nature immunology

Article

https://doi.org/10.1038/s41590-026-02479-6

A unique CD4$^+$ T cell subset expressing granzyme K is regulated by transcription factor EOMES and important for T cell-mediated intestinal inflammation

Received: 27 October 2025

Accepted: 25 February 2026

Published online: 02 April 2026

Check for updates

Tian Xie[1,2,3], Yizhou Du[2], Qihan Wang[2], Hao Zhang[1,2], Kun Wei[1], Xinxin Chi[1], Xue Bai[1,5], Yujie Fu[1,2], Zhilin Peng[2], Yicheng Zhu[2,4], Qiuyan Lan[2] & Chen Dong[2]

CD4$^+$ helper T (T_H) cells consist of multiple functional subsets defined by specific effector cytokines and transcription factors. Recently, single-cell transcriptomic analyses have revealed possible existence of additional populations. Here we identify a unique CD4$^+$ T cell subset in mouse and human colitis characterized by high levels of granzyme K (*Gzmk*) expression, designated as T_HK cells. These cells exhibit unique transcriptional signatures, with minimal expression of classical T_H-defining factors but rather prominent *Eomesodermin* (*Eomes*) expression. Notably, T_HK cell differentiation is independent of T_H1, T_H2 and T_H17 lineages in colitis. EOMES is both necessary and sufficient for T_HK cell induction, by directly driving the expression of *Gzmk* and associated effector molecules. Genetic ablation of *Eomes* ameliorates intestinal immunopathology in a T cell-induced colitis model. The T_HK transcriptional program seems to be conserved across species and in diverse disease contexts. Our findings establish T_HK cells as a distinct T_H cell subtype, and the EOMES–T_HK axis may serve as a potential therapeutic target in inflammatory diseases.

CD4$^+$ helper T (T_H) cells produce cytokines and act as central modulators of adaptive immunity. They are traditionally categorized into distinct subsets such as T_H1, T_H2, T_H17 and follicular helper T (T_{FH}) cells. Each subset is defined by a unique combination of lineage-determining transcription factors and effector cytokines that collectively mediate specialized immune responses to certain pathogens[1].

In parallel, CD8$^+$ T cells constitute another major arm of adaptive immunity and are critical for eliminating infected or malignant cells. Their cytotoxicity depends on perforin, a pore-forming protein that delivers granzymes into target cells[2]. Among these, granzyme B (GZMB) is the most potent and best-characterized executioner, inducing target cell death through direct cleavage and activation of caspases[2,3]. Notably, recent studies have identified a distinct CD8$^+$ T cell subset marked by high GZMK expression[4–7]. Unlike conventional GZMB$^+$ CD8$^+$ T cells, GZMK$^+$ cells exhibit minimal cytotoxic activity and are instead associated with enhanced cytokine production, tissue residency and immunomodulatory functions[3,8].

The function of GZMK remains poorly understood, as its cytotoxic potential is less established than that of GZMB; however, emerging evidence suggests it may modulate inflammation by promoting

[1]Institute for Immunology and School of Basic Medical Sciences, Tsinghua University, Beijing, China. [2]Westlake University School of Medicine and New Cornerstone Laboratory, Hangzhou, China. [3]Peking University-Tsinghua University-National Institute of Biological Sciences Joint Graduate Program, Tsinghua University, Beijing, China. [4]Shanghai Immune Therapy Institute, Shanghai Jiao Tong University School of Medicine-affiliated Renji Hospital, Shanghai, China. [5]Present address: Department of Gastroenterology, Shengjing Hospital of China Medical University, Shenyang, China.
e-mail: dongchen@westlake.edu.cn

https://doi.org/10.1038/s41590-026-02479-6

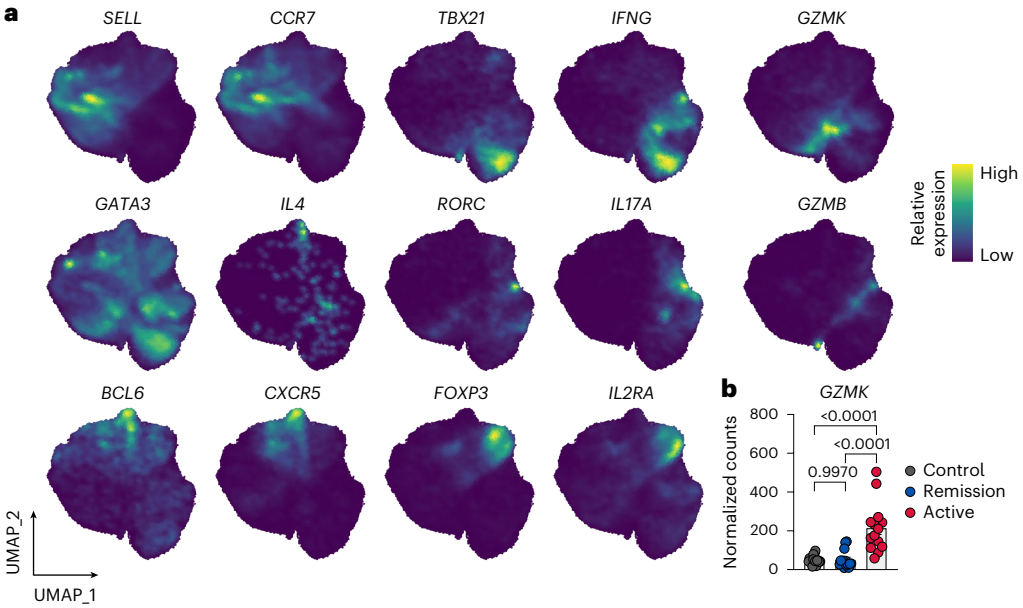


Fig. 1 | Identification of a *GZMK*-expressing CD4⁺ T cell subset in human intestinal inflammation. a, UMAP visualization of CD4⁺ T cells from an integrated scRNA-seq dataset of intestinal, peripheral blood and lymph node samples from individuals with IBD and healthy controls. Density plots show expression of the indicated genes (*SELL*, *CCR7*, *TBX21*, *IFNG*, *GATA3*, *IL4*, *RORC*, *IL17A*, *BCL6*, *CXCR5*, *FOXP3*, *IL2RA*, *GZMK* and *GZMB*). **b**, *GZMK* transcript levels in mucosal biopsies from patients with active UC ($n = 14$), UC in remission ($n = 14$) and healthy controls ($n = 16$) (GSE128682). Data are represented as mean ± s.e.m. Exact *P* values are shown in the graph. Statistical significance was determined by one-way analysis of variance (ANOVA) with Tukey's multiple comparisons test.

inflammatory responses through extracellular proteolytic activity, such as cleavage of complement components and protease-activated receptor 1 (PAR-1), to activate the complement cascade and induce inflammatory cytokine production[3–6,8].

GZMK expression has also been observed in CD4⁺ T cells, which are typically identified by single-cell RNA sequencing (scRNA-seq) and annotated as T_H1-like or cytotoxic CD4⁺ T cells[9–13]. Nevertheless, those cells have not been systematically evaluated, and their functional significance, transcriptional programs and developmental regulation remain to be defined.

GZMK⁺ CD4⁺ T cells with noncanonical identity in humans

To investigate the heterogeneity of CD4⁺ T cells in intestinal inflammation, we reanalyzed an integrated scRNA-seq dataset from patients with inflammatory bowel disease (IBD), including ulcerative colitis (UC) and Crohn's disease, as well as healthy individuals[14]. The dataset encompassed multiple tissue sources, including the small intestine, colon, peripheral blood and lymph nodes. Analysis of CD4⁺ T cells revealed canonical T_H cell subsets: naive and central memory T cells (expressing *SELL* and *CCR7*), T_H1 cells (expressing *TBX21* and *IFNG*), T_H2 cells (expressing *GATA3* and *IL4*), T_H17 cells (expressing *RORC* and *IL17A*), T_{FH} cells (expressing *BCL6* and *CXCR5*) and regulatory T (T_{reg}) cells (expressing *FOXP3* and *IL2RA*) (Fig. 1a). Notably, in addition to these known subsets, we discovered a CD4⁺ T cell population characterized by high levels of *GZMK* expression, which did not coincide with prominent expression of canonical CD4⁺ T cell lineage markers (Fig. 1a). Furthermore, this population exhibited minimal expression of *GZMB* (Fig. 1a). These transcriptional features suggest that *GZMK*^{high} CD4⁺ T cells represent a previously underappreciated subset with a unique molecular identity.

To assess the potential relevance of this population in intestinal inflammation, we analyzed bulk RNA-seq data (GSE128682) from mucosal biopsies of patients with active UC, UC in remission and healthy controls[15]. We observed a significant elevation of *GZMK* expression in the inflamed mucosa of patients with active UC compared to those in remission and healthy individuals (Fig. 1b). This finding suggests the presence of *GZMK*-expressing CD4⁺ cells in active disease.

Beyond the gut, we consistently observed a *GZMK*^{high} CD4⁺ T cell population in human pancancer scRNA-seq data[16], which also lacked canonical CD4⁺ T cell lineage markers (Extended Data Fig. 1), suggesting that this population may represent a broadly existing, functionally distinct state in humans across inflammatory and neoplastic contexts.

Characterization of *Gzmk*-expressing CD4⁺ T cells in murine colitis

To investigate the development and function of *GZMK*-expressing CD4⁺ T cells in vivo, we next turned to a mouse colitis model. We integrated public scRNA-seq datasets (CRA016814 and GSE235664)[17,18] of colonic CD4⁺ T cells from *Rag1*$^{-/-}$ mice adoptively transferred with naive CD4⁺ T cells (a colitis model that recapitulates key features of human IBD). Unsupervised clustering revealed six transcriptionally distinct populations (Fig. 2a). Cluster 0 expressed canonical T_H1 genes (*Tbx21*, *Il12rb2* and *Ifng*), whereas cluster 2 displayed a memory- and stem-like signature marked by high expression of *Il7r*, *Slamf6*, *Tcf7* and *Klf2*. Cluster 3 represented a proliferating population characterized by high *Mki67* expression. Cluster 4 expressed hallmark genes of T_H17 (*Rorc*, *Il17a* and *Il17f*) and T_{FH} (*Bcl6*) lineages and cluster 5 was defined by high expression of the T_{reg} cell master regulator *Foxp3* (Fig. 2b). Of note, cluster 1 emerged as a unique population distinguished by substantial expression of *Gzmk*, with minimal expression of other lineage-defining transcription factors (Fig. 2b,c).

To validate these findings experimentally, we utilized a novel *Gzmk-P2A-CreERT2-T2A-tdTomato* (*Gzmk*-tdTomato) reporter mouse strain. Flow cytometric analysis of intestinal lymphocytes from steady-state mice revealed tdTomato signals in subsets of αβ T cells (including CD4⁺, CD8αα⁺ and CD8αβ⁺ subsets), and γδ T cells within the small intestinal intraepithelial and lamina propria (LP) compartments. In contrast, *Gzmk*-tdTomato⁺ CD4⁺ T cells were rare in colonic intraepithelial and LP compartments, as well as in mesenteric lymph nodes (mLNs) (Extended Data Fig. 2).

https://doi.org/10.1038/s41590-026-02479-6

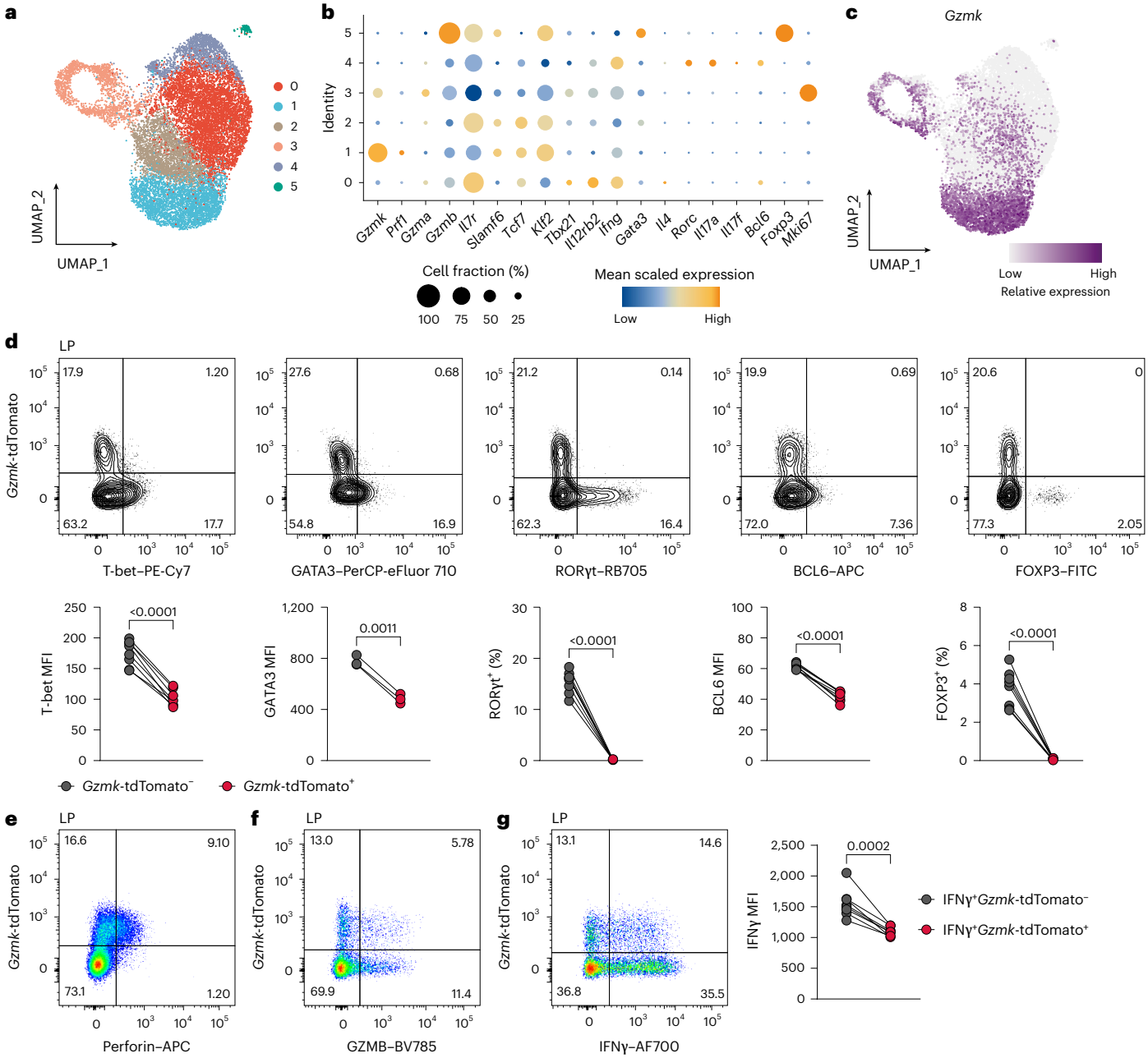


Fig. 2 | A distinct *Gzmk*-expressing CD4$^+$ T cell population in murine colitis. a, UMAP visualization of colonic CD4$^+$ T cells from the integration of two independent scRNA-seq datasets (CRA016814 and GSE235664) using the *Rag1*$^{-/-}$ adoptive transfer model of colitis. Clusters were identified by unsupervised clustering. **b**, Dot plot showing average expression (color) and percentage of expressing cells (dot size) for selected genes across six clusters. **c**, Feature plot displaying *Gzmk* expression in the UMAP. **d**, Flow cytometry analysis of *Gzmk*-tdTomato expression in LP CD4$^+$ T cells of *Rag1*$^{-/-}$ colitis mice adoptively transferred with naive CD4$^+$ T cells from *Gzmk*-tdTomato reporter mice. Representative plots (*y* axis, *Gzmk*-tdTomato; *x* axis, T-bet, GATA3, RORγt, BCL6 or FOXP3) (top). Quantification of MFI of T-bet ($n = 8$), GATA3 ($n = 3$) and BCL6 ($n = 8$), and frequency of RORγt$^+$ ($n = 8$) and FOXP3$^+$ ($n = 8$) cells in *Gzmk*-tdTomato$^+$ versus *Gzmk*-tdTomato$^-$ cells (bottom). **e**–**g**, Flow cytometry analysis of perforin (**e**), GZMB (**f**) and IFNγ (**g**) in LP CD4$^+$ T cells. Representative plots (*y* axis, *Gzmk*-tdTomato; *x* axis, target protein) (left). In **g**, right: MFI of IFNγ in IFNγ$^+$*Gzmk*-tdTomato$^-$ and IFNγ$^+$*Gzmk*-tdTomato$^+$ subsets ($n = 8$). Data are representative of at least two independent experiments. Exact *P* values are shown in each graph. Statistical significance was assessed by two-tailed paired *t*-test.

We next examined *Gzmk* expression by CD4$^+$ T cells during colitis. Naive CD4$^+$ T cells isolated from *Gzmk*-tdTomato reporter mice were adoptively transferred into *Rag1*$^{-/-}$ recipients to induce colitis. In line with the scRNA-seq data, approximately 20% of CD4$^+$ T cells in colonic LP and mLNs exhibited tdTomato fluorescence, confirming *Gzmk* expression during intestinal inflammation (Fig. 2d and Extended Data Fig. 3a). Flow cytometric analysis revealed that *Gzmk*-tdTomato$^+$ cells exhibited minimal or no expression of T-bet, GATA3, RORγt, BCL6 or FOXP3, further supporting that they were distinct from classical T_H and T_{reg} cell lineages (Fig. 2d and Extended Data Fig. 3a). For simplicity, we named this *Gzmk*^{high} subset T_HK cells.

