Figure 2E). Similar to the pattern of individual ACRs and genes, many TFs were used by multiple CD8 T cell subsets. This analysis revealed both known and novel TFs for each differentiation state. For example, GATA3 and FOXP1 were most active in naive T cells and have been reported to preserve naive T cell quiescence in mice (Wang et al., 2013; Wei et al., 2016). We also found several novel TFs with the predicted activity in naive T cells, including SOX4, KLF3, KLF12, and HSF2. CM was predicted to heavily use NFκB family members; EM1 ranked the highest for RORA and RORC; bZIP family TFs were predicted to have high activity across non-naive subsets. EMRA had high scores for TBX21, IKZF1, TP53, and BACH1, along with several zinc finger TFs not previously reported in T cells. $PD1^{+}CD39^{+}$ had the highest score for E2F2, whereas the closely related E2F4 was the highest in naive. E2F4 is a transcriptional repressor required to engage and maintain cell cycle arrest in G0/G1 (Trimarchi and Lees, 2002). In contrast, E2F2 triggers entry into cell cycle. Divergent activity scores in naive and $PD1^{+}CD39^{+}$ cells suggest that these cells are at opposite ends of the proliferative spectrum. Thus, by integrating epigenetic and transcriptional data, we identified known and novel TFs predicted to regulate different stages of human CD8 T cell differentiation.

Molecular trajectories in human CD8 T cell differentiation identify coordinated epigenetic and transcriptional control

We next investigated how gene expression and chromatin accessibility changed along a putative differentiation trajectory. Since multiple models of human CD8 T cell differentiation have been proposed (Ahmed et al., 2009; Restifo and Gattinoni, 2013), we took a data-driven approach to order the CD8 T cell subsets, first by using the Pearson correlation of each subset relative to the naive subset based on distal chromatin accessibility. This approach revealed a predicted developmental relationship: naive → $SCM\text{-}R3^{+}$ → CM → EM1 → EM2 → EMRA (Figure 3A), consistent with PC1 above (Figure 1) and a previously proposed model (Restifo and Gattinoni, 2013). A second approach using pseudotime resulted in the same trajectory (Figures 3B and 3C) although other models, including those with branches, cannot be ruled out. Indeed, we excluded $SCM\text{-}R3^{-}$ and $PD1^{+}CD39^{+}$ because these subsets may represent alternative differentiation branches (Pauken et al., 2016; Chen et al., 2019; Yao et al., 2016). Using this proposed order, the relative change of each ACR or gene was determined between each subset along the trajectory; ACRs or genes that followed the same pattern were considered a module (Figure 3D).

Analyzing the relative cumulative change of ACRs and genes along this trajectory revealed two major epigenetic and transcriptional inflection points: (1) between naive and SCM-R3+ and (2) between CM and EM1 (Figures 3E and 3F). These inflection points were also apparent in the pseudotime analysis (Figures 3B and 3C). Approximately, 82% of ACRs and 79% of genes that changed were captured in the first six modules that had a change at one or both of these points (Figures 3G and 3H). This transcriptional and epigenetic regulation in human CD8 T cell differentiation was exemplified by *GZMB*, a key

(D–F) (D) Analysis schematic. Relative cumulative change determined from differential analysis of ACRs (E) or genes (F) that follow the same pattern. The number of genes or ACRs per module is indicated on the plot and reflected by color as indicated by heat scale. Top enriched transcription factor (TF) motifs or TF families in ACR modules are indicated in (E).
(G and H) Cumulative percent of total changed ACRs (G) or genes (H) in each module.
(I and J) ATAC-seq signal tracks and RNA expression for (I) *GZMB* and (J) *LEF1*. The number of ACRs per module associated with *GZMB* or *LEF1* is summarized on top. RNA module of *GZMB* and *LEF1* is indicated on top.

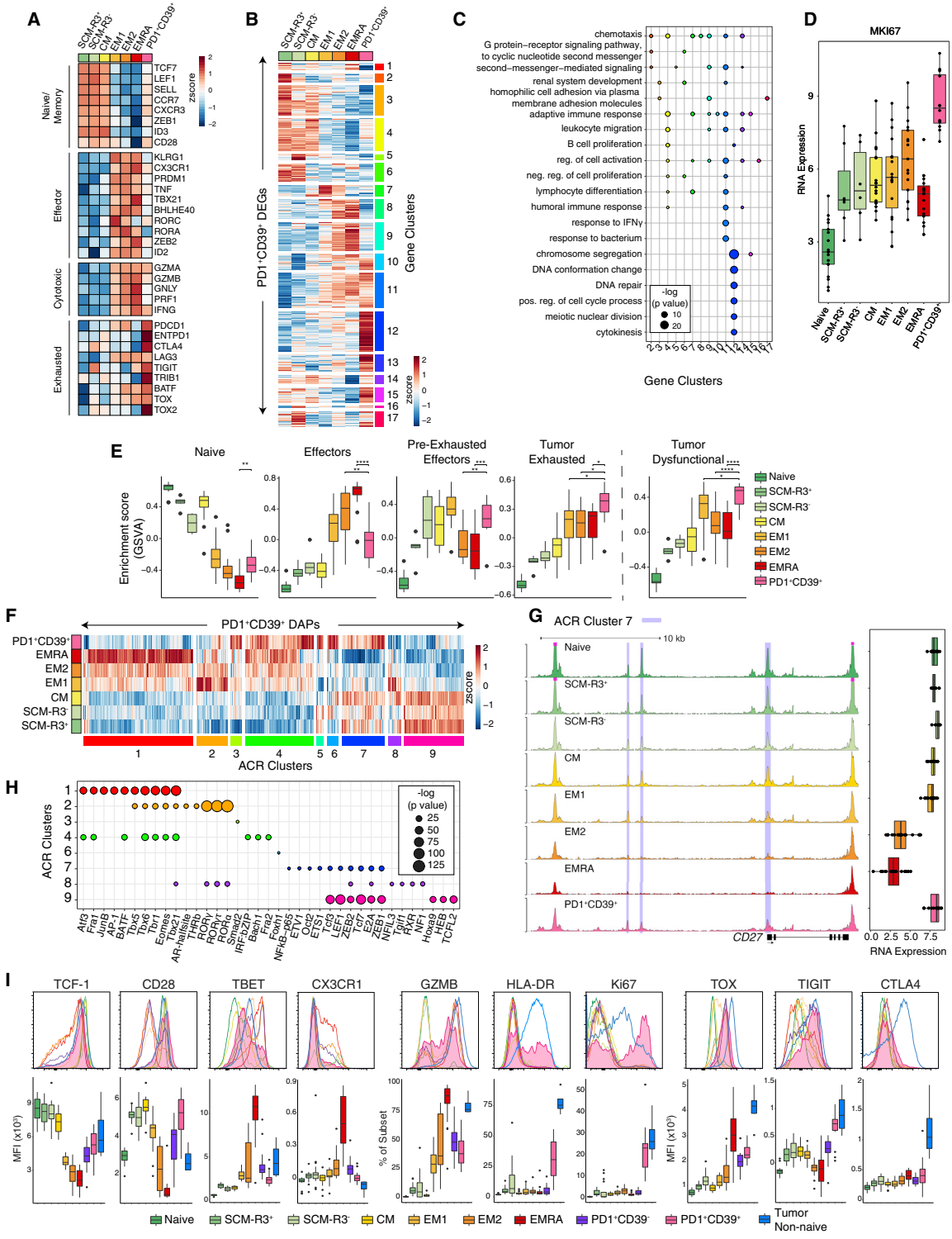


Figure 4. PD1+CD39⁺ CD8 T cells in the blood of healthy donors have a program of exhausted CD8 T cells

(A) RNA expression of T cell genes.

(B) Hierarchical clustering of DEGs from pairwise comparisons to the PD1⁺CD39⁺ subset. Gene clusters are indicated by the number and color block.

(legend continued on next page)

effector gene, and *LEF1*, a critical naive and memory TF. The *GZMB* locus contained ACRs that increased in accessibility at both inflection points (Figure 3I). All 28 ACRs in the *LEF1* locus decreased in accessibility between naive and SCM-R3$^+$ and from CM to EM1 (Figure 3J). Defining ACR modules revealed potential regulatory TFs (Figure 3E), including a central role for NFκB family TFs in SCM-R3$^+$ and CM CD8 T cell subsets and ROR TFs in EM subsets. The Fli1 motif was enriched in ACRs that lose accessibility in the naive to SCM-R3$^+$ transition confirming in humans the role of Fli1 in restraining effector CD8 T cell differentiation discovered in mice (Chen et al., 2021). Thus, these analyses identified a putative differentiation trajectory of human CD8 T cell populations and defined transcriptional and epigenetic modules that may regulate differentiation and/or subset-specific biologic functions. Moreover, we uncovered two major inflection points in gene expression and chromatin accessibility and identified TFs that may have key roles in transitions between CD8 T cell subsets.

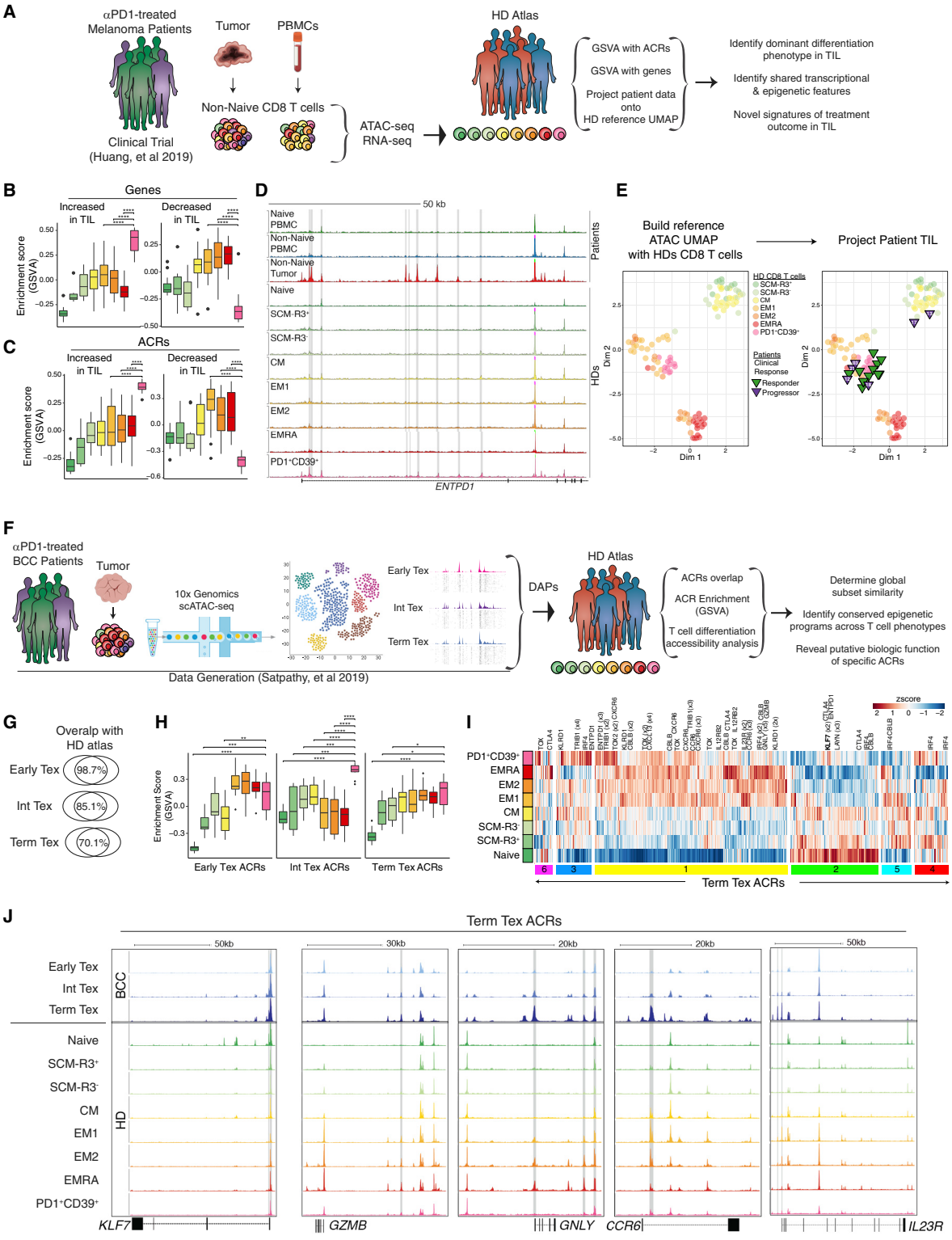
PD1$^+$CD39$^+$ CD8 T cells from the blood of HDs have characteristics of Tex CD8 T cells

One goal of this HD atlas is the investigation of the underlying molecular framework of disease-relevant CD8 T cell subsets in healthy subjects. For example, although Tex have been described in chronic infections and cancer (McLane et al., 2019; Thommen and Schumacher, 2018), their molecular regulation in humans remains poorly understood. To test whether PD1$^+$CD39$^+$ from HD had features of exhaustion, we first examined the expression of key naive/memory-, effector-, cytotoxic-, or exhaustion-associated genes (Figure 4A). PD1$^+$CD39$^+$ had increased expression of genes encoding inhibitory receptors as well as exhaustion-associated genes, such as *TRIB1* (Rome et al., 2020) and *TOX* (Alfei et al., 2019; Khan et al., 2019; Scott et al., 2019; Seo et al., 2019; Yao et al., 2016). The PD1$^+$CD39$^+$ subset also had intermediate expression of naive, memory, and effector genes, such as *TCF7*, *CCR7*, *TBX21*, and *GZMB* (Figure 4A), consistent with the data from Tex in mice (McLane et al., 2019). We further investigated the relationship between PD1$^+$CD39$^+$ and the other human CD8 T cell subsets using unbiased clustering (Figure 4B). Cluster 12 was specific to PD1$^+$CD39$^+$ and contained genes related to cell cycle, including *MKI67* (Figures 4B–4D). Data from mice demonstrate extensive proliferation of progenitor and/or intermediate Tex, giving rise to post-mitotic yet transiently Ki67$^+$ terminal Tex (Beltra et al., 2020; Paley et al., 2012; Blackburn et al., 2008; Im et al., 2016; Utzschneider et al., 2016; Wu et al., 2016). The cycling Tex subset in mice can be found in circulation and may be analogous to PD1$^+$CD39$^+$ CD8 T cells in the HD blood. This observation is consistent with recent studies in human cancer, including the identification of proliferating progenitor Tex in melanoma (Li et al., 2018). To directly compare PD1$^+$CD39$^+$ CD8 T cells from the blood of HD with human tumor infiltrating lymphocytes (TIL), we used existing single-cell RNA-seq data (Li et al., 2018; Guo et al., 2018). Of all subsets in the HD blood, PD1$^+$CD39$^+$ cells had the highest enrichment for tumor Tex gene sets (Figure 4E). Thus, PD1$^+$CD39$^+$ CD8 T cells from the blood of HDs have a distinct transcriptional signature, including the enrichment for exhaustion-associated genes, and transcriptional evidence of recent proliferation—characteristics of an intermediate Tex population.

