HHS Public Access

Author manuscript

Nat Immunol. Author manuscript; available in PMC 2017 August 20.

Published in final edited form as:

Nat Immunol. 2017 April ; 18(4): 422–432. doi:10.1038/ni.3688.

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

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], and **Gene W. Yeo**[1,4,5,7]

[1]Department of Cellular and Molecular Medicine, University of California, San Diego, CA, USA

[2]Department of Medicine, University of California, San Diego, CA, USA

[3]Division of Biological Sciences, University of California, San Diego, CA, USA

[4]Institute for Genomic Medicine, University of California, San Diego, CA, USA

[5]Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

SUMMARY

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 controlling CD8$^+$ T lymphocyte fate specification. These findings suggest a model of terminal effector cell differentiation initiated by an early burst of transcriptional activity and subsequently refined by epigenetic silencing of transcripts associated with memory lymphocytes, highlighting the power and necessity of single-cell approaches.

INTRODUCTION

Heterogeneity of cell fate is a hallmark of T lymphocyte responses to microbial infection. During an immune response to a microbial infection, responding naïve T lymphocytes give

Correspondence should be addressed to G.W.Y. (geneyeo@ucsd.edu) or J.T.C. (changj@ucsd.edu).
[6]These authors contributed equally to this work.
[7]These authors jointly directed this work.

AUTHOR CONTRIBUTIONS

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

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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 divided into two subsets, central memory T (T_{CM}) cells and effector memory T (T_{EM}) cells, distinguished by differences in their expression of homing and cytokine receptors such as L-selectin (CD62L) and CCR7, proliferative capacity, and anatomical localization[2]. A third subset of memory cells, tissue-resident memory T cells (T_{RM}), do not circulate, but instead remain in peripheral tissues after pathogen clearance[3].

Transcriptional profiling approaches have greatly advanced our understanding of the molecular regulation of T lymphocyte fate specification[4,5]. By comparing the gene expression of $CD8^+$ T lymphocytes during the course of microbial infections, these studies have identified many transcription factors and pathways that play a role in the specification of terminal differentiation versus long-lived memory (reviewed in [6]). However, most prior studies have been conducted on bulk populations of cells, thereby masking potential heterogeneity among individual cells. We previously sought to address these limitations by applying single-cell qRT-PCR analyses to interrogate the gene expression patterns of single $CD8^+$ T lymphocytes responding to bacterial infection *in vivo*[7]. Although we identified dynamic changes in gene expression in individual cells during differentiation, pre-selection of previously known genes for analysis precluded the discovery of novel genes and molecular pathways.

Single-cell RNA sequencing (scRNA-seq) has recently emerged as a powerful tool that has substantially advanced our understanding of diverse biologic processes, including development[8], $CD4^+$ T_H17 cell pathogenicity[9], and innate immune responses[10]. In the current study, we applied a scRNA-seq approach to analyze transcriptome-wide changes in individual $CD8^+$ T cells as they differentiated *in vivo*. We observed remarkable transcriptional heterogeneity among lymphocytes that had undergone their first division, revealing two distinct subpopulations that were distinguished in their expression of hundreds of genes involved in diverse biological functions, including cell cycle regulation, metabolism, effector function, and fate specification. The expression of many transcription factors previously implicated in effector and memory cell differentiation, along with chromatin regulators, was markedly increased in cells differentiating along the terminal effector pathway and extinguished by the peak of the adaptive immune response. This initial transcriptional program was subsequently refined by selective epigenetic repression of molecular determinants associated with memory cell differentiation. By contrast, induction of the memory program appeared to be associated 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 $CD8^+$ T lymphocyte fate specification 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], P14 CD8$^+$ T lymphocytes, which have transgenic expression of T cell antigen receptors (TCRs) that recognize an immunodominant epitope of lymphocytic choriomeningitis virus (LCMV), were adoptively transferred into congenic wild-type recipients one day prior to intraperitoneal (i.p.) infection with LCMV-Armstrong. On days 2, 4, and 7 of infection, activated P14 CD8$^+$ T lymphocytes (CD44^{hi}) were enriched from spleens of infected mice using a magnetic bead-based approach and sorted using flow cytometry. On day 42 of infection, T_{CM} (CD44^{hi}CD62L^{hi}) and T_{EM} (CD44^{hi}CD62L^{lo}) cells were similarly isolated; naïve P14 CD8$^+$ T cells (CD44^{lo}CD62L^{hi}) were also included in our analysis. For certain time points (days 2 and 4 post-infection), high numbers of P14 cells were adoptively transferred into recipient mice[4,7,11] to enable isolation of sufficient numbers of cells for scRNA-seq analysis; high cell transfer can alter the magnitude and kinetics of the immune response[12] and, therefore, represents a caveat of the study.

The C_1 Single-Cell Auto Prep system (Fluidigm) was used to perform PCR amplification of full-length, polyadenylated transcripts[13], followed by preparation and sequencing of single-cell cDNA libraries (Fig. 1a). 10 to 20 million 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 were performed on separate sequencing plates to ensure reproducibility and absence of batch effects (Supplementary Fig. 1d–f). After undertaking these quality control measures, we included 288 single-cell libraries divided into 224 unique sequencing samples and 32 pairs of duplicates for further in-depth analyses.

We detected over 6000 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 previously linked to CD8$^+$ T cell differentiation with bulk population analyses suggested patterns of expression consistent with their previously assigned roles in this process. For instance, genes associated with effector cell differentiation and function, such as *Batf*, *Id2*, and *Gzmb,* were highly expressed in cells isolated at days 4 and 7 post-infection (Fig. 1b,c), time points at which terminally differentiated effector cells are known to predominate. Conversely, memory-associated genes such as *Tcf7* and *Il7r* were highly expressed in T_{CM} and T_{EM} cells (Fig. 1b,c). These findings demonstrate that the expression of previously described effector- and memory-associated genes can be readily detected in the expected lymphocyte subsets using a single-cell approach.

Importantly, our single-cell analysis also revealed patterns of gene expression that were not previously discernable using bulk population analyses. For instance, 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 bulk population analyses (Fig. 1b,c). Thus, single-cell transcriptomic analyses capture the remarkable 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 t-distributed Stochastic Neighborhood Embedding (tSNE) clustering analysis to visualize individual $CD8^+$ T lymphocytes isolated at all time points in an unbiased manner (Fig. 2a). In tSNE analysis, naïve cells (gray), T_{CM} cells (purple), and T_{EM} cells (green) each formed distinct clusters (Fig. 2a), suggestive of unique molecular homogeneity within each population (Supplementary Table 1). Similarly, most Day 4 (orange) and Day 7 (yellow) cells formed their own separate clusters; however, a few cells from each time point grouped near the naïve and T_{CM} populations. Strikingly, unsupervised tSNE analysis revealed two distinct subpopulations among single $CD8^+$ T lymphocytes that had undergone their first cell division (red, Division 1) (Fig. 2a). Importantly, it should be emphasized that Division 1 cells were isolated on the basis of CFSE dilution (2^{nd} CFSE peak) and not phenotypic cell surface marker expression, other than high expression of the activation marker CD44 to ensure that all sorted cells had been activated *in vivo*. One subpopulation of Division 1 cells (Fig. 2a inset, red cells) appeared to be most similar to Day 4 and Day 7 effector $CD8^+$ T lymphocytes, whereas the other subpopulation of Division 1 cells (Fig. 2a inset, blue cells) appeared to be most similar to T_{CM} and naïve cells (Fig. 2a). The two Division 1 subpopulations were provisionally designated $Div1_{TE}$ and $Div1_{MEM}$ (Fig. 2a inset) on the basis of their similarities with terminal effector and memory cell populations, respectively.

Differential gene expression analysis revealed that 930 genes distinguished $Div1_{TE}$ and $Div1_{MEM}$ cells, with 903 more highly expressed in $Div1_{TE}$ cells and 27 more highly expressed in $Div1_{MEM}$ cells (Fig. 2b and Supplementary Table 2). Gene ontology analysis revealed that genes upregulated by $Div1_{TE}$ cells were enriched for 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 expressed higher levels of transcription factors, cytokine receptors, and signaling molecules previously associated with terminal effector cell differentiation and metabolic reprogramming[6] (Fig. 2c).