Phenotypic characterization revealed that the majority of perforin-expressing CD4$^+$ T cells were *Gzmk*-tdTomato$^+$, though only a subset of *Gzmk*-tdTomato$^+$ cells expressed perforin (Fig. 2e and Extended Data Fig. 3b). Among cells expressing *Gzmk*-tdTomato or GZMB, a small subset was double-positive for both granzymes, suggesting partial overlap in their expression patterns (Fig. 2f and Extended Data Fig. 3c). Furthermore, while a proportion

https://doi.org/10.1038/s41590-026-02479-6

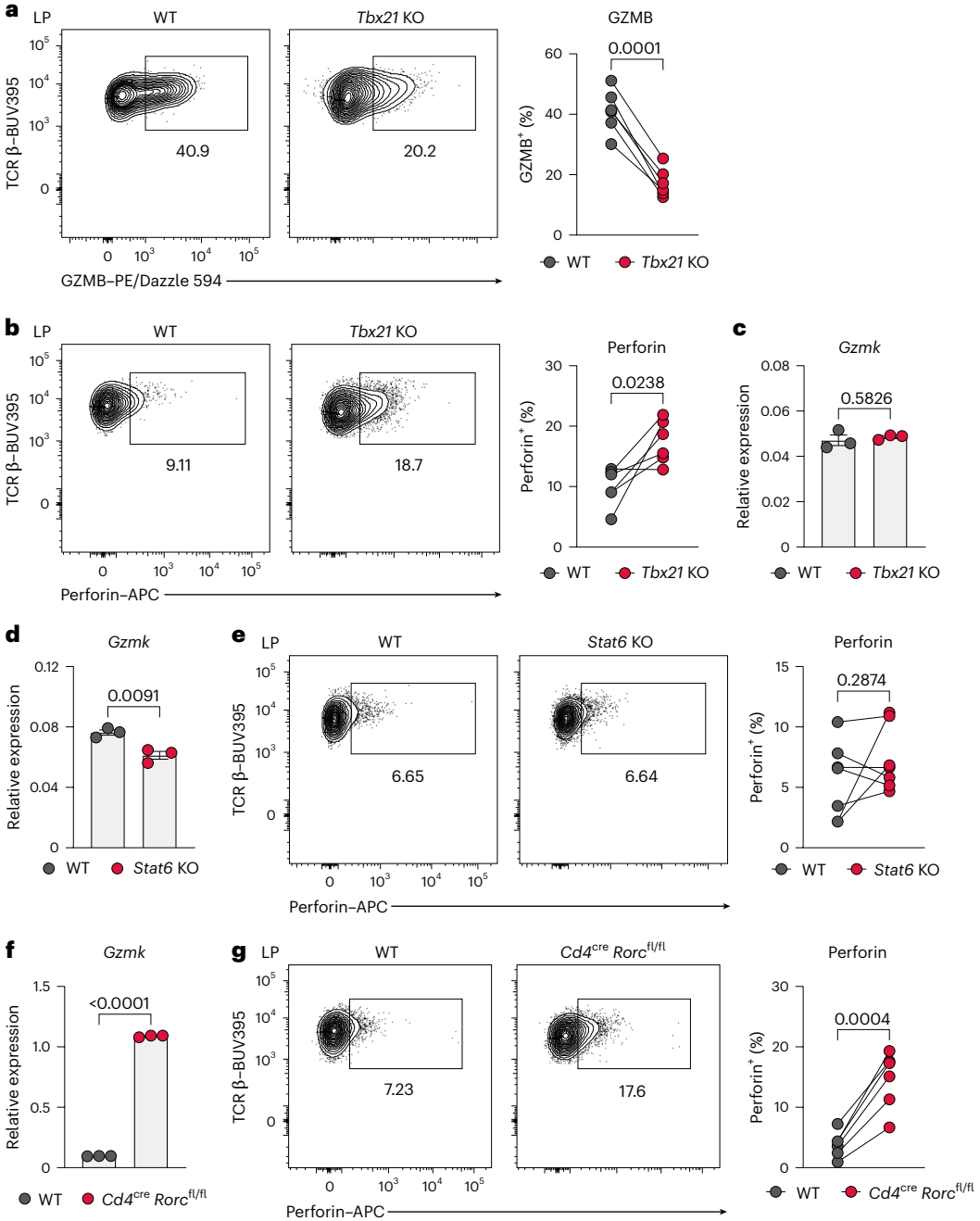


Fig. 3 | T_HK cell differentiation is independent of T_H1, T_H2 and T_H17 programs. **a**,**b**, Flow cytometry analysis of GZMB (**a**) and perforin (**b**) in LP CD4$^+$ T cells from *Rag1*$^{-/-}$ recipients co-transferred with wild-type (WT) (CD45.1$^+$) and *Tbx21*$^{-/-}$ (CD45.2$^+$) naive CD4$^+$ T cells ($n = 6$). Representative plots (left). Frequency of GZMB$^+$ or perforin$^+$ cells among donor-derived populations (right). KO, knockout. **c**, *Gzmk* mRNA expression in sorted WT and *Tbx21*$^{-/-}$ CD4$^+$ T cells from the colonic LP, measured by RT–qPCR ($n = 3$). **d**, *Gzmk* mRNA expression in sorted WT and *Stat6*$^{-/-}$ CD4$^+$ T cells from the colonic LP, measured by RT–qPCR ($n = 3$). **e**, Flow cytometry analysis of perforin in LP CD4$^+$ T cells from recipients of WT and *Stat6*$^{-/-}$ cells ($n = 7$). Representative plots (left). Frequency of perforin$^+$ cells (right). **f**, *Gzmk* mRNA expression in sorted WT and *Cd4*^{cre} *Rorc*$^{fl/fl}$ CD4$^+$ T cells from the colonic LP, measured by RT–qPCR ($n = 3$). **g**, Flow cytometry analysis of perforin in LP CD4$^+$ T cells from recipients of WT and *Cd4*^{cre} *Rorc*$^{fl/fl}$ cells ($n = 6$). Representative plots (left). Frequency of perforin$^+$ cells (right). Data are shown as mean ± s.e.m. and are representative of two independent experiments. Exact *P* values are shown in each graph. Statistical significance was assessed by two-tailed paired *t*-test (**a**,**b**,**e**,**g**) or two-tailed unpaired *t*-test (**c**,**d**,**f**).

of *Gzmk*-tdTomato$^+$ cells produced IFNγ, the mean fluorescence intensity (MFI) of IFNγ in these *Gzmk*-tdTomato$^+$ cells was slightly lower than that in IFNγ-producing *Gzmk*-tdTomato$^-$ cells (Fig. 2g and Extended Data Fig. 3d). ScRNA-seq data from colitis mice showed that, unlike *Gzmk* and *Prf1*, which were predominantly expressed in cluster 1 (T_HK), *Gzmb* and *Ifng* were more broadly distributed across clusters, with the highest levels observed in cluster 5 (T_{reg}) and cluster 0 (T_H1) cells, respectively (Extended Data Fig. 3e).

To investigate the functional properties of T_HK cells, we performed in vitro cytotoxicity assays using sorted *Gzmk*-tdTomato$^+$ and *Gzmk*-tdTomato$^-$ CD4$^+$ T cells isolated from LP of colitic mice. These cells were co-cultured with the mastocytoma cell line P815. While the majority of *Gzmk*-tdTomato$^+$ cells retained tdTomato expression and *Gzmk*-tdTomato$^-$ cells remained tdTomato-negative after co-culture (Extended Data Fig. 3f), both populations exhibited comparable cytotoxic activity against target cells in the presence of anti-CD3

(Extended Data Fig. 3g), indicating a cytotoxic potential for, but not unique to, T_HK cells.

T_HK cell differentiation independent of canonical lineages

We next asked whether T_HK cell differentiation relies on the established T_H cell developmental pathways. To test this, we first assessed the role of T_H1 lineage commitment in T_HK cell generation by co-transferring wild-type (CD45.1$^+$) and *Tbx21*-deficient (CD45.2$^+$) naive CD4$^+$ T cells into *Rag1*$^{-/-}$ recipient mice. Four weeks later, we analyzed donor-derived T cells in LP and mLNs. Notably, although *Tbx21* deficiency led to a marked reduction in GZMB expression (Fig. 3a and Extended Data Fig. 4a), the frequencies of perforin-expressing CD4$^+$ T cells were significantly increased (Fig. 3b and Extended Data Fig. 4b). As currently there is no good staining antibody for murine GZMK, we sorted CD45.1$^+$ and CD45.2$^+$ CD4$^+$ T cells from LP for quantitative PCR with reverse transcription PCR (RT–qPCR) analysis, and found that *Gzmk* expression was preserved in *Tbx21*-deficient cells (Fig. 3c). This indicates that *Gzmk* induction occurs independently of the T_H1 program.

To determine whether T_HK cell development depends on T_H2 or T_H17 lineages, we examined *Gzmk* expression in *Stat6*$^{-/-}$ and *Cd4*^{cre} *Rorc*$^{fl/fl}$ CD4$^+$ T cells using a similar co-transfer system. Notably, *Gzmk* and perforin expression was largely unaffected in *Stat6*-deficient cells, indicating that STAT6, an obligatory transcription factor for T_H2 differentiation, is not required for T_HK cell differentiation (Fig. 3d,e and Extended Data Fig. 4c). In contrast, *Gzmk* and perforin expression was significantly elevated in *Rorc*-deficient CD4$^+$ T cells (Fig. 3f,g and Extended Data Fig. 4d), suggesting that RORγt may restrain T_HK cell differentiation.

In addition, *Bcl6* deficiency resulted in reduced *Gzmk* and *Prf1* expression in colonic CD4$^+$ T cells (Extended Data Fig. 4e). Given that CD4$^+$ T cells expressing *Bcl6* in lymphoid follicles can differentiate into multiple effector subsets in LP under inflammatory conditions while losing *Bcl6* expression[17,18], this finding raises the possibility that T_HK cells may also arise from such a progenitor population.

Previous studies reported that CD4$^+$ T cells can be reprogrammed into CD4$^+$CD8α$^+$ cytotoxic T lymphocytes, which upregulate *Runx3* and downregulate ThPOK (*Zbtb7b*)[19,20]. We examined *Runx3* and *Zbtb7b* expression in scRNA-seq data from colitis mice and found that neither gene was specifically or substantially altered in cluster 1 (T_HK) (Extended Data Fig. 5a,b). Similarly, CD8α expression was not elevated in *Gzmk*-tdTomato$^+$ CD4$^+$ T cells compared to other CD4$^+$ T cell populations (Extended Data Fig. 5c,d). Furthermore, we co-transferred wild-type and *Runx3*-deficient (*Vav1*^{cre} *Runx3*$^{fl/fl}$) naive CD4$^+$ T cells into *Rag1*$^{-/-}$ recipient mice. Although *Runx3* deficiency led to a marked reduction in CD8α and GZMB expression in CD4$^+$ T cells from LP of colitis mice (Extended Data Fig. 5e,f), *Gzmk* expression was not significantly altered (Extended Data Fig. 5g), supporting that T_HK cells are distinct from CD4$^+$CD8α$^+$ T cells.

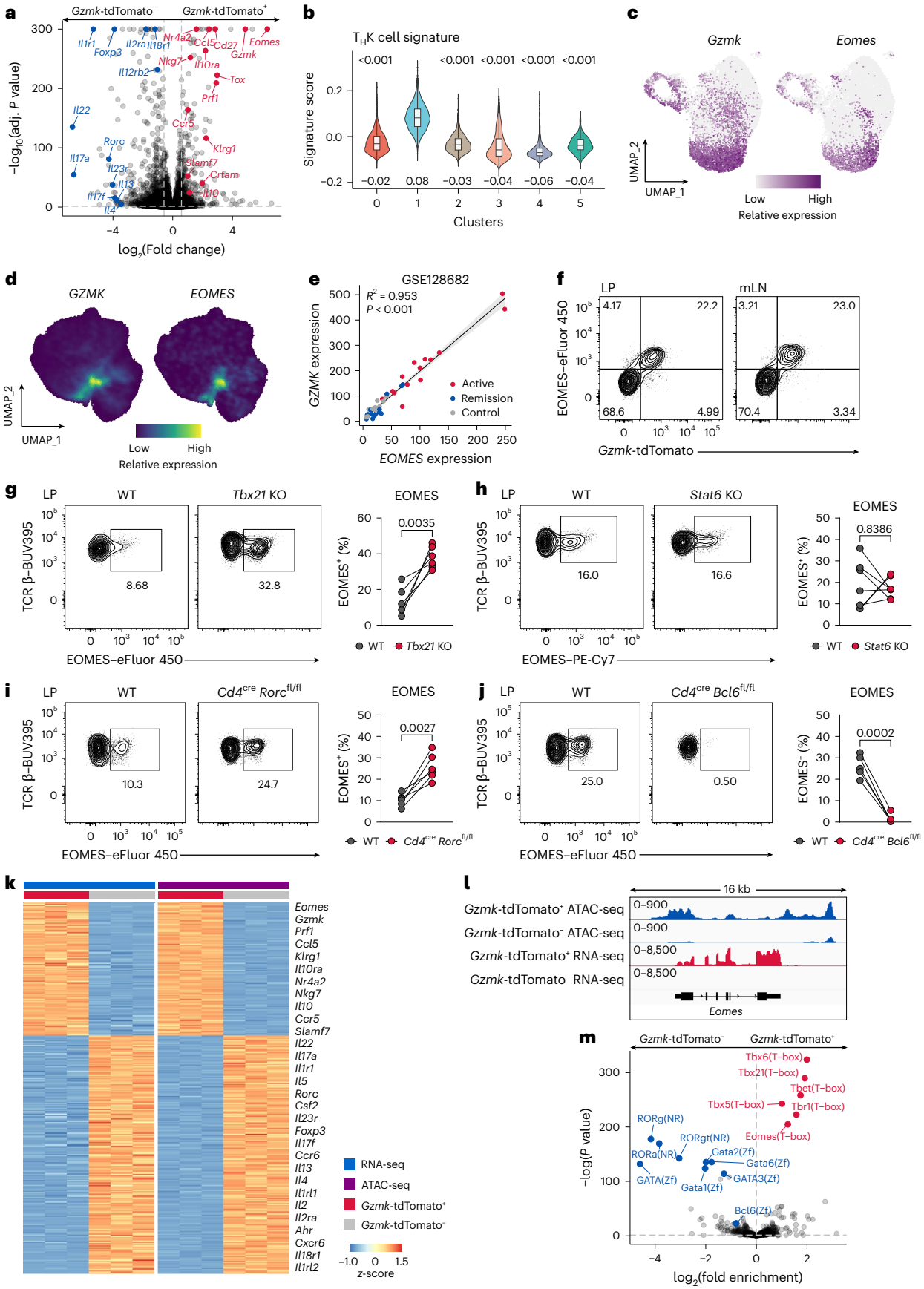
Eomes is transcriptionally and epigenetically associated with T_HK cells

Given that T_HK cell differentiation in vivo occurs independently of the canonical lineage-defining transcription factors of T_H1, T_H2 and T_H17 cells, we sought to identify the key transcriptional regulator associated with their unique cell identity. Bulk RNA-seq of sorted *Gzmk*-tdTomato$^+$ and *Gzmk*-tdTomato$^-$ CD4$^+$ T cells from LP of colitic mice revealed a distinct transcriptional signature in *Gzmk*-expressing cells (Fig. 4a). We defined the top 100 most differentially expressed genes as the 'T_HK cell signature' and applied it to score individual cells in the scRNA-seq data (Fig. 2a and Extended Data Fig. 6a). Cluster 1 exhibited the highest enrichment score, confirming its identity as the T_HK subset (Fig. 4b). Within this signature, *Eomes* emerged as one of the most significantly upregulated genes (Fig. 4a), implicating it as a key candidate regulator of T_HK cell identity.

This association was further supported by multimodal evidence. *Gzmk* and *Eomes* displayed nearly identical expression patterns across CD4$^+$ T cell clusters in mouse colitis scRNA-seq data (Fig. 4c and Extended Data Fig. 6b). A similar coexpression pattern was also observed in human CD4$^+$ T cell data (Fig. 4d and Extended Data Fig. 6c). In addition, correlation analysis of bulk RNA-seq data from patients with UC and healthy controls (GSE128682) further confirmed that *EOMES* displayed the strongest positive correlation with *GZMK* expression (Pearson $R^2 = 0.953$, $P < 0.001$) (Fig. 4e and Extended Data Fig. 6d). Consistently, flow cytometry further confirmed notable coexpression of *Gzmk*-tdTomato and EOMES proteins in CD4$^+$ T cells from LP and mLNs of colitic mice (Fig. 4f). Time-course analysis of LP and mLNs at days 7, 14, 21, 28 and 35 after naive CD4$^+$ T cell transfer showed that EOMES$^+$*Gzmk*-tdTomato$^+$ T_HK cells were nearly undetectable at day 7 post-transfer; however, this population became detectable by day 14 and reached a relatively stable frequency by approximately 3–4 weeks (Extended Data Fig. 6e). In contrast, T_H1 (T-bet$^+$) and T_H17 (RORγt$^+$) cells were already present at day 7 (Extended Data Fig. 6e). Consistent with the regulation of *Gzmk*, EOMES expression did not require T-bet, STAT6 or RORγt during colitis (Fig. 4g–i and Extended Data Fig. 6f–h); however, EOMES expression was also significantly reduced upon *Bcl6* deficiency (Fig. 4j and Extended Data Fig. 6i).