We next analyzed the chromatin landscape of PD1$^+$CD39$^+$ CD8 T cells (Figure 4F). Three ACR clusters (clusters 3, 5, and 6) were most accessible in PD1$^+$CD39$^+$ CD8 T cells; cluster 4 was shared between EM1, EM2, EMRA, and PD1$^+$CD39$^+$, but there was a striking lack of overlap for the PD1$^+$CD39$^+$ subset with the ACR cluster most accessible in EMRA (cluster 1). On the other hand, cluster 7 was accessible in SCM and CM as well as in PD1$^+$CD39$^+$ cells. Cluster 7 may represent a shared epigenetic program of durability or survival. Indeed, one gene with several ACRs contained in cluster 7 was *CD27* (Figure 4G), a costimulatory receptor associated with T cell activation, differentiation, and survival (Croft, 2014). Collectively, chromatin regions accessible in PD1$^+$CD39$^+$ were enriched in T-box motifs, a subset of AP-1 family members, as well as NFκB, ETS, ZEB1/2, E2A, TCF7, and LEF1 motifs, a combination that suggests both longevity and activation (Figure 4H). Therefore, the epigenetic landscape of PD1$^+$CD39$^+$ T cells in the blood of HDs exhibits a unique combination of ACRs including elements shared with long-lived subsets (SCM-R3$^+$, SCM-R3$^-$, CM) and distinct from terminal effectors (EMRA).

To examine whether these transcriptional and epigenetic features of PD1$^+$CD39$^+$ CD8 T cells translated to protein expression, we analyzed the expression of key molecules by flow cytometry and used non-naive CD8 T cells from melanoma tumors as a control (Figure 4I). PD1$^+$CD39$^-$ CD8 T cells were also included as a representative of recently activated T cells. Consistent with gene expression data, PD1$^+$CD39$^+$ CD8 T cells from HD blood expressed intermediate TCF1 and TBET, high CD28, and moderate GZMB, TIGIT, CTLA4, and TOX. TOX was also expressed by EMRA and PD1$^+$CD39$^-$ CD8 T cells (Figure 4I), in agreement with recent studies in humans and mice (Sekine et al., 2020; Khan et al., 2019) and consistent with a potential role for this TF in repetitively stimulated CD8 T cells as well as Tex. Furthermore, PD1$^+$CD39$^+$ CD8 T cells had the highest frequency of Ki67$^+$ cells among CD8 T cell subsets in the blood, and some of these cells expressed HLA-DR (Figure 4I). To

(C) Gene ontology of gene clusters in (B).
(D) RNA expression of *MKI67*.
(E) Gene set variation analysis (GSVA) for gene sets indicated on top. Gene sets to the left of the dotted line are derived from scRNA-seq from patients with non-small cell lung cancer (Guo et al., 2018); the tumor dysfunctional set on the right side is derived from patients with melanoma tumors (Li et al., 2018). p values are from two-tailed *t*-tests with Benjamini-Hochberg correction comparing PD1$^+$CD39$^+$ with EM1, EM2, and EMRA: *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.
(F) Hierarchical clustering of DAPs from pairwise comparisons to PD1$^+$CD39$^+$ samples.
(G) ATAC-seq signal tracks and RNA expression of *CD27*. Highlighted ACRs are in cluster 7 in (F).
(H) TF motif enrichment in ACR clusters in (F). No TF motifs were significantly enriched in cluster 5.
(I) Flow cytometry analysis. Cells were first gated as live CD8$^+$ singlets. Tumor samples were gated on live single non-naive CD8 T cells defined by excluding CD45RA$^+$CD27$^+$ cells. Top, representative histograms from 18 HDs and 4–5 melanoma tumors. Bottom, median fluorescence intensity (MFI) or percent positive as indicated. See also Figure S2.



(legend on next page)

determine whether PD1$^+$CD39$^+$ CD8 T cells were functionally exhausted, we measured cytokine production after T cell receptor (TCR) stimulation *in vitro*. Few PD1$^+$CD39$^+$ CD8 T cells expressed IFNγ or TNF, in contrast to the robust production of these effector molecules by EM1, EM2, and EMRA subsets, and PD1$^+$CD39$^-$ T cells (Figures S2A and S4B). The few PD1$^+$CD39$^+$ CD8 T cells that were IFNγ^+ or TNF$^+$ had a significantly lower MFI, indicating lower protein production (Figure S2C). These data show that PD1$^+$CD39$^+$ CD8 T cells in the blood of HDs have transcriptional, epigenetic, and protein features consistent with Tex and are distinct from classical memory and effector memory CD8 T cell subsets.

Application of the HD T cell atlas reveals biologic patterns and Tex-specific ACRs in TILs

One major application of this epigenomic atlas is to provide a framework for analyzing independent datasets. T cell RNA-seq and/or ATAC-seq data from disease settings can be "projected" onto the HD atlas to define cellular states and provide insights into the underlying biology. To test this idea, we used two cancer datasets and investigated whether new insights could be gained.

First, we used an anti-PD1 neoadjuvant/adjuvant melanoma dataset (Huang et al., 2019) to determine whether the HD epigenomic atlas could correctly identify predominant TIL differentiation state(s) (Figure 5A). We sorted non-naive CD8 T cells from tumor and PMBCs after treatment and performed RNA-seq and ATAC-seq (Table S1). We first compared genes that were increased or decreased in the tumor with the transcriptional profiles of each HD subset. Consistent with the results in Figure 4, blood-derived PD1$^+$CD39$^+$ T cells from the HD atlas had the greatest enrichment of TIL-specific genes compared to other HD subsets (Figure 5B). Comparing the chromatin regions that were differentially accessible in CD8 TIL revealed that the epigenetic landscape of CD8 TILs was also most similar to that of PD1$^+$CD39$^+$ CD8 T cells (Figure 5C). For example, prominent ACRs in the *ENTPD1* locus (encoding CD39) in TILs were found only in PD1$^+$CD39$^+$ CD8 T cells in the HD atlas (Figure 5D). These results indicate that TILs from post anti-PD-1-treated human melanoma tumors not only are transcriptionally similar to PD1$^+$CD39$^+$ Tex in HD blood but also share epigenetic features including ACRs that may control key genes including inhibitory receptors.

We next sought to define the predominant differentiation state of CD8 TIL from each individual patient. We built a Unifold Manifold Approximation Projection (UMAP) from the HD non-naive CD8 T cell subset atlas using distal ACRs and projected the ATAC-seq data from melanoma CD8 TILs into this UMAP (Figure 5E). Assigning each TIL sample to the nearest HD subset centroid revealed a significant number of TIL samples (16/18; p value = 0.001) that overlapped with the region populated by PD1$^+$CD39$^+$ CD8 T cells from the HD atlas (Figure 5E), identifying melanoma TIL as Tex Tex (Huang et al., 2017, 2019; Kamphorst et al., 2017; Li et al., 2018). Although TIL may exhibit some effector-like phenotypic properties (e.g., expression of cytotoxic molecules), this projection of TIL signatures into the HD atlas pointed to the distinction between more effector-like HD subsets (e.g., EMRA and EM2) and PD1$^+$CD39$^+$ Tex cells.

Notably, TILs from two patients (13 and 17) who experienced tumor recurrence mapped closest to SCM and CM HD subsets in HD UMAP space, prompting us to further compare all clinical responder and non-responder patients. This analysis revealed 37 differentially expressed genes in 5 of 6 non-responders including an increase in quiescence-associated genes *CCR7* and *LEF1* (Figure S3A). Patients 13 and 17 had the highest expression of these two genes. We then tested for potential signatures of clinical response versus non-response. Two signatures that emerged were cytotoxicity, consistent with previous studies (Rooney et al., 2015; Fridman et al., 2012), and WNT signaling. WNT signaling has a negative association with T cell inflammation in melanoma and other cancers (Spranger et al., 2015; Luke et al., 2019; Li et al., 2019), and, in mice, WNT signaling can block effector T cell differentiation through LEF1 and TCF7 activity (Gattinoni et al., 2009). Four out of six progressors had decreased cytotoxicity signatures, including patients 13 and 17, and 5 of 6 progressors had increased signatures of WNT signaling (Figure S3B). Previous clinical analysis of this cohort found a strong association between positive clinical response and pathologic assessment of "brisk" immune cell infiltration. Although there was no association between the cytotoxicity signature and brisk infiltration, patients with tumors that were not brisk (i.e., had few infiltrating immune cells) had transcriptional evidence of increased WNT signaling in TILs (Figure S3C). Thus, the HD epigenomic atlas demonstrated that TILs were similar to PD1$^+$CD39$^+$ Tex. This approach also identified cytotoxic and WNT signaling signatures associated with a positive or negative clinical response, respectively.

We next examined a scATAC-seq dataset of CD8 TILs isolated from anti-PD1-treated BCC (Satpathy et al., 2019; Figure 5F; Data S1). Three clusters of Tex CD8 T cells were previously defined in the BCC TIL: early, intermediate, and terminal Tex. It

Figure 5. Application of HD T cell atlas identifies cellular phenotypes and conserved T cell epigenetic programs in CD8 TIL

(A) Analysis schematic.

(B and C) Single sample enrichment (GSVA) of (B) ACRs or (C) genes increased or decreased by fold change > 2 in tumors compared to PBMCs in non-naive CD8 T cells after treatment.

(D) ATAC-seq signal tracks of *ENTPD1*.

(E) UMAP created using top 2k distal differentially accessible peaks (DAPs) from pairwise comparisons of included HD CD8 T cell subsets; tumor non-naive CD8 T cells of patients with melanoma overlaid.

(F) Analysis schematic.

(G) Percent overlap of ACRs identified in scATAC-seq from BCC TIL Tex DAPs compared to the HD epigenetic atlas.

(H) ACR set enrichment (GSVA) of Tex BCC TIL ACR sets calculated for each HD subset. p values are from two-tailed *t*-tests with Benjamini-Hochberg correction comparing PD1$^+$CD39$^+$ with every other subset: * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001.

(I) Median accessibility and hierarchical clustering of the terminal Tex BCC TIL ACRs in HD subsets.

(J) ATAC-seq signal tracks of indicated genes. Tex BCC-specific ACRs are highlighted in gray. Tracks are group scaled per dataset. See also Figure S3.