By contrast, $Div1_{MEM}$ cells expressed higher levels of several factors that have been previously associated with memory lymphocytes[6] including *Il7r*, *S1pr1*, and *Klf2* (Fig. 2d), in addition to previously unappreciated molecules such as the anti-proliferative gene *Btg1*[14] (Supplementary Table 2). Notably, however, the majority of transcription factors associated with memory lymphocytes were either expressed at comparable levels (*Lef1*, *Bach2*[15], and *Tcf7)* by $Div1_{TE}$ and $Div1_{MEM}$ cells (Fig. 2e), or at higher levels (*Eomes*, *Id3*, and *Foxo1*) by $Div1_{TE}$ cells (Fig. 2c). Taken together, these data demonstrate that $CD8^+$ T lymphocytes that have undergone their first cell division exhibit marked transcriptional heterogeneity that was not previously discernible[7], with the vast majority of differentially expressed genes upregulated by $Div1_{TE}$ cells, including many transcripts previously associated with memory lymphocytes.

Identifying 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. 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 *bona fide* memory (T_{CM} and T_{EM}) and terminal effector cell populations ('fate' classifier, Fig. 3c,d), would enable us to identify cells in intermediate states of differentiation as they progressed towards a terminally differentiated versus long-lived memory fate.

Using both early state and fate classifiers, 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 cells and Day 40 memory cells, was deployed into the 'past' to analyze the same Day 4 cells. Remarkably, the early state and fate classifiers agreed on the identity of intermediate cells (Fig. 3f), yet used largely non-overlapping sets of genes to predict pre-terminal effector or pre-memory cell states (Supplementary Tables 4 and 5). Thus, intermediate states of differentiation can be readily predicted using supervised binary classifiers trained on the preceding or subsequent states.

Identifying 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 differentially expressed between $Div1_{TE}$ and $Div1_{MEM}$ cells and that between terminal effector and memory cells. From the sets of 930 genes differentially expressed between $Div1_{TE}$ and $Div1_{MEM}$ cells (Supplementary Table 2) and 834 genes differentially expressed between terminal effector and memory cells (Supplementary Table 6), only 115 genes were shared (Fig. 4a,b). We next selected genes common only to $Div1_{TE}$ and effector cells or common only to $Div1_{MEM}$ and memory cells to identify genes encoding regulators of terminal effector or memory cell differentiation, respectively.

This analysis 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 these genes 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 or memory cell differentiation, based on gene expression of these putative regulators in Day 4 cells classified as being in either pre-memory or pre-effector states of differentiation (Fig. 4d and Supplementary Fig. 2). We observed that genes expressed in cells differentiating along the terminal effector pathway tended to exhibit a marked increase at the first division followed by a rapid decline (Fig. 4d), raising the possibility of epigenetic repression in differentiating effector cells. We selected *Ezh2*, a catalytic subunit of the Polycomb Repressive Complex 2

(PRC2) complex that mediates gene repression by mediating histone H3 trimethylation at lysine 27 (H3K27me3)[16], for further analysis and functional validation, given prior studies suggesting a critical role in $CD4^+$ T cell differentiation and function[17,18,19].

Ezh2, along with genes encoding other PRC2 components *Eed*, *Suz12*, and *Set*, was highly expressed in $Div1_{TE}$ cells relative to $Div1_{MEM}$ cells (Fig. 5a), suggesting a role of Ezh2 in regulating terminal effector cell differentiation. Consistent with this finding, we observed that $CD8^+$ T cells that had undergone their first division exhibited a bimodal pattern of Ezh2 protein expression (Fig. 5b), with high and low levels in putative 'pre-effector' IL-2Rα^{hi}CD62L^{lo} and 'pre-memory' IL-2Rα^{lo}CD62L^{hi} cells (Fig. 5c), respectively[7]. 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 transgenic mice and adoptively transferred Ezh2-deficient and control P14 $CD45.1^+$ $CD8^+$ T cells (Supplementary Fig. 3a) into wild-type CD45.2 recipient mice that were subsequently infected with LCMV-Armstrong. Compared to control cells, Ezh2-deficient $CD8^+$ T cells were substantially reduced by days 5 and 7 post-infection (Fig. 5e, Supplementary Fig. 3b) and exhibited an impaired capacity to secrete inflammatory cytokines (Fig. 5f,g). Importantly, the absence of effector cells was not due to the 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 post-infection (Fig. 5h) and preferentially affected early 'effector-like,' but not 'memory-like,' cells in an *in vitro* model of $CD8^+$ T cell differentiation[20,21] (Fig. 5i–k and Supplementary Fig. 3e and 4). Taken together, these findings suggest that Ezh2 plays a critical role in regulating terminal effector cell differentiation.

Epigenetic silencing of memory-associated determinants

Because the PRC2 complex mediates transcriptional repression, we hypothesized that high expression of Ezh2 in $Div1_{TE}$ cells catalyzes repressive H3K27me3 marks on a set of key genes, thereby promoting terminal effector differentiation. Focusing on the ~6000 genes detected in our scRNA-seq analysis, we mapped H3K27me3 peaks derived from chromatin immunoprecipitation sequencing (ChIP-seq) analysis performed on naïve, terminal effector (KLRG1^{hi}CD44^{hi}), and total memory (CD44^{hi}) $CD8^+$ T cells (Yu et al., manuscript under consideration). Relative to promoter regions in naïve cells, those in terminal effector cells exhibited significant gains in H3K27me3 coverage that correlated with reduced gene expression, whereas promoter regions in memory cells exhibited considerable losses in H3K27me3 coverage (Fig. 6a–c and Supplementary Tables 8 and 9), suggesting that epigenetic silencing may be more crucial for the differentiation of terminal effector cells compared to that of memory cells.

We next investigated Ezh2 binding and H3K27me3 levels in differentiating $CD8^+$ T cells by ChIP-seq (Fig. 6d,e), identifying 1564 genes bound by Ezh2 and 261 genes marked by H3K27me3 (Supplementary Table 10). Gene ontology analysis revealed that Ezh2 gene targets were enriched for processes involving transcriptional regulation, cytoskeletal protein binding, phosphoinositide binding, Wnt receptor signaling, and apoptosis (Supplementary Table 11). We observed that genes with expression that was reduced in Day 4 cells relative

to $Div1_{TE}$ cells were more likely to exhibit Ezh2 binding than those with increased expression (Fig. 6e). Comparison of the expression patterns of H3K27me3-marked and unmarked genes during terminal effector and memory cell differentiation revealed that H3K27me3-marked genes tended to exhibit more repression during terminal effector cell differentiation (distributions are shifted in Fig. 7a, top, and Supplementary Fig. 5a) than during memory cell differentiation (distributions are not different in Fig. 7a, bottom, and Supplementary Fig. 5b,c). These findings raised the intriguing possibility that the same set of genes were being selectively repressed during differentiation of terminal effector but not memory cells, due, in part, to differential expression of Ezh2 in cells progressing along these disparate pathways.

We investigated this possibility by analyzing a subset of Ezh2 target genes whose expression decreased during effector but not memory cell differentiation. This analysis revealed that many genes previously linked to memory, but not effector differentiation, exhibited significant Ezh2 association relative to input DNA control (Fig. 7b and Supplementary Table 10). These Ezh2 gene targets included memory-associated transcription factors, including *Tcf7* and *Eomes;* molecules that mediate TGF-β signaling, including *Smad2*, which has been implicated in $CD8^+$ T cell fate decisions[22,23,24]; metabolic regulators such as the branched chain aminotransferase isoenzyme *Bcat1*[25]; factors that control T cell survival and homing, including *Klf2*[26,27,28]; and a recently discovered regulator of mitochondrial fusion, *Opa1*, that plays a critical role in differentiating memory $CD8^+$ T lymphocytes[29]. Several of these genes, like *Foxo1* and *Tcf7*, underwent rapid decreases in expression following the first division; others, like *Eomes* and *Id3*, underwent an initial increase at the first division followed by a rapid decline, suggesting a possible role for these memory-associated genes in effector differentiation (Fig. 7c, orange line). By contrast, analysis of these same genes in differentiating memory cells revealed a distinct expression pattern characterized by modest increases at the first division followed by stable or increased levels as the cells became mature T_{CM} or T_{EM} (Fig. 7c, purple and green lines). A similar pattern of expression was observed for memory-associated genes, such as *Il7r*, *Lef1*, and *Bcl2,* that were not targeted by Ezh2 (Fig. 7d). Lastly, we observed that Ezh2 deficiency was associated with reduced H3K27me3 coverage and increased mRNA expression of many genes, including a number of memory-associated genes such as *Eomes*, *Tcf7*, and *Klf2* (Fig. 7e–h, Supplementary Fig. 6–8, and Supplementary Tables 12 and 13), consistent with a role for H3K27me3-mediated transcriptional repression by Ezh2. Thus, unique expression patterns of memory-associated genes in differentiating terminal effector and memory cells may result, in part, from the presence or absence of epigenetic repression owing to distinct levels of Ezh2 expression in these cells.

