nature immunology

Early transcriptional and epigenetic regulation of CD8+ T cell differentiation revealed by single-cell RNA sequencing

Boyko Kakaradov[1,6], Janilyn Arsenio[2,6], Christella E Widjaja[2,6], Zhaoren He[1], Stefan Aigner[1], Patrick J Metz[2], Bingfei Yu[3], Ellen J Wehrens[3], Justine Lopez[2], Stephanie H Kim[2], Elina I Zuniga[3], Ananda W Goldrath[3], John T Chang[2,7] & Gene W Yeo[1,4,5,7]

During microbial infection, responding CD8+ T lymphocytes differentiate into heterogeneous subsets that together provide immediate and durable protection. To elucidate the dynamic transcriptional changes that underlie this process, we applied a single-cell RNA-sequencing approach and analyzed individual CD8+ T lymphocytes sequentially throughout the course of a viral infection *in vivo*. Our analyses revealed a striking transcriptional divergence among cells that had undergone their first division and identified previously unknown molecular determinants that controlled the fate specification of CD8+ T lymphocytes. Our findings suggest a model for the differentiation of terminal effector cells initiated by an early burst of transcriptional activity and subsequently refined by epigenetic silencing of transcripts associated with memory lymphocytes, which highlights the power and necessity of single-cell approaches.

Heterogeneity of cell fate is a hallmark of the responses of T lymphocytes to microbial infection. During an immune response to a microbial infection, responding naive T lymphocytes give rise to terminal effector cells that mediate acute host defense and self-renewing memory cells that provide long-lived protection. Terminally differentiated effector T cells are characterized by high expression of the killer lectin-like receptor KLRG1 and low expression of the interleukin 7 receptor (IL-7R)[1]. Circulating memory T cells can be categorized into two subsets, central memory T cells (T_{CM} cells) and effector memory T cells (T_{EM} cells), distinguished by differences in their expression of homing and cytokine receptors such as L-selectin (CD62L) and CCR7, their proliferative capacity and their anatomical localization[2]. A third subset of memory cells, tissue-resident memory T cells (T_{RM} cells), do not circulate but instead remain in peripheral tissues after pathogen clearance[3].

Transcriptional profiling approaches have greatly advanced understanding of the molecular regulation of the fate specification of T lymphocytes[4,5]. Through comparison of the gene expression of CD8+ T lymphocytes during the course of microbial infections, such studies have identified many transcription factors and pathways with a role in the specification of terminal differentiation versus long-lived memory[6]. However, most studies have been conducted on bulk populations of cells, which masks the potential heterogeneity among individual cells. Published work has sought to address these limitations through the use of single-cell quantitative RT-PCR analysis to investigate the gene-expression patterns of single CD8+ T lymphocytes responding to bacterial infection *in vivo*[7]. Although dynamic changes in gene expression in individual cells during differentiation have been identified, pre-selection of genes already known to encode products involved in differentiation for analysis precluded the discovery of as-yet unknown genes and molecular pathways.

Single-cell RNA sequencing (scRNA-seq) has emerged as a powerful tool that has substantially advanced the understanding of diverse biological processes, including development[8], the pathogenicity of the T_H17 subset of helper T cells[9], and innate immune responses[10]. In our study here, we applied a scRNA-seq approach to analyze transcriptome-wide changes in individual CD8+ T cells as they differentiated *in vivo*. We observed unexpected and substantial transcriptional heterogeneity among lymphocytes that had undergone their first division, which revealed two distinct subpopulations distinguished by their expression of hundreds of genes encoding products involved in diverse biological functions, including cell-cycle regulation, metabolism, effector function and fate specification. The expression of many transcription factors that have been linked to effector and memory cell differentiation, along with chromatin regulators, was markedly upregulated in cells differentiating along the terminal effector pathway and was extinguished by the peak of the adaptive immune response. That initial transcriptional program was subsequently refined by selective epigenetic repression of molecular determinants associated with memory-cell differentiation. In contrast, induction of the memory program seemed to be associated

[1]Department of Cellular and Molecular Medicine, University of California, San Diego, California, USA. [2]Department of Medicine, University of California, San Diego, California, USA. [3]Division of Biological Sciences, University of California, San Diego, California, USA. [4]Institute for Genomic Medicine, University of California, San Diego, California, USA. [5]Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. [6]These authors contributed equally to this work. [7]These authors jointly directed this work. Correspondence should be addressed to G.W.Y. (geneyeo@ucsd.edu) or J.T.C. (changj@ucsd.edu).

Received 9 July 2016; accepted 13 January 2017; published online 20 February 2017; doi:10.1038/ni.3688

with more nuanced alterations in the expression levels of a few specific genes. Together these findings provide unexpected new insights into the tightly coupled transcriptional and epigenetic mechanisms underlying the fate specification of CD8$^+$ T lymphocytes and highlight the power and necessity of single-cell approaches.

RESULTS

scRNA-seq of CD8$^+$ T cells differentiating *in vivo*

To investigate transcriptional changes in individual CD8$^+$ T lymphocytes responding to microbial infection *in vivo*[11], we adoptively transferred P14 CD8$^+$ T lymphocytes, which have transgenic expression of T cell antigen receptors that recognize an immunodominant epitope of lymphocytic choriomeningitis virus (LCMV), into congenic wild-type recipient mice 1 d before infection of the host mice with the Armstrong strain of LCMV. On days 2, 4 and 7 after infection, activated P14 CD8$^+$ T lymphocytes (CD44^{hi}) were enriched from spleens of infected mice by a magnetic-bead-based approach and were sorted by flow cytometry. On day 42 of infection, T_{CM} (CD44^{hi}CD62L^{hi}) cells and T_{EM} (CD44^{hi}CD62L^{lo}) cells were isolated in a similar way; naive P14 CD8$^+$ T cells (CD44^{lo}CD62L^{hi}) were also included in our analysis. For certain time points (days 2 and 4 after infection), large numbers of P14 cells were adoptively transferred into recipient mice[4,7,11] to enable the isolation of sufficient numbers of cells for scRNA-seq analysis; transfer of a large number of cells can alter the magnitude and kinetics of the immune response[12] and therefore represents a caveat of this study.

We performed PCR amplification of full-length, polyadenylated transcripts[13], followed by preparation and sequencing of single-cell cDNA libraries (**Fig. 1a**). 10×10^6 to 20×10^6 reads per cell were achieved (**Supplementary Fig. 1a**), with slight variations among populations, with 60–90% uniquely mapped reads (**Supplementary Fig. 1b**,**c**). At least two technical replicates for each cell population

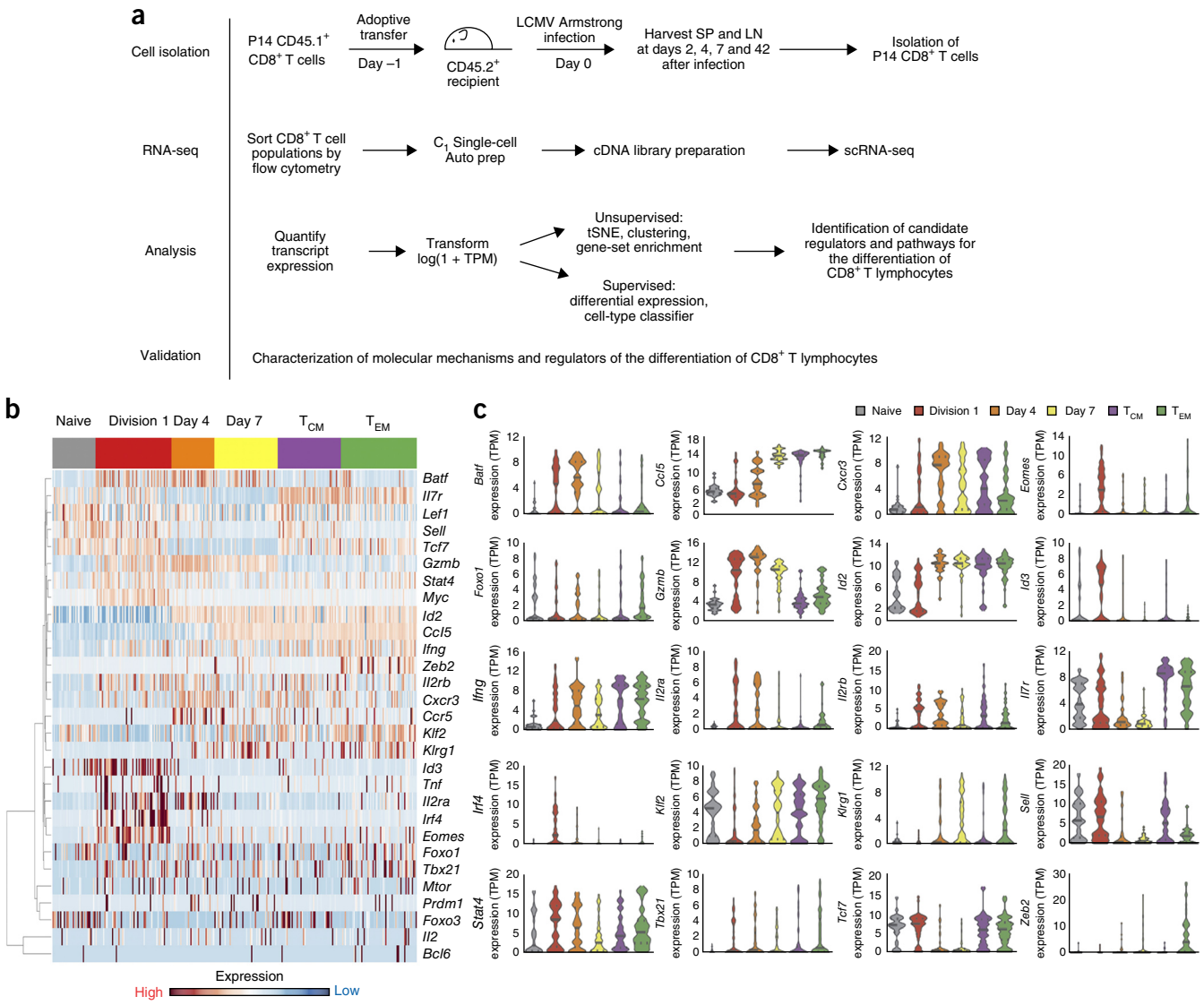


Figure 1 scRNA-seq analysis of CD8$^+$ T lymphocytes responding to viral infection. (**a**) Experimental and analytical approaches. SP, spleen; LN, lymph nodes. (**b**) Single-cell expression of selected genes (right margin) encoding products associated with the differentiation of CD8$^+$ T cells, by cells obtained from the spleen of wild-type recipient mice as in **a** (top row), with sorting before and on days 2, 4, 7 and 42 after infection for analysis of naive cells, division 1 cells, day 4 cells, day 7 cells, T_{CM} cells and T_{EM} cells (colors (above plot) match those in **c**); left margin (brackets), hierarchical clustering. (**c**) Expression of selected genes among those in **b** (vertical axes) by cells as in **b** (key), presented as transcripts per million reads (TPM) in 'violin plots'. Data are representative of at least two experiments per time point.

were assessed on separate sequencing plates to ensure reproducibility and an absence of batch effects (**Supplementary Fig. 1d–f**). After undertaking these quality-control measures, we included 288 single-cell libraries separated into 224 unique sequencing samples and 32 pairs of duplicates in further in-depth analyses.