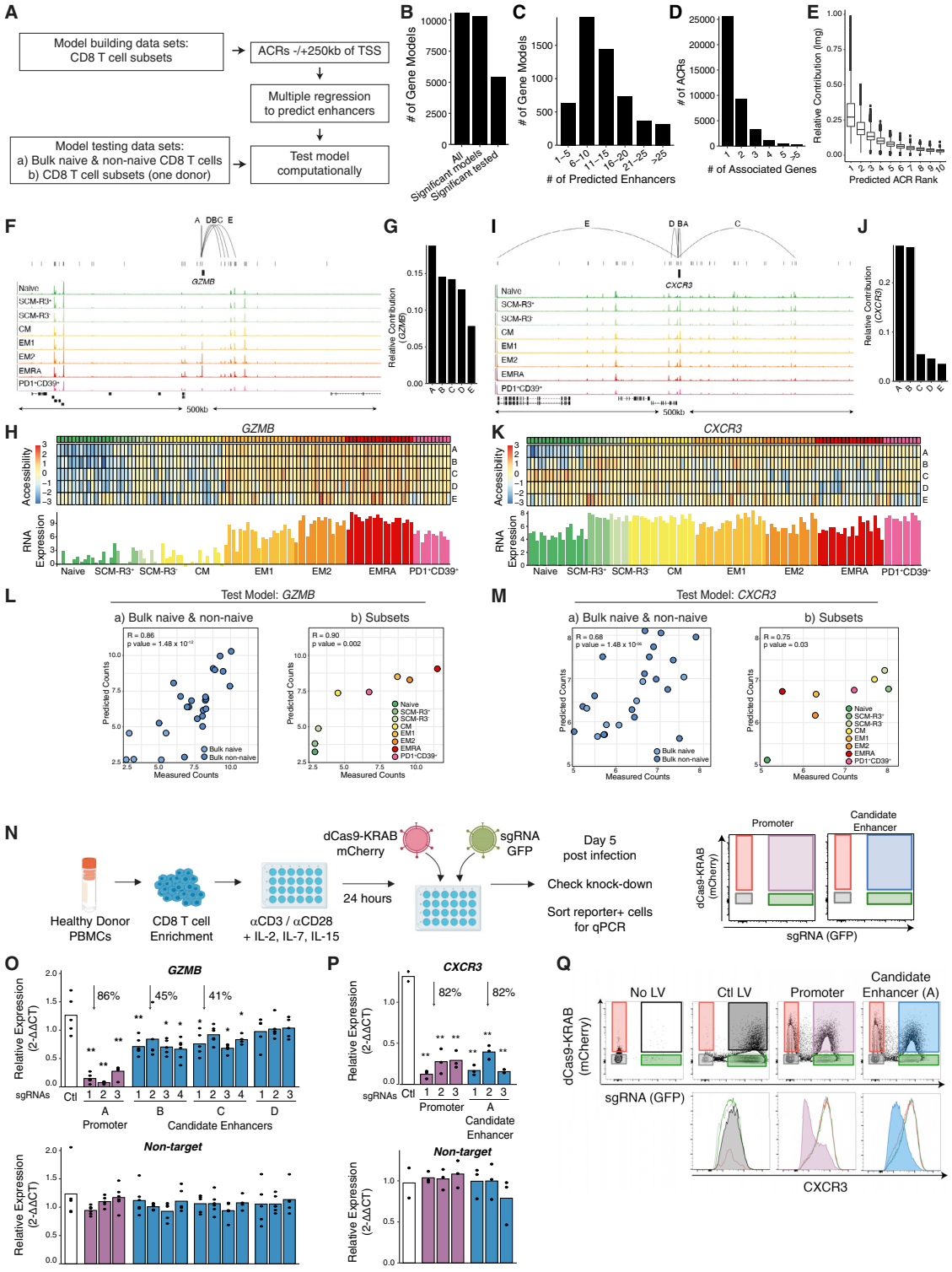


Figure 6. Predicting *cis* gene regulatory elements in human CD8 T cells

(A) Analysis schematic. Multiple regression was used to predict *cis*-regulatory elements that control gene expression.

(B) Summary of gene models: number of genes tested, significant gene models, and those that produced significant correlation in held-out testing datasets.

(legend continued on next page)

is unclear if these scATAC-seq Tex profiles are also found in HD CD8 T cells. To investigate this question, we identified DAPs from these 3 Tex scATAC-seq clusters and compared them to those of the HD CD8 T cell atlas. First, we determined what fraction of ACRs from each BCC subset overlapped with that in the HD atlas. Most ACRs from the early Tex cluster (98.7%) were also present in the HD data; 85.1% of the intermediate Tex ACR overlapped, whereas the BCC terminal Tex cluster had 70.1% overlap (Figure 5G). Global comparison revealed the enrichment of the BCC Tex intermediate ACRs in the HD $PD1^+CD39^+$ subset (Figure 5H), consistent with $PD1^+CD39^+$ CD8 T cells from the HD blood being analogous to intermediate Tex population discovered in mice (Beltra et al., 2020; Zander et al., 2019; Hudson et al., 2019). Unlike BCC intermediate Tex, ACRs from BCC terminal Tex had more broad overlap with HD atlas populations (Figure 5H), indicating that many of these Tex ACRs were not exhaustion specific but also used in other CD8 T cells from HDs. To deconvolve the mixture of non-coding elements, we projected the ACRs from BCC terminal Tex across the HD CD8 T cell subsets (Figure 5I). This analysis identified the subsets of terminal Tex ACRs that were accessible in naive cells (cluster 2); EM1, EM2, and EMRA (cluster 1); or the $PD1^+CD39^+$ HD subset (cluster 3)—revealing a composite picture of the underlying biology in BCC-infiltrated CD8 T cells. ACRs in cluster 3 contained four ACRs near *TRIB1*, consistent with a role for TRIB1 in Tex differentiation in mice (Rome et al., 2020) and expression of this gene in HD $PD1^+CD39^+$ CD8 T cells (Figure 4A). Cluster 1, accessible in effector-like HD subsets, included several ACRs located near effector-related genes, such as *GZMB, GNLY, CCR6*, and *IL23R* (Figures 5I and 5J). This cluster may represent non-coding regulatory regions used by effector and Tex. Indeed, studies in mice have demonstrated effector-like properties within the mouse equivalent of the terminal Tex subset (Beltra et al., 2020). ACRs from BCC terminal Tex that were accessible in HD naive CD8 T cells (cluster 2), such as the *KLF7* locus (Figures 5K and 5L), may reflect new T cell priming as reported in this BCC immunotherapy setting (Yost et al., 2019) or point to regulatory programs that contribute to active repression of effector mechanisms used in both naive CD8 T cells and Tex, such as *LAYN* (Zheng et al., 2017). Thus, application of the HD atlas to this BCC dataset not only confirmed the ability of this HD CD8 T cell atlas to correctly identify known biology, including the similarity of the BCC Tex intermediate population with the HD $PD1^+CD39^+$ subset, but also provided a resource to deconvolve overlapping modules of ACRs used by BCC terminal Tex.

Identifying functional enhancers in human CD8 T cells and application of the HD T cell ACR atlas to association with immune disease-associated genetic variants

We next integrated the transcriptional and epigenetic data from the HD T cell atlas to predict the function of non-coding elements. We applied a machine learning approach to predict which ACRs function as *cis*-regulatory elements in controlling gene expression across CD8 T cell subsets. The cohort was divided into two sets: one set to build gene regulation models and a second set to test these models (Figure 6A). Using multiple regression, we identified ACRs within 250 kb upstream and downstream of the TSS that best explained gene expression and then determined the relative contribution of each predicted *cis*-element (Figure 6A). We generated models for 10,338 genes and then tested how well these models predicted gene expression with the held-out dataset. Approximately, half of all models predicted gene expression that significantly correlated with measured gene expression (Figure 6B). Most genes were predicted to have 6–10 enhancers that contributed to their regulation (Figure 6C), and most ACRs were predicted to regulate only 1 gene (Figure 6D). The top predicted regulatory region contributed an average of 25% to gene expression (Figure 6E). However, distinct patterns could be observed for individual genes. For example, regulation of *GZMB* expression was distributed across several ACRs, suggesting additive and complex control that would allow fine tuning of expression across CD8 T cells subsets (Figures 6F–6H). In contrast, for *CXCR3*, the top two ACRs were predicted to control over 50% of gene expression variance (Figures 6I–6K). Gene regulation models for genes of potential translational interest, such as *LEF1*, *IFNG*, *PRF1*, and *PDCD1*, are shown in Figures S4A–S4P. Thus, by integrating transcriptional and epigenetic information, we identified specific ACRs controlling expression of individual genes in human CD8 T cell subsets. These models constructed from HD T cell data accurately predicted expression patterns of *LEF1*, *IFNG*, *PRF1*, and *PDCD1* in the blood and CD8 TILs of patients with melanoma (Figure S4Q). These results demonstrated that gene regulation models built from the HD T cell atlas can be used to predict gene expression patterns in an unrelated dataset from a disease setting.

The data described above provided a unique opportunity to ask whether these predicted *cis*-regulatory regions have relationships to genetic associations in human disease. Most disease-associated SNPs reside in non-protein-coding regions, and the mechanisms by which these SNPs result in biologic consequence often remain poorly defined. We therefore asked whether

(C) The number of predicted enhancers per gene.
(D) The number of genes associated with each ACR.
(E) Relative importance of predicted enhancers by rank.
(F–M) Model building and testing results shown for two genes: *GZMB* and *CXCR3*. (F and I) ATAC-seq signal tracks of genomic test regions: top five predicted enhancers are noted on top. (G and J) Relative contribution of top five predicted enhancers. (H and K) Chromatin accessibility of top ACRs and gene expression for each sample (column represents one donor). (L and M) Predicted compared to measured gene expression from two testing datasets as indicated. Pearson correlation as shown.
(N) Experimental strategy using CRISPRi to validate candidate enhancers. Relative expression determined by quantitative PCR from sorted cells for *GZMB* (O) and *CXCR3* (P) or non-target gene. Each bar represents a different guide. Significance for each guide is determined by two-tailed *t*-test: *$p < 0.05$, **$p < 0.01$. Average percentage decrease across significant individual guides to a single target compared to control is shown. Bar represents mean of 3–5 independent experiments shown as points.
(Q) Flow cytometry as indicated with reporter expression (dCas9-KRAB-mCherry and sgRNA-GFP) in the top row and protein of target gene, *CXCR3*, in the bottom row. See also Figures S4 and S5.

cis-regulatory ACRs identified in diverse human HD CD8 T cell subsets could help provide insights into immune-related GWAS SNPs. GWAS SNPs are usually defined using SNP arrays, and most reported SNPs are sentinel SNPs that indicate a causal SNP within a certain linkage disequilibrium (LD) region. We therefore investigated a 25kb window around each GWAS SNP (Zhu et al., 2004) and found 2,997 SNP windows overlapping with ACRs in the HD T cell atlas (Figure S5A). In some settings, SNPs are found in cell type-specific regions of open chromatin (Farh et al., 2015), for example, asthma-associated SNPs in Th2-specific enhancer elements (Seumois et al., 2014). Indeed, the HD atlas revealed T cell subset-biased patterns of ACR-associated SNPs that may implicate distinct subsets in different immune-related diseases. While non-naive CD4 T cells had broad potential involvement, we identified disease-biased associations particularly of EM1, EM2, and PD1$^+$CD39$^+$ CD8 T cells (Figure S5B). Next, we searched within the immune disease-associated SNP windows for ACRs with predicted regulatory function and found 2,896 such ACRs. For example, the top predicted ACR for controlling *IFNG* expression pattern was located within a genomic window that overlapped with several SNPs associated with ulcerative colitis (UC), psoriasis, and ankylosing spondylitis (AS) (Figure S5C). This result suggested that the causal SNP(s) responsible for these GWAS-identified SNPs were located in or near an enhancer that positively regulates *IFNG* in more effector-like CD8 T cells (EM1, EM2, and EMRA). Furthermore, the *IL23R*/*IL12RB2* locus contained 24 SNPs associated with more than 12 autoimmune diseases. The top four ACRs that collectively control 88% of *IL23R* gene expression in CD8 T cells reside within these SNP regions, and one ACR, highlighted in red, was also predicted to control *IL12RB2* gene expression (Figure S5D). The IL-12/23 pathway regulates the differentiation of Th17 CD4 T cells and IL-17 production in CD8 T cells; IL-17 has a pathogenic role in several mouse models of autoimmune diseases. SNPs within these predicted regulatory elements may alter the expression level of IL23R and ultimately the amount of IL-17 produced. These analyses demonstrate how genome-wide annotation of *cis*-regulatory regions across multiple CD8 T cell differentiation states could be applied to provide the context for genetic disease associations.

Functional validation of causal T cell *cis*-regulatory elements controlling gene expression using CRISPRi

Finally, we sought to validate our *in silico* predictions with functional perturbation of ACRs in primary human CD8 T cells. We employed CRISPR interference (CRISPRi) using single guide RNAs (sgRNAs) to target catalytically inactive Cas9 (dCas9) fused with a transcriptional repressor domain (Krüppel-associated box; KRAB) to specific ACR elements (Gilbert et al., 2013; Fulco et al., 2016; Thakore et al., 2016; Klann et al., 2017; Gasperini et al., 2019). We used this approach to test ACRs that were predicted to highly contribute to control of gene expression. Human CD8 T cells were stimulated *in vitro* and co-transduced with lentiviruses (LVs) expressing dCas9-KRAB (mCherry) and a sgRNA (GFP; Figure 6N). If the ACR targeted by the sgRNA acted as a regulatory element, the expression of the predicted gene target would change in cells transduced with both LVs (dCas9-KRAB-mCherry$^+$ and sgRNA-GFP$^+$) but not in the singly transduced (mCherry only or GFP only) cells.

We used this system to test the function of predicted ACRs for *GZMB* and *CXCR3*. In addition to the candidate enhancers, the promoter of each gene was targeted as a positive control. We tested the top four ACRs predicted to regulate the *GZMB* expression (Figure 6F). The ACR predicted to have the greatest impact on gene expression, *A*, is located in the promoter region. The other three (*D*, *B*, and *C*) lay upstream at −26, −31.5, and −36.6 kb, respectively, from the TSS (Figure 6F). Targeting the promoter-proximal ACR *A* nearly ablated gene expression but had no effect on a non-targeted control gene (Figure 6O). Targeting ACRs *B* and *C* resulted in 45% and 41% reductions in *GZMB* expression compared to control sgRNA, respectively (Figure 6O), whereas targeting ACR *D* had no significant effect. For *CXCR3*, there were two major predicted enhancers, ACR *B* and ACR *A* (Figure 6J). ACR *B* was not tested due to its proximity to the promoter. Targeting the *CXCR3* promoter resulted in substantial reduction in gene expression, as expected (Figure 6P). However, targeting a single candidate enhancer *B*, located ~4 kb upstream, also led to a similar decrease in the expression of *CXCR3* (Figure 6P). Furthermore, targeting the promoter or candidate enhancer led to the downregulation of the surface CXCR3 protein on mCherry$^+$GFP$^+$ cells (Figure 6Q), indicating that targeting this enhancer is sufficient to modulate both mRNA and protein expression. Thus, the HD T cell atlas identified *cis*-regulatory elements for *CXCR3* and *GZMB* that controlled gene expression.