Fig. 4 | *Eomes* is transcriptionally and epigenetically associated with T_HK cell identity. a, Volcano plot of differentially expressed genes in sorted *Gzmk*-tdTomato$^+$ versus *Gzmk*-tdTomato$^-$ CD4$^+$ T cells from LP of colitic mice (bulk RNA-seq). Differential expression was assessed using DESeq2 (Wald test, two-sided). *P* values were adjusted using the Benjamini–Hochberg procedure. Red dots, upregulated in *Gzmk*-tdTomato$^+$; blue dots, upregulated in *Gzmk*-tdTomato$^-$. Horizontal dashed line, adjusted $P = 0.05$; vertical dashed lines, $\log_2$(fold change) = $\pm\log_2(1.5)$. **b**, Gene set enrichment score of the 'T_HK cell signature' across CD4$^+$ T cell clusters (cluster 0, $n = 5{,}932$; cluster 1, $n = 3{,}502$; cluster 2, $n = 2{,}703$; cluster 3, $n = 1{,}703$; cluster 4, $n = 1{,}610$; and cluster 5, $n = 133$). Violin plots with embedded box plots showing median (center line) and 25th–75th percentiles (box bounds). Numbers above show adjusted *P* values (Wilcoxon rank-sum test, two-tailed, cluster 1 versus all others); numbers below show mean scores. **c**, Feature plots displaying *Gzmk* and *Eomes* expression in the UMAP from mouse colitis scRNA-seq data (Fig. 2a). **d**, Density plots showing *GZMK* and *EOMES* expression in the human IBD dataset (Fig. 1a). **e**, Correlation between *GZMK* and *EOMES* mRNA expression in bulk RNA-seq data from patients with UC and healthy controls (GSE128682). Shown is R^2 from Pearson correlation; two-sided *P* values are unadjusted. Points are colored by group. Black line shows linear regression fit. Gray shaded band indicates 95% CI. **f**, Flow cytometry analysis of *Gzmk*-tdTomato and EOMES in LP and mesenteric lymph node (mLN) CD4$^+$ T cells from *Rag1*$^{-/-}$ recipients transferred with naive CD4$^+$ T cells from *Gzmk*-tdTomato reporter mice. **g**–**j**, Flow cytometry analysis of EOMES expression in LP CD4$^+$ T cells from *Rag1*$^{-/-}$ recipients co-transferred with WT (CD45.1$^+$) and KO (CD45.2$^+$) naive CD4$^+$ T cells. *Tbx21*$^{-/-}$ ($n = 6$) (**g**), *Stat6*$^{-/-}$ ($n = 7$) (**h**), *Cd4*^{cre} *Rorc*$^{fl/fl}$ ($n = 6$) (**i**) and *Cd4*^{cre} *Bcl6*$^{fl/fl}$ ($n = 5$) (**j**). Representative plots (left). Frequency of EOMES$^+$ cells within donor-derived populations (right). **k**, Integrated heatmap of co-directional gene–peak pairs (adjusted $P < 0.05$ for both RNA-seq and ATAC-seq). Rows represent linked gene–peak pairs ranked by gene $\log_2$(fold change). Columns show *z*-scored $\log_2$(TPM + 1) values for RNA-seq (left) and *z*-scored $\log_2$(normalized counts + 1) for ATAC-seq peaks within ±100 kb of the transcription start site (TSS). TPM, transcripts per million. **l**, IGV tracks showing chromatin accessibility (blue) and gene expression (red) at the *Eomes* locus in *Gzmk*-tdTomato$^+$ versus *Gzmk*-tdTomato$^-$ cells. Because *Eomes* expression in *Gzmk*-tdTomato$^-$ cells was very low compared to *Gzmk*-tdTomato$^+$ cells, its signal fell below the visible threshold when the *y* axis was uniformly set to 0–8,500. **m**, Volcano plot of motif enrichment analysis. Shown are −log(*P* value) (natural log) calculated by HOMER using binomial test. Dashed gray lines, $P = 0.05$ and fold enrichment = 1. Motifs in red, enriched in *Gzmk*-tdTomato$^+$; blue, enriched in *Gzmk*-tdTomato$^-$. Data are representative of two (**g**–**j**) or three (**f**) independent experiments. Exact *P* values are shown in each graph. Statistical significance was assessed by two-tailed paired *t*-test. Adj., adjusted.

https://doi.org/10.1038/s41590-026-02479-6



https://doi.org/10.1038/s41590-026-02479-6

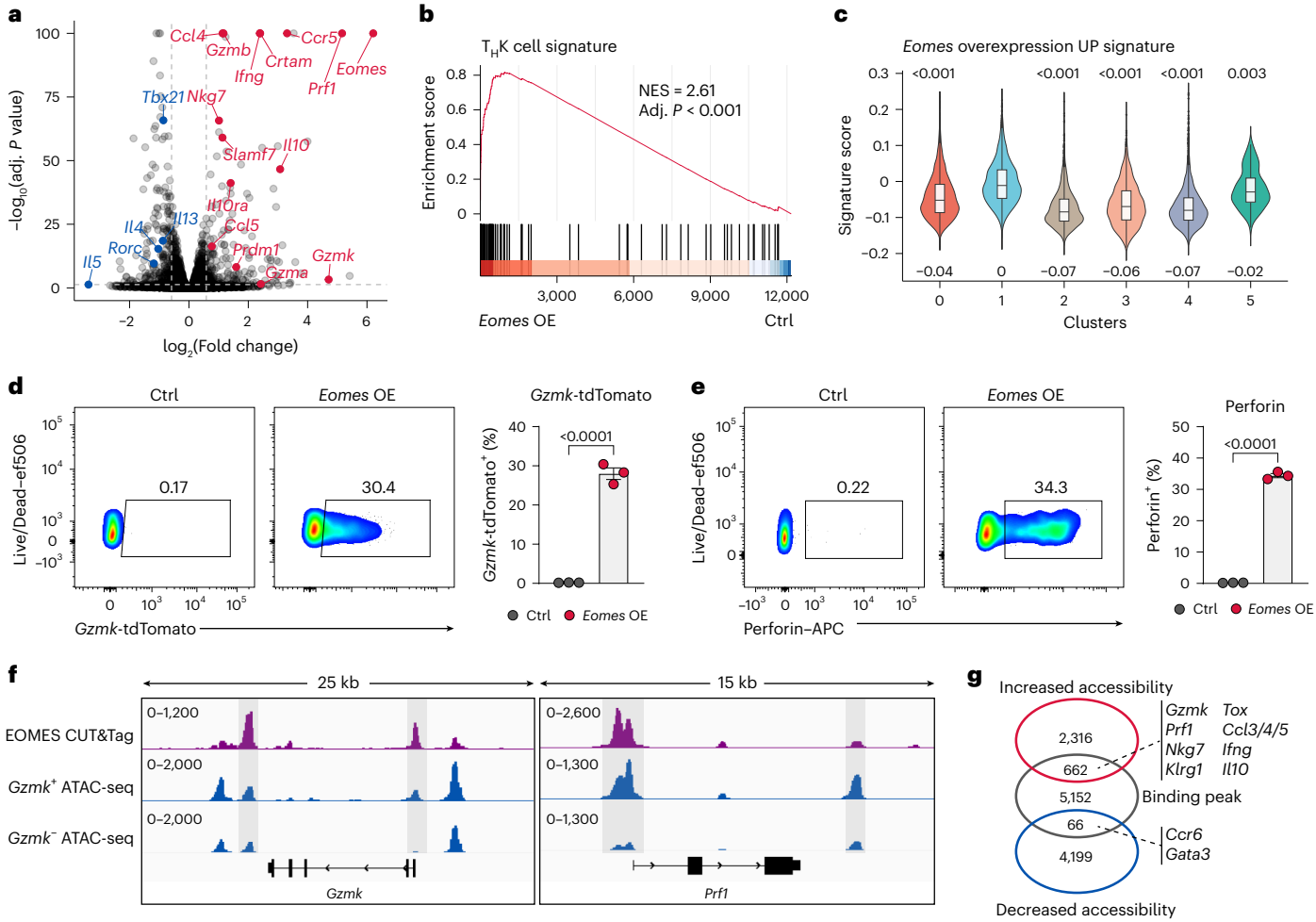


Fig. 5 | EOMES drives the T_HK program and constrains alternative T_H cell fates. a, Volcano plot of differentially expressed genes in CD4$^+$ T cells retrovirally transduced with *Eomes* versus control vector, activated under the nonpolarizing condition (bulk RNA-seq). Differential expression was assessed using DESeq2 (Wald test, two-sided). *P* values were adjusted using the Benjamini–Hochberg procedure. Red dots, upregulated after *Eomes* overexpression; blue dots, downregulated after *Eomes* overexpression. Horizontal dashed line, adjusted $P = 0.05$; vertical dashed lines, $\log_2(\text{fold change}) = \pm\log_2(1.5)$. **b**, GSEA of the '$T_H$K cell signature' in *Eomes*-overexpressing CD4$^+$ T cells compared to control. GSEA was performed using clusterProfiler (fgsea algorithm). *P* values were adjusted using the Benjamini–Hochberg procedure. Normalized enrichment score (NES) and adjusted *P* value are shown. **c**, Single-cell gene set enrichment score of the '*Eomes* overexpression UP' signature (top 100 upregulated genes) across CD4$^+$ T cell clusters (cluster 0, $n = 5{,}932$; cluster 1, $n = 3{,}502$; cluster 2, $n = 2{,}703$; cluster 3, $n = 1{,}703$; cluster 4, $n = 1{,}610$; and cluster 5, $n = 133$). Violin plots with embedded box plots showing median (center line) and 25th–75th percentiles (box bounds). Numbers above show adjusted *P* values (Wilcoxon rank-sum test, two-tailed, cluster 1 versus all others); numbers below show mean scores. **d**,**e**, Flow cytometry analysis of *Gzmk*-tdTomato (**d**) and perforin (**e**) in control or *Eomes*-overexpressing (OE) CD4$^+$ T cells activated under the nonpolarizing condition ($n = 3$). Representative plots (left). Frequency of *Gzmk*-tdTomato$^+$ and perforin$^+$ cells (right). **f**, IGV tracks showing EOMES CUT&Tag signal (purple), chromatin accessibility (blue) at the *Gzmk* and *Prf1* loci in *Gzmk*-tdTomato$^+$ versus *Gzmk*-tdTomato$^-$ CD4$^+$ T cells. **g**, Venn diagram showing overlap between EOMES-bound regions (CUT&Tag in *Eomes*-overexpressing CD4$^+$ T cells) and differentially accessible chromatin regions (ATAC-seq in *Gzmk*-tdTomato$^+$ versus *Gzmk*-tdTomato$^-$ CD4$^+$ T cells from colitic LP). Numbers indicate peak counts. Data are shown as mean ± s.e.m. and are representative of two independent experiments (**d**,**e**). Exact *P* values are shown in each graph. Statistical significance was assessed by two-tailed unpaired *t*-test (**d**,**e**).

To examine the epigenetic basis of this transcriptional program, we performed bulk ATAC-seq on the same sorted populations. Chromatin accessibility profiling revealed a distinct regulatory landscape in *Gzmk*-expressing cells, with increased chromatin accessibility at *cis*-regulatory elements near T_HK-associated genes, including *Eomes*, *Gzmk*, *Prf1*, *Cd27*, *Ccl5*, *Klrg1*, *Il10ra*, *Nr4a2*, *Nkg7*, *Il10*, *Ccr5* and *Slamf7* (Fig. 4k). Notably, the *Eomes* locus exhibited both elevated transcription and enhanced promoter/enhancer accessibility (Fig. 4l). In contrast, regulatory regions near genes of canonical CD4$^+$ T cell lineages, such as *Rorc*, *Il17a*, *Il17f*, *Il22*, *Csf2*, *Il23r*, *Il5*, *Il13*, *Il4* and *Il18r1*, were more accessible in *Gzmk*-tdTomato$^-$ cells (Fig. 4k). These findings demonstrate that *Gzmk*-expressing CD4$^+$ T cells possess a unique epigenetic program aligned with their distinct transcriptional identity.

Motif enrichment analysis of differentially accessible chromatin regions revealed that T-box family motifs, including the Eomes motif, were significantly enriched in regions with higher accessibility in *Gzmk*-tdTomato$^+$ cells (Fig. 4m). In contrast, regions with higher accessibility in *Gzmk*-tdTomato$^-$ cells were enriched for motifs of RORγ/RORα, Bcl6 and GATA family members, which are associated with T_H17, T_{FH} and T_H2 lineages, respectively (Fig. 4m).

EOMES drives T_HK fate and restricts alternative lineages

Having identified EOMES expression associated with T_HK cell identity, we next asked whether it is sufficient to induce the T_HK transcriptional program. We retrovirally overexpressed *Eomes* in CD4$^+$ T cells activated under the nonpolarizing condition and performed

bulk RNA-seq. Differential expression analysis revealed that *Gzmk* and *Prf1* were among the most significantly upregulated genes in *Eomes*-overexpressing cells (Fig. 5a). *Ifng* and *Gzmb* were also upregulated, albeit with lower fold changes (Fig. 5a). Notably, *Tbx21* expression was suppressed, suggesting that EOMES did not induce a canonical T_H1 program. In addition, signature genes of other helper T cell subsets, including *Rorc* (T_H17), and *Il5*, *Il4* and *Il13* (T_H2), were markedly downregulated (Fig. 5a). Gene set enrichment analysis (GSEA) confirmed strong enrichment of the 'T_HK cell signature' in *Eomes*-overexpressing cells (Fig. 5b). To assess this at single-cell resolution, we derived an '*Eomes* overexpression UP' gene signature, defined as the top 100 upregulated genes from RNA-seq of CD4$^+$ T cells with in vitro *Eomes* overexpression, and evaluated its enrichment across clusters in the mouse scRNA-seq data (Fig. 2a). Notably, cluster 1 (T_HK) exhibited the highest enrichment score (Fig. 5c), indicating that *Eomes* overexpression induces the T_HK transcriptional program.

We further validated these findings experimentally. RT–qPCR confirmed robust induction of *Gzmk* and *Prf1* in *Eomes*-overexpressing CD4$^+$ T cells under nonpolarizing conditions (Extended Data Fig. 7a), and flow cytometry showed strong upregulation of *Gzmk*-tdTomato and perforin compared to controls (Fig. 5d,e). Although *Gzmb* was also induced, its fold change was substantially lower than that of *Gzmk* (Extended Data Fig. 7a). Reanalysis of published RNA-seq data[21] from *Eomes*-overexpressing CD8$^+$ T cells also revealed a pronounced increase in *Gzmk* expression, with minimal impact on *Gzmb* (Extended Data Fig. 7b).

We next investigated whether EOMES influences the differentiation of canonical T_H cell subsets under polarizing conditions. First, we found that none of the established differentiation conditions for classical subsets (T_H1, T_H2, T_H17 and T_{reg} cells in vitro, and T_{FH} induced by ovalbumin (OVA) immunization in vivo) could effectively induce the expression of EOMES and *Gzmk*-tdTomato (Extended Data Fig. 7c). We further explored whether *Eomes* overexpression could suppress the differentiation of canonical CD4$^+$ T cells. Our results showed that while *Eomes* overexpression upregulated *Ifng* expression in the nonpolarizing condition (Fig. 5a), it markedly suppressed T-bet expression under T_H1-polarizing conditions (Extended Data Fig. 7d). In the T_H2 condition, *Eomes* overexpression modestly reduced GATA3 expression (Extended Data Fig. 7e). Previous studies have shown that *Eomes* strongly suppresses IL-5 production, while having minimal effects on IL-4 or IL-13 expression[22]. Under T_H17 and T_{reg} cell-polarizing conditions, *Eomes* overexpression markedly downregulated RORγt/IL-17A and FOXP3 expression, respectively (Extended Data Fig. 7f–h), consistent with previous reports[23,24].

Although *Bcl6* knockout reduced the expression of EOMES and *Gzmk* in CD4$^+$ T cells from colitis mice (Fig. 4j and Extended Data Figs. 4e and 6i), we further characterized the relationship between *Eomes* and T_{FH} cell differentiation. We overexpressed *Eomes* in OT-II cells, adoptively transferred them into wild-type recipients, and then immunized the mice with OVA. Seven days post-immunization, donor-derived *Eomes*-overexpressing cells exhibited significantly decreased frequencies of both PD-1$^+$CXCR5$^+$ and BCL6$^+$CXCR5$^+$ T_{FH} cells (Extended Data Fig. 7i,j), demonstrating a negative regulatory role of *Eomes* in T_{FH} cell differentiation. BLIMP1 is an antagonistic regulator of BCL6 and T_{FH} cell differentiation[25,26]. Our RNA-seq data revealed that *Eomes* overexpression upregulated *Prdm1* (encoding BLIMP1) expression (Fig. 5a). Furthermore, we found that the majority of EOMES$^+$ CD4$^+$ T cells from colitis mice highly expressed BLIMP1 (Extended Data Fig. 7k). Collectively, these findings indicate that T_HK cells are distinct from T_{FH} cells.

EOMES binds regulatory elements of T_HK genes

To determine whether EOMES directly regulates the T_HK transcriptional program, we performed CUT&Tag for HA-tagged EOMES overexpressed in CD4$^+$ T cells under nonpolarizing conditions. Genomic annotation of EOMES-occupied regions revealed that 37.38% in introns, 28.04% in distal intergenic regions, and 27.15% in promoters (Extended Data Fig. 8a), suggesting that EOMES acts upon both promoter-proximal and distal enhancer-like regulatory elements. Motif enrichment analysis of EOMES-bound regions showed significant enrichment of T-box motifs (including Eomes itself) and ETS-family motifs (for example, ETS1, ERG and Fli1) (Extended Data Fig. 8b). Notably, EOMES binding signals were robustly enriched at the *Gzmk* and *Prf1* loci (Fig. 5f). We also examined published ChIP-seq data[21] from CD8$^+$ T cells and found that EOMES similarly occupied the *Gzmk* and *Prf1* loci in CD8$^+$ T cells (Extended Data Fig. 8c).

To assess the functional relevance of EOMES binding in the context of endogenous T_HK cell differentiation, we integrated EOMES CUT&Tag data with ATAC-seq profiles from *Gzmk*-tdTomato$^+$ and *Gzmk*-tdTomato$^-$ CD4$^+$ T cells isolated from LP of colitic mice. Of the 5,880 EOMES binding peaks identified, 662 overlapped with regions of increased chromatin accessibility in *Gzmk*-tdTomato$^+$ cells, whereas only 66 overlapped with regions showing reduced accessibility, indicating that EOMES may preferentially act as the epigenetic activator during T_HK cell differentiation (Fig. 5g). Notably, EOMES peaks that colocalized with accessible chromatin in *Gzmk*-tdTomato$^+$ cells were located near T_HK-associated genes, including *Gzmk*, *Prf1*, *Nkg7*, *Klrg1*, *Tox*, *Ccl3/4/5*, *Ifng* and *Il10*. In contrast, EOMES binding at regions that were less accessible in *Gzmk*-tdTomato$^+$ cells occurred near genes associated with alternative lineages, such as *Ccr6* and *Gata3* (Fig. 5g). Notably, because EOMES CUT&Tag data came from an in vitro overexpression system while ATAC-seq profiles were from ex vivo-sorted LP CD4$^+$ T cells, their overlap is likely limited by differences in cellular contexts.