We detected over 6,000 genes with a mean expression of at least 1 transcript per million reads (TPM) per cell. Assessment at the single-cell level of a subset of genes encoding products previously linked to CD8$^+$ T cell differentiation by analyses of bulk populations suggested patterns of expression consistent with their previously assigned roles in this process. For example, genes encoding products associated with the differentiation and function of effector cells, such as *Batf*, *Id2* and *Gzmb*, had high expression in cells isolated at days 4 and 7 after infection (**Fig. 1b**,**c**), time points at which terminally differentiated effector cells are known to predominate. Conversely, genes encoding memory-associated products, such as *Tcf7* and *Il7r*, had high expression in T_{CM} cells and T_{EM} cells (**Fig. 1b**,**c**). These findings demonstrated that the expression of genes previously

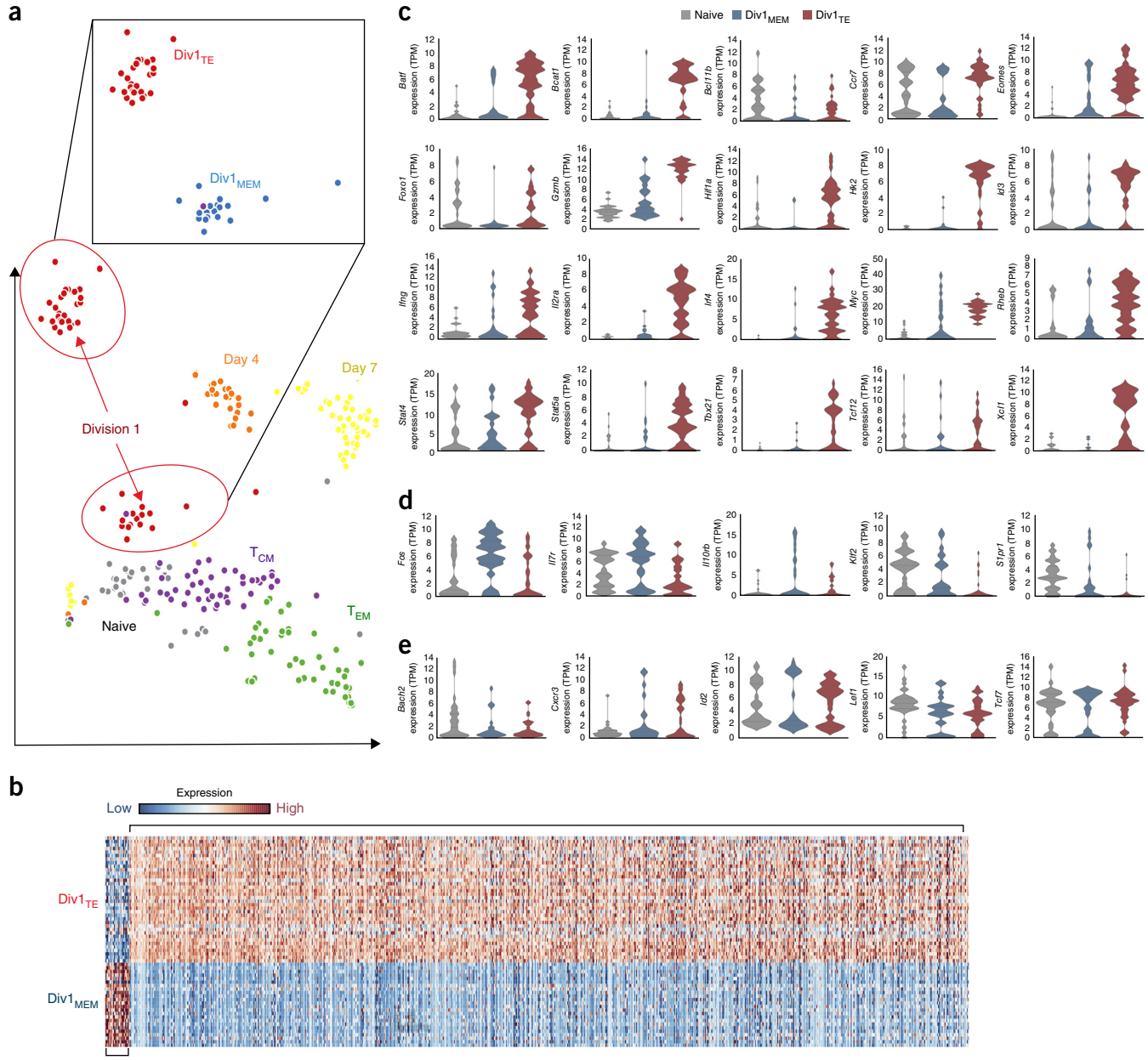


Figure 2 Cells that have undergone their first division exhibit transcriptional heterogeneity. (**a**) tSNE analysis of 6,000 genes expressed by single CD8$^+$ T cells (obtained as in **Fig. 1b**) among naive cells (gray), division 1 cells (red), day 4 cells (orange), day 7 cells (yellow), T_{CM} cells (purple), and T_{EM} cells (green); inset, tSNE clustering analysis of division 1 cells in the plot below, identifying $Div1_{TE}$ cells (red) and $Div1_{MEM}$ cells (blue). Each symbol represents an individual cell. (**b**) Differential expression of genes (one per column) by $Div1_{TE}$ cells (top plot) and $Div1_{MEM}$ cells (bottom plot) (one cell per row); brackets indicate genes with higher expression in $Div1_{TE}$ cells than in Div 1_{MEM} cells (top right) or lower expression in $Div1_{TE}$ cells than in $Div1_{MEM}$ cells (bottom left). (**c**–**e**) Expression of selected genes encoding products associated with the differentiation of CD8$^+$ T cells, by individual naive cells, $Div1_{TE}$ cells and $Div1_{MEM}$ cells (key), grouped as genes with higher expression in $Div1_{TE}$ cells than in Div 1_{MEM} cells (**c**), with lower expression in $Div1_{TE}$ cells than in $Div1_{MEM}$ cells (**d**) or with equal expression in $Div1_{TE}$ cells and $Div1_{MEM}$ cells (**e**) (presented as in **Fig. 1c**). Data are representative of two experiments.

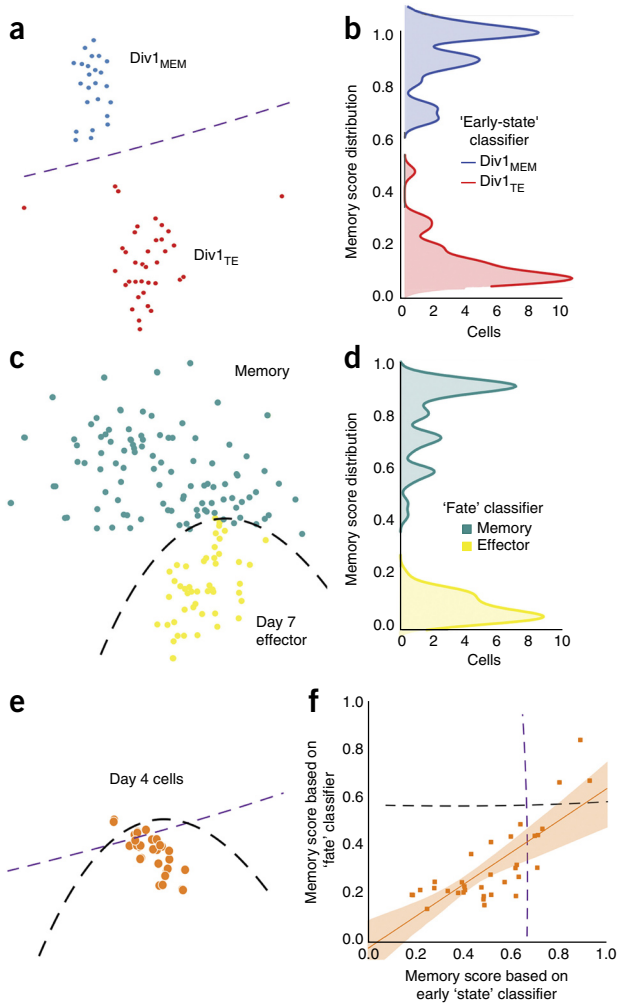


Figure 3 Generation and application of 'early-state' and 'fate' classifiers to predict the identity of cells in intermediate states of differentiation. (**a**,**c**) Classifier separation of $Div1_{MEM}$ cells (blue; $n = 24$ cells) from $Div1_{TE}$ cells (red; $n = 36$ cells) (**a**) and day 7 effector cells (yellow; $n = 48$ cells) from total memory cells (teal; $n = 96$ cells), which include T_{CM} cells ($n = 48$) and T_{EM} cells ($n = 48$) (**c**) (all cells obtained as in **Fig. 1**). (**b**,**d**) Cross-validated scores of division 1 $CD8^+$ T lymphocytes (**b**) and day 7 effector and memory $CD8^+$ T lymphocytes (**d**) on which the early-state classifier (**b**) and fate classifier (**d**) were trained; results are presented as kernel density histograms. (**e**) Application of early-state and fate classifiers to predict the fate of individual day 4 $CD8^+$ T lymphocytes ($n = 34$); black and purple dashed lines indicate the boundary between predicted memory-like or effector-like day 4 cells. (**f**) Prediction analysis of individual day 4 $CD8^+$ T lymphocytes as assessed in **e**, presented as the memory-score distribution (from 0 (effector) to 1 (memory)) of the early-state classifier (horizontal axis) versus that of final-fate classifier (vertical axis): dashed black and purple lines, fate classifier's 'decision boundary' between memory cells and day 7 effector cells, and orange line, early-state classifier's 'decision boundary' between $Div1_{MEM}$ cells and $Div1_{TE}$ cells (orange shaded area, 95% confidence interval, with the assumption of Gaussian error); both sets of lines are 'stylized estimates' of the real boundaries, which are complex 'piecewise-linear functions' and can be much more furrowed. Pearson correlation (linear), $r = 0.78$ ($P = 4.8 \times 10^{-8}$); Spearman correlation (monotonic), $r = 0.71$ ($P = 2.2 \times 10^{-6}$). Each symbol (**a**,**c**,**e**,**f**) represents an individual cell. Data are representative of two experiments.

described as encoding effector- and memory-associated products were readily detected in the expected lymphocyte subsets by a single-cell approach.

Notably, our single-cell analysis also revealed patterns of gene expression that were not previously discernable by analyses of bulk populations. For example, heterogeneous expression of many genes, including *Tbx21*, *Gzmb*, *Id3*, *Il7r*, *Il2ra*, *Sell*, *Eomes* and *Irf4*, was observed among individual cells derived from the same time point that would otherwise have been interpreted as identical by analyses of bulk populations (**Fig. 1b**,**c**). Thus, our single-cell transcriptomic analysis captured the substantial heterogeneity in gene expression exhibited by individual $CD8^+$ T lymphocytes throughout their differentiation in response to microbial infection.

Molecular heterogeneity among division 1 $CD8^+$ T cells

We performed unsupervised tSNE ('t-distributed stochastic neighborhood embedding') clustering analysis to visualize, in an unbiased manner, the gene-expression patterns (**Fig. 2a**) of individual $CD8^+$ T lymphocytes isolated at all time points after infection (as in **Fig. 1**). In this tSNE analysis, naive cells, T_{CM} cells and T_{EM} cells each formed distinct clusters (**Fig. 2a**), suggestive of unique molecular homogeneity within each population (**Supplementary Table 1**). Similarly, most cells at day 4 and day 7 formed their own separate clusters; however, a few cells from each time point grouped near the naive cell and T_{CM} cell populations (**Fig. 2a**). Strikingly, unsupervised tSNE analysis revealed two distinct subpopulations among single $CD8^+$ T lymphocytes that had undergone their first cell division (division 1) (**Fig. 2a**). We emphasize that the division 1 cells were isolated on the basis of their dilution of the division-tracking dye CFSE (the second CFSE peak) and not by their phenotypic expression of cell-surface markers, other than by high expression of the activation marker CD44 to ensure that all sorted cells had been activated *in vivo*. One subpopulation of division 1 cells seemed to be most similar to day 4 and day 7 effector $CD8^+$ T lymphocytes, whereas the other subpopulation of division 1 cells seemed to be most similar to T_{CM} cells and naive cells (**Fig. 2a**). We provisionally designated these two division 1 (Div1) subpopulations '$Div1_{TE}$' and '$Div1_{MEM}$' on the basis of their similarities to terminal effector-cell populations ('TE') and memory-cell populations ('MEM'), respectively.