DISCUSSION

We constructed a transcriptional and epigenetic atlas of T cell differentiation from HDs to identify the molecular programs that control human T cell differentiation state, fate, and function. This atlas enabled us to define relationships between CD8 T cell subsets and identify underlying epigenetic and transcriptional mechanisms associated with key transitions in differentiation. These studies also provided insights into the biology of PD1$^+$CD39$^+$ Tex in HDs. We applied this atlas to the analysis of three disease datasets, validating the utility of this resource and gaining new insights into disease-specific biology. Lastly, we identified *cis*-regulatory elements and validated predicted enhancers for *CXCR3* and *GZMB* using CRISPRi, providing a guide for future non-coding genome engineering of T cells.

One unique feature of this atlas is the inclusion of a putatively exhausted CD8 T cell population. T cell exhaustion represents a considerable barrier to successfully treat cancer and chronic viral infections. Co-expression of PD1 and CD39 is a hallmark of Tex in diseases (Gupta et al., 2015; Canale et al., 2018), but this population had not been examined in HDs. Although HD PD1$^+$CD39$^+$ CD8 T cells were most similar to TILs compared to other CD8 T cell subsets, they also displayed gene expression and chromatin accessibility features that partially overlapped with memory and effector CD8 T cells. This observation is consistent with Tex in mice that employ transcriptional modules of both memory (e.g., *TCF7*) and effector (e.g., *GZMB*) CD8 T cells (McLane et al., 2019). The blood-derived PD1$^+$CD39$^+$ subset represents a proliferative intermediate Tex population similar to that described in humans and mice during chronic infection or cancer (Beltra et al., 2020; Zander et al., 2019;

Hudson et al., 2019; Li et al., 2018). In HDs, the ontogeny and functional role of these Tex are unknown, although one possibility is that this subset contains cells specific for persisting viruses such as EBV, HSV-1/2, VZV, anelloviruses, or others. Tex in HDs could also reflect cells with self-reactivity, consistent with signatures of exhaustion in human autoimmune diseases (McKinney et al., 2015). T cell exhaustion, as an alternative fate to deletion for autoreactive T cells, could enable the host to maintain a greater diversity of T cell specificities. Future studies will be necessary to further investigate the biology of these Tex in HDs, but their presence in HDs indicates that T cell exhaustion is not restricted to pathogenic situations of chronic infection or cancer.

A major goal of generating this HD T cell atlas was to develop a molecular framework from canonical human T cell subsets that could provide insights into other human T cell datasets. We applied this HD epigenomic atlas to two cancer and one autoimmune-related dataset. First, we analyzed scATAC-seq data from BCC that included 3 populations of Tex CD8 T cells. $PD1^{+}CD39^{+}$ T cells contained ACRs that significantly overlapped with the intermediate BCC Tex population, but the other two BCC Tex populations also contained chromatin accessibility shared with other HD T cell subsets. We distinguished which epigenomic programs might be co-opted from normal effector-like and naive T cell programs from those that were unique to terminal Tex. Furthermore, these results reinforced the concept that many transcriptional and epigenetic modules are shared among different cell types. Second, we applied this HD atlas to published data from a melanoma PD1 immunotherapy trial (Huang et al., 2019). Previous analyses of this cohort defined broad immune signatures of anti-PD1 response or failure but only identified potential treatment resistance mechanisms in a minority of patients. Although the patient number was small, we identified treatment-resistant patients whose TILs clustered with quiescent memory T cell subsets (CM and SCM). Increased expression of naive and memory genes, *CCR7* and *LEF1*, and the evidence of WNT signaling were prominent in progression patients. WNT signaling has been implicated in negative outcomes in cancer through direct effects on cancer cells (Jung and Park, 2020) and antigen-presenting cells that then limit T cell activation (Spranger et al., 2015; Spranger et al., 2017; Luke et al., 2019; Li et al., 2019). Here, we provide evidence for a potential direct role in CD8 T cells. This signature may be specific for earlier stage patients compared to those with metastatic melanoma (stage IV) since, in the latter, increased CM CD8 T cells may be associated with response to anti-PD1 blockade (Krieg et al., 2018). The immune mechanisms necessary to prevent recurrence after surgical resection in the patients studied here may be distinct from those required to control advanced disease, although reinvigorated Tex have been implicated in clinical response in other studies (Huang et al., 2017). These data point to a potential direct effect of WNT signaling on CD8 T cells in melanoma and highlight the utility of applying a high-resolution T cell atlas to deconvolute complex T cell differentiation signatures to reveal specific biological modules and pathways.

Genome engineering for cellular and gene therapy is a clinical reality for cancer and other diseases. It is now possible to envision, rather than only targeting protein-coding genes in T cells, targeting non-coding elements including *cis*-regulatory enhancers to achieve the T cell state-specific regulation of gene expression or tailored differentiation trajectories. Indeed, we now provide proof-of-concept for epigenomic engineering and a landscape map of non-coding elements with their linkage to gene expression control. Strategies that manipulate the non-coding genome have several advantages. First, altering regulatory elements that enhance or suppress transcription would allow fine tuning of target gene expression. For example, although genetic loss of PD1 has an initial benefit during chronic infection, permanent loss of PD1 also limits the durability of the response in mice (Odorizzi et al., 2015). Similarly, in human patients with cancer, CRISPR-engineered CAR T cells lacking PD1 were less abundant at later time points (Stadtmauer et al., 2020), likely due to effects on progenitor CD8 T cells during activation (Chen et al., 2019; Johnnidis et al., 2021) and/or T cell memory (Pauken et al., 2020). These observations suggest that reducing but not ablating PD1 expression through targeting T cell subset-biased enhancers may be advantageous. Similar principles likely exist for other relevant genes. Second, regulatory elements can control expression of multiple genes directly or indirectly by altering spatial genome organization (Mumbach et al., 2016, 2017). Thus, targeting a limited number of enhancers may have broad transcriptional effects. Third, using enhancers known to regulate gene expression in a particular environment could provide context-specific modulation. For example, patients with melanoma with WNT-expressing tumors could be given CAR T cells that re-deploy LEF1- or TCF7-containing enhancers downstream of WNT signaling to drive transcription of effector genes such as *GZMB* or *IFNG*. Context-specific enhancers may also allow control of gene expression only in selected T cell subsets. For example, in mice, an exhaustion-specific *cis*-regulatory element upstream of *Pdcd1* controls PD1 expression in Tex but not in effector CD8 T cells (Sen et al., 2016). Here, we identify cell state-specific ACRs in humans including validated enhancers for *GZMB* and *CXCR3*. This concept of state-specific *cis*-regulatory elements allows one to envision genetic engineering strategies that impact a gene of interest only in the relevant T cell subset. This HD atlas provides a road map for designing such strategies.

Here, we provide a transcriptional and chromatin accessibility map across human T cell differentiation states and demonstrate application of this atlas to investigate molecular programs in health and disease, including the functional annotation and manipulation of non-coding genomic regions. Connecting specific chromatin regions to gene expression opens the door for genetic engineering of human T cells to achieve targeted biological outcomes or changes in T cell fate, enabling more optimal T cell-based therapies for a range of diseases.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.immuni.2022.02.004.

ACKNOWLEDGMENTS

We thank the members of the Wherry Lab, especially A. Greenplate and J. Wu. Emily Duffner assisted in cloning. Tumor collection was supported by P50-CA174523. This work was supported by T32 CA009140 and a Cancer Research Institute-Mark Foundation Fellowship (J.R.G.), NIH grant CA234842 (Z.C.), NHLBI grant 1R38HL143613 (D.A.O.), NIH CA230157, the Tara Miller Award, the Parker Institute Bridge Scholar Award (A.C.H.), and the Parker Institute for Cancer Immunotherapy, the Allen Institute for Immunology, and Stand up to Cancer and NIH grants AI155577, AI115712, AI117950, AI108545, AI082630, CA210944, and U19AI149680 (to E.J.W.). We also thank the Penn Center for AIDS Research/Human Immunology Core (P30-AI045008/P30-CA016520). Work in the Wherry Lab was supported by the Parker Institute for Cancer Immunotherapy.

AUTHOR CONTRIBUTIONS

J.R.G. and E.J.W. conceived and designed the experiments and analyses. J.R.G. performed FACS and prepared sequencing libraries with help from S.G. and Z.C. J.R.G. performed CRISPRi experiments with help from E.F. J.R.G., J.S., and E.F. designed CRISPRi experiments. J.R.G. analyzed data with help from S.M., D.A.O., H.H., and A.E.B. A.C.H. and T.C.M. designed the clinical trial at Penn; A.C.H., T.C.M., X.X., G.C.K., L.C., M.C., L.G., L.M.S., and R.K.A. implemented this clinical trial; and T.C.M. was the trial principal investigator. R.M.Y., C.H.J., and N.P.W. obtained older HD samples from the National Institute on Aging. J.R.G. and E.J.W. wrote the manuscript.

DECLARATION OF INTERESTS

E.J.W. is a member of the Parker Institute for Cancer Immunotherapy, which supported the study. E.J.W. is an advisor for Merck, Marengo, Janssen, Related Sciences, Synthekine, and Surface Oncology. E.J.W. is a founder of Surface Oncology, Danger Bio, and Arsenal Biosciences. E.J.W. has a patent on the PD1 pathway. R.M.Y. and C.H.J. are inventors on patents and/or patent applications licensed to Novartis Institutes of Biomedical Research and receive license revenue from such licenses. C.H.J. is a scientific founder of Tmunity Therapeutics and DeCart Therapeutics for which he has founders stock but no income. T.C.M. received honorarium for advisory board service from Merck and BMS.

Received: May 17, 2021
Revised: September 27, 2021
Accepted: February 7, 2022
Published: March 8, 2022

REFERENCES

Abdelsamed, H.A., Moustaki, A., Fan, Y., Dogra, P., Ghoneim, H.E., Zebley, C.C., Triplett, B.M., Sekaly, R.P., and Youngblood, B. (2017). Human memory CD8 T cell effector potential is epigenetically preserved during in vivo homeostasis. J. Exp. Med. *214*, 1593–1606.

Ahmed, R., Bevan, M.J., Reiner, S.L., and Fearon, D.T. (2009). The precursors of memory: models and controversies. Nat. Rev. Immunol. *9*, 662–668.

Akondy, R.S., Fitch, M., Edupuganti, S., Yang, S., Kissick, H.T., Li, K.W., Youngblood, B.A., Abdelsamed, H.A., McGuire, D.J., Cohen, K.W., et al. (2017). Origin and differentiation of human memory CD8 T cells after vaccination. Nature *552*, 362–367.

Alfei, F., Kanev, K., Hofmann, M., Wu, M., Ghoneim, H.E., Roelli, P., Utzschneider, D.T., von Hoesslin, M., Cullen, J.G., Fan, Y., et al. (2019). TOX reinforces the phenotype and longevity of exhausted T cells in chronic viral infection. Nature *571*, 265–269.

Araki, Y., Wang, Z., Zang, C., Wood, W.H., III, Schones, D., Cui, K., Roh, T.Y., Lhotsky, B., Wersto, R.P., Peng, W., et al. (2009). Genome-wide analysis of histone methylation reveals chromatin state-based regulation of gene transcription and function of memory CD8+ T cells. Immunity *30*, 912–925.

Beltra, J.C., Manne, S., Abdel-Hakeem, M.S., Kurachi, M., Giles, J.R., Chen, Z., Casella, V., Ngiow, S.F., Khan, O., Huang, Y.J., et al. (2020). Developmental relationships of four exhausted CD8+ T cell subsets reveals underlying transcriptional and epigenetic landscape control mechanisms. Immunity *52*, 825–841.e8.

Bengsch, B., Ohtani, T., Khan, O., Setty, M., Manne, S., O'Brien, S., Gherardini, P.F., Herati, R.S., Huang, A.C., Chang, K.M., et al. (2018). Epigenomic-guided mass cytometry profiling reveals disease-specific features of exhausted CD8 T cells. Immunity *48*, 1029–1045.e5.

Best, J.A., Blair, D.A., Knell, J., Yang, E., Mayya, V., Doedens, A., Dustin, M.L., and Goldrath, A.W.; Immunological Genome Project Consortium (2013). Transcriptional insights into the CD8(+) T cell response to infection and memory T cell formation. Nat. Immunol. *14*, 404–412.