Fig. 6 | *Eomes* is required for T_HK cell differentiation and intestinal immunopathology. a, Volcano plot of differentially expressed genes in *Eomes* cKO (*Cd4*^{cre} *Eomes*$^{fl/fl}$) versus WT donor CD4$^+$ T cells isolated from colonic LP of colitic mice (bulk RNA-seq). Differential expression was assessed using DESeq2 (Wald test, two-sided). *P* values were adjusted using the Benjamini–Hochberg procedure. Red dots, upregulated in *Eomes* cKO; blue dots, downregulated. Horizontal dashed line, adjusted $P = 0.05$; vertical dashed lines, $\log_2$(fold change) = $\pm\log_2(1.5)$. **b**, GSEA of the 'T_HK cell signature' in *Eomes* cKO versus WT CD4$^+$ T cells. GSEA was performed using clusterProfiler (fgsea algorithm). *P* values were adjusted using the Benjamini–Hochberg procedure. NES and adjusted *P* values are shown. **c**, Single-cell gene set enrichment score of the '*Eomes* cKO DOWN' signature (top 50 downregulated genes) across CD4$^+$ T cell clusters (cluster 0, $n = 5{,}932$; cluster 1, $n = 3{,}502$; cluster 2, $n = 2{,}703$; cluster 3, $n = 1{,}703$; cluster 4, $n = 1{,}610$; and cluster 5, $n = 133$). Violin plots with embedded box plots showing median (center line) and 25th–75th percentiles (box bounds). Numbers above show adjusted *P* values (Wilcoxon rank-sum test, two-tailed, cluster 1 versus all others); below shows mean scores. **d**, Venn diagram integrating genes upregulated in *Gzmk*-tdTomato$^+$ cells, induced by *Eomes* overexpression, suppressed in *Eomes* cKO cells, and bound by EOMES (CUT&Tag). **e**, Venn diagram integrating genes upregulated in *Gzmk*-tdTomato$^-$ cells, suppressed by *Eomes* overexpression, induced in *Eomes* cKO cells, and bound by EOMES (CUT&Tag). **f**–**h**, Disease severity in *Rag1*$^{-/-}$ recipients of naive CD4$^+$ T cells from *Eomes* cKO ($n = 6$) or control (*Eomes*$^{fl/fl}$) ($n = 6$) donors. Body weight change (**f**); colon length at end point (**g**); histological inflammation score; scale bars, 500 µm (**h**). **i**, Total number of donor-derived CD4$^+$ T cells in colonic LP at experimental end point. **j**–**l**, Flow cytometry analysis of effector molecules in donor-derived CD4$^+$ T cells from colonic LP ($n = 6$). Representative plots showing EOMES and *Gzmk*-tdTomato expression, with quantification of the frequency of *Gzmk*-tdTomato$^+$ cells (**j**). Representative plots (left) and frequencies (right) of perforin$^+$ (**k**) and GZMB$^+$ (**l**) cells. Data are shown as mean ± s.e.m. and are representative of three independent experiments. Exact *P* values are shown in each graph. Statistical significance was assessed by two-way ANOVA with Bonferroni's post hoc test (**f**) and two-tailed unpaired *t*-test (**g**–**l**).

https://doi.org/10.1038/s41590-026-02479-6

Eomes is required for T_HK cell differentiation in colitis

To determine whether *Eomes* is necessary for T_HK cell development in vivo, we isolated naive CD4$^+$ T cells from *Cd4*^{cre} *Eomes*$^{fl/fl}$ mice (*Eomes* cKO) and co-transferred them with CD45.1$^+$ wild-type control cells into *Rag1*$^{-/-}$ recipients, enabling assessment of the cell-intrinsic requirement for *Eomes*. Donor-derived wild-type and *Eomes* cKO CD4$^+$ T cells were sorted from the colonic LP of colitic mice for bulk RNA-seq. Transcriptomic analysis revealed that *Eomes* cKO cells exhibited significantly reduced expression of T_HK-associated genes, including *Gzmk*, *Prf1*, *Nkg7* and *Slamf7* (Fig. 6a). GSEA confirmed strong downregulation of the 'T_HK cell signature' in the *Eomes* cKO population (Fig. 6b). To define the *Eomes*-dependent transcriptional program, we generated an '*Eomes* cKO DOWN' gene signature comprising the top 50 genes downregulated in *Eomes* cKO versus wild-type CD4$^+$ T cells, and evaluated its enrichment across clusters in the scRNA-seq data of colonic CD4$^+$ T cells from colitic mice (Fig. 2a). This gene signature showed the highest enrichment in cluster 1 (T_HK) (Fig. 6c), suggesting that T_HK cells

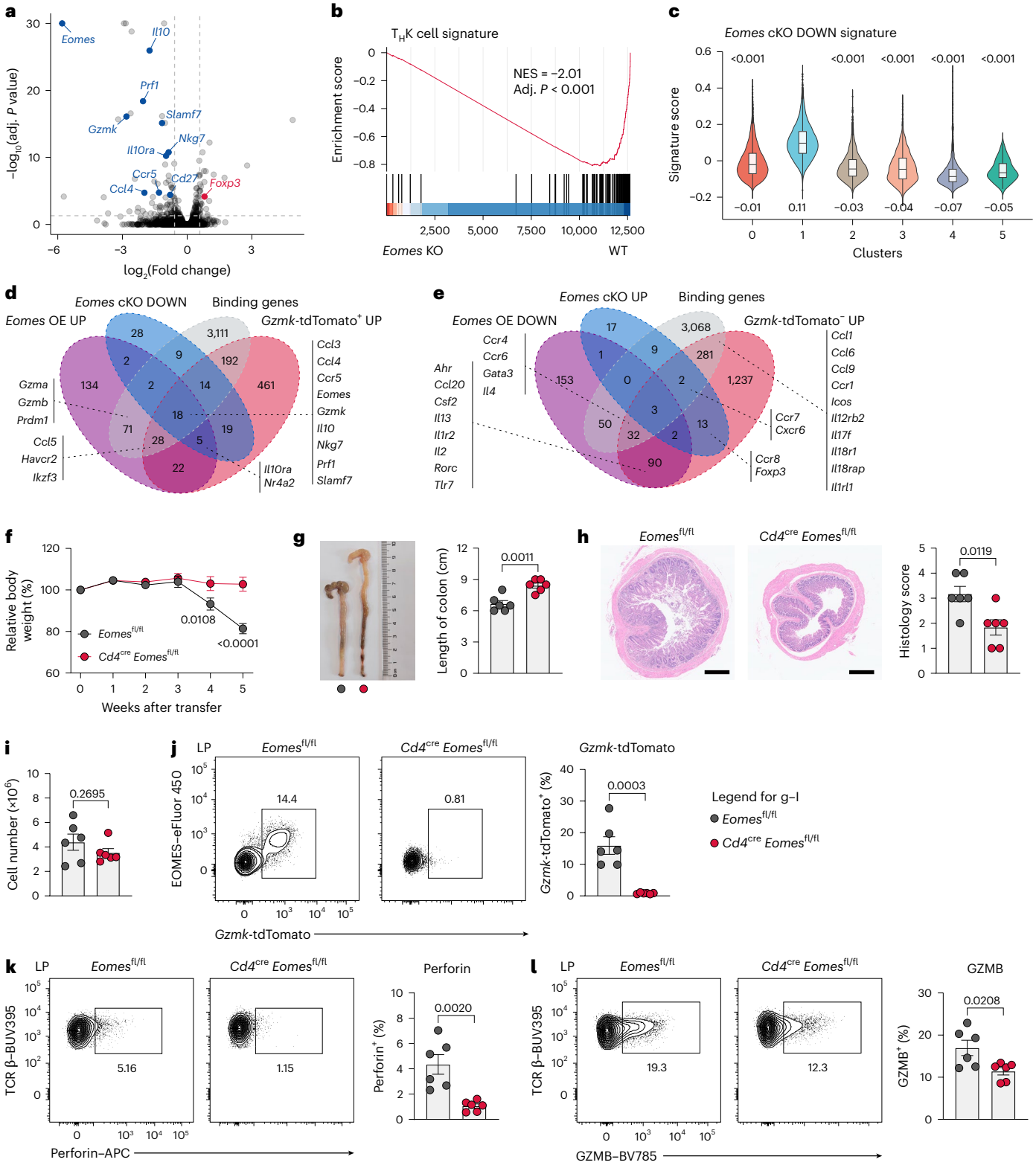


exhibit the strongest transcriptional dependence on *Eomes* among all CD4$^+$ subsets analyzed.

To identify high-confidence direct target genes of EOMES, we integrated differentially expressed genes from three RNA-seq datasets: (1) *Gzmk*-tdTomato$^+$ versus *Gzmk*-tdTomato$^-$ CD4$^+$ T cells; (2) *Eomes*-overexpressing versus control CD4$^+$ T cells; and (3) *Eomes* cKO versus wild-type CD4$^+$ T cells, together with EOMES binding genes identified in CUT&Tag. This analysis identified 18 genes consistently upregulated in *Gzmk*-tdTomato$^+$ CD4$^+$ T cells, induced by *Eomes* overexpression, suppressed in *Eomes* cKO cells, and bound by EOMES. These include *Ccl3*, *Ccl4*, *Ccr5*, *Eomes*, *Gzmk*, *Il10*, *Nkg7* and *Prf1* (Fig. 6d), forming a core set of direct EOMES target genes in T$_H$K cells. In contrast, integration of genes downregulated upon *Eomes* overexpression, upregulated in *Gzmk*-tdTomato$^-$ cells, upregulated in *Eomes* cKO cells, and EOMES binding genes yielded only three overlapping genes (Fig. 6e). Notably, several differentially expressed genes associated with alternative lineages, including *Rorc*, *Ahr*, *Il13*, *Csf2* and *Foxp3*, lacked direct EOMES binding (Fig. 6e), suggesting they are indirectly regulated.

Eomes-dependent T$_H$K program drives intestinal inflammation

To evaluate the functional relevance of T$_H$K cells in disease pathogenesis, we assessed the impact of *Eomes* ablation in CD4$^+$ T cells on the development of intestinal inflammation. Naive CD4$^+$ T cells from *Eomes* cKO mice or control *Eomes*$^{fl/fl}$ mice, which were crossed with *Gzmk*-tdTomato reporter mice, were transferred into *Rag1*$^{-/-}$ recipients separately. Mice receiving *Eomes* cKO CD4$^+$ T cells exhibited significantly attenuated colitis compared to those receiving control cells, as evidenced by reduced weight loss, less severe colonic shortening and lower histological inflammation scores (Fig. 6f–h). At the experimental end point, the total numbers of donor *Eomes* cKO and control CD4$^+$ T cells in LP were comparable (Fig. 6i), indicating that *Eomes* deficiency did not impair T cell expansion or tissue homing under these conditions.

Flow cytometric analysis revealed that *Eomes* deficiency nearly abolished *Gzmk*-tdTomato expression (Fig. 6j and Extended Data Fig. 9a) and markedly reduced perforin levels in donor CD4$^+$ T cells (Fig. 6k and Extended Data Fig. 9b). GZMB expression was modestly decreased but still present (Fig. 6l and Extended Data Fig. 9c), and T-bet expression remained largely unchanged (Extended Data Fig. 9d,e), with a slight reduction in IFNγ production observed in mLNs (Extended Data Fig. 9f,g). Notably, the frequencies of RORγt$^+$ and IL-17A$^+$ CD4$^+$ T cells were increased in the colonic LP and mLNs, and FOXP3$^+$ CD4$^+$ T cells were elevated in LP (Extended Data Fig. 9h–m) of mice receiving *Eomes*-deficient cells, suggesting a fate shift toward alternative lineages in the absence of T$_H$K differentiation. *Eomes* deficiency did not affect the frequencies of CD8α$^+$ cells among CD4$^+$ T cells in LP (Extended Data Fig. 9n), further indicating that T$_H$K cells are distinct from CD4$^+$CD8α$^+$ T cells.

To determine whether GZMK mediates the pathogenic function of T$_H$K cells, we performed adoptive transfer experiments using *Gzmk*$^{-/-}$ CD4$^+$ T cells. In contrast to *Eomes* cKO, *Gzmk* deletion did not ameliorate the disease, as recipients exhibited comparable weight loss and colon shortening (Extended Data Fig. 9o,p). These results indicate that T$_H$K-mediated immunopathology may depend on a coordinated, potentially redundant effector program downstream of EOMES.

The T$_H$K program is conserved across species and diseases

To assess the conservation of the T$_H$K program in humans, we performed differential gene expression analysis comparing *GZMK*^{high} CD4$^+$ T cell subsets to other CD4$^+$ subsets in published scRNA-seq datasets of IBD (Extended Data Fig. 10a; data from Fig. 1a) and pancancer (Extended Data Fig. 10b; data from Extended Data Fig. 1). In addition to *GZMK* and *EOMES*, many of the most significantly upregulated genes in human *GZMK*^{high} cells correspond to T$_H$K signature genes defined in our mouse colitis model (Extended Data Fig. 10a,b). We next converted the genes significantly upregulated in *Gzmk*-tdTomato$^+$ CD4$^+$ T cells from murine colitis (bulk RNA-seq, Fig. 4a) into their human orthologs, and overlapped them with the upregulated genes in human *GZMK*^{high} subsets (Extended Data Fig. 10c). Notably, the overlapping genes included core T$_H$K-associated factors, such as *EOMES*, *GZMK*, *CCR5*, *PRF1*, *NKG7* and *SLAMF7*, which were consistently and significantly upregulated in each dataset (Extended Data Fig. 10a–c). These findings suggest that the T$_H$K transcriptional program is evolutionarily conserved between mice and humans.

We further investigated whether T$_H$K cells also arise beyond the colitis model. We analyzed published scRNA-seq datasets of CD4$^+$ T cells from multiple murine disease contexts: Hepa1-6 tumor-infiltrating conventional CD4$^+$ T cells (GSE285225; Extended Data Fig. 10d)[27], CD4$^+$ T cells in the central nervous system of mice with experimental autoimmune encephalomyelitis (EAE) (GSE156196; Extended Data Fig. 10e) and splenic CD4$^+$ T cells from mice chronically infected with lymphocytic choriomeningitis virus clone 13 (LCMV cl13) (GSE201730; Extended Data Fig. 10f)[28]. In all these settings, we consistently identified a distinct *Gzmk*^{high} CD4$^+$ T cell subset that co-expressed high levels of *Eomes* (Extended Data Fig. 10d–f). Moreover, this *Gzmk*^{high} subset exhibited significantly higher enrichment scores for the 'T$_H$K cell signature' compared to other CD4$^+$ subsets (Extended Data Fig. 10g–i).

Although technical variations and differences in non-*Gzmk*^{high} cell composition across datasets could affect differential expression results, we found that many T$_H$K signature genes defined in our mouse colitis model were significantly upregulated in *Gzmk*^{high} populations (Extended Data Fig. 10j–l). Intersecting the upregulated genes from all four disease models (colitis, tumor, EAE and LCMV cl13), we found that *Eomes*, *Gzmk* and *Prf1* were consistently and significantly upregulated across all datasets, whereas additional T$_H$K-associated genes (for example, *Ccl5*, *Ccr5*, *Nkg7*, *Il10ra* and *Slamf7*) were consistently upregulated in some of these models (Extended Data Fig. 10m).

Collectively, these results indicate that T$_H$K cells are broadly present across diverse disease settings, including inflammatory (colitis), autoimmune (EAE), tumoral and chronic infectious (LCMV cl13) contexts, and that their core transcriptional program is conserved; however, they may also exhibit context-dependent transcriptional adaptations, reflecting their functional plasticity in different microenvironments.

Discussion

Effector T$_H$ cells are traditionally defined by lineage-specific transcription factors and cytokines. Here, we identify a distinct CD4$^+$ T cell subset, T$_H$K cells, characterized by coexpression of EOMES and GZMK, with a unique transcriptomic profile and a critical role in intestinal inflammation. T$_H$K cells are transcriptionally and developmentally distinct from classical T$_H$ lineages, differentiating independently of T-bet, RORγt or STAT6. EOMES is both necessary and sufficient for establishing T$_H$K identity, directly binding to and activating key genes such as *Gzmk* and *Prf1*. Reanalysis of published RNA-seq and ChIP-seq data[21] further reveals that EOMES binds the *Gzmk* locus and drives its expression in CD8$^+$ T cells, indicating that the EOMES–GZMK axis is a conserved regulatory module across T cell lineages.

Although T$_H$K cells can express IFNγ, they are not T$_H$1 cells, as they express low levels of T-bet and T$_H$1 signature genes such as *Il12rb2* and *Il18r1* (Fig. 4a). *Gzmk* and *Eomes* are expressed independently of T-bet, and *Eomes* overexpression suppresses T-bet. Nevertheless, T-bet may contribute at certain stages to regulating other T$_H$K effectors, particularly IFNγ and GZMB. While some EOMES$^+$ CD4$^+$ T cells express IL-10 and exhibit type 1 regulatory T (Tr1)-like properties[29–33], Tr1 development depends on T-bet[29,34], underscoring a distinct T$_H$K differentiation pathway. As T-box family members, EOMES and T-bet play nonredundant roles in defining T$_H$K versus T$_H$1 identities but may

share partially redundant functions in shaping CD4$^+$ T cell effector programs, warranting further investigation.

Our results showed that *Bcl6* deletion reduced *Gzmk* and EOMES expression. Previous studies indicate that *Bcl6*-expressing CD4$^+$ T cells act as precursors in transfer colitis, giving rise to T_H1, T_H17 and CD4$^+$CD8α$^+$ effectors in the inflamed LP while losing *Bcl6* expression[17,18]. These effector populations are also reduced upon *Bcl6* deletion[17,18]. Therefore, T_HK cells likely represent an additional effector fate derived from these precursors.

A key question is whether T_HK cells are cytotoxic CD4$^+$ T cells. Although previous studies describe cytotoxic CD4$^+$ subsets, often defined by GZMB or cytotoxic surface markers[35–37], they are typically EOMES-negative or only partially positive[36,37], distinguishing them from the EOMES$^+$ T_HK cells. Our results further confirm that T_HK cells are distinct from CD4$^+$CD8α$^+$ cytotoxic T cells. While T_HK cells express *Gzmb* and *Prf1* and exhibit ex vivo cytotoxicity, non-T_HK CD4$^+$ T cells also display cytotoxic activity, indicating that cytotoxicity is not unique to T_HK cells. Moreover, GZMK's function seems to be amplifying inflammation via extracellular proteolysis, such as cleaving PAR-1 and complement components[4,5,8], rather than direct cytotoxicity, which remains less validated. Thus, the 'K' in T_HK denotes both specific GZMK expression and a context-dependent potential for killing.