Gene-expression analysis revealed that differences in the expression of 930 genes distinguished $Div1_{TE}$ cells from $Div1_{MEM}$ cells, with 903 genes having higher expression in $Div1_{TE}$ cells than in $Div1_{MEM}$ cells and 27 genes having higher expression in $Div1_{MEM}$ cells than in $Div1_{TE}$ cells (**Fig. 2b** and **Supplementary Table 2**). Gene-ontology analysis revealed that the set of genes upregulated by $Div1_{TE}$ cells showed enrichment for those encoding products involved in diverse molecular and cellular processes involving transcription, protein transport, ribosome biogenesis, cell division and mRNA processing, among others (**Supplementary Table 3**). Moreover, $Div1_{TE}$ cells had higher expression of genes encoding transcription factors, cytokine receptors and signaling molecules previously associated with the differentiation and metabolic reprogramming of terminal effector cells[6] (**Fig. 2c**).

In contrast, $Div1_{MEM}$ cells had higher expression of genes encoding several factors that have been previously associated with memory lymphocytes[6], including *Il7r*, *S1pr1* and *Klf2* (**Fig. 2d**), in addition to genes encoding molecules not previously associated with such cells, including *Btg1*, which encodes an anti-proliferative molecule[14] (**Supplementary Table 2**). Notably, however, many genes encoding transcription factors associated with memory lymphocytes either had similar expression in $Div1_{TE}$ cells and $Div1_{MEM}$ cells (*Lef1*, *Bach2* (ref. 15) and *Tcf7*) (**Fig. 2e**) or had higher expression in $Div1_{TE}$ cells than in $Div1_{MEM}$ cells (*Eomes*, *Id3* and *Foxo1*) (**Fig. 2c**). Together these data demonstrated that $CD8^+$ T lymphocytes that had undergone their first cell division exhibited considerable transcriptional

heterogeneity that was not previously discernible[7], with the vast majority of differentially expressed genes upregulated by $Div1_{TE}$ cells, including many transcripts encoding products previously associated with memory lymphocytes.

Identification of cells in intermediate stages of differentiation

The observation that two distinct subpopulations of cells that had undergone their first division *in vivo* could be discerned by virtue of disparate gene-expression patterns suggested that these two subpopulations might represent cells that had already begun to diverge in fate. We sought to determine whether we could systematically predict the identity of cells in subsequent, intermediate stages of differentiation (analyzing cells obtained as in **Fig. 1**). We hypothesized that using two distinct supervised classifiers, one trained on the two division 1 subpopulations ('early-state' classifier; **Fig. 3a**,**b**) and the other trained on true memory cell populations (T_{CM} cells and T_{EM} cells) and terminal effector-cell populations ('fate' classifier; **Fig. 3c**,**d**), would enable us to identify cells in intermediate states of differentiation as they progressed toward a terminally differentiated fate versus a long-lived memory fate.

Using both the early-state classifier and fate classifier, we then developed a 'future-past' computational approach to independently predict the identity of cells in intermediate states of differentiation (**Fig. 3e**). The early-state classifier, trained on division 1 cells, was deployed into the 'future' on day 4 cells at intermediate states of differentiation, whereas the fate classifier, trained on day 7 effector

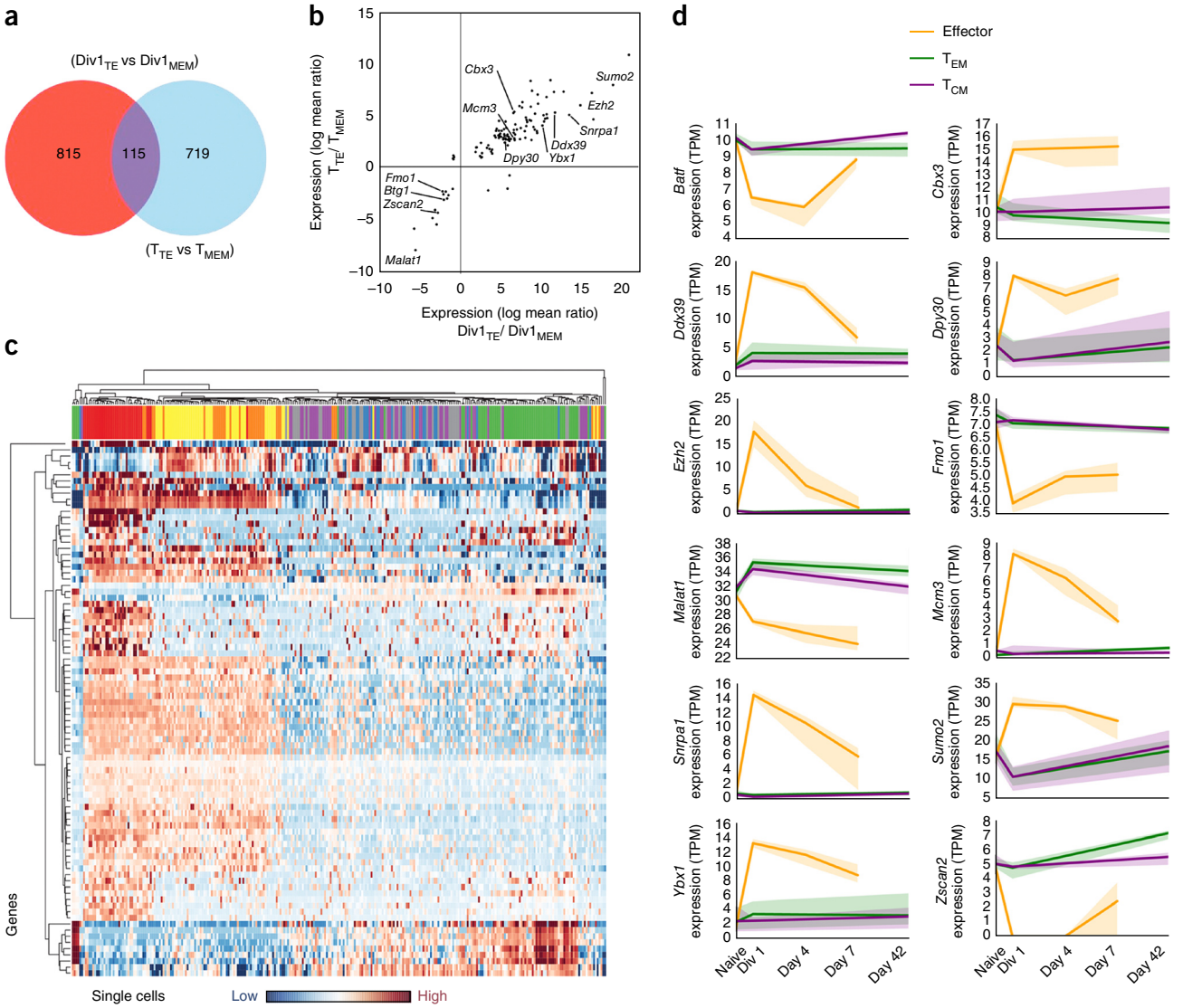


Figure 4 Identification of putative regulators of the differentiation of $CD8^+$ T lymphocytes. (**a**) Overlap (middle) of genes expressed differentially by $Div1_{TE}$ cells relative to their expression by $Div1_{MEM}$ cells (left) and those expressed differentially by terminally differentiated effector T cells (T_{TE} cells) relative to their expression by memory T cells (T_{MEM} cells) (right) (all cells obtained as in **Fig. 1**). (**b**) Plotting of the difference in expression of genes expressed differentially by effector cells versus memory cells (as in **a**; T_{TE}/T_{MEM}) against that of genes expressed differentially by $Div1_{TE}$ cells versus $Div1_{MEM}$ cells ($Div1_{TE}/Div1_{MEM}$). Pearson correlation, $r = 0.78$ ($P = 3.6 \times 10^{-13}$). Each symbol represents an individual gene. (**c**) Differential expression of 89 common genes (one per row) by all $CD8^+$ T lymphocyte populations of naive cells (gray), $Div1_{TE}$ cells (red), $Div1_{MEM}$ cells (blue), day 4 cells (orange), day 7 cells (yellow), T_{CM} cells (green) and T_{EM} cells (purple) (one cell per column); left margin (brackets), hierarchical clustering. (**d**) Temporal expression patterns of selected genes across inferred paths of differentiation for effector cells (orange), T_{CM} cells (purple) and T_{EM} cells (green); shaded areas, 95% confidence interval (bootstrapped from all possible single-cell expression trajectories). Data are representative of two experiments.

cells and day 40 memory cells, was deployed into the 'past' to analyze the same day 4 cells. Notably, the early-state classifier and fate classifier agreed on the identity of intermediate cells (**Fig. 3f**), yet they used largely non-overlapping sets of genes for the prediction of pre-terminal effector or pre-memory cell states (**Supplementary Tables 4** and **5**). Thus, intermediate states of differentiation could be readily predicted through the use of supervised binary classifiers trained on the preceding or subsequent states.

Identification of regulators of CD8$^+$ T cell differentiation

We next sought to identify previously unknown regulators of CD8$^+$ T cell differentiation by searching for commonality between the set of genes expressed differentially by Div1$_{TE}$ cells relative to their expression in Div1$_{MEM}$ cells and by terminal effector cells relative to their expression in memory cells (analyzing cells obtained as in **Fig. 1**). Among the set of 930 genes expressed differentially by Div1$_{TE}$ cells versus Div1$_{MEM}$ cells (**Supplementary Table 2**) and the set of 834 genes expressed differentially by terminal effector cells versus memory cells (**Supplementary Table 6**), only 115 genes were shared by both sets (**Fig. 4a**,**b**). We next selected genes common only to Div1$_{TE}$ cells and effector cells or common only to Div1$_{MEM}$ cells and memory cells to identify genes encoding regulators of terminal-effector-cell differentiation or memory-cell differentiation, respectively.

The analysis described above yielded 89 putative regulators of CD8$^+$ T cell fate specification (**Supplementary Table 7**). We visualized their temporal expression patterns by clustering single-cell expression of the genes encoding these putative regulators across all time points (**Fig. 4c**). Because application of the 'future-past' binary classifiers enabled us to predict the identity of cells at intermediate time points, we were able to infer pathways of terminal-effector-cell or memory-cell differentiation on the basis of the expression of genes encoding these putative regulators in day 4 cells classified as being in either a pre-memory state or a pre-effector state of differentiation (**Fig. 4d** and **Supplementary Fig. 2**). We observed that genes expressed in cells differentiating along the terminal-effector-cell pathway tended to exhibit a substantial increase in expression at the first division, followed by a rapid decrease in expression (**Fig. 4d**); this raised the possibility of epigenetic repression in differentiating effector cells. We selected *Ezh2*, which encodes a catalytic subunit of the Polycomb complex PRC2 that mediates gene repression by mediating trimethylation of histone H3 at Lys27 (H3K27me3)[16], for further analysis and functional validation, given the results of published studies suggesting a critical role for Ezh2 in the differentiation and function of CD4$^+$ T cells[17–19].

Ezh2, along with genes encoding other components of the PRC2 (*Eed*, *Suz12* and *Set*), had higher expression in Div1$_{TE}$ cells than in Div1$_{MEM}$ cells (**Fig. 5a**), suggestive of a role for Ezh2 in regulating the differentiation of terminal effector cells. Consistent with that finding, we observed that CD8$^+$ T cells that had undergone their first division exhibited a bimodal pattern in their expression of Ezh2 protein (**Fig. 5b**), with high expression and low expression in putative 'pre-effector' IL-2Rα^{hi}CD62L^{lo} cells and 'pre-memory' IL-2Rα^{lo}CD62L^{hi} cells, respectively[7] (**Fig. 5c**). Moreover, the kinetics of Ezh2 expression during differentiation at the protein level paralleled that at the mRNA level (**Fig. 5d**).