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

observation that these effector-associated genes were not targeted by Ezh2 or marked by H3K27me3 suggests that regulation of these genes during differentiation may be due to alternative mechanisms. Taken together, these findings suggest a model of terminal effector cell differentiation initiated by a rapid burst of transcriptional activity that includes upregulation of genes that promote the effector and memory fates as well as chromatin regulators, followed by subsequent epigenetic repression of the memory program.

DISCUSSION

In the present study, we sought to discover novel molecular determinants and gain new insights into the molecular regulation of $CD8^+$ T lymphocyte fate specification by performing scRNA-seq on 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 differentially expressed between these two subpopulations, provisionally termed $Div1_{TE}$ and $Div1_{MEM}$ cells. The vast majority (97%) of these genes were more highly expressed by $Div1_{TE}$ cells, with diverse functions that spanned cell cycle regulation, transcription, translation, metabolism, and differentiation. Unexpectedly, transcription factors linked to both effector and memory differentiation were highly upregulated in $Div1_{TE}$ cells, suggesting that memory-associated transcription factors may play a transient but important role in terminal effector cell differentiation. Moreover, the patterns of differentially expressed genes between $Div1_{TE}$ and $Div1_{MEM}$ cells were unique, in that genes more highly expressed by $Div1_{TE}$ cells were largely undetectable in $Div1_{MEM}$ cells. By contrast, genes more highly expressed by $Div1_{MEM}$ cells were also expressed by $Div1_{TE}$ cells, albeit at slightly lower levels. These dichotomous patterns suggest that the earliest steps of terminal effector cell differentiation are associated with a profound transcriptional burst involving marked upregulation of hundreds of genes, whereas induction of the memory program may be associated with more nuanced alterations in the expression levels of a few specific genes.

Based on 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. Application of ‘future-past’ binary classifiers enabled us to predict the identity of cells at intermediate time points and thereby infer pathways of terminal effector or memory cell differentiation. Visualization of the trajectories of individual genes suggested distinct patterns of expression between differentiating effector and memory cells as well as between differentiating T_{CM} and T_{EM} cells. Although it has been previously appreciated that T_{CM} and T_{EM} cells are molecularly distinct[30,31], the ontology of these cells remains poorly understood[32,33]. Our data suggest the possibility that $Div1_{MEM}$ cells may represent a common progenitor of both circulatory memory subsets, but it remains unknown when differentiating T_{CM} and T_{EM} cells diverge in fate. Future studies with more precise time points are likely to provide additional insight into this question and may also elucidate whether T_{RM} cells are derived from $Div1_{MEM}$ cells, as it has been recently shown that T_{RM} and T_{CM} cells share a common clonal origin[34].

What factors control the remarkable 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 IL-2Rα and IFN-γR during mitosis[7,11,35], for example, could promote IL-2 and IFN-γ signaling and result in the increased expression of *Il2ra*, *Stat5a*, and *Tbx21* observed in $Div1_{TE}$ cells. Increased expression of mediators of metabolic programming in $Div1_{TE}$ cells, moreover, is consistent with the asymmetric mitotic distribution of Myc, mTOR, and phosphatidylinositol-3-OH kinase signaling pathways that has been recently reported[36,37,38]. Finally, our recent demonstration that activated $CD8^+$ T cells deficient in atypical protein kinase c (PKC), a central regulator of asymmetric division, give rise to daughter cells with an effector-like transcriptional signature[39] supports a possible role for asymmetric division in mediating the transcriptional heterogeneity in Division 1 cells observed in the current study.

We sought to uncover new candidate regulators of differentiation by searching for commonality between the set of genes differentially expressed amongst $Div1_{MEM}$ and $Div1_{TE}$ cells and that amongst terminal effector and memory cell subsets. This approach yielded 89 candidate molecular determinants, whose functions spanned regulation of proliferation, chromatin structure, transcription, and energy metabolism. We demonstrated that one candidate, *Ezh2*, played a critical role in effector differentiation *in vivo*, thus demonstrating the success of our experimental and computational approaches in discovering functionally important regulators of $CD8^+$ T cell differentiation. Consistent with our findings, Ezh2 was recently shown to regulate human effector $CD8^+$ T cell polyfunctionality and survival through H3K27me3-mediated repression of pro-apoptotic genes as well as components of the Notch signaling pathway[40].

A role for epigenetic regulation of $CD8^+$ T cell fate determination has been increasingly appreciated, with prior studies showing the importance of DNA methylation and histone modifications in this process[41,42,43]. Recent studies have examined the overall epigenetic landscapes of naïve, effector, and memory $CD8^+$ T cells, demonstrating significant differences in permissive H3K4me3 and repressive H3K27me3 deposition among cell subsets and during differentiation[44,45]. Our study extends these observations by demonstrating that transcription factors that promote alternative fates may be differentially targeted 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 suggest that repression of memory-associated genes may serve to enforce the terminal effector differentiation program set into motion by effector-associated genes. However, it remains possible that repressed memory-associated genes may remain in a poised, bivalent ($H3K4me3^+H3K27me3^+$) state[44,45], thereby conferring effector cells with a certain degree of plasticity.

In summary, our data suggest a model of terminal effector cell differentiation initiated by a rapid and profound transcriptional burst and refined by epigenetic silencing of transcripts associated with memory lymphocytes. By contrast, induction of the memory transcriptional program appears to occur in a distinct subpopulation of differentiating lymphocytes and is associated with more nuanced, gradual increases in the expression levels of a few specific genes. Together, these findings suggest that closely linked transcriptional and epigenetic

mechanisms together control $CD8^+$ T lymphocyte fate specification and underscore the power and necessity of single-cell approaches in future studies.

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 to 8 weeks old. Recipient mice were male, 6 to 8 weeks old. For infection experiments, no randomization or blinding was used and no animals were excluded from analysis.

Antibodies and flow cytometry

The following antibodies were purchased from Biolegend, eBiosciences, or Sigma-Aldrich: CD8α (53–6.7), CD45.1 (A20), CD62L (MEL-14), KLRG1 (2F1), CD44 (1M7), IL-2Rα (PC61), V_a2 (B20.1), $V_β8.1/8.2$ (KL16–133.18), IL-7Rα (A7R34), T-bet (4B10), IRF4 (IRF4.3E4), Granzyme B (GB11), IFN-γ (XMG1.2), TNF (MP6-XT22). Ezh2 (11/Ezh2) antibody was purchased from BD Pharmingen. Annexin V Apoptosis Detection Kit and Mito Flow were purchased from Biolegend and Cell Technology. Biotinylated H-$2D^b$ gp33 monomer (NIH Tetramer 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 prior to staining with intracellular antibodies. All samples were analyzed on an Accuri C6, FACS Aria II, or FACS Canto (BD Biosciences).

Adoptive cell transfer and virus 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 (i.p.) 1 day later with 2×10^5 plaque forming units (pfu) per mouse of LCMV-Armstrong. Splenocytes were isolated from recipient mice at 7 d post-infection (n = 4) and splenocytes and lymph nodes were harvested at 42 d (n = 40) post-infection. For the isolation of $CD8^+$ T cells at 4 d post 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) prior to adoptive transfer into recipient mice (n = 24) and harvested 2 d following LCMV infection.

Cell culture and differentiation

Splenocytes obtained from P14 mice were activated with gp_{33-41} 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 (WTA) of up to 96 single cells simultaneously. After cell isolation, 2.5×10^5 to 2×10^6 FACS-sorted P14 $CD8^+$ T cells 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) WTA 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.