Functionally, the EOMES-driven T_HK program is critical for intestinal immunopathology. Although *Eomes* has been linked to inflammation[30,38], most studies lack CD4$^+$ T cell-intrinsic genetic evidence. To address this, we transferred sorted naive CD4$^+$ T cells, restricting deletion to conventional CD4$^+$ T cells; however, while loss of the entire T_HK program by *Eomes* deletion ameliorated disease, *Gzmk* deletion alone did not provide significant protection. This suggests that targeting EOMES or its core transcriptional network may be more effective therapeutically than inhibiting individual downstream effectors, as EOMES regulates multiple effectors including perforin and GZMB; however, because *Eomes* deficiency may indirectly alter the differentiation or function of other CD4$^+$ T cell subsets, the observed protection cannot be solely attributed to T_HK loss, highlighting the need for direct validation of T_HK-intrinsic pathogenicity.

Notably, T_HK cells are not restricted to colitis. Our analysis of published scRNA-seq datasets revealed a conserved T_HK transcriptional signature in human IBD and pancancer contexts, as well as in murine models of tumor, EAE and LCMV cl13 infection. GZMK$^+$EOMES$^+$ CD4$^+$ T cells have also been reported in other diseases, such as IgG4-related disease[10]. These findings indicate that T_HK cells are evolutionarily conserved and broadly deployed across pathological contexts, warranting further investigation into their functional roles.

T_HK cell differentiation mechanisms warrant further investigation. The upstream signals that initiate T_HK differentiation, including cytokines, chronic antigen exposure and tissue-derived cues, remain to be defined. Although previous studies have reported that IL-10 and IL-27 can be required for optimal EOMES expression in CD4$^+$ T cells under specific conditions[29,32,39], whether and how they directly drive T_HK differentiation remains unclear.

Although EOMES$^+$ CD4$^+$ T cells are often associated with proinflammatory or effector functions in diverse settings[24,32,38–48], some studies suggest they can also exhibit immunosuppressive features, as indicated by IL-10 expression[29–33]. This functional duality implies that, like T_H17 cells[49], T_HK cells may also adopt either inflammatory or regulatory roles depending on the microenvironment. Whether such heterogeneity exists within T_HK cells and how it is regulated remain key questions for future study.

In summary, we have identified T_HK cells as a distinct CD4$^+$ T cell subset defined by a core EOMES–GZMK module. This population exhibits unique transcriptional and epigenetic characteristics, and plays a critical role in driving intestinal inflammation. T_HK cells are conserved across species and broadly present in diverse disease contexts. Our work expands the functional taxonomy of T_H cells and highlights the EOMES–T_HK axis as a compelling therapeutic target in chronic inflammatory diseases.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-026-02479-6.

References

1. Dong, C. Cytokine regulation and function in T cells. *Annu. Rev. Immunol.* **39**, 51–76 (2021).
2. Voskoboinik, I., Whisstock, J. C. & Trapani, J. A. Perforin and granzymes: function, dysfunction and human pathology. *Nat. Rev. Immunol.* **15**, 388–400 (2015).
3. Aubert, A., Jung, K., Hiroyasu, S., Pardo, J. & Granville, D. J. Granzyme serine proteases in inflammation and rheumatic diseases. *Nat. Rev. Rheumatol.* **20**, 361–376 (2024).
4. Donado, C. A. et al. Granzyme K activates the entire complement cascade. *Nature* **641**, 211–221 (2025).
5. Lan, F. et al. GZMK-expressing CD8(+) T cells promote recurrent airway inflammatory diseases. *Nature* **638**, 490–498 (2025).
6. Guo, C. L. et al. Granzyme K(+)CD8(+) T cells interact with fibroblasts to promote neutrophilic inflammation in nasal polyps. *Nat. Commun.* **15**, 10413 (2024).
7. Guo, X. et al. Contrasting cytotoxic and regulatory T cell responses underlying distinct clinical outcomes to anti-PD-1 plus lenvatinib therapy in cancer. *Cancer Cell* **43**, 248–268.e9 (2025).
8. Guo, C. L., Wang, C. S., Wang, X. H., Yu, D. & Liu, Z. GZMK(+)CD8(+) T cells: multifaceted roles beyond cytotoxicity. *Trends Immunol.* **46**, 562–572 (2025).
9. Kwek, S. S. et al. Identification and regulation of circulating tumor-TCR-matched cytotoxic CD4$^+$ lymphocytes by KLRG1 in bladder cancer. *JCI Insight* **10**, e177373 (2025).
10. Koga, R. et al. Granzyme K- and amphiregulin-expressing cytotoxic T cells and activated extrafollicular B cells are potential drivers of IgG4-related disease. *J. Allergy Clin. Immunol.* **153**, 1095–1112 (2024).
11. Aoyagi, R. et al. Single-cell transcriptomics reveals granzyme K-expressing cytotoxic Tfh cells in tertiary lymphoid structures in IgG4-RD. *J. Allergy Clin. Immunol.* **153**, 513–520.e10 (2024).
12. Elyahu, Y. et al. Aging promotes reorganization of the CD4 T cell landscape toward extreme regulatory and effector phenotypes. *Sci. Adv.* **5**, eaaw8330 (2019).
13. Magen, A. et al. Intratumoral dendritic cell-CD4(+) T helper cell niches enable CD8(+) T cell differentiation following PD-1 blockade in hepatocellular carcinoma. *Nat. Med.* **29**, 1389–1399 (2023).
14. Nie, H. et al. Single-cell meta-analysis of inflammatory bowel disease with scIBD. *Nat. Comput. Sci.* **3**, 522–531 (2023).
15. Fenton, C. G., Taman, H., Florholmen, J., Sorbye, S. W. & Paulssen, R. H. Transcriptional signatures that define ulcerative colitis in remission. *Inflamm. Bowel Dis.* **27**, 94–105 (2021).
16. Zheng, L. et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* **374**, abe6474 (2021).
17. Bai, X. et al. Reciprocal regulation of T follicular helper cells and dendritic cells drives colitis development. *Nat. Immunol.* **25**, 1383–1394 (2024).
18. Li, Y. et al. Stem-like T cells are associated with the pathogenesis of ulcerative colitis in humans. *Nat. Immunol.* **25**, 1231–1244 (2024).

19. Reis, B. S., Rogoz, A., Costa-Pinto, F. A., Taniuchi, I. & Mucida, D. Mutual expression of the transcription factors Runx3 and ThPOK regulates intestinal CD4$^+$ T cell immunity. *Nat. Immunol.* **14**, 271–280 (2013).
20. Mucida, D. et al. Transcriptional reprogramming of mature CD4(+) helper T cells generates distinct MHC class II-restricted cytotoxic T lymphocytes. *Nat. Immunol.* **14**, 281–289 (2013).
21. Li, J., He, Y., Hao, J., Ni, L. & Dong, C. High levels of eomes promote exhaustion of anti-tumor CD8(+) T cells. *Front. Immunol.* **9**, 2981 (2018).
22. Endo, Y. et al. Eomesodermin controls interleukin-5 production in memory T helper 2 cells through inhibition of activity of the transcription factor GATA3. *Immunity* **35**, 733–745 (2011).
23. Ichiyama, K. et al. Transcription factor Smad-independent T helper 17 cell induction by transforming-growth factor-β is mediated by suppression of Eomesodermin. *Immunity* **34**, 741–754 (2011).
24. Lupar, E. et al. Eomesodermin expression in CD4$^+$ T cells restricts peripheral Foxp3 induction. *J. Immunol.* **195**, 4742–4752 (2015).
25. Johnston, R. J. et al. Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. *Science* **325**, 1006–1010 (2009).
26. Sun, Q. et al. BCL6 promotes a stem-like CD8(+) T cell program in cancer via antagonizing BLIMP1. *Sci. Immunol.* **8**, eadh1306 (2023).
27. Tan, S. N. et al. Regulatory T cells converted from Th1 cells in tumors suppress cancer immunity via CD39. *J. Exp. Med.* **222**, e20240445 (2025).
28. Zander, R., Khatun, A., Kasmani, M. Y., Chen, Y. & Cui, W. Delineating the transcriptional landscape and clonal diversity of virus-specific CD4(+) T cells during chronic viral infection. *eLife* **11**, e80079 (2022).
29. Zhang, P. et al. Eomesodermin promotes the development of type 1 regulatory T (T(R)1) cells. *Sci. Immunol.* **2**, eaah7152 (2017).
30. Gruarin, P. et al. Eomesodermin controls a unique differentiation program in human IL-10 and IFN-γ coproducing regulatory T cells. *Eur. J. Immunol.* **49**, 96–111 (2019).
31. Bonnal, R. J. P. et al. Clonally expanded EOMES(+) Tr1-like cells in primary and metastatic tumors are associated with disease progression. *Nat. Immunol.* **22**, 735–745 (2021).
32. Zhang, P. et al. Eomesodermin$^+$ CD4$^+$ T cells are critical for curative immunotherapy outcomes. *Immunity* **58**, 3024–3039.e7 (2025).
33. Thelen, B. et al. Eomes is sufficient to regulate IL-10 expression and cytotoxic effector molecules in murine CD4(+) T cells. *Front. Immunol.* **14**, 1058267 (2023).
34. Ansaldo, E. et al. T-bet-expressing Tr1 cells driven by dietary signals dominate the small intestinal immune landscape. *Proc. Natl Acad. Sci. USA* **123**, e2520747122 (2026).
35. Devarajan, P. et al. Cytotoxic CD4 development requires CD4 effectors to concurrently recognize local antigen and encounter type I IFN-induced IL-15. *Cell Rep.* **42**, 113182 (2023).
36. Marshall, N. B. et al. NKG2C/E marks the unique cytotoxic CD4 T cell subset, ThCTL, generated by influenza infection. *J. Immunol.* **198**, 1142–1155 (2017).
37. Takeuchi, A. et al. CRTAM determines the CD4$^+$ cytotoxic T lymphocyte lineage. *J. Exp. Med.* **213**, 123–138 (2016).
38. Mazzoni, A. et al. Eomes controls the development of Th17-derived (non-classic) Th1 cells during chronic inflammation. *Eur. J. Immunol.* **49**, 79–95 (2019).
39. Curran, M. A. et al. Systemic 4-1BB activation induces a novel T cell phenotype driven by high expression of Eomesodermin. *J. Exp. Med.* **210**, 743–755 (2013).
40. Raveney, B. J. et al. Eomesodermin-expressing T-helper cells are essential for chronic neuroinflammation. *Nat. Commun.* **6**, 8437 (2015).
41. Stienne, C. et al. Foxo3 transcription factor drives pathogenic T helper 1 differentiation by inducing the expression of Eomes. *Immunity* **45**, 774–787 (2016).
42. Zhang, C. et al. Extrapituitary prolactin promotes generation of Eomes-positive helper T cells mediating neuroinflammation. *Proc. Natl Acad. Sci. USA* **116**, 21131–21139 (2019).
43. Raveney, B. J. E. et al. Involvement of cytotoxic Eomes-expressing CD4(+) T cells in secondary progressive multiple sclerosis. *Proc. Natl Acad. Sci. USA* **118**, e2021818118 (2021).
44. Chen, S. et al. Eomesodermin expression in CD4(+) T-cells associated with disease progression in amyotrophic lateral sclerosis. *CNS Neurosci. Ther.* **30**, e14503 (2024).
45. Joulia, E. et al. Eomes-dependent mitochondrial regulation promotes survival of pathogenic CD4$^+$ T cells during inflammation. *J. Exp. Med.* **221**, e20230449 (2024).
46. Kanazawa, T. et al. Pathogenic potential of Eomesodermin-expressing T-helper cells in neurodegenerative diseases. *Ann. Neurol.* **95**, 1093–1098 (2024).
47. Elyahu, Y. et al. CD4 T cells acquire Eomesodermin to modulate cellular senescence and aging. *Nat. Aging* **5**, 1970–1982 (2025).
48. Roessner, P. M. et al. EOMES and IL-10 regulate antitumor activity of T regulatory type 1 CD4(+) T cells in chronic lymphocytic leukemia. *Leukemia* **35**, 2311–2324 (2021).
49. Schnell, A., Littman, D. R. & Kuchroo, V. K. T(H)17 cell heterogeneity and its role in tissue inflammation. *Nat. Immunol.* **24**, 19–29 (2023).

https://doi.org/10.1038/s41590-026-02479-6

Methods

Mice

C57BL/6J (cat. no. 000664), CD45.1 (cat. no. 002014), *Rag1*$^{-/-}$ (cat. no. 002216), *Tbx21*$^{-/-}$ mice (cat. no. 004648), *Stat6*$^{-/-}$ (cat. no. 005977), *Cd4*^{cre} (cat. no. 022071), *Vav1*^{cre} (cat. no. 008610), BLIMP1-EYFP (cat. no. 008828) and $OVA_{323-339}$-specific T cell receptor transgenic OT-II (cat. no. 004194) mice were originally from The Jackson Laboratory. *Gzmk*$^{-/-}$ mice (cat. no. T029683) were originally from GemPharmatech. *Runx3*$^{fl/fl}$ mice were provided by Y. Sun (Beijing Institute of Genomics, Chinese Academy of Sciences)[50] and crossed with *Vav1*^{cre} mice to generate *Vav1*^{cre} *Runx3*$^{fl/fl}$ mice. *Gzmk-2A-CreERT2-2A-tdTomato* mice (referred to as *Gzmk*-tdTomato, cat. no. NM-KI-241127) were purchased from Shanghai Model Organisms Center. *Eomes*$^{fl/fl}$, *Rorc*$^{fl/fl}$ and *Bcl6*$^{fl/fl}$ mice, as previously described[21,51,52], were crossed with *Cd4*^{cre} mice to generate T cell-specific knockout lines. All mice were housed in specific-pathogen-free conditions at ~22 °C with humidity of 40–70% and a light–dark cycle of 12 h and were used and maintained under the guidelines of the Institutional Animal Care and Use Committee of Westlake University.

T cell-induced colitis

$CD4^+$ T cells were isolated from the spleens and lymph nodes of donor mice. $CD4^+CD25^-CD44^-CD62L^+$ naive T cells were subsequently purified by fluorescence-activated cell sorting on the FACSAria sorter, achieving a post-sort purity of >98%. A total of 1.5–2 × 10^6 sorted naive T cells were intravenously injected into 8–12-week-old male *Rag1*$^{-/-}$ recipient mice to induce colitis. Body weight was monitored weekly. At ~4 weeks or other designated time points post-transfer, mice were killed for sample collection; colon tissues and mLNs were collected for downstream analyses.

Isolation of colon lymphocytes

Colon tissues were collected, opened longitudinally and cleared of fecal content. They were then incubated in RPMI 1640 medium containing 5 mM EDTA (Invitrogen, cat. no. 15575020) and 1 mM dithiothreitol (DTT; Sigma-Aldrich, cat. no. D9779) at 37 °C for 30 min with constant shaking to remove epithelial cells. After two washes with RPMI 1640 containing 2 mM EDTA, the tissues were minced into small pieces (~2 mm) and digested enzymatically in RPMI 1640 containing 0.5 mg ml^{-1} collagenase D (Roche, cat. no. 11088858001), 1 mg ml^{-1} dispase (Invitrogen, cat. no. 17105041) and 4 μg ml^{-1} DNase I (Roche, cat. no. 11284932001) at 37 °C for 30 min with gentle shaking. The resulting digest was filtered through a 100-μm nylon cell strainer. To enrich for LP lymphocytes, the cell suspension was resuspended in 37% Percoll (Cytiva, cat. no. 17089109) and carefully layered onto a 70% Percoll solution in a 15-ml tube. After centrifugation, lymphocytes were collected from the 37%/70% interface, washed and resuspended in complete medium for downstream applications.

Histological analysis

Distal colon segments were fixed in 4% paraformaldehyde (Leagene, cat. no. DF0135), embedded in paraffin, and sectioned at a thickness of 5 μm. Tissue sections were stained with hematoxylin and eosin. Histopathological scoring (0–4) was performed in a blinded manner on the basis of the severity of inflammatory cell infiltration, colon wall thickening and loss of crypts and goblet cells.

EOMES-expressing vector construction and transfection

The coding sequence for mouse EOMES was cloned into the RVKM retroviral vector, which contains an internal ribosome entry site (IRES)-driven blue fluorescent protein reporter. To generate a C-terminally tagged version, a sequence encoding a glycine–serine linker followed by an HA epitope tag was inserted in-frame immediately upstream of the stop codon. The inserted nucleotide sequence is shown below, with the HA-coding portion underlined: GGAGGAGGCGGATCAGGAGGAGGCGGATCA<u>TACCCATACGACGTACCAGATTACGCT</u>. For retrovirus production, each plasmid was co-transfected with the packaging plasmid pCL-ECO into 293T cells using the calcium phosphate method. Virus-containing supernatants were collected 48 h post-transfection.

T cell differentiation and culture

Naive $CD4^+$ T cells were activated in vitro using plate-bound anti-CD3 (5 μg ml^{-1}) and anti-CD28 (5 μg ml^{-1}). Under the nonpolarizing condition, no additional cytokines or blocking antibodies were added to the culture medium. For differentiation under T_H0 condition, cultures were supplemented with anti-IFNγ (10 μg ml^{-1}) and anti-IL-4 (10 μg ml^{-1}). For T_H1 polarization, cells were cultured in the presence of anti-IL-4 (10 μg ml^{-1}) and mIL-12 (10 ng ml^{-1}). For T_H2 polarization, cells were cultured with anti-IFNγ (10 μg ml^{-1}), IL-4 (20 ng ml^{-1}) and mIL-2 (10 U ml^{-1}). T_H17 differentiation was induced with human TGF-β (1 ng ml^{-1}), mIL-6 (20 ng ml^{-1}), mIL-1β (10 ng ml^{-1}), IL-23 (25 ng ml^{-1}), together with anti-IFNγ (10 μg ml^{-1}) and anti-IL-4 (10 μg ml^{-1}). For T_{reg} cell differentiation, cells were cultured with mIL-2 (10 U ml^{-1}) and human TGF-β (2 ng ml^{-1}), supplemented with anti-IFNγ (10 μg ml^{-1}) and anti-IL-4 (10 μg ml^{-1}). All antibodies, including anti-CD3, anti-CD28, anti-IFNγ and anti-IL-4, were purchased from BioXCell; IL-23 was obtained from R&D Systems; all other cytokines were from Peprotech.