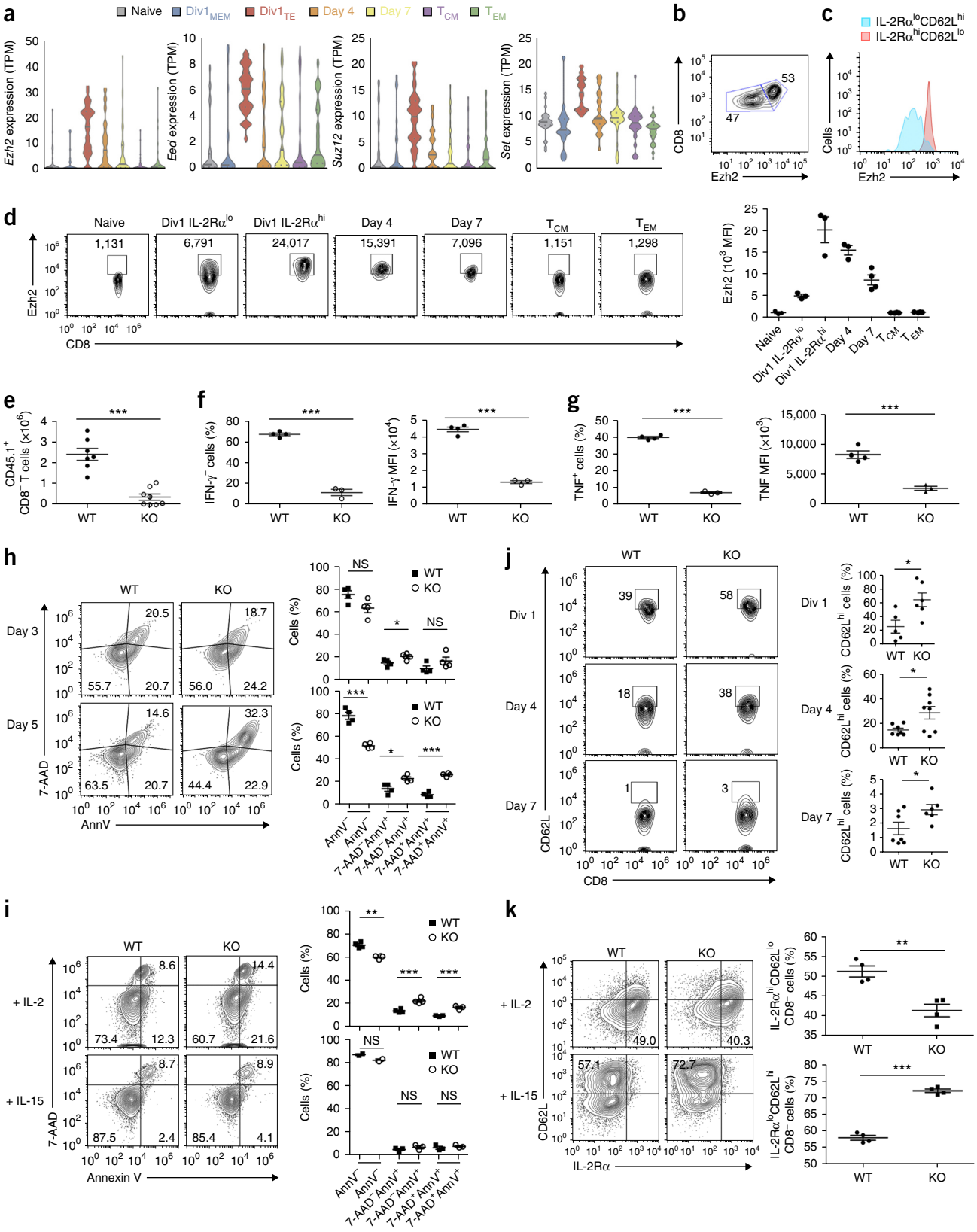
We next generated *Ezh2*$^{fl/fl}$*Cd4*^{Cre} P14 mice (in which *lox*P-flanked *Ezh2* alleles (*Ezh2*$^{fl/fl}$) are deleted by Cre recombinase expressed from the T cell–specific *Cd4* promoter) and adoptively transferred Ezh2-deficient (*Ezh2*$^{fl/fl}$*Cd4*^{Cre}) or control (*Ezh2*$^{fl/fl}$*Cd4*$^{+/+}$) P14 (CD45.1$^+$) CD8$^+$ T cells (**Supplementary Fig. 3a**) into wild-type (CD45.2$^+$) recipient mice that we subsequently infected with LCMV-Armstrong. Ezh2-deficient CD8$^+$ T cells were much less abundant by days 5 and 7 after infection than were control cells (**Fig. 5e** and **Supplementary Fig. 3b**) and exhibited an impaired ability to secrete inflammatory cytokines relative to that of control cells (**Fig. 5f**,**g**). Notably, the absence of effector cells (**Fig. 5e**) was not due to a failure of Ezh2-deficient CD8$^+$ T cells to undergo activation or proliferation (**Supplementary Fig. 3c**,**d**). However, Ezh2 deficiency was associated with increased apoptosis by day 5 after infection (**Fig. 5h**) and 'preferentially' affected early 'effector-like' cells, but not 'memory-like' cells, in an *in vitro* model of CD8$^+$ T cell differentiation[20,21] (**Fig. 5i–k** and **Supplementary Figs. 3e** and **4**). Together these findings suggested a critical role for Ezh2 in regulating the differentiation of terminal effector cells.

Epigenetic silencing of memory-associated determinants

Because the PRC2 complex mediates transcriptional repression, we hypothesized that the high expression of Ezh2 in Div1$_{TE}$ cells would catalyze repressive H3K27me3 marks on a set of key genes and thereby promote the differentiation of terminal effector cells. Focusing on the ~6,000 genes detected by our scRNA-seq analysis, we mapped H3K27me3 peaks derived from the analysis of naive, terminal effector (KLRG1^{hi}IL-7Rα^{lo}) CD8$^+$ T cells and total memory (CD44^{hi}) CD8$^+$ T cells by chromatin immunoprecipitation followed by sequencing

Figure 5 Ezh2 regulates effector CD8$^+$ T lymphocyte differentiation. (**a**) Expression of *Ezh2*, *Eed*, *Suz12* and *Set* (which encode components of the PRC2 complex) in single CD8$^+$ T lymphocytes (obtained as in **Fig. 1b**) separated as naive cells, Div1$_{MEM}$ cells, Div1$_{TE}$ cells, day 4 cells, day 7 cells, T$_{CM}$ cells and T$_{EM}$ cells (key) (presented as in **Fig. 1c**). (**b**) Flow cytometry of gated CD8$^+$ T lymphocytes (as in **a**) that had undergone their first cell division *in vivo*. Numbers adjacent to outlined areas indicate percent Ezh2^{lo} cells (left) or Ezh2^{hi} cells (right) among the CD8$^+$ T lymphocytes. (**c**) Flow cytometry of gated division 1 (Div1) CD8$^+$ IL-2Rα^{hi}CD62L^{lo} or IL-2Rα^{lo}CD62L^{hi} cells (key), assessing Ezh2 expression. (**d**) Flow cytometry (left) of naive cells, division 1 cells (gated on IL-2Rα^{lo} or IL-2Rα^{hi} cells), day 4 cells, day 7 cells, T$_{CM}$ cells and T$_{EM}$ cells (above plots) among CD8$^+$ T lymphocytes (as in **a**); right, summary of results at left. Numbers adjacent to outlined areas (left) indicate mean fluorescence intensity of Ezh2$^+$ events. (**e**) Absolute number of gated *Ezh2*$^{fl/fl}$*Cd4*$^{+/+}$ (WT) or *Ezh2*$^{fl/fl}$*Cd4*^{Cre} (KO) P14 CD8$^+$ T cells in the spleen of wild-type (CD45.2$^+$) recipient mice given adoptive transfer of *Ezh2*$^{fl/fl}$*Cd4*$^{+/+}$ or *Ezh2*$^{fl/fl}$*Cd4*^{Cre} P14 (CD45.1$^+$) CD8$^+$ T cells, followed by infection of the host mice with LCMV-Armstrong and analysis of cells 7 d after infection. (**f**,**g**) Flow cytometry of cells as in **e** 7 d after infection, showing the frequency of cells (left) with intracellular expression of IFN-γ (**f**) or TNF (**g**), among CD45.1$^+$ (donor) CD8$^+$ cells, and the mean fluorescence intensity (right) of intracellular IFN-γ (**f**) or TNF (**g**). (**h**) Flow cytometry (left) of gated CD45.1$^+$ (P14) CD8$^+$ cells as in **e**, analyzing staining of the DNA-intercalating dye 7-AAD and the apoptosis marker annexin V (AnnV) at 3 or 5 d after infection (left margin); right, summary of results at left. Numbers adjacent to outlined areas (left) indicate percent cells in each. (**i**) Flow cytometry (as in **h**; left) of gated IL-2Rα^{hi}CD62L^{lo} and IL-2Rα^{lo}CD62L^{hi} cells among cells as in **e**, assessed after 3 d of *in vitro* culture with IL-2 (top) or IL-15 (bottom); right, summary of results at left. (**j**) Flow cytometry (left) of division 1, day 4 and day 7 CD8$^+$ T cells as in **e**; right, summary of results at left. Numbers adjacent to outlined areas (left) indicate percent CD62L^{hi} cells among CD8$^+$ T cells. (**k**) Flow cytometry (left) of cells treated as in **i**; right, summary of results at left. Numbers in quadrants (left) indicate percent CD62L^{lo}IL-2Rα^{hi} cells (bottom right) or IL-2Rα^{lo}CD62L^{hi} cells (top left). Each symbol (**d**–**k**) represents an individual mouse; small horizontal lines indicate the mean (+ s.e.m.). NS, not significant ($P < 0.05$); *$P \geq 0.05$, **$P < 0.01$ and *** $P < 0.001$ (Student's two-tailed *t*-test). Data are representative of two experiments (**a**), two experiments with four (**b**,**c**) or three (**d**) mice in each (**b**–**d**) or four (**f**–**h**,**k**) or three (**i**) mice per genotype in each (**f**–**i**,**k**), or are pooled from two independent experiments (**e**,**j**).

(ChIP-seq)[22]. The promoter regions in terminal effector cells exhibited significant gains in H3K27me3 coverage that correlated with lower gene expression, whereas those in memory cells exhibited considerable losses in H3K27me3 coverage, relative to such coverage in naive cells (**Fig. 6a–c** and **Supplementary Tables 8** and **9**): among 1,234 genes with lower expression in terminal effector cells than in naive cells, 97 (7.86%) were marked by H3K27me3; among 627 genes with lower expression in memory cells than in naive cells, 11 (1.75%)



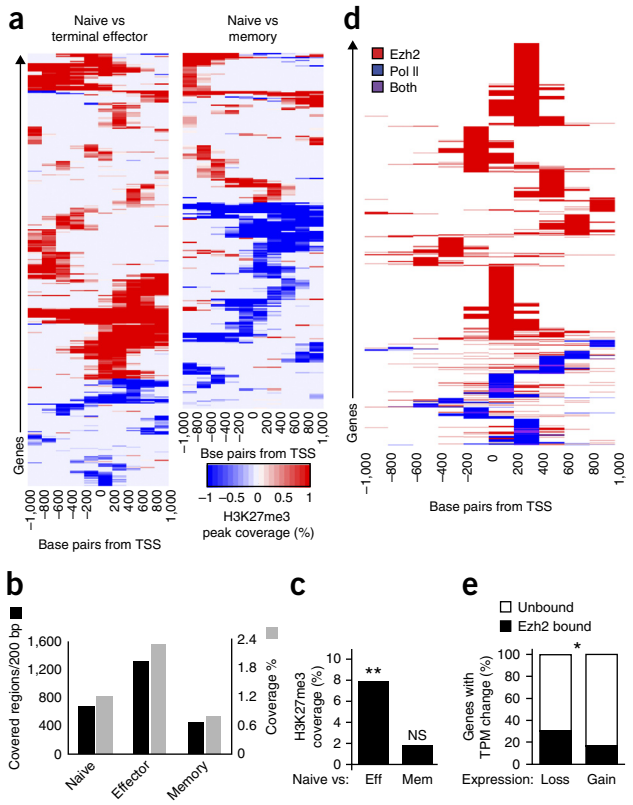


Figure 6 Increased epigenetic repression of genes expressed during the differentiation of terminal effector cells. (**a**) H3K27me3 peak coverage in various genes (rows) at various distances (columns) from the transcription start site (TSS), in naive cells versus terminal effector cells (left) or in naive cells versus memory cells (right); results are presented (key) as coverage gains (red) or losses (blue) as naive cells differentiate into those fates. (**b**) Peak H3K27me3 coverage around the transcription start site for the 6,000 genes detected in the scRNA-seq data set (**Fig. 1**), presented as quantification of covered regions (left vertical axis; black) or proportion of coverage (right vertical axis; gray). (**c**) H3K27me3 coverage of genes with lower expression in effector cells (Eff) or memory cells (Mem) than in naive cells. (**d**) ChIP analysis of the binding of Ezh2 or RNA polymerase II (PolII) or both (key) at proximal promoter regions (from −1 kb to +1 kb around the TSS) of the 6,000 genes detected by scRNA-seq (**Fig. 1**). (**e**) Frequency of genes with a change in TPM among those with lower expression (Loss; 1,492 genes) or higher expression (Gain; 219 genes) in day 4 cells than in $Div1_{TE}$ cells (horizontal axis; all cells obtained as in **Fig. 1**), with (Ezh2 bound) or without (Unbound) binding of Ezh2 (key). $^{*}P < 0.01$ and $^{**}P < 0.001$ (Fisher's exact test). Data are representative of three experiments (**a**–**c**) or two experiments (**d**,**e**).

were marked by H3K27me3. This suggested that epigenetic silencing might be more crucial for the differentiation of terminal effector cells than for that of memory cells.