Blackburn, S.D., Shin, H., Freeman, G.J., and Wherry, E.J. (2008). Selective expansion of a subset of exhausted CD8 T cells by alphaPD-L1 blockade. Proc. Natl. Acad. Sci. USA *105*, 15016–15021.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Canale, F.P., Ramello, M.C., Núñez, N., Araujo Furlan, C.L., Bossio, S.N., Gorosito Serrán, M., Tosello Boari, J., Del Castillo, A., Ledesma, M., Sedlik, C., et al. (2018). CD39 expression defines cell exhaustion in tumor-infiltrating CD8+ T cells. Cancer Res *78*, 115–128.

Chen, Z., Arai, E., Khan, O., Zhang, Z., Ngiow, S.F., He, Y., Huang, H., Manne, S., Cao, Z., Baxter, A.E., et al. (2021). In vivo CD8+ T cell CRISPR screening reveals control by Fli1 in infection and cancer. Cell *184*, 1262–1280.e22.

Chen, Z., Ji, Z., Ngiow, S.F., Manne, S., Cai, Z., Huang, A.C., Johnson, J., Staupe, R.P., Bengsch, B., Xu, C., et al. (2019). TCF-1-centered transcriptional network drives an effector versus exhausted CD8 T cell-fate decision. Immunity *51*, 840–855.e5.

Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, S.M., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet. *48*, 1193–1203.

Croft, M. (2014). The TNF family in T cell differentiation and function – unanswered questions and future directions. Semin. Immunol. *26*, 183–190.

Crompton, J.G., Narayanan, M., Cuddapah, S., Roychoudhuri, R., Ji, Y., Yang, W., Patel, S.J., Sukumar, M., Palmer, D.C., Peng, W., et al. (2016). Lineage relationship of CD8+ T cell subsets is revealed by progressive changes in the epigenetic landscape. Cell. Mol. Immunol. *13*, 502–513.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Farh, K.K.H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337–343.

Fridman, W.H., Pagès, F., Sautès-Fridman, C., and Galon, J. (2012). The immune contexture in human tumours: impact on clinical outcome. Nat. Rev. Cancer *12*, 298–306.

Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. Science *354*, 769–773.

Gasperini, M., Hill, A.J., McFaline-Figueroa, J.L., Martin, B., Kim, S., Zhang, M.D., Jackson, D., Leith, A., Schreiber, J., Noble, W.S., et al. (2019). A genome-wide framework for mapping gene regulation via cellular genetic screens. Cell *176*, 377–390.e19.

Gattinoni, L., Lugli, E., Ji, Y., Pos, Z., Paulos, C.M., Quigley, M.F., Almeida, J.R., Gostick, E., Yu, Z., Carpenito, C., et al. (2011). A human memory T cell subset with stem cell-like properties. Nat. Med. *17*, 1290–1297.

Gattinoni, L., Zhong, X.S., Palmer, D.C., Ji, Y., Hinrichs, C.S., Yu, Z., Wrzesinski, C., Boni, A., Cassard, L., Garvin, L.M., et al. (2009). Wnt signaling arrests effector T cell differentiation and generates CD8+ memory stem cells. Nat. Med. *15*, 808–813.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

Goronzy, J.J., and Weyand, C.M. (2019). Mechanisms underlying T cell ageing. Nat. Rev. Immunol. *19*, 573–583.

Grömping, U. (2006). Relative importance for linear regression in R: the package relaimpo. J. Stat. Softw. *17*, 1–27.

Guo, X., Zhang, Y., Zheng, L., Zheng, C., Song, J., Zhang, Q., Kang, B., Liu, Z., Jin, L., Xing, R., et al. (2018). Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat. Med. *24*, 978–985.

Gupta, P.K., Godec, J., Wolski, D., Adland, E., Yates, K., Pauken, K.E., Cosgrove, C., Ledderose, C., Junger, W.G., Robson, S.C., et al. (2015). CD39 expression identifies terminally exhausted CD8+ T cells. PLoS Pathog. *11*, e1005177.

Hahne, F., and Ivanek, R. (2016). Visualizing genomic data using Gviz and bioconductor. Methods Mol. Biol. *1418*, 335–351.

Haider, S., Waggott, D., Lalonde, E., Fung, C., Boutros, P.C., et al. (2019). bedr: Genomic Region Processing using Tools Such as 'BEDTools', 'BEDOPS' and 'Tabix' (R package version 1.0.7.). https://CRAN.R-project.org/package=bedr.

Hamann, D., Baars, P.A., Rep, M.H.G., Hooibrink, B., Kerkhof-Garde, S.R., Klein, M.R., and van Lier, R.A.W. (1997). Phenotypic and functional separation of memory and effector human CD8+ T cells. J. Exp. Med. *186*, 1407–1418.

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics *14*, 7.

He, B., Xing, S., Chen, C., Gao, P., Teng, L., Shan, Q., Gullicksrud, J.A., Martin, M.D., Yu, S., Harty, J.T., et al. (2016). CD8+ T cells utilize highly dynamic enhancer repertoires and regulatory circuitry in response to infections. Immunity *45*, 1341–1354.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Huang, A.C., Orlowski, R.J., Xu, X., Mick, R., George, S.M., Yan, P.K., Manne, S., Kraya, A.A., Wubbenhorst, B., Dorfman, L., et al. (2019). A single dose of neoadjuvant PD-1 blockade predicts clinical outcomes in resectable melanoma. Nat. Med. *25*, 454–461.

Huang, A.C., Postow, M.A., Orlowski, R.J., Mick, R., Bengsch, B., Manne, S., Xu, W., Harmon, S., Giles, J.R., Wenz, B., et al. (2017). T-cell invigoration to tumour burden ratio associated with anti-PD-1 response. Nature *545*, 60–65.

Hudson, W.H., Gensheimer, J., Hashimoto, M., Wieland, A., Valanparambil, R.M., Li, P., Lin, J.X., Konieczny, B.T., Im, S.J., Freeman, G.J., et al. (2019). Proliferating transitory T cells with an effector-like transcriptional signature emerge from PD-1+ stem-like CD8+ T cells during chronic infection. Immunity *51*, 1043–1058.e4.

Im, S.J., Hashimoto, M., Gerner, M.Y., Lee, J., Kissick, H.T., Burger, M.C., Shan, Q., Hale, J.S., Lee, J., Nasti, T.H., et al. (2016). Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. Nature *537*, 417–421.

Johnnidis, J.B., Muroyama, Y., Ngiow, S.F., Chen, Z., Manne, S., Cai, Z., Song, S., Platt, J.M., Schenkel, J.M., Abdel-Hakeem, M., et al. (2021). Inhibitory signaling sustains a distinct early memory CD8+T cell precursor that is resistant to DNA damage. Sci. Immunol. *6*, eabe3702.

Jung, Y.S., and Park, J.I. (2020). Wnt signaling in cancer: therapeutic targeting of Wnt signaling beyond β-catenin and the destruction complex. Exp. Mol. Med. *52*, 183–191.

Kaech, S.M., and Cui, W. (2012). Transcriptional control of effector and memory CD8+ T cell differentiation. Nat. Rev. Immunol. *12*, 749–761.

Kamphorst, A.O., Pillai, R.N., Yang, S., Nasti, T.H., Akondy, R.S., Wieland, A., Sica, G.L., Yu, K., Koenig, L., Patel, N.T., et al. (2017). Proliferation of PD-1+ CD8 T cells in peripheral blood after PD-1-targeted therapy in lung cancer patients. Proc. Natl. Acad. Sci. USA *114*, 4993–4998.

Kassambara, A. (2020). rstatix: Pipe-Friendly Framework for Basic Statistical Tests (R package version 0.5.0.). https://CRAN.R-project.org/package=rstatix.

Khan, O., Giles, J.R., McDonald, S., Manne, S., Ngiow, S.F., Patel, K.P., Werner, M.T., Huang, A.C., Alexander, K.A., Wu, J.E., et al. (2019). TOX transcriptionally and epigenetically programs CD8+ T cell exhaustion. Nature *571*, 211–218.

Klann, T.S., Black, J.B., Chellappan, M., Safi, A., Song, L., Hilton, I.B., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2017). CRISPR–Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. Nat. Biotechnol. *35*, 561–568.

Konopka. (2020). umap: Uniform Manifold Approximation and Projection (R package version 0.2.5.0.). https://CRAN.R-project.org/package=umap.

Krieg, C., Nowicka, M., Guglietta, S., Schindler, S., Hartmann, F.J., Weber, L.M., Dummer, R., Robinson, M.D., Levesque, M.P., and Becher, B. (2018). High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy. Nat. Med. *24*, 144–153.

Kulinskaya, E. (2008). On two-sided p-values for non-symmetric distributions, Preprint at ArXiv *0810* (2124).

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics Oxf. Engl. *28*, 882–883.

Li, H., van der Leun, A.M., Yofe, I., Lubling, Y., Gelbard-Solodkin, D., van Akkooi, A.C.J., van den Braber, M., Rozeman, E.A., Haanen, J.B.A.G., et al.

(2018). Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma. Cell *176*, 775–789.e18.

Li, X., Xiang, Y., Li, F., Yin, C., Li, B., and Ke, X. (2019). WNT/beta-catenin signaling pathway regulating T cell-inflammation in the tumor microenvironment. Front. Immunol. *10*, 2293.

Lun, and Risso. (2019). SingleCellExperiment: S4 Classes for Single Cell Data (R package version 1.4.1).

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Luke, J.J., Bao, R., Sweis, R.F., Spranger, S., and Gajewski, T.F. (2019). WNT/β-catenin pathway activation correlates with immune exclusion across human cancers. Clin. Cancer Res. *25*, 3074–3083.

Mahto. (2019). splitstackshape: Stack and Reshape Datasets After Splitting Concatenated Values (R package version 1.4.8.). https://CRAN.R-project.org/package=splitstackshape.

Mathewson, N.D., Ashenberg, O., Tirosh, I., Gritsch, S., Perez, E.M., Marx, S., Jerby-Arnon, L., Chanoch-Myers, R., Hara, T., Richman, A.R., et al. (2021). Inhibitory CD161 receptor identified in glioma-infiltrating T cells by single-cell analysis. Cell *184*, 1281–1298.e26.

McKinney, E.F., Lee, J.C., Jayne, D.R.W., Lyons, P.A., and Smith, K.G.C. (2015). T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. Nature *523*, 612–616.

McLane, L.M., Abdel-Hakeem, M.S., and Wherry, E.J. (2019). CD8 T cell exhaustion during chronic viral infection and cancer. Annu. Rev. Immunol. *37*, 457–495.

Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat. Methods *13*, 919–922.

Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., et al. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet. *49*, 1602–1612.

Neuwirth. (2014). RColorBrewer: ColorBrewer Palettes (R package version 1.1-2.). https://CRAN.R-project.org/package=RColorBrewer.

Odorizzi, P.M., Pauken, K.E., Paley, M.A., Sharpe, A., and Wherry, E.J. (2015). Genetic absence of PD-1 promotes accumulation of terminally differentiated exhausted CD8+ T cells. J. Exp. Med. *212*, 1125–1137.

Paley, M.A., Kroy, D.C., Odorizzi, P.M., Johnnidis, J.B., Dolfi, D.V., Barnett, B.E., Bikoff, E.K., Robertson, E.J., Lauer, G.M., Reiner, S.L., and Wherry, E.J. (2012). Progenitor and terminal subsets of CD8+ T cells cooperate to contain chronic viral infection. Science *338*, 1220–1225.

Pauken, K.E., Godec, J., Odorizzi, P.M., Brown, K.E., Yates, K.B., Ngiow, S.F., Burke, K.P., Maleri, S., Grande, S.M., Francisco, L.M., et al. (2020). The PD-1 pathway regulates development and function of memory CD8+ T cells following respiratory viral infection. Cell Rep *31*, 107827.

Pauken, K.E., Sammons, M.A., Odorizzi, P.M., Manne, S., Godec, J., Khan, O., Drake, A.M., Chen, Z., Sen, D.R., Kurachi, M., et al. (2016). Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. Science *354*, 1160–1165.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., and Vanderplas, J. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

Qu, K., Zaba, L.C., Giresi, P.G., Li, R., Longmire, M., Kim, Y.H., Greenleaf, W.J., and Chang, H.Y. (2015). Individuality and variation of personal regulomes in primary human T cells. Cell Syst *1*, 51–61.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics Oxf. Engl. *26*, 841–842.

Raivo Kolde. (2019). pheatmap: Pretty Heatmaps (R package version 1.0.12). https://CRAN.R-project.org/package=pheatmap.

Restifo, N.P., and Gattinoni, L. (2013). Lineage relationship of effector and memory T cells. Curr. Opin. Immunol. *25*, 556–563.

Rome, K.S., Stein, S.J., Kurachi, M., Petrovic, J., Schwartz, G.W., Mack, E.A., Uljon, S., Wu, W.W., DeHart, A.G., McClory, S.E., et al. (2020). Trib1 regulates T cell differentiation during chronic infection by restraining the effector program. J. Exp. Med. *217*, e20190888.

Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell *160*, 48–61.

Sallusto, F., Lenig, D., Förster, R., Lipp, M., and Lanzavecchia, A. (1999). Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. Nature *401*, 708–712.

Satpathy, A.T., Granja, J.M., Yost, K.E., Qi, Y., Meschi, F., McDermott, G.P., Olsen, B.N., Mumbach, M.R., Pierce, S.E., Corces, M.R., et al. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat. Biotechnol. *37*, 925–936.

Scharer, C.D., Bally, A.P., Gandham, B., and Boss, J.M. (2017). Cutting edge: chromatin accessibility programs CD8 T cell memory. J. Immunol. *198*, 2238–2243.