For the infection experiment, naive T cells were activated under T_H0 condition for 24 h before viral infection. At 24 h post-infection, the medium was replaced with the specified differentiation conditions for an additional 4 days before analysis, or cells were transferred into recipient mice for further experiments.

OVA immunization

One day after the adoptive transfer of OT-II $CD4^+$ T cells, mice were immunized subcutaneously near the inguinal lymph nodes with a 1:1 emulsion of chicken OVA (3 mg ml^{-1}, Sigma-Aldrich, cat. no. A5503) and Complete Freund's Adjuvant (Sigma-Aldrich, cat. no. F5881). Analysis was performed on day 7 post-immunization.

In vitro cytotoxicity assay

$CD4^+$ T cells were isolated from LP of colitic mice and used as effector cells. The mouse mastocytoma cell line P815 was labeled with CFSE (BD, cat. no. 565082) and used as target cells. Effector and target cells were co-cultured for 19 h in the presence of soluble anti-CD3 (0 or 1 μg ml^{-1}, BioXCell, cat. no. BE0002, clone 17A2) at the indicated effector-to-target (E:T) ratios. Specific cytotoxicity was assessed by flow cytometry using a Fixable Viability Dye eF506 (eBioscience, cat. no. 65-0866-18).

Flow cytometry analysis

Before surface staining, cells were stained with Fixable Viability Dye eF506 (eBioscience, cat. no. 65-0866-18) and blocked with CD16/CD32 antibodies (BD, cat. no. 553142, clone 2.4G2, AB_394656) to prevent nonspecific binding. Cells were first stained with various surface markers, then fixed and permeabilized using the FOXP3/Transcription Factor Staining Buffer Set (eBioscience, cat. no. 00-5523-00) for intracellular staining. If fluorescent protein signals were to be detected, cells were fixed with 2% paraformaldehyde for 20 min before permeabilization to preserve fluorescence integrity. For intracellular cytokine detection, cells were restimulated with phorbol 12-myristate 13-acetate (PMA; 50 ng ml^{-1}, Sigma-Aldrich, cat. no. P8139) and ionomycin (500 ng ml^{-1}, Sigma-Aldrich, cat. no. I0634) in the presence of GolgiStop (BD, cat. no. 554724) for 4.5 h. The following antibodies were used at a 1:400 dilution unless otherwise noted: BUV395 rat anti-mouse CD4 (BD, cat. no. 563790, clone GK1.5, AB_2738426); Brilliant Violet 785 anti-mouse CD4 (BioLegend, cat. no. 100552, clone RM4-5, AB_2563053); Brilliant Violet 421 anti-mouse CD45.2 (BioLegend, cat. no. 109832, clone 104, AB_2565511); Brilliant Violet 605 anti-mouse CD8a (BioLegend, cat.

no. 100744, clone 53-6.7, AB_2562609); Spark PLUS UV395 anti-mouse TCR β chain (BioLegend, cat. no. 109264, clone H57-597, AB_3662168); RB705 rat anti-mouse CD279 (PD-1) (BD, cat. no. 570566, clone 29F.1A12, AB_3685850); EOMES monoclonal antibody, eFluor 450 (eBioscience, cat. no. 48-4875-82, clone Dan11mag, AB_2574062); T-bet monoclonal antibody, PE-cyanine7 (eBioscience, cat. no. 25-5825-82, clone eBio4B10 (4B10), AB_11042699); RB705 mouse anti-T-bet (BD, cat. no. 570294, clone O4-46, AB_3685647); Gata3 monoclonal antibody, PerCP-eFluor 710 (eBioscience, cat. no. 46-9966-42, clone TWAJ, AB_10804487); RB705 mouse anti-mouse RORγt (BD, cat. no. 570259, clone Q31-378, AB_3685620); PE-CF594 mouse anti-Bcl-6 (BD, cat. no. 562401, clone K112-91, AB_11152084, 1:200); FOXP3 monoclonal antibody, FITC (eBioscience, cat. no. 11-5773-82, clone FJK-16s, AB_465243); FOXP3 monoclonal antibody, PerCP-eFluor 710 (eBioscience, cat. no. 46-5773-82, clone FJK-16s, AB_2811810); APC anti-mouse perforin antibody (BioLegend, cat. no. 154304, clone S16009A, AB_2721463); Brilliant Violet 785 anti-human/mouse granzyme B recombinant antibody (BioLegend, cat. no. 396438, clone QA18A02, AB_3106140); PE/Dazzle 594 anti-human/mouse granzyme B recombinant antibody (BioLegend, cat. no. 372216, clone QA16A02, AB_2728383); Alexa Fluor 700 rat anti-mouse IFNγ (BD, cat. no. 557998, clone XMG1.2, AB_396979); biotin rat anti-mouse CD185 (CXCR5) (BD, cat. no. 551960, clone 2G8, AB_394301, 1:50); PE/cyanine7 streptavidin (BioLegend, cat. no. 405206, 1:200). Flow cytometry was performed using an LSRFortessa (BD) instrument and data were analyzed with FlowJo software v.10.9.0.

ScRNA-seq data analysis

For human scRNA-seq data, preprocessed datasets from previous studies[14,16] were directly utilized for downstream analysis and visualization. Gene expression density mapping was performed using the Nebulosa package[53] (v.1.16.0). In the IBD dataset[14], *GZMK*^{high} cells were primarily localized within the originally annotated 'CD4$^+$ Temra' cluster. To refine this population, we re-subclustered this group and separated it into a major *GZMK*^{high} subset and a minor *GZMK*$^{low/neg}$ subset; the former was used for subsequent differential expression analysis. In the pancancer dataset[16], the originally annotated 'CD4.c12(GZMK$^+$ T_{EM})' cluster was directly designated as the *GZMK*^{high} population for comparative analysis.

For mouse colitis data, Seurat objects from two published datasets (CRA016814 and GSE235664)[17,18] were subset to include only wild-type cells, merged, and then processed through a standard workflow using Seurat[54] (v.5.2.1). This included normalization, selection of highly variable features, and scaling. Principal-component analysis was performed, followed by batch effect correction with harmony[55] (v.1.2.3).

For additional mouse datasets, including tumor-infiltrating conventional CD4$^+$ T cells from Hepa1-6-bearing mice (GSE285225)[27], CNS-infiltrating CD4$^+$ T cells from EAE mice (GSE156196) and GP66$^+$ splenic CD4$^+$ T cells from mice infected with LCMV cl13 (GSE201730), data were processed through the standard Seurat workflow. After unsupervised clustering, the cluster exhibiting the highest *Gzmk* expression was designated as the *Gzmk*^{high} subset for downstream differential expression analysis. Differential expression analysis was performed using Seurat's FindMarkers function.

All data were visualized using UMAP. To evaluate coordinated gene expression patterns in mouse datasets, module scores for predefined gene sets were calculated using the AddModuleScore function in Seurat.

RNA extraction and gene expression analysis

T cells were restimulated with PMA (50 ng ml^{-1}) and ionomycin (500 ng ml^{-1}) for 2 h before RNA extraction. Total RNA was extracted using TRIzol reagent (Magen, cat. no. R4801) according to the manufacturer's instructions. RNA concentration and purity were determined, and complementary DNA was synthesized from qualified RNA samples using the HiScript III All-in-one RT SuperMix Perfect (Vazyme, cat. no. R333). Quantitative PCR with reverse transcription (RT–qPCR) was performed using SYBR Green Master Mix Low ROX (Yeasen, cat. no. 11202ES08), with *Actb* serving as the endogenous reference control. The following primer sequences were used for RT–qPCR: *Gzmk* (F: 5′-TGGCTGGCGTTTATATGTCTTC-3′, R: 5′-GCTGCGGTACTGGATGGAC-3′); *Prf1* (F: 5′-AGCACAAGTTCGTGCCAGG-3′, R: 5′-GCGTCTCTCATTAGGGAGTTTTT-3′); *Gzmb* (F: 5′-TGTGAAGCCAGGAGATGTGT-3′, R: 5′-TCAGCTCAACCTCTTGTAGC-3′); and *Actb* (F: 5′-TGGAATCCTGTGGCATCCATGAAAC-3′, R: 5′-TAAAACGCAGCTCAGTAACAGTCCG-3′).

For RNA-seq analysis, total RNA was sent to OE Biotech Co. for library preparation with the VAHTS Universal V10 RNA-seq Library Prep kit (Vazyme, cat. no. NRM606) and subsequent sequencing on the DNBSEQ-T7 platform, which generated 150-bp paired-end reads. Approximately 20 million raw reads were obtained per sample.

RNA-seq data analysis

Raw paired-end FASTQ reads were quality controlled and adaptor-trimmed using fastp[56] (v.0.23.4) to obtain high-quality clean reads. The clean reads were then aligned to the mm39 (GRCm39) reference genome using HISAT2 (v.2.2.1)[57] with the strand-specificity option set to RF. The resulting alignments were sorted and indexed using SAMtools[58] (v.1.18). Gene-level read counts were quantified with featureCounts (Subread[59], v.2.0.6) in paired-end mode. TPM values were calculated to normalize for gene length and sequencing depth. Differential expression analysis was performed using DESeq2 (v.1.46.0)[60]. GSEA was conducted using the clusterProfiler[61] (v.4.12.0). For visualization, strand-specific BigWig files were generated from the aligned BAM files using bamCoverage (deepTools[62], v.3.5.4) with RPKM normalization, read extension and a 10-bp bin size. Group-average signal tracks were generated from these files using bigwigAverage (deepTools[62], v.3.5.4).

ATAC-seq and data analysis

Approximately 100,000 cells were sorted per sample for ATAC-seq. Libraries were prepared using the Hyperactive ATAC-Seq Library Prep kit (Vazyme, cat. no. TD711) according to the manufacturer's instructions, followed by sequencing on the Illumina NovaSeq X Plus platform to generate 150-bp paired-end reads. Approximately 40 million raw reads were generated per sample.

Raw reads were quality controlled and adaptor-trimmed using fastp[56] (v.0.24.0) with default parameters. Trimmed reads were aligned to the mm39 (GRCm39) reference genome using Bowtie2[63] (v.2.5.4) with the options –very-sensitive, -X 2000, –no-mixed, –no-discordant and –no-unal. Post-alignment processing included removal of PCR duplicates using Picard MarkDuplicates (v.3.1.0), exclusion of mitochondrial reads (chrM) and filtering for properly paired reads with mapping quality (MAPQ) ≥ 30 using SAMtools[58] view (v.1.21). Reads overlapping ENCODE blacklist regions were removed using bedtools[64] intersect -v (v.2.31.1), and the final filtered BAM files were sorted and indexed with SAMtools[58]. Peak calling was performed on individual samples using MACS2 (v.2.2.9.1) with the following parameters: -f BAMPE, -q 0.01, –nomodel, –shift -100, –extsize 200, –keep-dup auto, -g mm and –call-summits. BigWig files were generated from the final BAM files using bamCoverage (deepTools[62], v.3.5.6) with RPKM normalization, an extension size of 200 bp, a bin size of 10 bp and smoothing over 50 bp (–smoothLength 50). For group-level analysis, BigWig files from samples within the same experimental group were averaged using bigwigAverage (deepTools[62], v.3.5.6) across all genomic positions with a bin size of 10 bp to generate consensus signal tracks for downstream visualization. Downstream analysis was performed with the DiffBind package (v.3.16.0) to identify differentially accessible peaks. Motif enrichment analysis was performed using HOMER[65] (v.5.1) to compare peaks with increased accessibility in the '*Gzmk*-tdTomato$^+$' group versus those with decreased accessibility (and vice versa) to identify

condition-specific enriched transcription factor binding motifs, with peak centers extended to 200 bp and mm39 as the reference genome.

Gene-to-peak association analysis and integrated visualization

Gene-to-peak associations were analyzed to link chromatin accessibility changes with gene expression. Differentially expressed genes (DEGs) and differentially accessible regions (DARs) were defined using adjusted *P* values < 0.05 and absolute $\log_2$(fold change) >1. Transcription start site (TSS) coordinates were obtained from the TxDb.Mmusculus.UCSC.mm39.knownGene annotation package (v.3.20.0). For each DEG, DARs within ±100 kb of its TSS were identified, and only pairs showing concordant directionality were retained; the nearest such DAR to the TSS was assigned to the gene. Gene–peak pairs were ordered by gene expression fold change for visualization. RNA-seq (TPM) and ATAC-seq (normalized counts) data were $\log_2$-transformed with a pseudocount of 1 and underwent row-wise *z*-score normalization for heatmap display.

CUT&Tag experiments

$CD4^+$ T cells overexpressing HA-tagged EOMES protein or empty vector were collected for CUT&Tag assays, with three biological replicates per sample. The experiments were performed using the Hyperactive Universal CUT&Tag Assay kit (Vazyme, cat. no. TD904-C2) and an antibody to HA-Tag (Cell Signaling Technology, cat. no. 3724), following the manufacturer's instructions. Sequencing was conducted on the Illumina NovaSeq X Plus platform to generate 150-bp paired-end reads, yielding approximately 20 million raw reads per sample.

CUT&Tag data analysis

Raw paired-end reads were processed using the same pipeline as described for ATAC-seq data, which included quality control and adaptor trimming with fastp[56] (v.0.24.0), alignment to the mm39 genome using Bowtie2[63] (v.2.5.4) (with parameters: –very-sensitive, -X 2000, –no-mixed, –no-discordant and –no-unal), duplicate removal with Picard MarkDuplicates (v.3.1.0), and filtering of mitochondrial and ENCODE blacklist regions using bedtools[64] intersect (v.2.31.1). Final BAM files were sorted and indexed. BigWig files were generated using bamCoverage (deepTools[62], v.3.5.6) with RPKM normalization, a bin size of 10 bp, and the –centerReads option. For each experimental group, three replicates were averaged using bigwigAverage (deepTools[62], v.3.5.6) with a bin size of 10 bp to generate consensus tracks. The control-averaged signal was then subtracted from the treatment sample using bigwigCompare (deepTools[62], v.3.5.6) with the same bin size; the resulting differential signal tracks were used for downstream visualization. Peak calling was performed on individual replicates using MACS2 (v.2.2.9.1) in BAMPE mode with the parameters: -q 0.01, –keep-dup auto, –call-summits and -g mm. Consensus peaks for downstream analysis were defined as regions identified in all three biological replicates using sequential bedtools[64] intersect operations. Genomic annotation of the consensus peaks was performed using ChIPseeker[66] (v.1.42.1) with the TxDb.Mmusculus.UCSC.mm39.knownGene annotation package (v.3.20.0), defining promoter regions as ±3 kb from the TSS. Motif enrichment analysis was performed on these consensus peaks using HOMER[65] (v.5.1) with a region size of 400 bp (−300 to +100 bp relative to the peak center) and the -len 8,10,12 option to test multiple motif widths.

Quantification and statistical analysis

No statistical methods were used to predetermine sample sizes but our sample sizes are similar to those reported in previous publications[17,18]. Data distribution was assumed to be normal but this was not formally tested. Statistical analyses were performed using GraphPad Prism (v.10.1.2). For experimental data, two-tailed paired or unpaired Student's *t*-tests, one-way ANOVA with Tukey's post hoc test or two-way ANOVA with Bonferroni's post hoc test were applied as appropriate. For next-generation sequencing data, R (v.4.4.3) was used for statistical analysis. Error bars represent the s.e.m. throughout the study. Animals were randomly assigned to experimental groups. The order of experimental conditions and data collection was randomized to minimize bias. Data collection and analysis were not performed blind to the conditions of the experiments. No animals or data points were excluded from the analyses.

Reporting summary

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

Data availability

The raw sequence data reported in this paper have been deposited in National Genomics Data Center under accession code PRJCA048888. The processed RNA-seq, ATAC-seq and CUT&Tag data of this study are available on figshare at https://doi.org/10.6084/m9.figshare.30477716 (ref. 67). Source data are provided with this paper.

Code availability

This article does not report any original code. All functionality used to analyze the sequencing data is publicly available via the software packages described in Methods.

References

50. Liu, Z. et al. Epigenetic reprogramming of Runx3 reinforces $CD8^+$ T-cell function and improves the clinical response to immunotherapy. *Mol. Cancer* **22**, 84 (2023).
51. Chi, X. et al. RORγt expression in mature T(H)17 cells safeguards their lineage specification by inhibiting conversion to T(H)2 cells. *Sci. Adv.* **8**, eabn7774 (2022).
52. Liu, X. et al. Genome-wide analysis identifies Bcl6-controlled regulatory networks during T follicular helper cell differentiation. *Cell Rep.* **14**, 1735–1747 (2016).
53. Alquicira-Hernandez, J. & Powell, J. E. Nebulosa recovers single-cell gene expression signals by kernel density estimation. *Bioinformatics* **37**, 2485–2487 (2021).
54. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
55. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
56. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
57. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
58. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
59. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
60. 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).
61. Xu, S. et al. Using clusterProfiler to characterize multiomics data. *Nat. Protoc.* **19**, 3292–3320 (2024).
62. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
63. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
64. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

65. 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).
66. Wang, Q. et al. Exploring epigenomic datasets by ChIPseeker. *Curr. Protoc.* **2**, e585 (2022).
67. Xie, T. Processed RNA-seq, ATAC-seq and CUT&Tag data. *figshare* https://doi.org/10.6084/m9.figshare.30477716 (2026).

Acknowledgements

We thank Y. Qiu (Peking University, Beijing) for providing *Stat6*$^{-/-}$ mice, F. Shao (National Institute of Biological Sciences, Beijing) for *Gzmk*$^{-/-}$ mice and Y. Sun (Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing) for *Runx3*$^{fl/fl}$ mice. We thank all Dong laboratory members for their help. This work was supported by grants from the National Natural Science Foundation of China (32550002 to C.D., 82402058 to Q.L. and 92574302 to Q.L.), the Zhejiang Provincial Natural Science Foundation of China (LD25C120002 to C.D.), the National Key R&D Program of China (2023YFC2508200 to H. Yuan) and the Construction Fund of Key Medical Disciplines of Hangzhou on Immune Diseases and Precision Therapy (2025HZGF01 to C.D.). C.D. was a New Cornerstone Investigator.