We next investigated binding of Ezh2 and H3K27me3 marks in differentiating $CD8^+$ T cells by ChIP-seq and found 1,564 genes bound by Ezh2 and 261 genes marked by H3K27me3 (**Fig. 6d** and **Supplementary Table 10**). Gene-ontology analysis revealed that the set of genes targeted by Ezh2 showed enrichment for those encoding products involved in processes involving transcriptional regulation, cytoskeletal protein binding, phosphoinositide binding, Wnt receptor signaling, and apoptosis (relative to their abundance in sets of genes encoding products involved in other biological processes) (**Supplementary Table 11**). We observed that among cells obtained as described initially above (**Fig. 1**), genes with expression that was lower in day 4 cells than in $Div1_{TE}$ cells were more likely to exhibit Ezh2 binding than those with higher expression in day 4 cells than in $Div1_{TE}$ cells (**Fig. 6e**). Comparison of the expression patterns of H3K27me3-marked and unmarked genes during terminal-effector-cell and memory-cell differentiation revealed that H3K27me3-marked genes tended to exhibit more repression during the differentiation of terminal effector cells (as indicated by shifting of distribution; **Fig. 7a**, top, and **Supplementary Fig. 5a**) than during the differentiation of memory cells (as indicated by no shifting of distribution; **Fig. 7a**, bottom, and **Supplementary Fig. 5b**,**c**). These findings raised the possibility that the same set of genes were being selectively repressed during the differentiation of terminal effector cells but not during the differentiation of memory cells, due in part to differential expression of Ezh2 in cells progressing along these disparate pathways.

We investigated that possibility by analyzing a subset of Ezh2 targets whose expression decreased during the differentiation of effector cells but not during the differentiation of memory cells. This analysis revealed that many genes encoding products that have been linked to the differentiation of memory cells but not to the differentiation of effector cells exhibited greater association with Ezh2 than that of the input DNA control (**Fig. 7b** and **Supplementary Table 10**). These Ezh2 targets included genes encoding memory-associated transcription factors, such as *Tcf7* and *Eomes;* molecules that mediate TGF-β signaling, such as *Smad2*, whose product has been linked to $CD8^+$ T cell fate 'decisions'[23–25]; metabolic regulators, such as *Bcat1*, which encodes a branched-chain aminotransferase isoenzyme[26]; and factors that control T cell survival and homing, such as *Klf2* (refs. 27–29); as well as *Opa1*, which encodes a regulator of mitochondrial fusion with a critical role in differentiating memory $CD8^+$ T lymphocytes[30] (**Fig. 7b** and **Supplementary Table 10**). Several of these genes, such as *Foxo1* and *Tcf7*, underwent a rapid decrease in expression following the first division; others, like *Eomes* and *Id3*, underwent an initial increase at the first division followed by a rapid decrease (**Fig. 7c**); this suggested a possible role for the products of these memory-cell-associated genes in the differentiation of effector cells. In contrast, analysis of those same genes in differentiating memory cells revealed a distinct expression pattern characterized by a modest increase at the first division followed by stable or increased expression as the cells became mature T_{CM} cells or T_{EM} cells (**Fig. 7c**). A similar pattern of expression was observed for memory-cell-associated genes, such as *Il7r*, *Lef1* and *Bcl2*, that were not targeted by Ezh2 (**Fig. 7d**). Finally, we observed that Ezh2 deficiency was associated with diminished H3K27me3 coverage and increased expression of mRNA from many genes, including a number of memory-cell-associated genes, such as *Eomes*, *Tcf7* and *Klf2* (**Fig. 7e**–**h**, **Supplementary Figs. 6**–**8** and **Supplementary Tables 12** and **13**), consistent with the proposal of a role for H3K27me3-mediated transcriptional repression by Ezh2. Thus, the unique expression patterns of memory-cell-associated genes in differentiating terminal effector cells and memory cells might have resulted in part from the presence or absence of epigenetic repression owing to distinct levels of Ezh2 expression in these cells.

In parallel with our analysis of memory-cell-associated genes, we also assessed the expression patterns of genes encoding products previously associated with effector-cell differentiation, including those encoding transcription factors (*Batf*, *Irf4* and *Tbx21*), signaling molecules (*Il2ra* and *Akt1*) and metabolic regulators (*Hif1a* and *Myc*), along inferred terminal-effector-cell or memory-cell pathways. The expression patterns of these genes in differentiating effector cells resembled those of memory-cell-associated genes targeted by Ezh2, with a substantial early increase at the first division followed by a rapid loss by day 4 after infection (**Fig. 7i**); in contrast, these

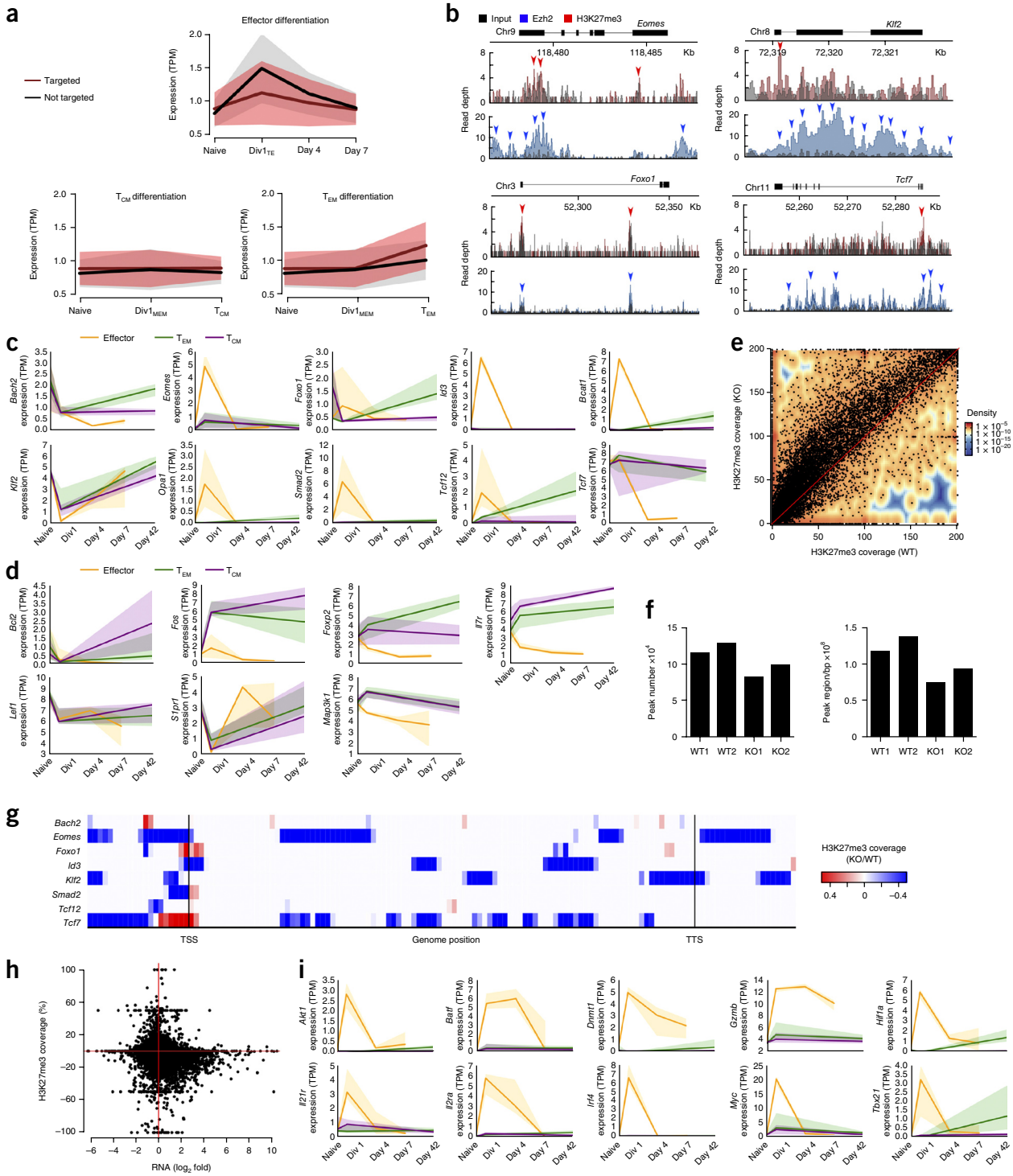


Figure 7 Ezh2 mediates the effector differentiation of CD8⁺ T lymphocytes through epigenetic repression. (**a**) Expression (averaged normalized TPM values) of genes encoding products in the three inferred differentiation pathways (effector, T_{CM} or T_{EM}; inferred as in **Fig. 4d**), classified as genes marked (Targeted) or not marked (Not targeted) by H3K27me3 (key), assessed in cells obtained as in **Figure 1** (horizontal axes). (**b**) ChIP-seq analysis of the binding of Ezh2 and H3K27me3 at *Eomes*, *Foxo1*, *Klf2* and *Tcf7* (gray indicates input); arrowheads indicate binding peaks for Ezh2 (red) or H3K27me3 (blue). (**c**,**d**) Temporal expression patterns (time relative to infection as in **Fig. 1**) of selected memory-cell-associated genes encoding products linked to CD8⁺ T cell differentiation, targeted by Ezh2 (**c**) or not (**d**), in inferred paths of differentiation for effector cells, T_{CM} cells and T_{EM} cells (key); shaded areas, 95% confidence interval (bootstrapped from all possible single-cell expression trajectories). (**e**) H3K27me3 coverage (key) for individual genes in *Ezh2*$^{fl/fl}$*Cd4*$^{+/+}$ (WT) and *Ezh2*$^{fl/fl}$*Cd4*^{Cre} (KO) cells. (**f**) H3K27me3 coverage in *Ezh2*$^{fl/fl}$*Cd4*$^{+/+}$ and *Ezh2*$^{fl/fl}$*Cd4*^{Cre} cells (two biologic replicates each (1, 2)), presented as quantification of H3K27me3 binding peaks (left) or regions (right). (**g**) H3K27me3 coverage of memory-cell-associated genes targeted by Ezh2, in *Ezh2*$^{fl/fl}$*Cd4*^{Cre} CD8⁺ T cells relative to that in *Ezh2*$^{fl/fl}$*Cd4*$^{+/+}$ CD8⁺ T cells (KO/WT). (**h**) H3K27me3 coverage (ratio as in **g**) plotted against RNA expression (of all genes) in *Ezh2*$^{fl/fl}$*Cd4*^{Cre} CD8⁺ T cells relative to that in *Ezh2*$^{fl/fl}$*Cd4*$^{+/+}$ CD8⁺ T cells. (**i**) Temporal expression patterns (as in **c**) of selected genes associated with effector differentiation. Data are representative of two experiments.

same genes were expressed at much lower levels during the process of memory differentiation (**Fig. 7i**). The observation that these effector-cell-associated genes were not targeted by Ezh2 or marked by H3K27me3 suggested that the regulation of these genes during differentiation might occur by alternative mechanisms. Together these findings suggested a model for the differentiation of terminal effector cells that is initiated by a rapid burst of transcriptional activity that includes the upregulation of genes encoding products that promote the effector and memory fates, as well as the upregulation of chromatin regulators, followed by subsequent epigenetic repression of the memory program.