Ezh2 and H3K27me3 chromatin immunoprecipitation (ChIP)-Seq

For Ezh2 and H3K27me3 ChIP seq, wild-type $CD8^+$ T cells (from n = 4 mice) were activated *in vitro* with plate-bound anti-CD3 and anti-CD28 antibodies for 4 days and sorted by flow cytometry to exclude dead cells. 4×10^6 $CD8^+$ T cells were crosslinked in 1% formaldehyde and ChIP performed using the EZ-Magna ChIP kit (Millipore) according to the manufacturer's instructions. Briefly, 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 million cells and 3 μg of antibody were used. Antibodies used were: rabbit anti-Ezh2 antibody (H-80; Santa Cruz Biotechnology), rabbit anti-H3K27me3 (Millipore), mouse anti-RNA polymerase II (Millipore), and mouse and rabbit normal IgG. 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 between wild-type and Ezh2-deficient cells, chromatin from 500,000 cells was used and amplified for 14 cycles (H3K27me3 IPs) or 17 cycles (IgG controls). Amplification yielded 200–600 fmoles per sample. Two hundred fmoles of each library were pooled, size-selected 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 post-infection, 5×10^4 P14 wild-type (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[46], using GENCODE GRCm38.p4 transcriptome as the reference, with the following parameter: −1 200 –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)[47]; poly-A and adaptor-trimmed with cutadapt (v1.8.1)[48]; quantified by Kallisto (v0.42.1)[49] to a reference transcriptome (Gencode vM3)[47] without bias correction; and aligned by STAR (v2.4.1b)[50] to the reference mouse genome (mm10)[51] 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)[52] algorithm via its Barnes-Hut approximation (bhSNE)[53]. 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[54,55,56,57]. In contrast, our total RNA sequencing data for each cell gave signal for over 22,000 genes (6000 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[57]. After dimensionality reduction, each point on the resulting 2-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 terminally differentiated effector (T_{TE}) and memory (T_{MEM}) T cell clusters. 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} (purple) and all T_{EM} (green) cells 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, 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: (1) Differential gene expression analysis. We performed differential gene expression 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[58,59] and can even be bimodal due to dropout[60], we used two non-parametric statistical tests for heterogeneity of expression: Mann Whitney Wilcoxon (MWW, also known as MWU)[61] rank-sum test which relies on a large sample to approximate normality, and Kolmogorov-Smirnov 2-sample (KS2) test[62] which finds the largest difference between the empirical cumulative distributions, even between two small samples such as our 1st division sub-types $Div1_{TE}$ ($n = 36$) and $Div1_{MEM}$ ($n = 24$). (2) Cell type classifier. We trained two binary T cell classifiers to identify gene expression signatures that not only differentiate the examined T cell sub-populations (like the differential gene expression described above) but can also be used to predict the ‘memory-‘ or ‘effector-ness’ of previously unseen cells. Each classifier constructed an independent ensemble of Extremely Randomized Trees[63]. Using the terminally differentiated effector and memory (T_{CM}, 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 10-fold 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%[64]. After both the fate and early state classifiers were trained on their respective subpopulations, they were both provided previously unseen intermediate Day 4 $CD8^+$ T cells. Their predicted ‘memory-ness’ scores were scatter-plotted and shown to correlate in 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[64]. The surprisingly small overlap in gene expression signatures between 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 timecourses for each lineage. Briefly, 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* based on our earlier work with similar timecourses of RT-

qPCR data[7]. Specifically, the 'effector' lineage starts from the naïve population, and progresses through the $Div1_{TE}$ subpopulation, then onto Day 4, and finally Day 7. In contrast, the 'effector memory' and 'central memory' lineages start from naïve, through the $Div1_{MEM}$ subpopulation, ending with T_{EM} and T_{CM} respectively. These bootstrapped trajectories were visually summarized by a Seaborn software package timeseries plot[65] which links the average expression for each population sample with a solid line segment and represents the 95% confidence interval by a shaded area around it.

H3K27me3 coverage data analysis

H3K27me3 ChIP data performed on $CD8^+$ T cells sorted at 8 days (effector cells) and 60 days (memory cells) post-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 fold-enrichment over input threshold as 4.0. To analyze coverage changes around the Transcriptional Start Site (TSS) for the 6000 expressed genes (Fig. 6a), the overlap of peaks and the 20 bins of 100 bp around TSS regions were calculated by BEDtools Intersect[66]. The coverage change was then calculated by deducting naïve 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 that have 3X read coverage over input was considered as significantly covered. In Fig. 6c, absolute TPM changes greater than 0.5 and absolute TSS ChIP coverage changes greater than 600bp were considered significant. Data was expressed as 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)[50] 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 HOMER[67] command 'makeTagDirectory' with 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[66]. The changes of coverage (Fig. 6a) were then calculated by deducting naive cell coverage around TSS from memory and effector cells, respectively. 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. naïve, memory

vs. naïve), genes were marked as increased or decreased if bulk gene expression changed more than 2-fold. Fisher's exact test was then performed to determine whether Ezh2- or H3K27me3-targeted genes negatively correlated with TPM changes. To be included as Ezh2 or H3K27me3 targets in Fig. 6 and 7, a gene must have >=100bp TSS region covered by Ezh2 peaks or >= 600bp TSS region covered by H3K27me3. In Fig. 6e, data were expressed as a percentage of unbound or Ezh2-bound genes in total genes that have either loss or gain of expression in Day 4 cells compared to $Div1_{TE}$ cells. In Fig. 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 most significant differentially expressed genes (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 H3K27me3- and Ezh2-targeted genes identified by ChIP (Fig. 6c,e). Differences with an associated P value < 0.05 were considered statistically significant.

Data availability

Source RNA-seq and ChIP-seq datasets are available at Gene Expression Omnibus, GSE89405 and GSE89036.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank members of the Chang and Yeo laboratories for discussions and critical reading of the manuscript. 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.). J.T.C. is a Howard Hughes Medical Institute Physician-Scientist Early Career Awardee. C.E.W. and P.J.M. were supported by NIH grant DK007202. We acknowledge the Sanford Consortium Stem Cell Genomics Core and Institute for Genomic Medicine Genomics Center for single-cell captures and sequencing.

References

1. Joshi NS, et al. Inflammation directs memory precursor and short-lived effector CD8(+) T cell fates via the graded expression of T-bet transcription factor. Immunity. 2007; 27:281–295. [PubMed: 17723218]
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. 1999; 401:708–712. [PubMed: 10537110]
3. Mueller SN, Mackay LK. Tissue-resident memory T cells: local specialists in immune defence. Nature reviews. Immunology. 2016; 16:79–89.
4. Best JA, et al. Transcriptional insights into the CD8(+) T cell response to infection and memory T cell formation. Nature immunology. 2013; 14:404–412. [PubMed: 23396170]
5. Kaech SM, Hemby S, Kersh E, Ahmed R. Molecular and functional profiling of memory CD8 T cell differentiation. Cell. 2002; 111:837–851. [PubMed: 12526810]

6. Chang JT, Wherry EJ, Goldrath AW. Molecular regulation of effector and memory T cell differentiation. Nature immunology. 2014; 15:1104–1115. [PubMed: 25396352]
7. Arsenio J, et al. Early specification of CD8+ T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses. Nature immunology. 2014; 15:365–372. [PubMed: 24584088]
8. Treutlein B, et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature. 2014; 509:371–375. [PubMed: 24739965]
9. Gaublomme JT, et al. Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity. Cell. 2015; 163:1400–1412. [PubMed: 26607794]
10. Shalek AK, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature. 2013; 498:236–240. [PubMed: 23685454]
11. Chang JT, et al. Asymmetric T lymphocyte division in the initiation of adaptive immune responses. Science. 2007; 315:1687–1691. [PubMed: 17332376]
12. Badovinac VP, Haring JS, Harty JT. Initial T cell receptor transgenic cell precursor frequency dictates critical aspects of the CD8(+) T cell response to infection. Immunity. 2007; 26:827–841. [PubMed: 17555991]
13. Ramskold D, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nature biotechnology. 2012; 30:777–782.
14. Rouault JP, et al. BTG1, a member of a new family of antiproliferative genes. The EMBO journal. 1992; 11:1663–1670. [PubMed: 1373383]
15. Roychoudhuri R, et al. BACH2 regulates CD8(+) T cell differentiation by controlling access of AP-1 factors to enhancers. Nature immunology. 2016; 17:851–860. [PubMed: 27158840]
16. Blackledge NP, Rose NR, Klose RJ. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. Nature reviews. Molecular cell biology. 2015; 16:643–649. [PubMed: 26420232]
17. DuPage M, et al. The chromatin-modifying enzyme Ezh2 is critical for the maintenance of regulatory T cell identity after activation. Immunity. 2015; 42:227–238. [PubMed: 25680271]
18. Su IH, et al. Polycomb group protein ezh2 controls actin polymerization and cell signaling. Cell. 2005; 121:425–436. [PubMed: 15882624]
19. Tumes DJ, et al. The polycomb protein Ezh2 regulates differentiation and plasticity of CD4(+) T helper type 1 and type 2 cells. Immunity. 2013; 39:819–832. [PubMed: 24238339]
20. Manjunath N, et al. Effector differentiation is not prerequisite for generation of memory cytotoxic T lymphocytes. The Journal of clinical investigation. 2001; 108:871–878. [PubMed: 11560956]
21. van der Windt GJ, et al. Mitochondrial respiratory capacity is a critical regulator of CD8+ T cell memory development. Immunity. 2012; 36:68–78. [PubMed: 22206904]
22. Ma C, Zhang N. Transforming growth factor-beta signaling is constantly shaping memory T-cell population. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112:11013–11017. [PubMed: 26283373]
23. Mackay LK, et al. T-box Transcription Factors Combine with the Cytokines TGF-beta and IL-15 to Control Tissue-Resident Memory T Cell Fate. Immunity. 2015; 43:1101–1111. [PubMed: 26682984]
24. Tinoco R, Alcalde V, Yang Y, Sauer K, Zuniga EI. Cell-intrinsic transforming growth factor-beta signaling mediates virus-specific CD8+ T cell deletion and viral persistence in vivo. Immunity. 2009; 31:145–157. [PubMed: 19604493]
25. Ananieva EA, Patel CH, Drake CH, Powell JD, Hutson SM. Cytosolic branched chain aminotransferase (BCATc) regulates mTORC1 signaling and glycolytic metabolism in CD4+ T cells. The Journal of biological chemistry. 2014; 289:18793–18804. [PubMed: 24847056]
26. Schober SL, 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. Journal of immunology. 1999; 163:3662–3667.
27. Skon CN, et al. Transcriptional downregulation of S1pr1 is required for the establishment of resident memory CD8+ T cells. Nature immunology. 2013; 14:1285–1293. [PubMed: 24162775]