Author contributions

T.X. and C.D. conceived of the study. T.X. performed the experiments, conducted bioinformatic analyses and drafted the paper. Y.D., Q.W., H.Z., K.W., X.C., X.B., Y.F., Z.P., Y.Z. and Q.L. provided experimental support. C.D. supervised the research and revised the paper.

Competing interests

The authors declare no competing interests.

Additional information

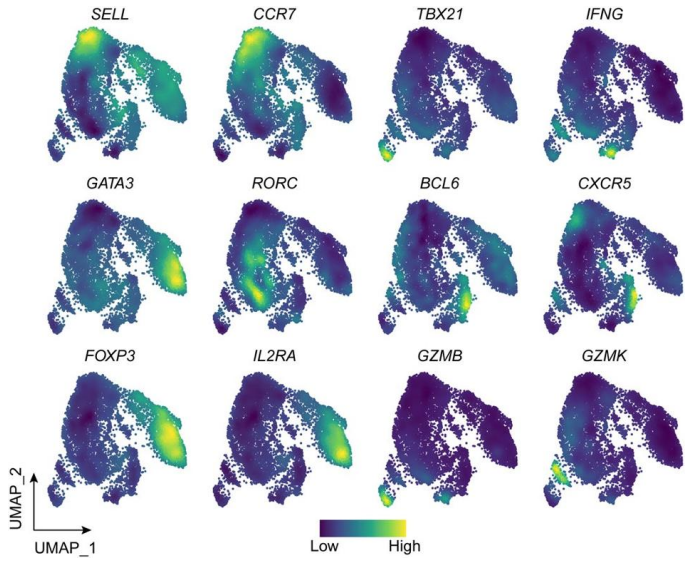
Extended data is available for this paper at https://doi.org/10.1038/s41590-026-02479-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41590-026-02479-6.

Correspondence and requests for materials should be addressed to Chen Dong.

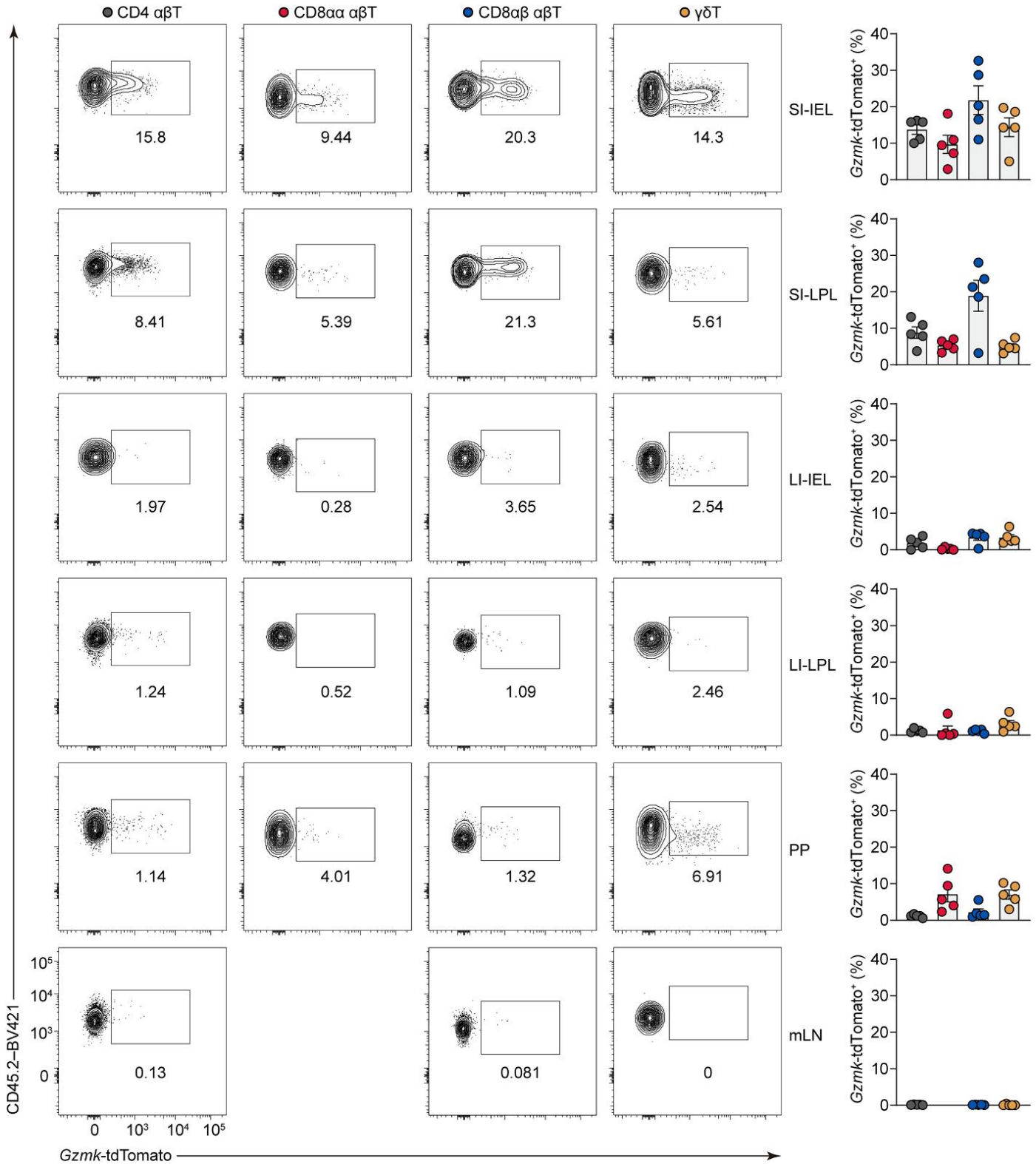
Peer review information *Nature Immunology* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: L. A. Dempsey, in collaboration with the *Nature Immunology* team.

https://doi.org/10.1038/s41590-026-02479-6



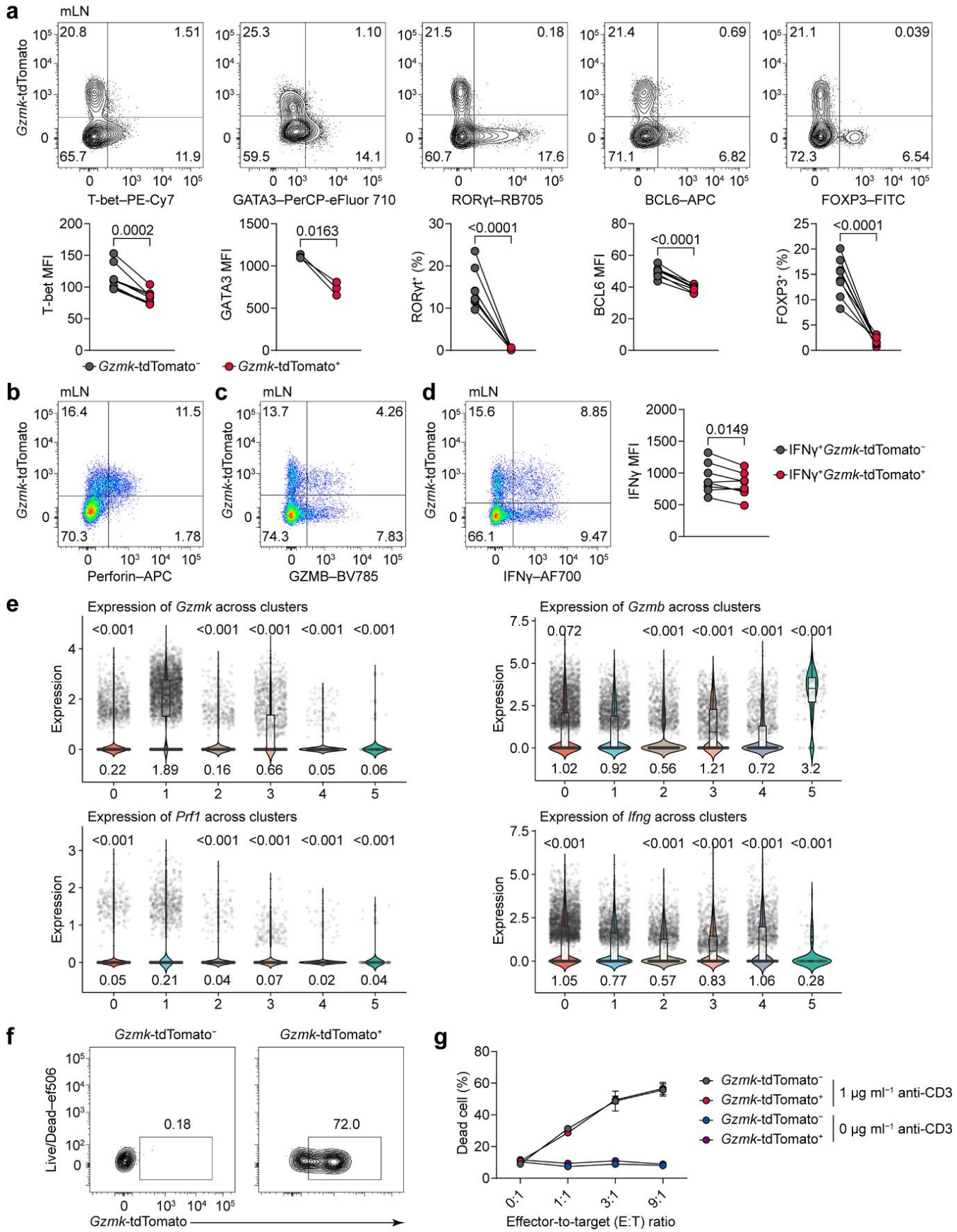
Extended Data Fig. 1 | A *GZMK*-high CD4$^+$ T cell population in integrated pan-cancer scRNA-seq data. UMAP visualization of CD4$^+$ T cells from a published integrated pan-cancer scRNA-seq dataset. Density plots show expression of *SELL*, *CCR7*, *TBX21*, *IFNG*, *GATA3*, *RORC*, *BCL6*, *CXCR5*, *FOXP3*, *IL2RA*, *GZMB*, and *GZMK*.

https://doi.org/10.1038/s41590-026-02479-6



Extended Data Fig. 2 | Analysis of *Gzmk*-tdTomato reporter expression across intestinal and lymphoid tissues in steady-state mice. Flow cytometric analysis of lymphocytes from small intestinal intraepithelial lymphocytes (SI-IEL), small intestinal lamina propria lymphocytes (SI-LPL), large intestinal IEL (LI-IEL), large intestinal LPL (LI-LPL), Peyer's patches (PP), and mesenteric lymph nodes (mLN) of *Gzmk*-tdTomato reporter mice at steady-state. Left panels: representative plots showing tdTomato signal in $CD4^+$, $CD8\alpha\alpha^+$, $CD8\alpha\beta^+$, and γδ T cell subsets. Right panels: frequency of $tdTomato^+$ cells within each subset ($n = 5$). Data are shown as mean ± s.e.m. and are representative of three independent experiments.

https://doi.org/10.1038/s41590-026-02479-6

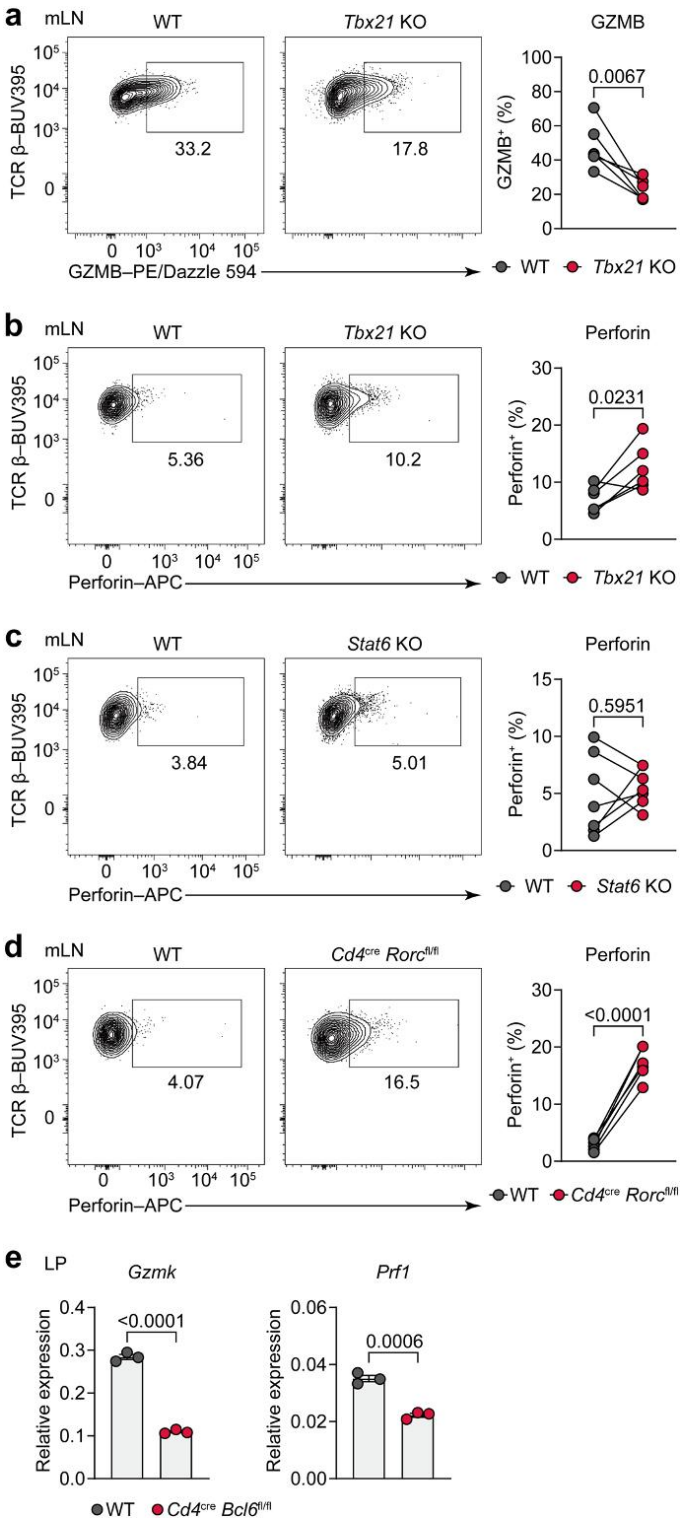


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41590-026-02479-6

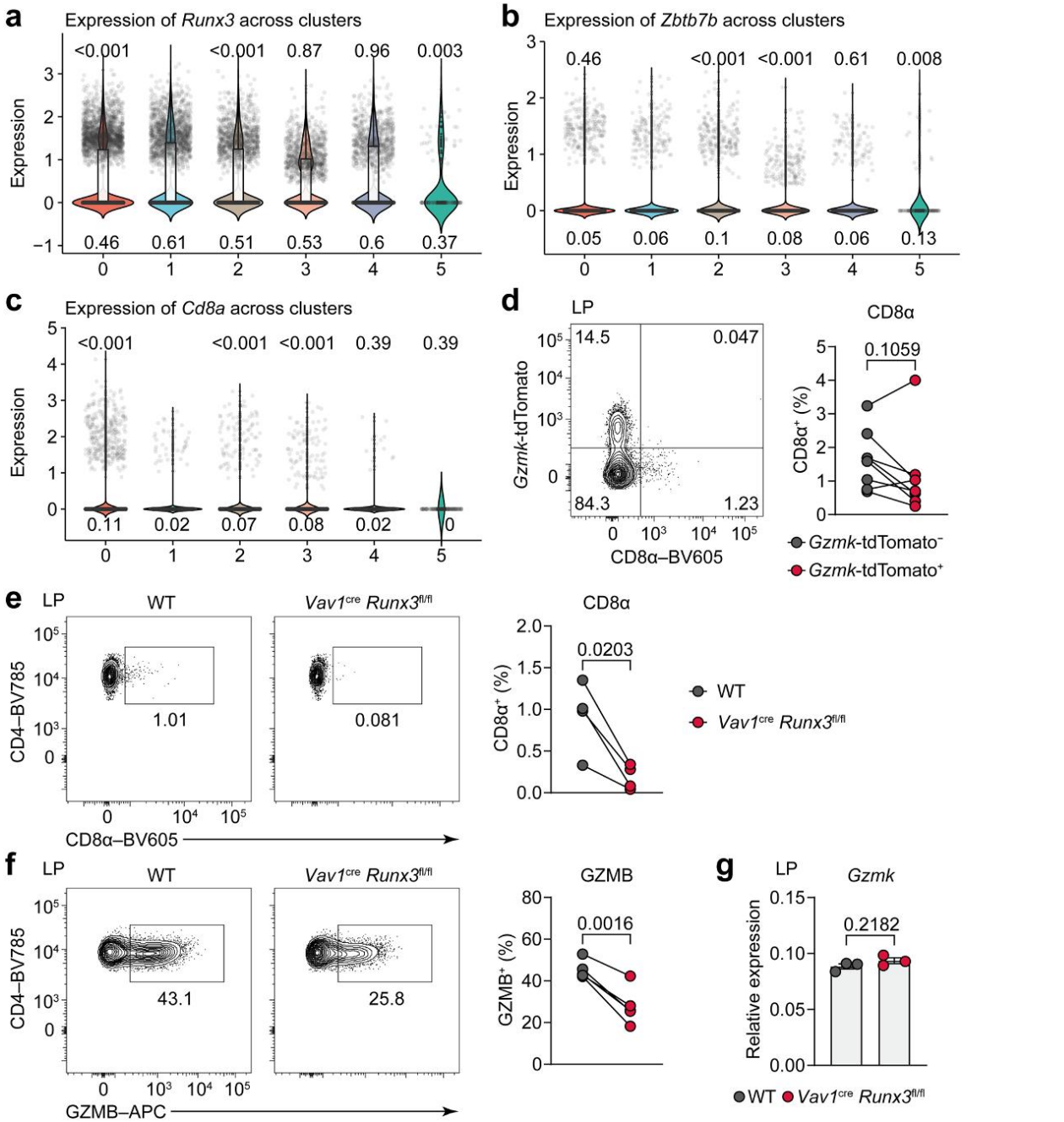
Extended Data Fig. 3 | Phenotypic and functional characterization of *Gzmk*-expressing CD4$^+$ T cells. a, Flow cytometry analysis of *Gzmk*-tdTomato expression in mesenteric lymph nodes (mLN) CD4$^+$ T cells of *Rag1*$^{-/-}$ colitis mice adoptively transferred with naive CD4$^+$ T cells from *Gzmk*-tdTomato reporter mice. Upper: representative plots (y-axis: *Gzmk*-tdTomato; x-axis: T-bet, GATA3, RORγt, BCL6, or FOXP3). Lower: quantification of MFI of T-bet ($n = 8$), GATA3 ($n = 3$), and BCL6 ($n = 8$), and frequency of RORγt$^+$ ($n = 8$) and FOXP3$^+$ ($n = 8$) cells in *Gzmk*-tdTomato$^+$ versus *Gzmk*-tdTomato$^-$ cells. **b**–**d**, Flow cytometry analysis of perforin (**b**), GZMB (**c**), and IFNγ (**d**) in mLN CD4$^+$ T cells. Left: representative plots (y-axis: *Gzmk*-tdTomato; x-axis: target protein). In **d**, right: mean fluorescence intensity (MFI) of IFNγ in IFNγ$^+$*Gzmk*-tdTomato$^-$ and IFNγ$^+$*Gzmk*-tdTomato$^+$ subsets ($n = 8$). **e**, Violin plots showing expression of *Gzmk*, *Prf1*, *Gzmb*, and *Ifng* across CD4$^+$ T cell clusters (cluster 0, $n = 5{,}932$; cluster 1, $n = 3{,}502$; cluster 2, $n = 2{,}703$; cluster 3, $n = 1{,}703$; cluster 4, $n = 1{,}610$; cluster 5, $n = 133$). Violin plots with embedded box plots showing median (center line) and 25th–75th percentiles (box bounds). Numbers above: adjusted *P* values (Wilcoxon rank-sum test, two-tailed, Cluster 1 versus all others); below: average expression levels. **f**, Proportion of *Gzmk*-tdTomato$^+$ cells among sorted populations after 19-h co-culture with P815 target cells. **g**, In vitro cytotoxicity assay measuring proportion of dead P815 cells at varying effector-to-target ratios, with or without anti-CD3 antibody ($n = 3$). Data are shown as mean ± s.e.m. and are representative of at least two independent experiments. Exact *P* values are shown in each graph. Statistical significance was assessed by two-tailed paired *t*-test (**a**,**d**).