DISCUSSION

In the present study, we sought to discover previously unknown molecular determinants and gain new insights into the molecular regulation of specification of $CD8^+$ T lymphocyte fates by performing scRNA-seq of antigen-specific $CD8^+$ T cells derived sequentially throughout the course of a viral infection *in vivo*. Our analyses revealed a striking transcriptional divergence among cells that had undergone their first division, with hundreds of genes expressed differentially by the two subpopulations identified, which we provisionally called '$Div1_{TE}$ cells' and '$Div1_{MEM}$ cells'. The vast majority (97%) of these genes had higher expression by $Div1_{TE}$ cells than by $Div1_{MEM}$ cells and encoded products with diverse functions that spanned cell-cycle regulation, transcription, translation, metabolism and differentiation. Unexpectedly, genes encoding transcription factors linked to the differentiation of both effector cells and memory cells were substantially upregulated in $Div1_{TE}$ cells relative to their expression in $Div1_{MEM}$ cells, which suggested that memory-cell-associated transcription factors might have a transient but important role in the differentiation of terminal effector cells. Moreover, the expression patterns of genes expressed differentially by $Div1_{TE}$ cells versus $Div1_{MEM}$ cells were unique, in that the genes with higher expression by $Div1_{TE}$ cells had largely undetectable expression in $Div1_{MEM}$ cells. In contrast, the genes with higher expression by $Div1_{MEM}$ cells were also expressed by $Div1_{TE}$ cells, albeit with slightly lower expression. These dichotomous patterns suggested that the earliest steps in the differentiation of terminal effector cells were associated with a profound transcriptional burst involving considerable upregulation of hundreds of genes, whereas induction of the memory program might be associated with more nuanced alterations in the expression of a few specific genes.

On the basis of their molecular similarities with effector and memory cells, we hypothesized that $Div1_{TE}$ and $Div1_{MEM}$ cells represented early differentiation states of these cellular subsets. The application of 'future-past' binary classifiers enabled us to predict the identity of cells at intermediate time points and thereby infer pathways for the differentiation of terminal effector cells or memory cells. Visualization of the 'trajectories' of individual genes suggested patterns of expression in differentiating effector cells that were distinct from those in memory cells, as well as in differentiating T_{CM} cells versus T_{EM} cells. Although it has been appreciated that T_{CM} cells and T_{EM} cells are molecularly distinct[31,32], the ontology of these cells remains poorly understood[33,34]. Our data suggest the possibility that $Div1_{MEM}$ cells might represent a common progenitor of both circulating memory subsets, but it remains unknown when differentiating T_{CM} cells and T_{EM} cells diverge in fate. Future studies with more precise time points should probably provide additional insight into this question and might also elucidate whether T_{RM} cells are derived from $Div1_{MEM}$ cells, as it has been shown that T_{RM} cells and T_{CM} cells share a common clonal origin[35].

Which factors control the substantial transcriptional divergence observed following the first cell division remains an open question. One contributing factor could be asymmetric division, an evolutionarily conserved mechanism that enables activated T lymphocytes to apportion certain determinants unequally to daughter cells during mitosis[11]. Asymmetric segregation of factors such as the cytokine receptors IL-2Rα and IFN-γR during mitosis[7,11,36], for example, could promote signaling via the cytokines IL-2 and IFN-γ and result in the increased expression of *Il2ra*, *Stat5a* and *Tbx21* observed in $Div1_{TE}$ cells. Increased expression of genes encoding mediators of metabolic programming in $Div1_{TE}$ cells, moreover, would be consistent with the asymmetric mitotic distribution of the Myc, mTOR and phosphatidylinositol-3-OH kinase signaling pathways that has been reported[37–39]. Finally, the demonstration that activated $CD8^+$ T cells deficient in the atypical protein kinase PKC, a central regulator of asymmetric division, give rise to daughter cells with an effector-like transcriptional signature[40] supports the possibility of a role for asymmetric division in mediating the transcriptional heterogeneity in division 1 cells observed in the our study here.

We sought to identify previously unknown candidate regulators of differentiation by searching for commonality between the set of genes expressed differentially by $Div1_{MEM}$ cells relative to their expression in $Div1_{TE}$ cells and those expressed differentially by terminal-effector-cell subsets relative to their expression in memory-cell subsets. This approach yielded 89 candidate molecular determinants whose functions spanned regulation of proliferation, chromatin structure, transcription, and energy metabolism. We showed that one candidate, *Ezh2*, encodes a product with a critical role in the differentiation of terminal effector cells *in vivo*, which demonstrated the success of our experimental and computational approaches in identifying functionally important regulators of the differentiation of $CD8^+$ T cells. Consistent with our findings, Ezh2 has been shown to regulate the polyfunctionality and survival of human effector $CD8^+$ T cells through H3K27me3-mediated repression of genes encoding pro-apoptotic molecules, as well as those encoding components of the Notch signaling pathway[41].

A role for epigenetic regulation of determination of the $CD8^+$ T cell fate has been increasingly appreciated, with published studies showing the importance of DNA methylation and histone modifications in this process[42–44]. Published studies have investigated the overall epigenetic landscapes of naive, effector and memory $CD8^+$ T cells and have demonstrated substantial differences among cell subsets and during differentiation in their deposition of permissive H3K4me3 marks and repressive H3K27me3 marks[45,46]. Our study has extended those observations by demonstrating that transcription factors that promote alternative fates might be targeted differentially by Ezh2 in a T cell state–specific manner. In differentiating terminal effector cells, transcription factors associated with the alternative memory fate were selectively targeted by Ezh2. These findings suggested that the repression of memory-cell-associated genes might serve to enforce the terminal-effector-cell-differentiation program set into motion by effector-cell-associated genes. However, it remains possible that repressed memory-cell-associated genes might remain in a poised, bivalent ($H3K4me3^+H3K27me3^+$) state[45,46] and might thereby confer on effector cells a certain degree of plasticity.

In summary, our data suggest a model for the differentiation of terminal effector cells initiated by a rapid and profound transcriptional burst and refined by epigenetic silencing of transcripts associated with memory lymphocytes. In contrast, induction of the memory transcriptional program seemed to occur in a distinct subpopulation of differentiating lymphocytes and was associated with more

nuanced, gradual increases in the expression of a few specific genes. Together our findings suggest that closely linked transcriptional and epigenetic mechanisms together control specification of the CD8$^+$ T lymphocyte fate and underscore the power and necessity of single-cell approaches in future studies.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank members of the Chang and Yeo laboratories for discussions and critical reading of the manuscript, and the Sanford Consortium Stem Cell Genomics Core and Institute for Genomic Medicine Genomics Center for single-cell captures and sequencing. Supported by the US National Institutes of Health (DK093507, OD008469, and AI095277 to J.T.C.; NS075449, HG004659 and MH107367 to G.W.Y.; AI072117 and AI096852 to A.W.G; AI081923 and AI113923 to E.I.Z.; DK007202 for C.E.W. and P.J.M.) and the Howard Hughes Medical Institute (J.T.C.).

AUTHOR CONTRIBUTIONS

B.K., Z.H. and G.W.Y. performed computational analysis; J.A., C.E.W., E.J.W., E.I.Z., B.Y. A.W.G., J.T.C. and G.W.Y. designed experiments and analyzed data; J.A., C.E.W., S.A., P.J.M., B.Y., J.L. and S.H.K. performed experiments; and B.K., J.A., C.E.W., Z.H., J.T.C. and G.W.Y. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Joshi, N.S. *et al.* Inflammation directs memory precursor and short-lived effector CD8$^+$ T cell fates via the graded expression of T-bet transcription factor. *Immunity* **27**, 281–295 (2007).
2. Sallusto, F., Lenig, D., Forster, R., Lipp, M. & Lanzavecchia, A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708–712 (1999).
3. Mueller, S.N. & Mackay, L.K. Tissue-resident memory T cells: local specialists in immune defence. *Nat. Rev. Immunol.* **16**, 79–89 (2016).
4. Best, J.A. *et al.* Transcriptional insights into the CD8$^+$ T cell response to infection and memory T cell formation. *Nat. Immunol.* **14**, 404–412 (2013).
5. Kaech, S.M., Hemby, S., Kersh, E. & Ahmed, R. Molecular and functional profiling of memory CD8 T cell differentiation. *Cell* **111**, 837–851 (2002).
6. Chang, J.T., Wherry, E.J. & Goldrath, A.W. Molecular regulation of effector and memory T cell differentiation. *Nat. Immunol.* **15**, 1104–1115 (2014).
7. Arsenio, J. *et al.* Early specification of CD8$^+$ T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses. *Nat. Immunol.* **15**, 365–372 (2014).
8. Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).
9. Gaublomme, J.T. *et al.* Single-cell genomics unveils critical regulators of Th17 cell pathogenicity. *Cell* **163**, 1400–1412 (2015).
10. Shalek, A.K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013).
11. Chang, J.T. *et al.* Asymmetric T lymphocyte division in the initiation of adaptive immune responses. *Science* **315**, 1687–1691 (2007).
12. Badovinac, V.P., Haring, J.S. & Harty, J.T. Initial T cell receptor transgenic cell precursor frequency dictates critical aspects of the CD8$^+$ T cell response to infection. *Immunity* **26**, 827–841 (2007).
13. Ramskold, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
14. Rouault, J.P. *et al.* BTG1, a member of a new family of antiproliferative genes. *EMBO J.* **11**, 1663–1670 (1992).
15. Roychoudhuri, R. *et al.* BACH2 regulates CD8$^+$ T cell differentiation by controlling access of AP-1 factors to enhancers. *Nat. Immunol.* **17**, 851–860 (2016).
16. Blackledge, N.P., Rose, N.R. & Klose, R.J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat. Rev. Mol. Cell Biol.* **16**, 643–649 (2015).
17. DuPage, M. *et al.* The chromatin-modifying enzyme Ezh2 is critical for the maintenance of regulatory T cell identity after activation. *Immunity* **42**, 227–238 (2015).
18. Su, I.H. *et al.* Polycomb group protein ezh2 controls actin polymerization and cell signaling. *Cell* **121**, 425–436 (2005).
19. Tumes, D.J. *et al.* The polycomb protein Ezh2 regulates differentiation and plasticity of CD4$^+$ T helper type 1 and type 2 cells. *Immunity* **39**, 819–832 (2013).
20. Manjunath, N. *et al.* Effector differentiation is not prerequisite for generation of memory cytotoxic T lymphocytes. *J. Clin. Invest.* **108**, 871–878 (2001).
21. van der Windt, G.J. *et al.* Mitochondrial respiratory capacity is a critical regulator of CD8$^+$T cell memory development. *Immunity* **36**, 68–78 (2012).
22. Yu, B. *et al.* Epigenetic landscapes reveal transcription factors regulating CD8$^+$ T cell differentiation. *Nat. Immunol.* (in the press).
23. Ma, C. & Zhang, N. Transforming growth factor-β signaling is constantly shaping memory T-cell population. *Proc. Natl. Acad. Sci. USA* **112**, 11013–11017 (2015).
24. Mackay, L.K. *et al.* T-box transcription factors combine with the cytokines TGF-β and IL-15 to control tissue-resident memory T cell fate. *Immunity* **43**, 1101–1111 (2015).
25. Tinoco, R., Alcalde, V., Yang, Y., Sauer, K. & Zuniga, E.I. Cell-intrinsic transforming growth factor-β signaling mediates virus-specific CD8$^+$ T cell deletion and viral persistence in vivo. *Immunity* **31**, 145–157 (2009).
26. Ananieva, E.A., Patel, C.H., Drake, C.H., Powell, J.D. & Hutson, S.M. Cytosolic branched chain aminotransferase (BCATc) regulates mTORC1 signaling and glycolytic metabolism in CD4$^+$ T cells. *J. Biol. Chem.* **289**, 18793–18804 (2014).
27. Schober, S.L. *et al.* Expression of the transcription factor lung Kruppel-like factor is regulated by cytokines and correlates with survival of memory T cells in vitro and in vivo. *J. Immunol.* **163**, 3662–3667 (1999).
28. Skon, C.N. *et al.* Transcriptional downregulation of S1pr1 is required for the establishment of resident memory CD8$^+$ T cells. *Nat. Immunol.* **14**, 1285–1293 (2013).
29. Yamada, T., Park, C.S., Mamonkin, M. & Lacorazza, H.D. Transcription factor ELF4 controls the proliferation and homing of CD8$^+$ T cells via the Kruppel-like factors KLF4 and KLF2. *Nat. Immunol.* **10**, 618–626 (2009).
30. Buck, M.D. *et al.* Mitochondrial dynamics controls T cell fate through metabolic programming. *Cell* **166**, 63–76 (2016).
31. Chtanova, T. *et al.* Identification of T cell-restricted genes, and signatures for different T cell responses, using a comprehensive collection of microarray datasets. *J. Immunol.* **175**, 7837–7847 (2005).
32. Willinger, T., Freeman, T., Hasegawa, H., McMichael, A.J. & Callan, M.F. Molecular signatures distinguish human central memory from effector memory CD8 T cell subsets. *J. Immunol.* **175**, 5895–5903 (2005).
33. Bouneaud, C., Garcia, Z., Kourilsky, P. & Pannetier, C. Lineage relationships, homeostasis, and recall capacities of central- and effector-memory CD8 T cells in vivo. *J. Exp. Med.* **201**, 579–590 (2005).
34. Wherry, E.J. *et al.* Lineage relationship and protective immunity of memory CD8 T cell subsets. *Nat. Immunol.* **4**, 225–234 (2003).
35. Gaide, O. *et al.* Common clonal origin of central and resident memory T cells following skin immunization. *Nat. Med.* **21**, 647–653 (2015).
36. Chang, J.T. *et al.* Asymmetric proteasome segregation as a mechanism for unequal partitioning of the transcription factor T-bet during T lymphocyte division. *Immunity* **34**, 492–504 (2011).
37. Lin, W.H. *et al.* Asymmetric PI3K signaling driving developmental and regenerative cell fate bifurcation. *Cell Rep.* **13**, 2203–2218 (2015).
38. Pollizzi, K.N. *et al.* Asymmetric inheritance of mTORC1 kinase activity during division dictates CD8$^+$ T cell differentiation. *Nat. Immunol.* **17**, 704–711 (2016).
39. Verbist, K.C. *et al.* Metabolic maintenance of cell asymmetry following division in activated T lymphocytes. *Nature* **532**, 389–393 (2016).
40. Metz, P.J. *et al.* Regulation of asymmetric division and CD8$^+$ T lymphocyte fate specification by protein kinase Czeta and protein kinase Clambda/iota. *J. Immunol.* **194**, 2249–2259 (2015).
41. Zhao, E. *et al.* Cancer mediates effector T cell dysfunction by targeting microRNAs and EZH2 via glycolysis restriction. *Nat. Immunol.* **17**, 95–103 (2016).
42. Araki, Y., Fann, M., Wersto, R. & Weng, N.P. Histone acetylation facilitates rapid and robust memory CD8 T cell response through differential expression of effector molecules (eomesodermin and its targets: perforin and granzyme B). *J. Immunol.* **180**, 8102–8108 (2008).
43. Youngblood, B. *et al.* Chronic virus infection enforces demethylation of the locus that encodes PD-1 in antigen-specific CD8$^+$ T cells. *Immunity* **35**, 400–412 (2011).
44. Chappell, C., Beard, C., Altman, J., Jaenisch, R. & Jacob, J. DNA methylation by DNA methyltransferase 1 is critical for effector CD8 T cell expansion. *J. Immunol.* **176**, 4562–4572 (2006).
45. Araki, Y. *et al.* 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 (2009).
46. Russ, B.E. *et al.* Distinct epigenetic signatures delineate transcriptional programs during virus-specific CD8$^+$ T cell differentiation. *Immunity* **41**, 853–865 (2014).