28. Yamada T, Park CS, Mamonkin M, Lacorazza HD. Transcription factor ELF4 controls the proliferation and homing of CD8+ T cells via the Kruppel-like factors KLF4 and KLF2. Nature immunology. 2009; 10:618–626. [PubMed: 19412182]
29. Buck MD, et al. Mitochondrial Dynamics Controls T Cell Fate through Metabolic Programming. Cell. 2016; 166:63–76. [PubMed: 27293185]
30. Chtanova T, et al. Identification of T cell-restricted genes, and signatures for different T cell responses, using a comprehensive collection of microarray datasets. Journal of immunology. 2005; 175:7837–7847.
31. Willinger T, Freeman T, Hasegawa H, McMichael AJ, Callan MF. Molecular signatures distinguish human central memory from effector memory CD8 T cell subsets. Journal of immunology. 2005; 175:5895–5903.
32. Bouneaud C, Garcia Z, Kourilsky P, Pannetier C. Lineage relationships, homeostasis, and recall capacities of central- and effector-memory CD8 T cells in vivo. The Journal of experimental medicine. 2005; 201:579–590. [PubMed: 15710650]
33. Wherry EJ, et al. Lineage relationship and protective immunity of memory CD8 T cell subsets. Nature immunology. 2003; 4:225–234. [PubMed: 12563257]
34. Gaide O, et al. Common clonal origin of central and resident memory T cells following skin immunization. Nature medicine. 2015; 21:647–653.
35. Chang JT, et al. Asymmetric proteasome segregation as a mechanism for unequal partitioning of the transcription factor T-bet during T lymphocyte division. Immunity. 2011; 34:492–504. [PubMed: 21497118]
36. Lin WH, et al. Asymmetric PI3K Signaling Driving Developmental and Regenerative Cell Fate Bifurcation. Cell reports. 2015; 13:2203–2218. [PubMed: 26628372]
37. Pollizzi KN, et al. Asymmetric inheritance of mTORC1 kinase activity during division dictates CD8(+) T cell differentiation. Nature immunology. 2016; 17:704–711. [PubMed: 27064374]
38. Verbist KC, et al. Metabolic maintenance of cell asymmetry following division in activated T lymphocytes. Nature. 2016; 532:389–393. [PubMed: 27064903]
39. Metz PJ, et al. Regulation of asymmetric division and CD8+ T lymphocyte fate specification by protein kinase Czeta and protein kinase Clambda/iota. Journal of immunology. 2015; 194:2249–2259.
40. Zhao E, et al. Cancer mediates effector T cell dysfunction by targeting microRNAs and EZH2 via glycolysis restriction. Nature immunology. 2016; 17:95–103. [PubMed: 26523864]
41. Araki Y, Fann M, Wersto R, Weng NP. 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). Journal of immunology. 2008; 180:8102–8108.
42. Youngblood B, et al. Chronic virus infection enforces demethylation of the locus that encodes PD-1 in antigen-specific CD8(+) T cells. Immunity. 2011; 35:400–412. [PubMed: 21943489]
43. Chappell C, Beard C, Altman J, Jaenisch R, Jacob J. DNA methylation by DNA methyltransferase 1 is critical for effector CD8 T cell expansion. Journal of immunology. 2006; 176:4562–4572.
44. 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. 2009; 30:912–925. [PubMed: 19523850]
45. Russ BE, et al. Distinct epigenetic signatures delineate transcriptional programs during virus-specific CD8(+) T cell differentiation. Immunity. 2014; 41:853–865. [PubMed: 25517617]
46. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nature biotechnology. 2016; 34:525–527.
47. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011; 27:863–864. [PubMed: 21278185]
48. Chen C, Khaleel SS, Huang H, Wu CH. Software for pre-processing Illumina next-generation sequencing short read sequences. Source code for biology and medicine. 2014; 9:8. [PubMed: 24955109]
49. Ntranos V, Kamath GM, Zhang JM, Pachter L, Tse DN. Fast and accurate single-cell RNA-seq analysis by clustering of transcript-compatibility counts. Genome biology. 2016; 17:112. [PubMed: 27230763]

50. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29:15–21. [PubMed: 23104886]
51. Mouse Genome Sequencing, C. et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002; 420:520–562. [PubMed: 12466850]
52. Maaten LHGE. Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research. 2008; 9:2579–2605.
53. Maaten L. Barnes-Hut-SNE. 2013:1301. arXIv.
54. Amir el AD, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nature biotechnology. 2013; 31:545–552.
55. Becher B, et al. High-dimensional analysis of the murine myeloid cell system. Nature immunology. 2014; 15:1181–1189. [PubMed: 25306126]
56. Bendall SC, et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science. 2011; 332:687–696. [PubMed: 21551058]
57. Cheng Y, Wong MT, van der Maaten L, Newell EW. Categorical Analysis of Human T Cell Heterogeneity with One-Dimensional Soli-Expression by Nonlinear Stochastic Embedding. Journal of immunology. 2016; 196:924–932.
58. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014; 15:550. [PubMed: 25516281]
59. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140. [PubMed: 19910308]
60. Fan J, et al. Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. Nature methods. 2016; 13:241–244. [PubMed: 26780092]
61. Li J, Tibshirani R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Statistical methods in medical research. 2013; 22:519–536. [PubMed: 22127579]
62. Stephens MA. EDF Statistics for Goodness of Fit and Some Comparisons. Journal of the American Statistical Association. 1974; 69:730–737.
63. Geurts PE, Wehenkel DL. Extremely Randomized Trees. Machine Learning. 2006; 62:3–42.
64. Pedregosa FV, Gramfor GA. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011; 12:2825–2830.
65. Setty M, et al. Wishbone identifies bifurcating developmental trajectories from single-cell data. Nature biotechnology. 2016; 34:637–645.
66. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis … [et al.]. 2014; 47:11 12 11–34.
67. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell. 2010; 38:576–589. [PubMed: 20513432]

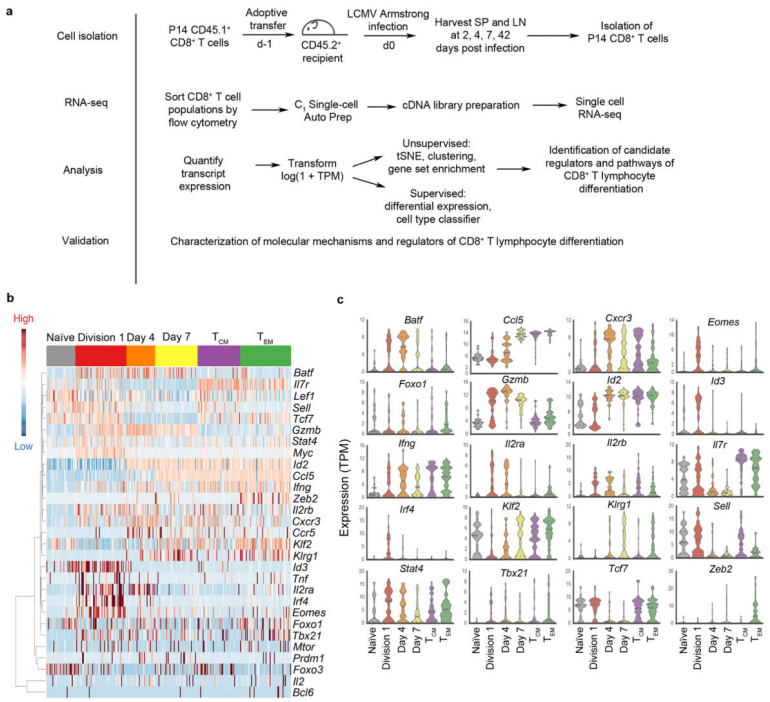

Figure 1.