Scharer, C.D., Barwick, B.G., Youngblood, B.A., Ahmed, R., and Boss, J.M. (2013). Global DNA methylation remodeling accompanies CD8 T cell effector function. J. Immunol. *191*, 3419–3429.

Scott, A.C., Dündar, F., Zumbo, P., Chandran, S.S., Klebanoff, C.A., Shakiba, M., Trivedi, P., Menocal, L., Appleby, H., Camara, S., et al. (2019). TOX is a critical regulator of tumour-specific T cell differentiation. Nature *571*, 270–274.

Scott-Browne, J.P., López-Moyado, I.F., Trifari, S., Wong, V., Chavez, L., Rao, A., and Pereira, R.M. (2016). Dynamic changes in chromatin accessibility occur in CD8+ T cells responding to viral infection. Immunity *45*, 1327–1340.

Scrucca, L., Fop, M., Murphy, T.B., and Raftery, A.E. (2016). mclust 5: clustering classification and density estimation using Gaussian finite mixture models. R J *8*, 289–317.

Sekine, T., Perez-Potti, A., Nguyen, S., Gorin, J.B., Wu, V.H., Gostick, E., Llewellyn-Lacey, S., Hammer, Q., Falck-Jones, S., Vangeti, S., et al. (2020). TOX is expressed by exhausted and polyfunctional human effector memory CD8+ T cells. Sci. Immunol. *5*, eaba7918.

Sen, D.R., Kaminski, J., Barnitz, R.A., Kurachi, M., Gerdemann, U., Yates, K.B., Tsao, H.W., Godec, J., LaFleur, M.W., Brown, F.D., et al. (2016). The epigenetic landscape of T cell exhaustion. Science *354*, 1165–1169.

Seo, H., Chen, J., González-Avalos, E., Samaniego-Castruita, D., Das, A., Wang, Y.H., López-Moyado, I.F., Georges, R.O., Zhang, W., Onodera, A., et al. (2019). TOX and TOX2 transcription factors cooperate with NR4A transcription factors to impose CD8+ T cell exhaustion. Proc. Natl. Acad. Sci. USA *116*, 12410–12415.

Seumois, G., Chavez, L., Gerasimova, A., Lienhard, M., Omran, N., Kalinke, L., Vedanayagam, M., Ganesan, A.P., Chawla, A., Djukanović, R., et al. (2014). Epigenomic analysis of primary human T cells reveals enhancers associated with TH2 memory cell differentiation and asthma susceptibility. Nat. Immunol. *15*, 777–788.

Spranger, S., Bao, R., and Gajewski, T.F. (2015). Melanoma-intrinsic beta-catenin signalling prevents anti-tumour immunity. Nature *523*, 231–235.

Spranger, S., Dai, D., Horton, B., and Gajewski, T.F. (2017). Tumor-residing Batf3 dendritic cells are required for effector T cell trafficking and adoptive T cell therapy. Cancer Cell *31*, 711–723.e4.

Stadtmauer, E.A., Fraietta, J.A., Davis, M.M., Cohen, A.D., Weber, K.L., Lancaster, E., Mangan, P.A., Kulikovskaya, I., Gupta, M., Chen, F., et al. (2020). CRISPR-engineered T cells in patients with refractory cancer. Science *367*, eaba7365.

Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics *19*, 477.

Tarumoto, Y., Lu, B., Somerville, T.D.D., Huang, Y.H., Milazzo, J.P., Wu, X.S., Klingbeil, O., El Demerdash, O., Shi, J., and Vakoc, C.R. (2018). LKB1, salt-inducible kinases, and MEF2C are linked dependencies in acute myeloid leukemia. Mol. Cell *69*, 1017–1027.e6.

Thakore, P.I., Black, J.B., Hilton, I.B., and Gersbach, C.A. (2016). Editing the epigenome: technologies for programmable transcription and epigenetic modulation. Nat. Methods *13*, 127–137.

Thommen, D.S., and Schumacher, T.N. (2018). T cell dysfunction in cancer. Cancer Cell *33*, 547–562.

Trimarchi, J.M., and Lees, J.A. (2002). Sibling rivalry in the E2F family. Nat. Rev. Mol. Cell Biol. *3*, 11–20.

Ucar, D., Márquez, E.J., Chung, C.H., Marches, R., Rossi, R.J., Uyar, A., Wu, T.-C., George, J., Stitzel, M.L., Palucka, A.K., et al. (2017). The chromatin accessibility signature of human immune aging stems from CD8 + T cells. J. Exp. Med. *214*, 3123–3144.

Utzschneider, D.T., Charmoy, M., Chennupati, V., Pousse, L., Ferreira, D.P., Calderon-Copete, S., Danilo, M., Alfei, F., Hofmann, M., Wieland, D., et al. (2016). T cell Factor 1-expressing memory-like CD8(+) T cells sustain the immune response to chronic viral infections. Immunity *45*, 415–427.

Vella, L.A., Buggert, M., Manne, S., Herati, R.S., Sayin, I., Kuri-Cervantes, L., Bukh Brody, I., O'Boyle, K.C., Kaprielian, H., Giles, J.R., et al. (2019). T follicular helper cells in human efferent lymph retain lymphoid characteristics. J. Clin. Invest. *129*, 3185–3200.

Venables, W.N., and Ripley, B.D. (2002). Modern Applied Statistics with S, Fourth Edition (Springer).

Wang, Y., Misumi, I., Gu, A.D., Curtis, T.A., Su, L., Whitmire, J.K., and Wan, Y.Y. (2013). GATA-3 controls the maintenance and proliferation of T cells downstream of TCR and cytokine signaling. Nat. Immunol. *14*, 714–722.

Wei, H., Geng, J., Shi, B., Liu, Z., Wang, Y.H., Stevens, A.C., Sprout, S.L., Yao, M., Wang, H., and Hu, H. (2016). Cutting edge: Foxp1 controls naive CD8+ T cell quiescence by simultaneously repressing key pathways in cellular metabolism and cell cycle progression. J. Immunol. *196*, 3537–3541.

Wickham, H. (2007). Reshaping data with the reshape Package. J. Stat. Softw. *21*, 1–20.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, H., Hester, H., et al. (2019). Welcome to the tidyverse. J. Open Source Softw. *4*, 1686.

Wong, N., Liu, W., and Wang, X. (2015). WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. Genome Biol *16*, 218.

Wu, T., Ji, Y., Moseman, E.A., Xu, H.C., Manglani, M., Kirby, M., Anderson, S.M., Handon, R., Kenyon, E., Elkahloun, A., et al. (2016). The TCF1-Bcl6 axis counteracts type I interferon to repress exhaustion and maintain T cell stemness. Sci Immunol *1*, eaai8593.

Yao, C., Sun, H.W., Lacey, N.E., Ji, Y., Moseman, E.A., Shih, H.Y., Heuston, E.F., Kirby, M., Anderson, S., Cheng, J., et al. (2016). Single-cell RNA-seq reveals TOX as a key regulator of CD8+ T cell persistence in chronic infection. Nat Immunol *20*, 890–901.

Yoshida, H., Lareau, C.A., Ramirez, R.N., Rose, S.A., Maier, B., Wroblewska, A., Desland, F., Chudnovskiy, A., Mortha, A., Dominguez, C., et al. (2019). The cis-regulatory atlas of the mouse immune system. Cell *176*, 897–912.e20.

Yost, K.E., Satpathy, A.T., Wells, D.K., Qi, Y., Wang, C., Kageyama, R., McNamara, K.L., Granja, J.M., Sarin, K.Y., Brown, R.A., et al. (2019). Clonal replacement of tumor-specific T cells following PD-1 blockade. Nat. Med. *25*, 1251–1259.

Yu, B., Zhang, K., Milner, J.J., Toma, C., Chen, R., Scott-Browne, J.P., Pereira, R.M., Crotty, S., Chang, J.T., Pipkin, M.E., et al. (2017). Epigenetic landscapes reveal transcription factors that regulate CD8+ T cell differentiation. Nat. Immunol. *18*, 573–582.

Zander, R., Schauder, D., Xin, G., Nguyen, C., Wu, X., Zajac, A., and Cui, W. (2019). CD4+ T cell help is required for the formation of a cytolytic CD8+ T cell subset that protects against chronic infection and cancer. Immunity *51*, 1028–1042.e4.

Zhang, K., Wang, M., Zhao, Y., and Wang, W. (2019). Taiji: system-level identification of key transcription factors reveals transcriptional waves in mouse embryonic development. Sci. Adv. *5*, eaav3262.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zheng, C., Zheng, L., Yoo, J.K., Guo, H., Zhang, Y., Guo, X., Kang, B., Hu, R., Huang, J.Y., Zhang, Q., et al. (2017). Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. Cell *169*, 1342–1356.e16.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. *10*, 1523.

Zhu, X., Zhang, S., Kan, D., and Cooper, R. (2004). Haplotype block definition and its application. Pac. Symp. Biocomput. 152–163.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-CD27 Brilliant Violet 785	Biolegend	Clone O323, cat# 302832, RRID: AB_2562674
Anti-CD11a Brilliant Violet 650	BD Biosciences	Clone HI111, cat# 563934, RRID: AB_2738493
Anti-CD45RA Brilliant Violet 605	Biolegend	Clone HI100, cat# 304134, RRID: AB_2563814
Anti-PD1 Brilliant Violet 421	Biolegend	Clone E12.2.H7, cat # 329920, RRID: AB_10960742
Anti-CCR7 FITC	BD Biosciences	Clone 150503, cat# 561271, RRID: AB_10561679
Anti-CD45d PE-Cy7	Biolegend	Clone 9F10, cat# 304314, RRID: AB_10643278
Anti-CD4 PE-Cy5.5	Invitrogen	Clone S3.5, cat# MHCD0418, RRID: AB_10376013
Anti-CD3 PE-Cy5	Biolegend	Clone HIT3a, cat# 300309, RRID: AB_314045
Anti-CD95 PE/Dazzle	Biolegend	Clone DX2, cat# 305634, RRID: AB_2564221
Anti-CD39 PE	Biolegend	Clone A1, cat# 328208, RRID: AB_940429
Anti-CD8 APC/Fire 750	Biolegend	Clone RPA-T8, cat# 301066, RRID: AB_2629695
Anti-CXCR3 Alexa 647	Biolegend	Clone G0257, cat # 353712, RRID: AB_10962948
Anti-CD25 Brilliant Violet 650	Biolegend	Clone BC96, cat# 302634, RRID: AB_2563807
Anti-CD38 Brilliant Blue 515	BD Biosciences	Clone HIT1, cat# 564498, RRID: AB_2744374
Anti-ICOS PE-Cy7	Biolegend	Clone C398.4D, cat# 313520, RRID: AB_10643411
Anti-CD127 PE/Dazzle	Biolegend	Clone A019D5, cat# 351335, RRID: AB_2563636
Anti-CCR7 PE	Biolegend	Clone G043H7, cat# 353203, RRID: AB_10916391
Anti-CD3 Alexa 700	BD Biosciences	Clone SP34-2, cat# 557917, RRID: AB_396938
Anti-CXCR5 Alexa 647	Biolegend	Clone J252D4, cat# 356906, RRID: AB_2561815
Anti-CD45RA Brilliant Ultraviolet 395	BD Biosciences	Clone HI100, cat# 740298, RRID: AB_2740037
Anti-CD8 Brilliant Ultraviolet 496	BD Biosciences	Clone RPA-T8, cat# 612942, RRID: AB_2744460
Anti-CD38 Brilliant Ultraviolet 661	BD Biosciences	Clone HIT2, cat# 565069, RRID: AB_2744377
Anti-CD27 Brilliant Ultraviolet 737	BD Biosciences	Clone L128, cat# 612830, RRID: AB_2744350
Anti-CD3 Brilliant Ultraviolet 805	BD Biosciences	Clone UCHT1, cat# 612896, RRID: AB_2739277
Anti-CXCR3 Brilliant Violet 421	Biolegend	Clone G025H7, cat# 353716, RRID: AB_2561448
Anti-CD49d Brilliant Violet 480	BD Biosciences	Clone 9F10, cat# 566134, RRID: AB_2739533
Anti-PD1 Brilliant Violet 605	Biolegend	Clone E12.2.H7, cat# 563245, RRID: AB_2738091
Anti-CD95 Brilliant Violet 650	Biolegend	Clone DX2, cat# 305642, RRID: AB_2632622
Anti-CD39 Brilliant Violet 711	Biolegend	Clone A1, cat# 328228, RRID: AB_2632894
Anti-CD4 Brilliant Violet 750	BD Biosciences	Clone SK3, cat# 566355, RRID: AB_2744426
Anti-CD28 Brilliant Violet 786	BD Biosciences	Clone CD28.2, cat# 740996, RRID: AB_2740619
Anti-CD127 Brilliant Blue 700	BD Biosciences	Clone HIL-7R-M21, cat# 566398, RRID: AB_2744279
Anti-CX3CR1 biotin	Biolegend	Clone 2A9-1, cat# 341617, RRID: AB_2616937
Anti-TOX PE	Miltenyi	Clone REA473, cat# 130-120-716, RRID: AB_2801780
Anti-EOMES PE-eF610	Invitrogen	Clone WD1928, cat# 61-4877-42, RRID: AB_2574616
Anti-CTLA4 PE-Cy5	BD Biosciences	Clone BNI3, cat# 561717, RRID: AB_10893816
Anti-GZMB PE-Cy5.5	Invitrogen	Clone GB11, cat# GRB18, RRID: AB_2536541
Anti-Tbet PE-Cy7	Biolegend	Clone 4B10, cat# 644824, RRID: AB_2561761
Anti-TCF7 Alexa 647	Cell Signaling Technology	Clone C63D9, cat# 6709S, RRID: AB_2797631
Anti-Ki67 Alexa 700	BD Biosciences	Clone B56, cat# 561277, RRID: AB_10611571
Anti-TIGIT APC/Fire 750	Biolegend	Clone VSTM3, cat# 372708, RRID: AB_2632755
Anti-CD57 biotin	Biolegend	Clone HCD57, discontinued, RRID: AB_1083981
Anti-HLA-DR	Thermo Fischer	Clone TU36, cat# MHLDR18, RRID: AB_10372966
Streptavidin Brilliant Blue 790	BD Biosciences	Custom build