ONLINE METHODS

Mice. All animal work was approved by the Institutional Animal Care and Use Guidelines of the University of California, San Diego. All mice were bred and housed in specific pathogen-free conditions. Wild-type C57BL/6J and *Ezh2*$^{fl/fl}$ mice were obtained from the Jackson Laboratory. *Ezh2*$^{fl/fl}$ mice were crossed with P14 *Cd4*^{Cre} mice. Donor mice were male or female, 6–8 weeks old. Recipient mice were male, 6–8 weeks old. For infection experiments, no randomization or blinding was used and no animals were excluded from analysis.

Antibodies and flow cytometry. Antibodies to the following were purchased from Biolegend and were used at 1:100 dilution: CD8α (53-6.7), CD45.1 (A20), CD62L (MEL-14), KLRG1 (2F1), CD44 (1M7), IL-2Rα (PC61), $V_α2$ (B20.1), $V_β8.1/8.2$ (KL16-133.18), IL-7Rα (A7R34), T-bet (4B10), IRF4 (IRF4.3E4), IFN-γ (XMG1.2), TNF (MP6-XT22). Antibody to Granzyme B (GB11) was purchased from Life Technologies and was used at 1:100 dilution. Antibody to Ezh2 (11/Ezh2) was purchased from BD Pharmingen and was used at 1:20 dilution. Annexin V Apoptosis Detection Kit and Mito Flow were purchased from Biolegend and Cell Technology. Biotinylated H-2D^b gp33 monomer (NIH Tetramer Core Facility) was conjugated to streptavidin-PE (Prozyme) to generate H-2D^b gp33 tetramer for flow cytometry analysis. For intracellular detection of IFN-γ and TNF, CD8$^+$ T cells were stimulated *ex vivo* with LCMV gp33-41 peptide (KAVYNFATM) (GenScript) in the presence of brefeldin A (Sigma) for 6 h at 37 °C; cells were stained with surface antibodies and then fixed in 4% paraformaldehyde (Electron Microscopy Services) and permeabilized before staining with intracellular antibodies. All samples were analyzed on an Accuri C6, FACSAria II, or FACSCanto (BD Biosciences).

Adoptive cell transfer and viral infection. 5×10^3 P14 CD45.1$^+$ CD8$^+$ T cells were adoptively transferred into congenic wild-type CD45.2$^+$ recipient mice, followed by intraperitoneal infection 1 d later with 2×10^5 plaque-forming units per mouse of LCMV-Armstrong. Splenocytes were isolated from recipient mice at 7 d after infection and splenocytes and lymph nodes were harvested at 42 d after infection. For the isolation of CD8$^+$ T cells at 4 d after infection, 5×10^4 P14 CD8$^+$ T cells per mouse were adoptively transferred into 24 recipient mice. For the isolation of CD8$^+$ T cells that had undergone their first cell division, 2×10^6 P14 CD8$^+$ T cells were first labeled with carboxyfluorescein diacetate succinimidyl ester (CFSE) before adoptive transfer into recipient mice ($n = 24$) and were harvested 2 d following LCMV infection.

Cell culture and differentiation. Splenocytes obtained from P14 mice were activated with gp33-41 peptide (500 ng/ml). After 1 d of activation, P14 CD8$^+$ T cells were isolated using the CD8$^+$ T Isolation Kit (Miltenyi). CD8$^+$ T cells were then cultured with IL-2 (10 U/ml) or IL-15 (15 ng/ml) (PeproTech) for an additional 3 d.

Single-cell transcriptome amplification and RNA-sequencing. The C_1 Single-Cell Auto Prep System (Fluidigm) was used to perform whole-transcriptome amplification of up to 96 single cells simultaneously. After cell isolation, 2.5×10^5 to 2×10^6 P14 CD8$^+$ T cells sorted by flow cytometry were loaded onto the C_1 Single-Cell Auto Prep mRNA Array IFC for single-cell capture on chip. Live/dead stain (Invitrogen) was included to exclude dead cells. Viable single cells captured on chip were manually imaged. Cell lysis and RT-PCR were performed on chip. SMARTer chemistry (Clontech) whole-transcriptome amplification was performed according to the manufacturer's instructions. Illumina Nextera XT single-cell complementary DNA (cDNA) libraries were generated according to the manufacturer's instructions (Illumina). Quality control measures of the single-cell cDNA libraries were performed on the 2100 Bioanalyzer (Agilent Technologies), Qubit 3.0 Fluorometer (Thermo Fisher Scientific), and MiSeq Sequencing System (Illumina). Single-cell cDNA libraries were sequenced (paired-end 100 or single-end 100) on the HiSeq2500 Sequencing System at the UCSD Institute for Genomics Medicine (IGM) Center.

ChIP-Seq analysis of Ezh2 and H3K27me3. For ChIP-seq analysis of Ezh2 and H3K27me3, wild-type CD8$^+$ T cells (from $n = 4$ mice) were activated *in vitro* with 5 μg/ml plate-bound anti-CD3 (145-2C11) and anti-CD28 (37.51, Bio X Cell) for 4 d and were sorted by flow cytometry to exclude dead cells. 4×10^6 CD8$^+$ T cells were crosslinked in 1% formaldehyde and ChIP was performed using the EZ-Magna ChIP kit (Millipore) according to the manufacturer's instructions. In brief, nuclear extracts were prepared and chromatin sheared to an average size of 300 bp using a Covaris E220 hydro-shearing instrument. For each immunoprecipitation (IP), chromatin from 1×10^6 cells and 3 μg of antibody were used. Antibodies used were as follows: rabbit anti-Ezh2 (H-80, Santa Cruz Biotechnology), rabbit anti-H3K27me3 (07-449, Millipore), mouse anti-RNA polymerase II (05-623, Millipore), mouse normal IgG (12-371, Millipore) and rabbit normal IgG (026102, Life Technologies). Sequence-indexed libraries were prepared from immunoprecipitated DNA and input controls (1%) using the NEB Next ChIP Library Preparation Reagent Set (NEB), according to the manufacturer's instructions. Library amplification by PCR used 10 cycles for pol II IPs, 12 cycles for input controls and H3K27me3 IPs, 14 cycles for Ezh2 IPs, and 17 cycles for IgG controls. For the H3K27me3 coverage comparison of wild-type versus Ezh2-deficient cells, chromatin from 500,000 cells was used and was amplified for 14 cycles (H3K27me3 IPs) or 17 cycles (IgG controls). Amplification yielded 200–600 fmoles per sample. 200 fmol of each library was pooled, selected by size to 250–650 bp on a PippinPrep instrument (Sage Science), and sequenced to a depth of 30 million reads (50 nt SE) on an Illumina HighSeq4000 instrument.

Bulk-cell RNA-seq. For isolation of CD8$^+$ T cells at 4 d after infection, 5×10^4 P14 wild-type CD8$^+$ T cells ($n = 4$) or Ezh2-deficient CD8$^+$ T cells ($n = 8$) were adoptively transferred into recipient mice and sorted by flow cytometry. mRNA-stranded cDNA libraries were generated and sequenced on an Illumina HighSeq4000 instrument. The bulk samples were processed with Kallisto[47], using GENCODE GRCm38.p4 transcriptome as the reference, with the following parameter: -l 200 –s 20 – single. The read count of each transcript derived from Kallisto was summed according to gene names and normalized to a 1,000,000 read-count of all genes in total for each sample. Differentially expressed genes were calculated by the assumption that the read count of each gene follows a Poisson distribution. The *P* value threshold was Bonferroni-corrected and ranged from 5^{-5} to 5^{-6} for selection of differentially expressed genes for gene-ontology analyses, depending on the gene number in each set.