Single-cell RNA-seq analysis of CD8$^+$ T lymphocytes responding to viral infection. (**a**) Experimental and analytical approaches. (**b**) Single-cell expression of selected genes previously associated with CD8$^+$ T cell differentiation among cells from the following populations: naïve cells (gray), Division 1 (red, CD44^{hi}), Day 4 (orange), Day 7 (yellow), central memory (T_{CM}) (green), and effector memory (T_{EM}) cells. (**c**) Expression of selected genes previously associated with CD8$^+$ T cell differentiation measured in (**b**), presented as violin plots, ordered alphabetically. TPM, transcripts per million reads.

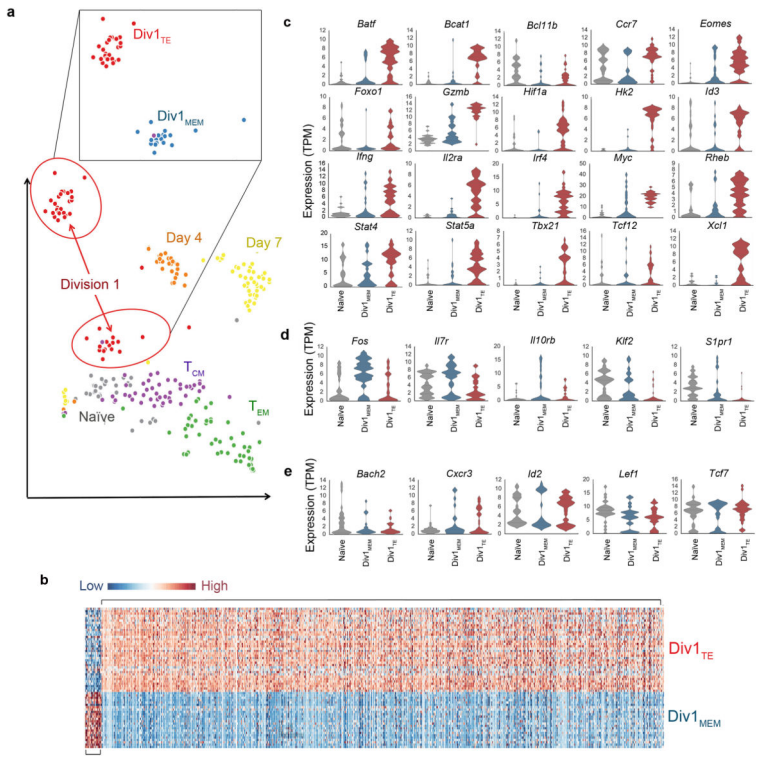

Figure 2.

Cells that have undergone their first division exhibit transcriptional heterogeneity. (**a**) t-distributed Stochastic Neighborhood Embedding (tSNE) of 6000 expressed genes from single $CD8^+$ T cells from the following populations: naïve (gray), Division 1 (red), Day 4 (orange), Day 7 (yellow), central memory (T_{CM}) (green) and effector memory (T_{EM}) cells. Each circle represents a single cell. Inset, separation of Division 1 cells by tSNE clustering analysis into two distinct subpopulations, labeled $Div1_{TE}$ (red) and $Div1_{MEM}$ (blue) cells. TE, terminal effector; MEM, memory. (**b**) Differential expression of genes among $Div1_{TE}$ (top) and $Div1_{MEM}$ (bottom) cells presented as a heatmap. (**c**–**e**) Expression of selected genes previously associated with $CD8^+$ T cell differentiation among individual naïve (gray), $Div1_{TE}$ (red), and $Div1_{MEM}$ (blue) cells, presented as violin plots, ordered alphabetically. Higher expression in $Div1_{TE}$ relative to Div 1_{MEM} cells (**c**), lower expression in $Div1_{TE}$ relative to $Div1_{MEM}$ cells (**d**), or equal expression in $Div1_{TE}$ and $Div1_{MEM}$ cells (**e**).

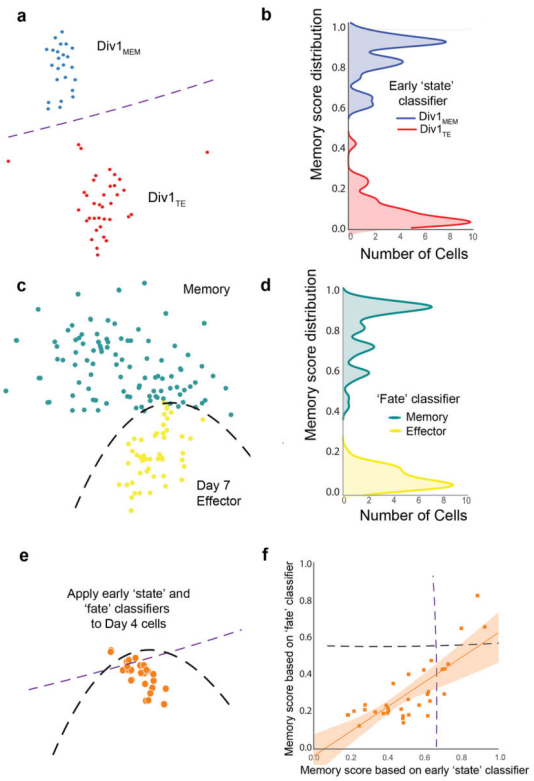

Figure 3.

Generation and application of early 'state' and 'fate' classifiers to predict the identity of cells in intermediate states of differentiation. Early state and fate classifiers learn differences in the gene expression signatures of early memory-like cells ($Div1_{MEM}$) versus early effector-like cells ($Div1_{TE}$) identified in Fig. 2a and Day 7 effector cells versus memory cells, respectively. (**a**, **c**) Schematic representation of Extra Trees Classifier (ETC) that separates Division 1 lymphocyte clusters ($Div1_{MEM}$, blue, n=24; $Div1_{TE}$, red, n=36) (**a**) and Day 7 effector, yellow, n=48 versus total memory cells, teal, n=96, including T_{CM} cells (n=48) and T_{EM} cells (n= 48) (**c**). (**b, d**) Kernel density histograms of cross-validated scores on Division 1 $CD8^+$ T lymphocytes (**b**) and Day 7 effector and memory $CD8^+$ T lymphocytes (**d**) from which early state and fate classifiers were trained, respectively. (**e**) Schematic representation of applying early state and fate classifiers to predict the fate of individual Day 4 $CD8^+$ T lymphocytes, n=34. The 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 measured by (**e**). Memory score distribution of early state classifier (x-axis, 0=effector to 1=memory) versus memory score distribution of final fate classifier (y-axis, 0=effector to 1=memory). Squares represent individual Day 4 $CD8^+$ T lymphocytes. Early state and fate classifier scores correlate well in both linear (Pearson: r=0.78, $p=4.8 \times 10^{-8}$) and monotonic sense (Spearman: r=0.71, $p=2.2 \times 10^{-6}$). The dashed black and purple lines indicate the fate classifier's decision boundary between memory and Day 7 effector cells. The orange line indicates the early state classifier's decision boundary between $Div1_{MEM}$ and $Div1_{TE}$ cells. Both of these lines are stylized estimates of the real decision boundaries, which are complex piecewise-linear functions and

can be much more furrowed. The orange shaded area around the linear regression line indicates the 95% confidence interval assuming Gaussian error.

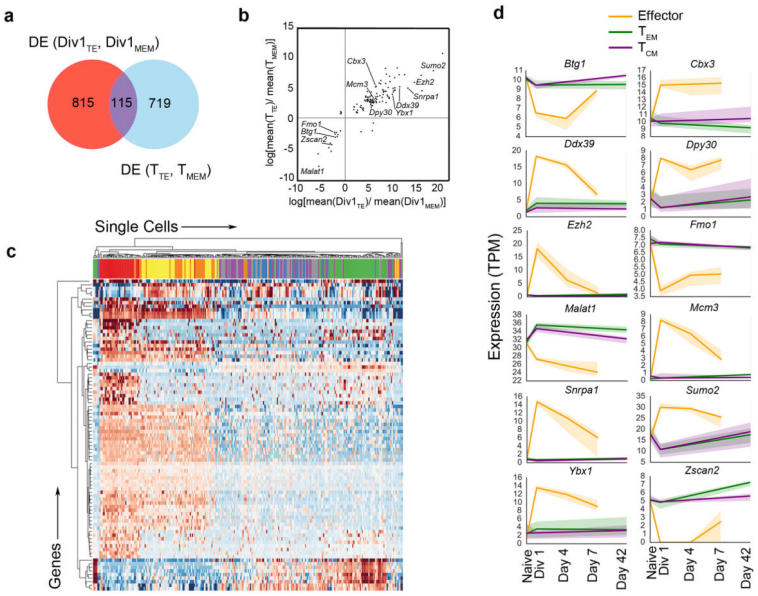

Figure 4.