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Anti-IFNg Alexa 700	Biolegend	Clone B27, cat# 506516, RRID:AB_961351
Anti-TNF PE-Cy7	Biolegend	Clone MAb11, cat# 502930, RRID:AB_2204079
Anti-IL-2 Brilliant Violet 750	BD Biosciences	Clone MQ1-17H12,cat# 566361, RRID:AB_2739710
Biological samples		
Human PBMCs	This paper	Table S1
Melanoma biopsies (single cell suspension)	This paper	Table S1
Critical commercial assays		
Takara Pico Input SMARTer Stranded Total RNA-Seq Kit	Takara	cat# #634413
Nextera DNA Library Preparation Kit	Illumina	cat # FC-121-1031
Deposited data		
Raw sequencing data	This paper	Deposited in GEO: GSE179613
Experimental models: Cell lines		
HUMAN HEK293T	ATCC	CRL-3216
Oligonucleotides		
CRISPRi guides (Table S4)		N/A
Primers for qPCR (Table S5)		N/A
Recombinant DNA		
LRG 2.1T	Tarumoto et al. 2018	Addgene #108098
psPAX2		Addgene #12260
pED9x	This study	Addgene #163956
Software and algorithms		
Genome hg19	UCSC	http://hgdownload.cse.ucsc.edu/goldenpath/hg19/
Blacklisted regions	UCSC	http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/wgEncodeDacMapabilityConsensusExcludable.bed
bowtie2/2.1.0	(Langmead and Salzberg, 2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
samtools/1.1	http://www.htslib.org/doc/#publications	http://samtools.sourceforge.net/
picard tools/1.141	Broad Institute	http://broadinstitute.github.io/picard/
MACS2/2.1.1.20160309	(Zhang et al. 2008)	https://github.com/taoliu/MACS/wiki
BEDTools/2.15	(Quinlan and Hall, 2010)	https://bedtools.readthedocs.io/en/latest/
FastQC/0.11.2	NA	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
HOMER/4.10.3	(Heinz et al. 2010)	http://homer.ucsd.edu/homer/
IGV/2.4.16	Broad Institute	http://software.broadinstitute.org/software/igv/
STAR/2.5.2a	(Dobin et al. 2013)	https://github.com/alexdobin/STAR
PORT/0.8.4-beta	http://bioinf.itmat.upenn.edu/benchmarking/rnaseq/port/index.php	https://github.com/itmat/Normalization
Flowjo/10.5.3	Tree Star	https://www.flowjo.com/
Python/2.7.5	Python Software Foundation	https://www.python.org/
R/3.5.1	The R Foundation	https://www.r-project.org/
Taiji/0.2	(Zhang et al. 2019)	https://taiji-pipeline.github.io/
Metascape	(Zhou et al. 2019)	https://metascape.org
scikit-learn/0.21.3	(Pedregosa et al. 2011)	N/A
DESeq2/1.22.2	(Love et al., 2014)	Bioconductor
umap/0.2.5.0	Konopka (2020)	CRAN
sva/3.30.1	(Leek et al., 2012)	Bioconductor
pheatmap/1.0.12	Raivo Kolde (2019)	CRAN

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Gviz/1.26.5	(Hahne and Ivanek 2016)	Bioconductor
slingshot/1.0.0	(Street et al. 2018)	Bioconductor
MASS/ 7.3-51.5	(Venables and Ripley 2002)	CRAN
Bedr/1.0.7	Haider et al. (2019)	CRAN
relaimpo/ 2.2-3	(Grömping 2006)	CRAN
GSEABase/1.44.0	Morgan, M., Falcon, S. and Gentleman, R. GSEABase: Gene set enrichment data structures and methods. R package version 1.30.2.	Bioconductor
GSVA/1.30.0	(Hänzelmann et al., 2013)	Bioconductor
SingleCellExperiment/1.4.1	Lun and Risso (2019)	Bioconductor
mclust/ 5.4.5	(Scrucca et al., 2016)	CRAN
tidyverse/1.3.0	(Wickham et al., 2019)	CRAN
reshape2/1.4.3	(Wickham, 2007)	CRAN
splitstackshape/1.4.8	Mahto (2019)	CRAN
RColorBrewer/1.1-2	Neuwirth (2014)	CRAN
Rstatix/ 0.5.0	Kassambara (2020)	CRAN

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact, E. John Wherry (wherry@pennmedicine.upenn.edu).

Materials availability

The pED9x vector is available at Addgene #163956.

Data and code availability

All RNA-seq and ATAC-seq data generated in this study are deposited in GEO under GEO: GSE179613. The ATAC-seq processing script is provided in Data S2. The IQR and permutation code is available here: https://github.com/wherrylab/statistics_code/blob/master/MutualInformationMetricsForDiscreteCategoricalComparison.R. Other code can be made available upon reasonable request. No new algorithms were developed during this study.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Healthy donor human samples

Healthy donor peripheral blood mononuclear cells (PBMCs) were obtained by the University of Pennsylvania Human Immunology Core/CFAR Immunology Core or the National Institute of Aging from de-identified healthy donors. PBMCs were purified from whole blood or leukapheresis products by Ficoll-Hypaque density gradient centrifugation. Donors were self-identified as healthy.

Clinical trial human samples

Melanoma patient PBMC and tumor samples were collected as part of a phase 1b clinical trial (NCT02434354) which was a single institution investigator-initiated study sponsored by the University of Pennsylvania. The protocol and its amendments were approved by the Institutional Review Board at the University of Pennsylvania, and all patients provided written informed consent. All detailed methods regarding the trial, patients, and sample collection can be found in (Huang et al., 2019). In brief, all patients underwent a baseline pre-treatment biopsy, then received a neoadjuvant single flat dose of pembrolizumab 200 mg intravenously, followed by complete resection 3 weeks later. Patients also provided paired blood samples at the pre-treatment and post-treatment time points. After resection and on surgical recovery, patients continued to receive adjuvant pembrolizumab every 3 weeks for up to 1 year, or until the time of recurrence or any unacceptable treatment-related toxicity. Both biopsies and resection specimens were processed in the Department of Pathology and Laboratory Medicine, Hospital of the University of Pennsylvania.

METHOD DETAILS

Cell sorting for sequencing libraries

Cryopreserved PBMCs were thawed in RPMI-1640 media supplemented with 10% FBS, 1x non-essential amino acids (Gibco #11140050), and 10mM Hepes (Gibco # 15630080), 2mM L-glutamine (Gibco # 25030081), 100U/mL penicillin/streptomycin (Gibco # 15140122) (cRPMI). DNAase and MgCl2 were included for cryopreserved tumor samples. Cells were washed with 1× PBS and stained with an amine–reactive dye (Invitrogen) for 20 minutes at room temperature to assess cell viability, followed by an antibody cocktail in cRPMI for 45 minutes at room temperature. Samples were sorted on a BD FACSAria II machine into RPMI-1640 media supplemented with 50% FBS, 1% Hepes, 1% L-glutamine, 1% penicillin/streptomycin. A small aliquot of all sorted samples were run to check purity. Voltages on the machine were standardized using fluorescent targets and Spherotech rainbow beads (#URCP-50-2F). Not all T cell subsets were captured from each donor due to limitations in cell number, 2500-68000, average ~45000, cells were sorted per subset for each assay.

Flow cytometry for HD PBMCs

Cells were thawed in staining media (SM) consisting of PBS with 3% FCS, 5mM EDTA, and 1% penicillin/streptomycin. Cells were washed with 1× PBS and stained with an amine–reactive dye (Invitrogen #L34966) for 20 minutes to assess cell viability, followed by an antibody cocktail in SM for 45 minutes, then streptavidin-Brilliant Blue 790 (BD Biosciences) in SM for 20 minutes. Permeabilization was performed using the Foxp3 Fixation/Permeabilization Concentrate and Diluent kit (eBioscience #00-5521-00) for 20 minutes. Intracellular staining with antibody cocktails was done for 2 hours. All steps were performed at room temperature. Samples were run on a BD Symphony A5 instrument. Voltages on the machine were standardized using fluorescent targets and Spherotech rainbow beads (#URCP-50-2F). Data were analyzed with FlowJo software (version 10.5.3, TreeStar).

RNA-seq and ATAC-seq libraries preparation and sequencing

To extract RNA, sorted cells were resuspended in buffer RLT supplemented with beta-mercaptoethanol and processed with a Qiagen RNeasy Plus Micro Kit (#74034) per manufacturer's instructions. Total RNA libraries were prepared using a Takara Pico Input SMARTer Stranded Total RNA-Seq Kit (#634413). Extracted RNA and libraries were assessed for quality on an Agilent TapeStation 2200 instrument (#5067-5579 & #5067-5580, #5067-5592 & #5067-5593, respectively).

ATAC libraries were generated as described with minor changes (Buenrostro et al., 2013). Briefly, nuclei from sorted cells were isolated using a lysis solution composed of 10mM Tris-HCl, 10mM NaCl, 3mM $MgCl_2$, and 0.1% IGEPAL CA-630 (melanoma patients) or 0.1% Tween 20 (healthy donors). Immediately following cell lysis, nuclei were pelleted in Eppendorf DNA LoBind 1.5ml tubes (Fisher #22431021) and resuspended in TD Buffer with Tn5 transposase (Illumina #FC-121-1031). Transposition reaction was performed at 37°C for 30 minutes. DNA fragments were purified from enzyme solution using Qiagen MinElute Enzyme Reaction Cleanup Kit (#28204). Libraries were barcoded (Nextera Index Kit, Illumina # FC-121-1012) and amplified with NEBNext High Fidelity PCR Mix (New England Biolabs # M0541L). Library quality was assessed using a TapeStation instrument (#5067-5584 & # 5067-5585). RNA and ATAC libraries were quantified using a KAPA Library Quantification Kit (#KK4824) and sequenced on an Illumina NextSeq 550 instrument.

In vitro stimulation assay

Cryopreserved PBMCs were thawed in cRPMI, and CD8 T cells were isolated per manufacturer's instructions (Stemcell #17953). CD8 T cells were stimulated with plate-bound anti-CD3 (10ug/mL, UCHT1 clone, Biolegend #300402) and soluble anti-CD28 (2ug/mL, CD28.2 clone, Biolegend #302902) in cRPMI for 4 hours, then an additional 5 hours with BFA (Biolegend #420601) and Monensin (Biolegend #420701) at $1.5x10^6$/well in a 24-well plate in a 37°C incubator. Flow cytometry was performed as described above, except Cytofix/Cytoperm (BD Bioscience #554714) was used for fixation and permeabilization.

CRISPR design, cloning, and virus prep

After genomic regions of interest were identified, sgRNAs were designed in Benchling using the hg19 genome and spCas9 species specifications (Table S4). Guides were selected based on even distribution throughout the region of interest and off-target score as previously described (Hsu et al., 2013). The control guide is a non-targeting sgRNA with no perfect match in the human genome. Guides with TTTT were excluded because of decreased binding efficiency (Wong et al., 2015). Three or four sgRNA were selected for each target region (enhancer or promoter). Bases CACCG were added to the 5' end of the forward strand of the sgRNA and bases AAAC were added to the 5' end as well as a cytosine to the 3' end of the reverse strand of the sgRNA to ensure cloning to the cleaved lentiviral vector. Using these designs, oligos of the forward and reverse strands of the sgRNAs were ordered at a 10 nmole scale from Eurofins. The forward and reverse oligos were phosphorylated and annealed. Designed sgRNAs were then cloned into the BsmBI-digested lentiviral vector LRG 2.1T (Addgene #108098) (Tarumoto et al., 2018), after which they were transformed into Stbl3 chemically competent E. coli cells. After overnight growth, individual colonies were picked, cultured, and plasmid DNA was extracted using the QIAprep Spin Miniprep kit (#27104) per manufacturer's instructions. Extracted DNA was then sent for Sanger sequencing verification. Extracted plasmid DNA, VsVG (Addgene #14888), and psPax2 (Addgene #12260) were transfected into HEK 293T cells with polyethylenimine in CST media (Gibco #A1048501) supplemented with 1x non-essential amino acids (Gibco #11140050), 10mM

Hepes (Gibco # 15630080), 2mM L-glutamine (Gibco # 25030081), and 100U/mL penicillin/streptomycin (Gibco # 15140122) (cCST). Resulting virus was collected and filtered (0.45uM PVDF filter).