Single-cell RNA-seq data pre-processing. Single-cell mRNA sequencing data from 256 CD8$^+$ T cells were processed with a bioinformatics pipeline focusing on quality control (QC) and robust expression quantification. For each cell, raw RNA-seq reads were: checked for quality metrics with fastqc (v0.10.1)[48]; poly-A and adaptor-trimmed with cutadapt (v1.8.1)[49]; quantified by Kallisto (v0.42.1)[50] to a reference transcriptome (Gencode vM3)[48] without bias correction; and aligned by STAR (v2.4.1b)[51] to the reference mouse genome (mm10)[52] with default parameters for quality control and downstream analysis. Next, the transcript per million (TPM) outputs of Kallisto for all cells were combined into a cell-by-gene expression matrix (C = 288 cells = rows, G = 22425 genes = columns) by summing the expression values for all quantified transcripts of a given gene. Finally, the TPM value for each cell *c* and gene *g* was natural log-transformed to yield a normalized expression value: $EXPR_{c,g} = \ln(1 + TPM_{c,g})$.

Dimensionality reduction and cell-heterogeneity visualization. To reduce the dimensionality of the cell-by-gene expression matrix EXPR and visualize the diversity of gene expression among CD8$^+$ T cells of different subtypes in a 2-dimensional scatter plot, we applied the t-distributed stochastic neighborhood embedding (tSNE) algorithm[53] via its Barnes-Hut approximation (bhSNE)[54]. tSNE is an unsupervised technique based on a non-convex objective which solves the so-called crowding problem, and has been successfully used to visualize millions of single-cell cytometry measurements where the original dimension is D = ~40 (approximately)[55–58]. In contrast, our total RNA-sequencing data for each cell gave signal for over 22,000 genes (6,000 of which had a mean expression over all cells greater than 1 TPM). Therefore, we first applied standard principal-component analysis (PCA) to reduce the dimensionality down to D = 10, and only then applied bhSNE to visualize in D = 2 (with perplexity = 30 and theta = 0.75 parameters). This composition of transformations is standard practice and results in a dimensionality reduction that is invariant to reflection[58]. After

doi:10.1038/ni.3688

dimensionality reduction, each point on the resulting two-dimensional scatter plot was colored by the stage of its corresponding T cell population. Since we observed two distinct clusters of division 1 T cells (red dots) in our tSNE plot (**Fig. 2a**), we re-colored those cells distinctly for the inset scatterplot according to their proximity to the centroids of the clusters of terminally differentiated effector T cells (T_{TE} cells) and memory T cells (T_{MEM} cells). Specifically, the proposed $Div1_{MEM}$ cells (inset) were re-colored blue because they were closer (in tSNE space) to the overall centroid of all T_{CM} cells (purple) and all T_{EM} cells (green) than to that of all day 7 cells (yellow). The remaining $Div1_{TE}$ cells (**Fig. 2a**, inset) remained red because they were closer (in tSNE space) to the centroid of all day 7 cells (yellow) than to that of the memory T cells (purple and green).

Gene-ontology (GO) analysis. Generated gene lists were uploaded to DAVID for analysis. Default background and default threshold were used and GOTERM_BP_FAT, GOTERM_MF_FAT, SP_PIR_KEYWORDS, UP_SEQ_FEATURE were chosen for target categories.

Supervised analysis of gene expression data. In contrast to the unsupervised dimensionality reduction (PCA and tSNE) and hierarchical clustering methods which are blind to the cell type labels, we also applied two supervised methods which utilize the extra information to give more interpretable results.

For differential gene expression analysis, we performed this analysis between all pairs of T cell sub-populations from two non-overlapping sets of rows in the log-transformed expression matrix EXPR. Since single-cell gene expression does not conform to the usual negative binomial distribution[59,60] and can even be bimodal due to dropout[61], we used two non-parametric statistical tests for heterogeneity of expression: the Mann-Whitney-Wilcoxon (MWW; also known as MWU) rank-sum test[62], which relies on a large sample to approximate normality; and the Kolmogorov-Smirnov two-sample (KS2) test[63], which finds the largest difference between the empirical cumulative distributions, even between two small samples such as our first division subtypes $Div1_{TE}$ ($n = 36$) and $Div1_{MEM}$ ($n = 24$).

For the cell type classifier, we trained two binary T cell classifiers to identify gene-expression signatures that not only distinguish the T cell subpopulations examined (such as the differential gene expression described above) but also can be used to predict the 'memory-ness' or 'effector-ness' of previously unseen cells. Each classifier constructed an independent ensemble of extremely randomized trees[64]. Using the terminally differentiated effector and memory (T_{CM} and T_{EM}) populations, we built a training set for a fate classifier for $CD8^+$ T cells. Using the newly observed segregation of daughter T cells into $Div1_{TE}$ and $Div1_{MEM}$ subpopulations after the first division, we built a second training set for another early state classifier. Both classifiers were provided their respective training sets and evaluated using tenfold cross-validation. A receiver operating characteristic (ROC) curve was computed by combining the predictions on each 10% held-out test set while training on the remaining 90% (ref. 65). After both the fate classifier and early-state classifier were trained on their respective subpopulations, they were both provided previously unseen intermediate day 4 $CD8^+$ T cells. Their predicted 'memory-ness' scores were scatterplotted and were shown to correlate (**Fig. 3f**). For each T cell, its 'effector-ness' scores is 1 minus the 'memory-ness' score and is redundant for this analysis. The signature genes for each classifier were selected from all G = 22,425 genes by their GINI score[65]. The surprisingly small overlap in gene-expression signatures for the two classifiers was computed to contrast with their seeming agreement in their 'memory-ness' score predictions.

Temporal expression trajectories through inferred lineage paths. To understand the temporal dynamics of expression for key genes along the effector and memory lineages, we constructed hypothetical differentiation time courses for each lineage. In brief, we sampled with replacement 50 cells from each population and constructed all trajectories through the cross-product of populations ordered in a particular lineage. These orders were determined a priori on the basis of published work with similar time courses of RT-qPCR data[7]. Specifically, the 'effector' lineage starts from the naive population, and progresses through the $Div1_{TE}$ subpopulation, then onto day 4, and finally to day 7. In contrast, the 'effector memory' and 'central memory' lineages start from naive, through the $Div1_{MEM}$ subpopulation, ending with T_{EM} cells and T_{CM} cells, respectively. These bootstrapped trajectories were visually summarized by a Seaborn software package time-series plot[66], which links the average expression for each population sample with a solid line segment and presents the 95% confidence interval as a shaded area around it.

H3K27me3-coverage data analysis. Data from H3K27me3 ChIP performed on $CD8^+$ T cells sorted at 8 d (effector cells) and 60 d (memory cells) after infection were mapped to mm10 reference genome with STAR (v2.4.1b) with the following options: outSAMunmapped None; outFilterMultimapNmax 10; outFilterMultimapScoreRange 1; limitOutSJoneRead 1; outReadsUnmapped Fastx; and all other options as default. ChIP peaks were called by 'Homer findPeaks'-style histone command with Poisson *P* value cutoff as 0.1% and enrichment over input threshold as 4.0 ('fold' values). To analyze coverage changes around the transcriptional start site (TSS) for the 6,000 expressed genes (**Fig. 6a**), the overlap of peaks and the 20 bins of 100 bp around TSS regions were calculated by BEDtools Intersect[67]. The coverage change was then calculated by deducting naive cell coverage around TSS from memory and effector cells, respectively. Reads intensity around TSS (**Fig. 6b**) was calculated by the sum of the total reads that were located in the TSS region, normalized for both the input reads that were located in TSS regions and the total number of reads obtained for each sample. In a similar way, the read intensity of each TSS region was derived, and any region with 3× read coverage over input was considered significantly covered. In **Figure 6c**, absolute TPM changes greater than 0.5 and absolute TSS ChIP coverage changes greater than 600 bp were considered significant. Data are presented as the ratio of H3K27me3-marked genes over total genes with decreased expression.

Ezh2 ChIP data analysis. All ChIP data were mapped to mm10 reference genome with STAR (v2.4.1b)[51] with the following options: outSAMunmapped None; outFilterMultimapNmax 10; outFilterMultimapScoreRange 1; limitOutSJoneRead 1; outReadsUnmapped Fastx; and all other options as default. ChIP and input data were then converted into tag directory with the HOMER command[68] 'makeTagDirectory' with the following options: keepOne; tbp 1; normGC; iterNorm 0.01. H3K27me3 ChIP regions were called by HOMER 'findPeaks' command, using input as background with following options: size 200; minDist 1000; L 0**;** and all other options as default. Ezh2 and PolII ChIP peaks were identified by using the HOMER software package 'findPeaks' command using input as background with following options: size 100; and all other options as default. In order to get the coverage of TSS of the 6000 expressed genes (**Fig. 6a**,**d**), the overlap of peaks and the 20 bins of 100 bp around TSS regions were calculated by BEDtools[67]. The changes in coverage (**Fig. 6a**) were then calculated by deducting naive cell coverage around TSS from memory cells or effector cells. In order to quantify the impact of H3K27me3 and Ezh2 on gene expression in memory and effector differentiation (**Fig. 6c**,**e**), artificial bulk gene expression TPM was calculated from single-cell data. In the two sets of comparisons (effector vs. naive, memory vs. naive), genes were marked as increased or decreased if bulk gene expression changed more than twofold. Fisher's exact test was then performed to determine whether genes targeted by Ezh2 or H3K27me3 negatively correlated with TPM changes. To be included as targets of Ezh2 or H3K27me3 in **Figures 6** and **7**, genes were required to have ≥100-bp TSS region covered by Ezh2 peaks or ≥600-bp TSS region covered by H3K27me3. In **Figure 6e**, data are presented as a percentage of unbound or Ezh2-bound genes among total genes with either loss or gain of expression in day 4 cells relative to that in $Div1_{TE}$ cells. In **Figure 7**, normalized TPM was calculated to ensure that all genes had mean expression TPM = 1 across all single-cell samples.

Statistical analysis. Pearson correlation and Spearman correlation were used to assess the significance of memory score prediction by supervised classifiers (**Fig. 3f**). Pearson correlation was used to determine the genes with the most significantly differential expression (**Fig. 4b**). Student's unpaired *t*-test was used for comparisons involving two groups (**Fig. 5d-k**). Fisher's exact test was used for comparisons of genes targeted by Ezh2 or H3K27me3, identified by ChIP (**Fig. 6c**,**e**). Differences with an associated *P* value of <0.05 were considered significant.

Data availability. Source RNA-seq and ChIP-seq data sets are available with GEO accession codes GSE89405 and GSE89036.

47. Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
48. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
49. Chen, C., Khaleel, S.S., Huang, H. & Wu, C.H. Software for pre-processing Illumina next-generation sequencing short read sequences. *Source Code Biol. Med.* **9**, 8 (2014).
50. Ntranos, V., Kamath, G.M., Zhang, J.M., Pachter, L. & Tse, D.N. Fast and accurate single-cell RNA-seq analysis by clustering of transcript-compatibility counts. *Genome Biol.* **17**, 112 (2016).
51. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
52. Mouse Genome Sequencing, C. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
53. Maaten, L.H.G.E. Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
54. Maaten, L. Barnes-Hut-SNE. *arXIv* https://arxiv.org/abs/1301.3342 (2013).
55. Amir el, A.D. *et al.* viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* **31**, 545–552 (2013).
56. Becher, B. *et al.* High-dimensional analysis of the murine myeloid cell system. *Nat. Immunol.* **15**, 1181–1189 (2014).
57. Bendall, S.C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696 (2011).
58. Cheng, Y., Wong, M.T., van der Maaten, L. & Newell, E.W. Categorical analysis of human T cell heterogeneity with one-dimensional soli-expression by nonlinear stochastic embedding. *J. Immunol.* **196**, 924–932 (2016).
59. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
60. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
61. Fan, J. *et al.* Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods* **13**, 241–244 (2016).
62. Li, J. & Tibshirani, R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Stat. Methods Med. Res.* **22**, 519–536 (2013).
63. Stephens, M.A. EDF statistics for goodness of fit and some comparisons. *J. Am. Stat. Assoc.* **69**, 730–737 (1974).
64. Geurts, P.E. D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **62**, 3–42 (2006).
65. Pedregosa, F.V. G; Gramfor, A. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
66. Setty, M. *et al.* Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* **34**, 637–645 (2016).
67. Quinlan, A.R. BEDTools: The Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinformatics.* **47**, 11–34 (2014).
68. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

doi:10.1038/ni.3688