Identification of putative regulators of CD8$^+$ T lymphocyte differentiation. (**a**) Venn diagram of differential expression (DE) of genes between Division 1 cells (Div1$_{TE}$, Div1$_{MEM}$) (red), and DE of genes between effector and memory cells (blue). (**b**) Change in mean expression of DE genes in memory cells compared to effector cells (y-axis) versus the change in mean expression of DE genes in Div1$_{MEM}$ cells compared to Div1$_{TE}$ cells. Pearson correlation r=0.78, p=3.6 × 10^{-13}. Circles represent individual genes (**c**) Differential expression of 89 common genes (rows) clustered across all CD8$^+$ T lymphocyte populations. Each column represents a single cell. Naïve cells (gray), Div1$_{TE}$ (red), Div1$_{MEM}$ (blue), Day 4 (orange), Day 7 (yellow), central memory (T$_{CM}$) (green), and effector memory (T$_{EM}$) cells. (**d**) Temporal expression patterns of selected genes across inferred paths of differentiation for effector (orange), T$_{CM}$ (purple), and T$_{EM}$ cells (green). Shaded areas indicate around the lines indicate the 95% confidence interval bootstrapped from all possible single-cell expression trajectories.

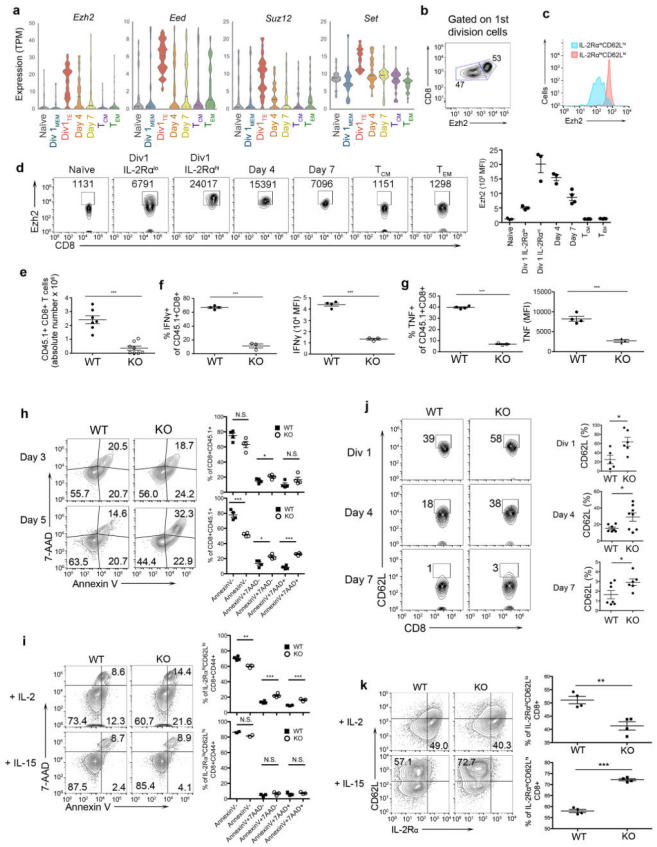

Figure 5.

Ezh2 regulates effector CD8$^+$ T lymphocyte differentiation. (**a**) Expression of PRC2 complex genes *Ezh2*, *Eed*, *Suz12*, and *Set*, in single CD8$^+$ T lymphocytes in naïve (gray), Div1$_{MEM}$ (blue), Div1$_{TE}$ (red), Day 4 (orange), Day 7 (yellow), T_{CM} (purple), and T_{EM} (green) populations, presented as violin plots. (**b**) Ezh2 protein expression in gated CD8$^+$ T lymphocytes that have undergone their first cell division *in vivo*. (**c**) Ezh2 protein expression in gated Division 1 CD8$^+$ IL-2Rα^{hi}CD62L^{lo} (red) and IL-2Rα^{lo}CD62L^{hi} (blue) cells. (**d**) Ezh2 expression in naïve, Division 1 (gated on IL-2Rα^{lo} or IL-2Rα^{hi} cells), Day 4, Day 7, T_{CM}, and T_{EM} CD8$^+$ T lymphocytes, presented as flow cytometry plots (left) or bar graphs (right). Numbers in FACS plots indicate mean fluorescence intensity of Ezh2$^+$ events. (**e**) Absolute numbers of gated *Ezh2*$^{fl/fl}$*Cd4*$^{+/+}$ ('WT') and *Ezh2*$^{fl/fl}$*Cd4*^{Cre} ('KO') P14 cells adoptively transferred into recipient mice subsequently infected with LCMV and analyzed in the spleen 7 d post-infection. Data are pooled from two independent experiments. (**f**, **g**) Flow cytometry analysis (left) and mean fluorescence intensity (right) of intracellular IFNγ (**f**) and TNF (**g**) in cells at 7 d post-infection. (**h**) 7-AAD and Annexin V expression in gated P14 CD8$^+$CD45.1$^+$ WT and KO cells, analyzed at 3 or 5 d post-infection, presented as flow cytometry analysis (left) and dot plots (right). (**i**) Analysis of 7-AAD and Annexin V expression in gated IL-2Rα^{hi}CD62L^{lo} and IL-2Rα^{lo}CD62L^{hi} WT or KO cells at 3 d following *in vitro* culture with IL-2 (top) or IL-15 (bottom), respectively, presented as flow cytometry analysis (left) and dot plots (right). (**j**) Analysis of CD62L expression by Division 1, Day 4, and Day 7 WT and KO CD8$^+$ T cells responding to LCMV infection, presented as

flow cytometry analysis (left) and dot plots (right). (**k**) Proportion of IL-2Rα^{hi}CD62L^{lo} and IL-2Rα^{lo}CD62L^{hi} WT or KO cells at 3 d following *in vitro* culture with IL-2 (top) or IL-15 (bottom), respectively, presented as flow cytometry analysis (left) and dot plots (right). ** $p < 0.01$, *** $p < 0.001$, N.S. not significant (Student's two-tailed *t*-test). Data are representative of 2 independent experiments with 3 (**d**) or 4 (**b**, **c**) WT mice each; 2 independent experiments with 4 individual WT and KO mice each (**e–g**); 3 individual WT and KO mice (**h, i**); 6 individual WT and KO mice (**j**). Mean and S.E.M. are indicated (**d–k**).