In vitro CRISPRi assay

Cryopreserved PBMCs were obtained from the University of Pennsylvania Human Immunology Core/CFAR Immunology Core as described above. To decrease donor-specific effects and increase consistency, a mix of multiple donors was used for each experiment so that no more than 20% of the total cells were from any one donor. Cells were thawed into cCST media. CD8 T cells were isolated per manufacturer's instructions (Stemcell #17953) and stimulated with anti-CD3/anti-CD28 Dynabeads (Gibco #11131D) at a bead:cell 3:1 ratio with 10ng/mL IL-2 (Peprotech #200-02), 5ng/mL IL-7 (Peprotech #200-07), and 5ng/mL IL-15 (Peprotech #200-15) in 1.5mL of cCST media at $1.5x10^6$/well in a 24-well plate in a 37°C incubator. After 30 hours, 1.25mL of media was removed from each well and replaced with 1mL dCas9-KRAB-mCherry and 0.5mL sgRNA plus fresh cytokines and polybrene (8μg/ml). Plates were centrifuged at 2000xg for 75 minutes at 37°C, then returned to the incubator. The next day, 1.5mL of media was removed from each well and replaced with 1.5mL fresh cCST media plus cytokines. Cell were expanded as necessary until day 5 post-infection when they were harvested for FACS and flow cytometry.

For qPCR analysis, cells were sorted on a BD FACSAria II machine as single cells that were dCas-KRAB-mCherry$^+$ sgRNA-GFP$^+$, dCas-KRAB-mCherry$^+$ sgRNA-GFP$^-$, dCas-KRAB-mCherry$^-$ sgRNA-GFP$^+$, dCas-KRAB-mCherry$^-$ sgRNA-GFP$^-$ into cCST with 50% FCS in 1.5mL Eppendorf DNA LoBind tubes (Fisher # 22431021). To extract RNA, sorted cells were spun down and resuspended in buffer RLT supplemented with b-mercaptoethanol and processed with a Qiagen RNeasy Plus Micro Kit (#74034) per manufacturer's instructions. RNA quantity and quality were checked on a Thermo Scientific Nanodrop 2000c. Equal amounts of RNA were used as input for cDNA synthesis, performed using the Applied Biosystem High-Capacity cDNA Reverse Transcription Kit (Thermo #4368814) per manufacturer's guidelines. qPCR reactions used iTaq Universal SYBR Green Supermix (Bio-Rad #1725121) and primers listed in Table S5 and were run on an Applied Biosystems QuantStudio 6 Flex. All reactions were performed in triplicate. For flow cytometry analysis, cells were stained for 30 mins at room temperature with anti-CXCR3 BV421(clone G025H7, Biolegend #353716). Samples were run on a BD Symphony A5 instrument. Voltages on the machine were standardized using fluorescent targets and Spherotech rainbow beads (#URCP-50-2F). Data were analyzed with FlowJo software (version 10.5.3, TreeStar).

QUANTIFICATION AND STATISTICAL ANALYSIS

RNA-seq data processing and analysis

FASTQ files were aligned using STAR 2.5.2a with the hg19 human reference genome. Aligned files were processed using PORT (https://github.com/itmat/Normalization). Batch correction for sample location and process group was done with using the Combat function in the sva R package. PCA was done with the prcomp function in R using the top twenty percent most variable genes. PC1 and PC3 are shown in Figures 1F and S1B; PC2 from RNA-seq did not map to any known biologic or technical variable and may capture an element of unknown subject immunological history. UMAP analysis was performed using the umap function from the R package umap. Differentially expressed genes (DEGs) were identified with DESeq2 (DESeq function) using adjusted p value < 0.05; genes were first filtered on minimum expression (median reads per group ≤ 5). For pairwise comparisons between HD CD8 T cell subsets, donor type, sample location, and sample preparation group were included in the model formula to adjust for statistical confounding; for older versus young donor analysis, sample location and sample preparation group were included; in the melanoma cohort, sample preparation group was included. All plots with RNA expression are shown as normalized, $\log_2$ transformed, and batch corrected.

ATAC-seq data processing and analysis

The script used for processing raw ATAC-seq FASTQ provided in Data S2. In brief, samples were aligned to the hg19 human reference genome with Bowtie2. Unmapped, unpaired, and mitochondrial reads were removed using samtools. ENCODE Blacklist regions were removed. PCR duplicates were removed using Picard. Peak calling was performed with MACS2 with a FDR q-value $<$ 0.001. A union peak list of each data set was created by combining all peaks in all samples, merging overlapping peaks using bedtools *merge*, and keeping peaks that were called in more than one sample. The number of reads in each peak was determined with bedtools *coverage*. Peaks were annotated using Homer. Batch correction for sample location and process group was done with Combat. PCA was done with the prcomp function in R using the top twenty percent most variable ACRs. UMAP analysis was performed using the umap function from the R package, umap. Differentially accessible peaks (DAPs) were identified with the R package DESeq2 (DESeq function) using adjusted p value < 0.05. For pairwise comparisons between HD CD8 T cell subsets, donor type, sample location, and sample preparation group were included in the model formula to adjust for statistical confounding; for older versus young donor analysis, sample location and sample preparation group were included; in the melanoma cohort, sample preparation group and donor were included. Motif enrichment was performed with Homer using the union peak list as background. For peak set enrichment, peak names between experiment and peak set of interest were unified using custom R scripts; enrichment scores were calculated using the "gsva" method in the GSVA R package. ATAC-seq signal tracks were generated with the gviz R package. ATAC signal tracks are generated using gviz with bigwigs normalized for library size and group scaled across all samples within the dataset, peaks out of range are indicated with a contrast color on top.

Clustering

Biclustering was performed using the SpectralBiclustering function from sklearn using differentially expressed genes from all pairwise comparisons (adjusted p-value $\leq$ 0.05, fold change > 3, median reads per group > 25) with k, the number of gene and sample clusters, set to the number of cell subsets included in the analysis, eight. For visualization, gene clusters were ordered based on median gene expression in the naïve CD8 T cell subset; sample clusters were ordered based on median expression of ranked gene clusters. All other clustering was performed with pheatmap from the pheatmap R package using Euclidean for distance and ''complete'' for the clustering method (which are the default methods). The k for row/column cluster identification was chosen based on the biologic question as indicated.

Gene ontology and gene set enrichment analysis

Gene ontology (biological processes) analysis was performed using Metascape (metascape.org) using all expressed genes as the background gene list. Gene set enrichment was performed with GSEA software (https://www.gsea-msigdb.org), or scores for individual samples were calculated with the ''gsva'' method in the GSVA R package – as indicated.

Measuring the informativeness of transcriptomic and epigenomic data for inferring cell subtypes (IQR analysis)

First, sample clusters from RNA-seq or ATAC-seq (proximal ACRs or distal ACRs) were identified using hierarchical clustering (k = 12, the number of sorted subsets). We then quantified the association between cell subsets (by sorted phenotype) with each set of sample clusters (generated separately from RNA-seq gene expression, ATAC-seq proximal ACR, or ATAC-seq distal ACR data) using information quality ratio (IQR). IQR is a metric which quantifies mutual dependence of one variable (cell subset) based on a second variable (clusters) and is mathematically defined as the ratio of mutual information divided by joint entropy between these two variables. IQR can therefore be viewed as a form of normalized mutual information, where IQR = 0 implies that two variables are mutually independent whereas IQR = 1 implies that each variable perfectly predicts the other. To calculate mutual information and joint entropy, probability of each cell subtype/cluster combination was estimated by observed sample proportions.

To test whether observed cell subset/cluster IQRs were significantly different than background, we generated a null distribution of raw IQR values by permuting cell subtype labels and calculated one-tailed P-values based on the right tail of simulated raw IQR distribution (one-sample permutation test, N = $2^{*}10^{4}$ permutations). To test whether RNA-seq or ATAC-seq clusters (proximal ACRs or distal ACRs) more accurately predicted cell subsets, we generated a null distribution of IQR differences by permuting cell subtype labels across the combined dataset and calculated two-tailed P-values based on the left and right tails of simulated IQR difference distribution (two-sample permutation test, N=10^{4} permutations). For two-sided tests, we calculated a conditional two-sided p-value centered at the median as previously described (Kulinskaya, 2008).

This analysis was also performed to compare the sample clusters generated from bi-clustering with the sorted cell subset labels (Figure 2). Functions used for the IQR analysis are available: https://github.com/wherrylab/statistics_code/blob/master/MutualInformationMetricsForDiscreteCategoricalComparison.R.

Calculating TF activity (Taiji analysis)

The Taiji pipeline integrates diverse datasets to identify master regulators, including genome-wide expression profile and chromatin state. Herein, we implemented the pipeline described previously (http://wanglab.ucsd.edu/star/taiji). In brief, ATAC–seq peaks were called by MACS2 v.2.1.1 to annotate genome-wide regulatory elements and the regulatory elements are assigned to their nearest genes. Known transcription-factor motifs are scanned in the open chromatin region within each regulatory element to pinpoint the putative binding-sites. Transcription factors with putative binding-sites in promoters or enhancers are then linked to their target genes to form a network. As part of Taiji pagerank analysis, a personalized PageRank algorithm is used to assess the importance of transcription factors in the network and ranks are calculated for each transcription factor on the basis of epigenetic and RNA expression data. The normalized ranks are then compared across conditions by calculating fold change and the top transcription factors are chosen using a cut-off of 1.5× above the mean and having a gene expression of at least 50 normalized counts. These transcription factors are finally visualized in a heat map using the pheatmap R package.

Identification of epigenetic and transcriptional modules across differentiation trajectory

To identify genes and ACRs that exhibited the same pattern across CD8 differentiation, we first established a proposed trajectory, order of CD8 T cell subsets, using Pearson correlation and the pseudotime algorithm from the R package Slingshot: naïve → SCM-R3+ → CM → EM1 → EM2 → EMRA. For each gene and each ACR, differential gene expression analysis or accessibility was determined between each pairwise comparison along the trajectory (Naïve vs SCM-R3+, SCM-R3+ vs CM, CM vs EM1, EM1 vs EM2, EM2 vs EMRA), as described above – using the R package DESeq2 with adjusted p value < 0.05. If there was an increase in the pairwise comparison, the change between that comparison was scored as 1; if there was a decrease, the change was scored as -1; no statistical change was scored as 0. The relative gene expression and chromatin accessibility was determined by starting all genes at a baseline of 0, then adding these scores across the pairwise transitions as shown schematically in Figure 3D.

Predicting functional ACRs for target genes

To predict which ACR(s) regulate which gene(s), we performed multiple regression with gene expression level of a particular gene as outcome and ACRs within 250Kb of the corresponding gene as predictors (stepAIC in MASS R package) (Yoshida et al., 2019). We

tested genes that had a minimum median of 20 counts in at least one cell subset. We calculated the relative importance of each predicted peak (calc.relimp from the reliampo R package).

UMAP Analysis of melanoma TIL samples

To directly compare the melanoma TIL to the HD atlas, we first created a new union peak list by combining and merging (bedtools mergeBed) peak lists from each data set and calculated the number of reads in each peak (bedtools coverage). Batch correction was done with using the Combat function in the sva R package. To construct the HD reference UMAP, first we determined the top 5k (by adjusted p value) distal DAPs (> 2kb from nearest transcriptional start site, TSS) for each non-naïve HD CD8 T cell subset versus all as described above using with the R package DESeq2 (DESeq function). These features were then used to construct the UMAP (umap function, umap R package) with the HD non-naïve CD8 T cell subsets. The melanoma TIL non-naïve CD8 T cell samples were projected onto this UMAP using the predict function from base R.

To compare the location of each melanoma TIL sample to the HD subsets in UMAP, we first calculated the centroid of each HD subset. Next, the Euclidean distance between each TIL sample and each HD subset centroid was calculated; the TIL sample was assigned to the nearest HD subset. A binomial test was used to test the null hypothesis that melanoma TIL are just as likely to be near the $PD1^{+}CD39^{+}$ subset as all other subsets combined.

GWAS analysis

We used the GWAS catalog and EFO standardized term mapping downloaded from https://www.ebi.ac.uk/gwas/docs/file-downloads on 2/10/2021. The SNPs were filtered using the parental EFO ontology term "Immune system disorder", then a 25kb SNP window was generated by extending 12.5kb on either side. The R package, bedr, was used to determine the overlap between the SNP windows and HD T cell atlas ACRs; the R package, gviz, was used to visualize the genomic regions.

Quantitative PCR (qPCR)

For each sgRNA, three samples were sorted: two control samples (dCas-KRAB-mCherry^{+} sgRNA-GFP^{-}, dCas-KRAB-mCherry^{-} sgRNA-GFP^{+}) and one experimental sample (dCas-KRAB-mCherry^{+} sgRNA-GFP^{+}). For each sorted sample, the housekeeping cycle threshold (CT_{HG}) was calculated as the geometric mean of *RPL13A* and *TBP*. ΔCT was calculated by subtracting CT of the target or non-target gene (CT_{TG}) from CT_{HS} for each sorted sample. The control ΔCT (ΔCT_{CTL}) was calculated by taking the mean of control samples. Relative expression ($2^{-\Delta\Delta CT}$) was calculated as the difference between ΔCT from the experimental group (dCas-KRAB-mCherry^{+} sgRNA-GFP^{+}) (ΔCT_{EXP}) and ΔCT_{CTL}.