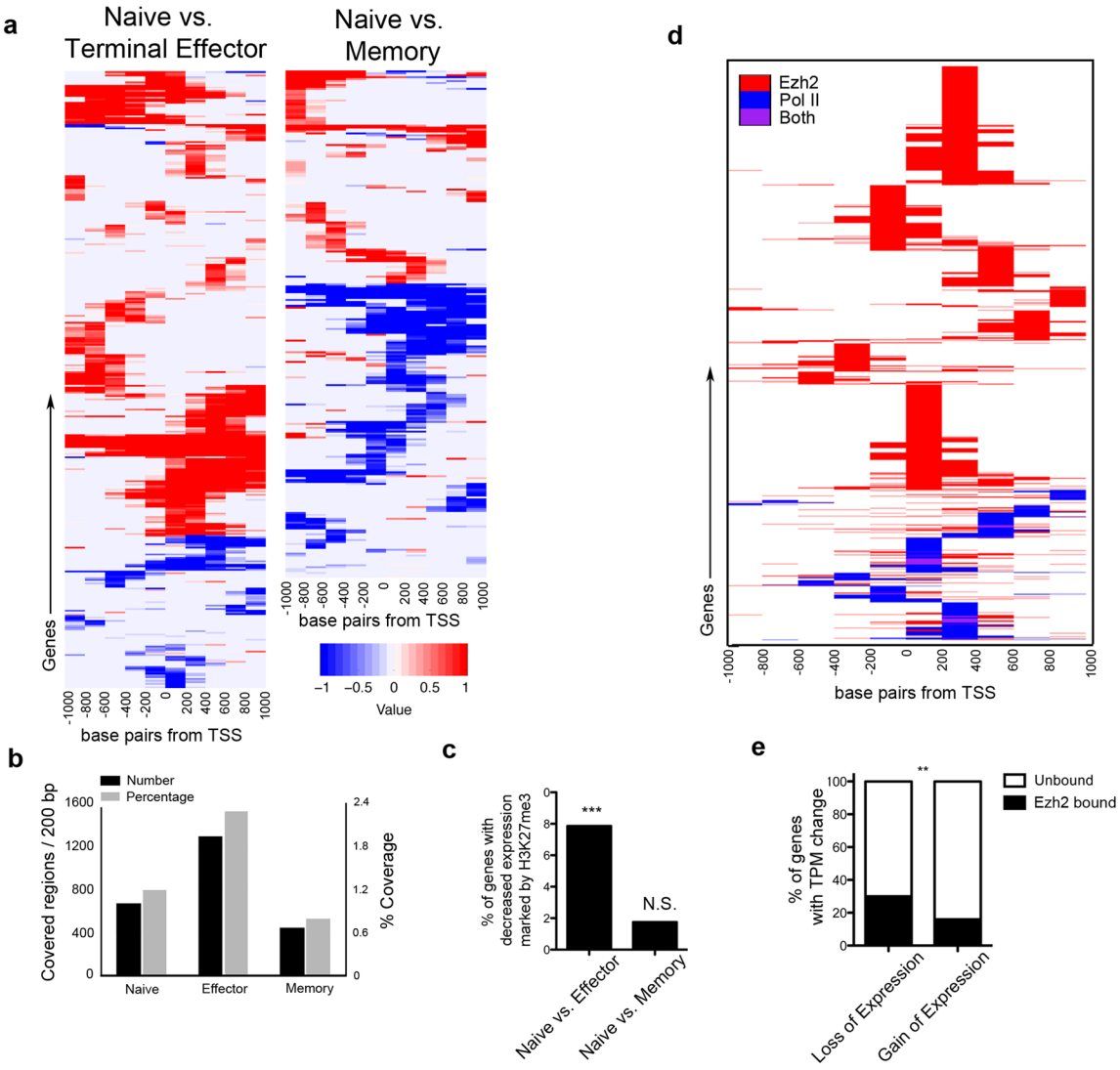

Figure 6.

Increased epigenetic repression in genes expressed during terminal effector differentiation. (**a**) Change in H3K27me3 coverage from naïve to effector (left) and naïve to memory (right). Red: H3K27me3 coverage gains; Blue: H3K27me3 coverage losses. (**b**) H3K27me3 peak coverage around TSS for the 6000 genes detected in single-cell RNA-seq dataset. (**c**) Analysis of H3K27me3 coverage in genes exhibiting reduced expression in effector or memory cells relative to naive cells. Comparing naïve to effector cells, 1234 genes have decreased expression with 97 genes marked by H3K27me3 (7.86%); comparing naïve to memory cells, 627 genes have decreased expression with 11 genes marked by H3K27me3 (1.75%). (**d**) ChIP analysis of Ezh2 (red) and Pol2 (blue) binding at proximal promoter regions (−1 and +1 kb of the TSS) of the 6000 genes detected by single-cell RNA-seq. (**e**) Presence or absence of Ezh2 binding to genes whose expression was reduced (left, total 1492 genes) or increased (right, total 219 genes) in Day 4 cells relative to $Div1_{TE}$ cells. Genes with reduced expression had 448 Ezh2-bound targets while genes with increased

expression had 34 Ezh2-bound targets. ** $p < 0.01$, *** $p < 0.001$, N.S. not significant (Fisher's Exact Test) (**c, e**).

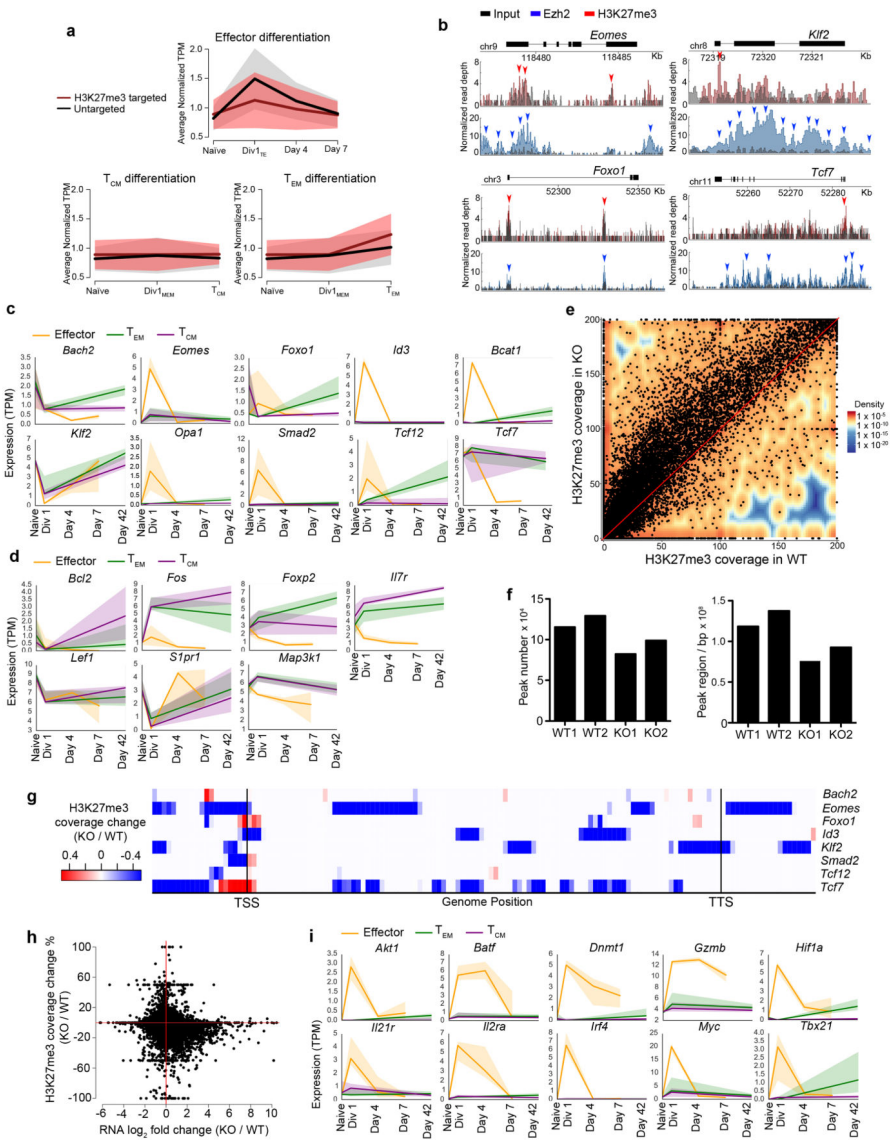

Figure 7.

Ezh2 mediates CD8$^+$ T lymphocyte effector differentiation through epigenetic repression. (**a**) Averaged normalized expression in the three inferred differentiation paths (effector, T_{CM}, T_{EM}) for genes marked or unmarked by H3K27me3. (**b**) ChIP-seq analysis of the binding of Ezh2 and H3K27me3 histone modifications at *Eomes*, *Foxo1*, *Klf2*, and *Tcf7*. Gray indicates input. Red (Ezh2) and blue (H3K27me3) arrows indicate binding peaks. (**c, d**) Temporal expression patterns of selected genes previously implicated in CD8$^+$ T cell differentiation, grouped alphabetically, across inferred paths of differentiation for effector (orange), T_{CM} (purple), and T_{EM} cells (green). Memory-associated genes targeted by Ezh2 (**c**) or not targeted by Ezh2 (**d**) are shown. Shaded areas around the lines indicate the 95% confidence interval bootstrapped from all possible single-cell expression trajectories. (**e**) Density dot plot representing H3K27me3 coverage for individual genes in wild-type ('WT') vs. Ezh2-deficient ('KO') cells. (**f**) H3K27me3 coverage, represented by peak numbers (top)

or regions (bottom) in KO compared to WT cells. Two biologic replicates were performed. (**g**) Heatmap showing changes in H3K27me3 coverage of Ezh2-targeted memory-associated genes in WT vs. KO $CD8^+$ T cells. (**h**) Dot plot showing relationship between H3K27me3 coverage changes (KO/WT) and gene expression changes (KO/WT) for individual genes in WT vs. KO $CD8^+$ T cells. (**i**) Temporal expression patterns, as in **c**, of selected genes associated with effector differentiation.