https://doi.org/10.1038/s41590-026-02479-6



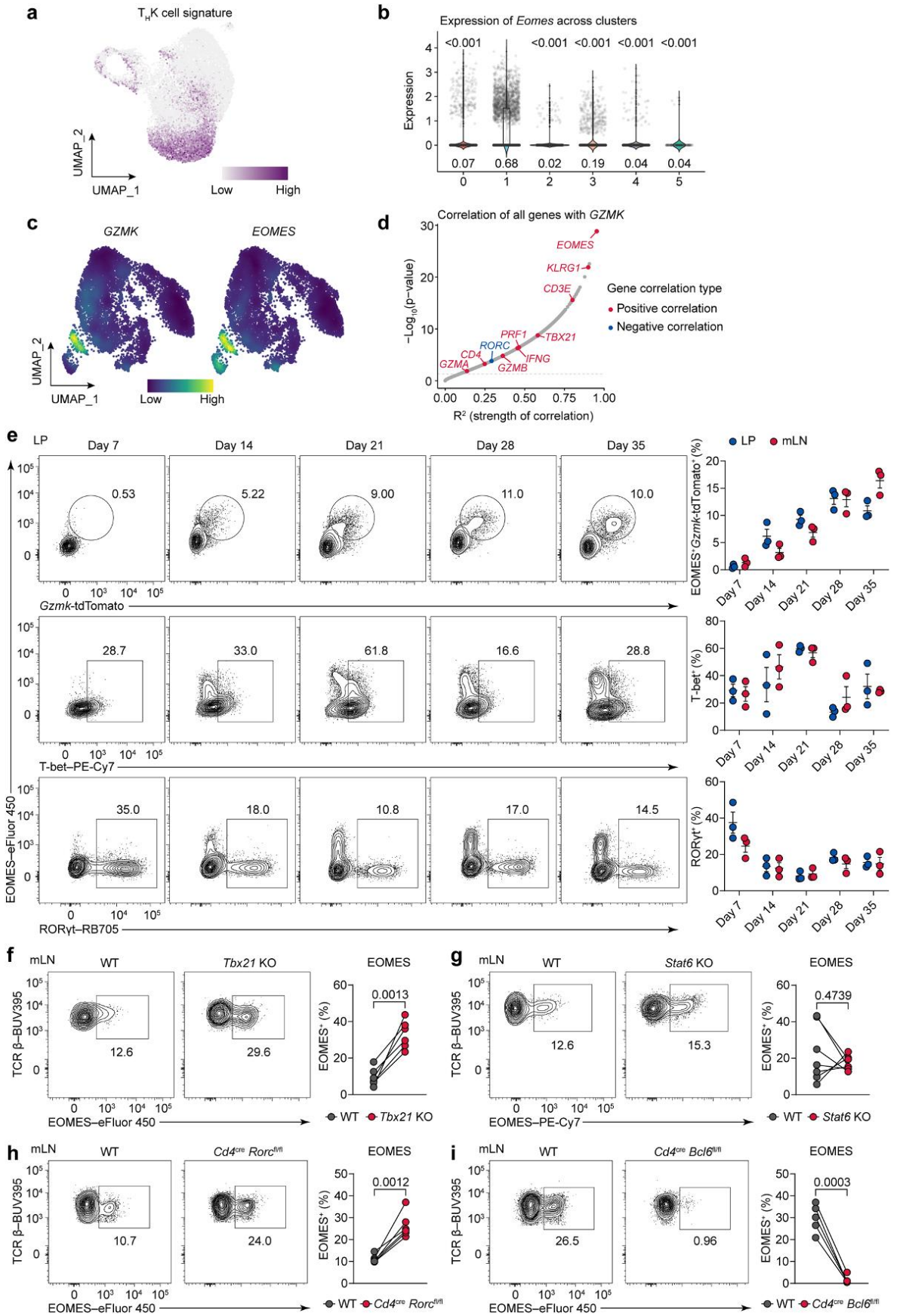
Extended Data Fig. 4 | T_HK cell differentiation is independent of T_H1, T_H2, and T_H17 programs. a,**b**, Flow cytometry analysis of GZMB (**a**) and perforin (**b**) in mesenteric lymph nodes (mLN) CD4$^+$ T cells from mice co-transferred with wild-type and *Tbx21*$^{-/-}$ cells ($n = 6$). Representative plots and quantification as in Fig. 3a, b. **c**,**d**, Flow cytometry analysis of perforin in mLN CD4$^+$ T cells from recipients of wild-type and *Stat6*$^{-/-}$ ($n = 7$) (**c**) or *Cd4*^{cre} *Rorc*$^{fl/fl}$ ($n = 6$) (**d**) cells; quantification as in Fig. 3e. **e**, *Gzmk* and *Prf1* mRNA expression in sorted wild-type and *Cd4*^{cre} *Bcl6*$^{fl/fl}$ CD4$^+$ T cells from the colonic LP, measured by RT–qPCR ($n = 3$). Data are shown as mean ± s.e.m. and are representative of two independent experiments. Exact *P* values are shown in each graph. Statistical significance was assessed by two-tailed paired *t*-test (**a**–**d**) or two-tailed unpaired *t*-test (**e**).

https://doi.org/10.1038/s41590-026-02479-6



Extended Data Fig. 5 | T_HK cells are distinct from CD4⁺CD8α⁺ cytotoxic T cells. **a**–**c**, Violin plots showing expression of *Runx3* (**a**), *Zbtb7b* (**b**) and *Cd8a* (**c**) across CD4⁺ T cell clusters (cluster 0, $n = 5{,}932$; cluster 1, $n = 3{,}502$; cluster 2, $n = 2{,}703$; cluster 3, $n = 1{,}703$; cluster 4, $n = 1{,}610$; cluster 5, $n = 133$). Violin plots with embedded box plots showing median (center line) and 25th–75th percentiles (box bounds). Numbers above: adjusted *P* values (Wilcoxon rank-sum test, two-tailed, Cluster 1 versus all others); below: average expression levels. **d**, Flow cytometric analysis showing CD8α expression in *Gzmk*-tdTomato⁺ CD4⁺ T cells compared to other CD4⁺ T cell populations in lamina propria (LP). **e**,**f**, Flow cytometry analysis of CD8α ($n = 4$) and GZMB ($n = 5$) in LP CD4⁺ T cells from recipients of wild-type and *Vav1*^{cre} *Runx3*$^{fl/fl}$ CD4⁺ T cells **g**, *Gzmk* mRNA expression in sorted wild-type and *Vav1*^{cre} *Runx3*$^{fl/fl}$ CD4⁺ T cells from the colonic LP, measured by RT–qPCR ($n = 3$). Data are shown as mean ± s.e.m. and are representative of two independent experiments. Exact *P* values are shown in each graph. Statistical significance was assessed by two-tailed paired *t*-test (**d**–**f**) or two-tailed unpaired *t*-test (**g**).

https://doi.org/10.1038/s41590-026-02479-6

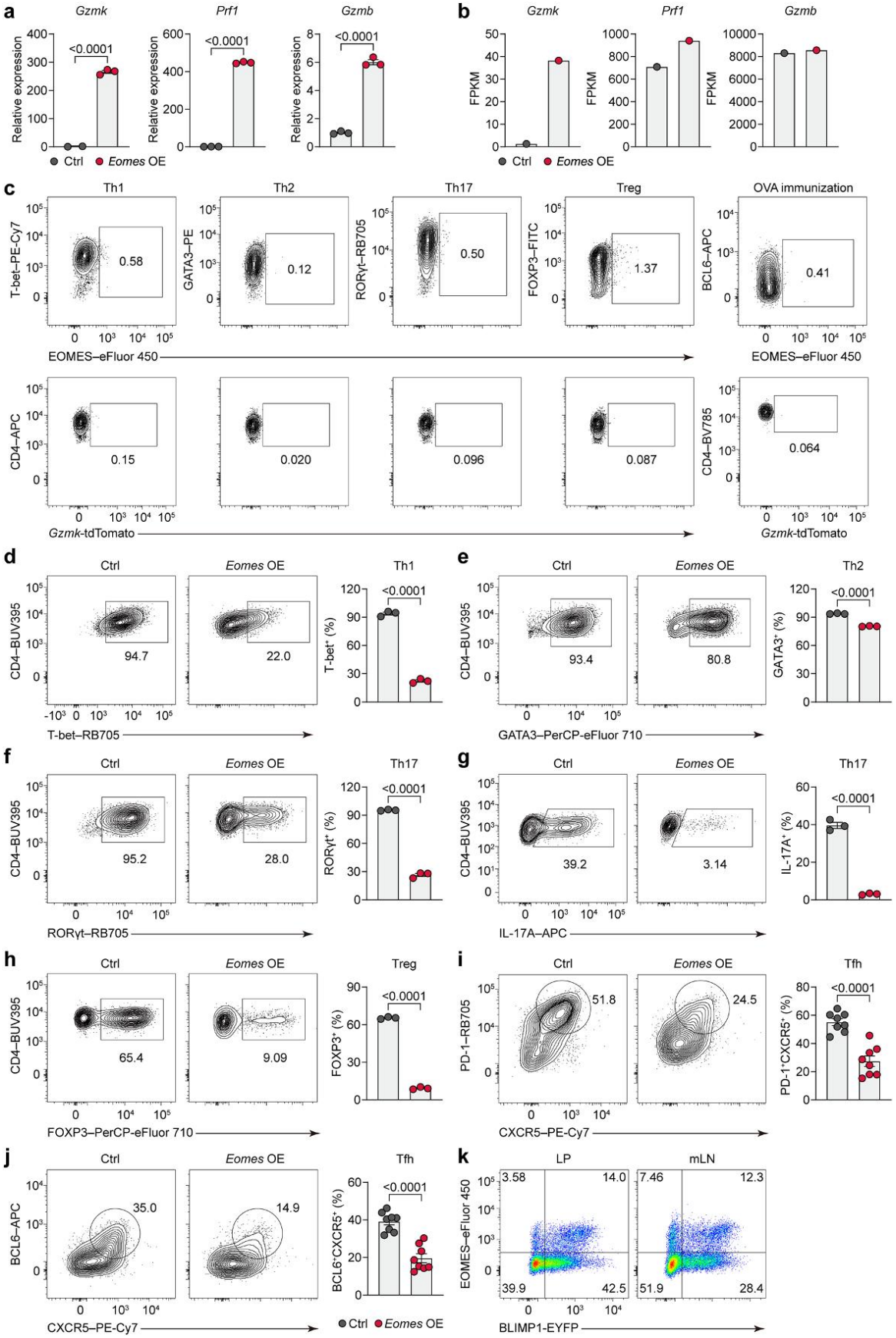


Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s41590-026-02479-6

Extended Data Fig. 6 | *Eomes* is transcriptionally and epigenetically associated with T_HK cell identity. a, UMAP embedding of scRNA-seq data from colitic $CD4^+$ T cells, colored by 'T_HK cell signature' enrichment score. **b**, *Eomes* expression across $CD4^+$ T cell clusters (cluster 0, $n = 5{,}932$; cluster 1, $n = 3{,}502$; cluster 2, $n = 2{,}703$; cluster 3, $n = 1{,}703$; cluster 4, $n = 1{,}610$; cluster 5, $n = 133$). Violin plots with embedded box plots showing median (center line) and 25th–75th percentiles (box bounds). Numbers above: adjusted *P* values (Wilcoxon rank-sum test, two-tailed, Cluster 1 versus all others); below: average expression levels. **c**, Density plots showing *GZMK* and *EOMES* expression in human pan-cancer dataset. **d**, Correlation analysis between *GZMK* and all other genes in bulk RNA-seq data from ulcerative colitis patients and healthy controls (GSE128682). Shown is R^2 from Pearson correlation; *P* values are unadjusted. **e**, Time-course analysis of EOMES, *Gzmk*-tdTomato, T-bet, and RORγt expression in $CD4^+$ T cells from the lamina propria (LP) and mesenteric lymph nodes (mLN) at days 7, 14, 21, 28, and 35 after naive $CD4^+$ T cell transfer. Left: Representative flow cytometry plots showing coexpression of EOMES and *Gzmk*-tdTomato (top), T-bet (middle), and RORγt (bottom) in LP. Right: Quantification of $EOMES^+$*Gzmk*-$tdTomato^+$, $T\text{-}bet^+$, and $ROR\gamma t^+$ $CD4^+$ T cells in LP and mLN ($n = 3$ for each time point). **f–i**, Flow cytometry analysis of EOMES expression in mLN $CD4^+$ T cells from *Rag1*$^{-/-}$ recipients co-transferred with wild-type ($CD45.1^+$) and knockout ($CD45.2^+$) naive $CD4^+$ T cells. *Tbx21*$^{-/-}$ ($n = 6$) (**f**), *Stat6*$^{-/-}$ ($n = 7$) (**g**), *Cd4*^{cre} *Rorc*$^{fl/fl}$ ($n = 6$) (**h**), and *Cd4*^{cre} *Bcl6*$^{fl/fl}$ ($n = 5$) (**i**). Left: representative plots. Right: frequency of $EOMES^+$ cells within donor-derived populations. Data are shown as mean ± s.e.m. and are representative of two independent experiments. Exact *P* values are shown in each graph. Statistical significance was assessed by two-tailed paired *t*-test (**f–i**).

https://doi.org/10.1038/s41590-026-02479-6

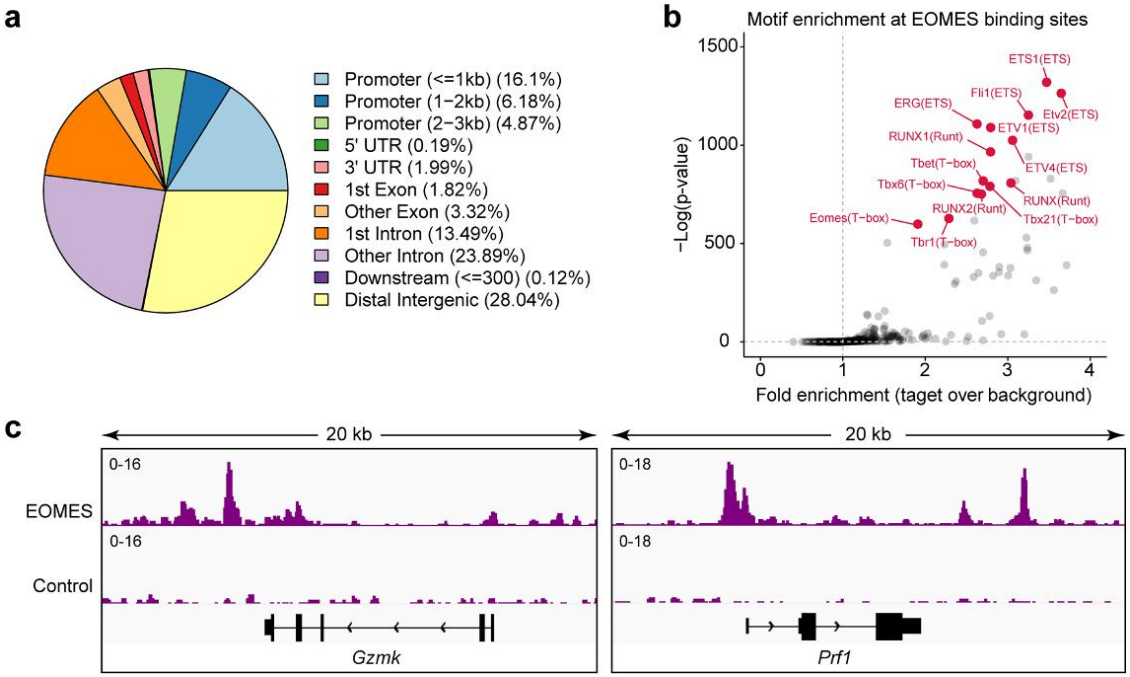


Extended Data Fig. 7 | See next page for caption.

https://doi.org/10.1038/s41590-026-02479-6

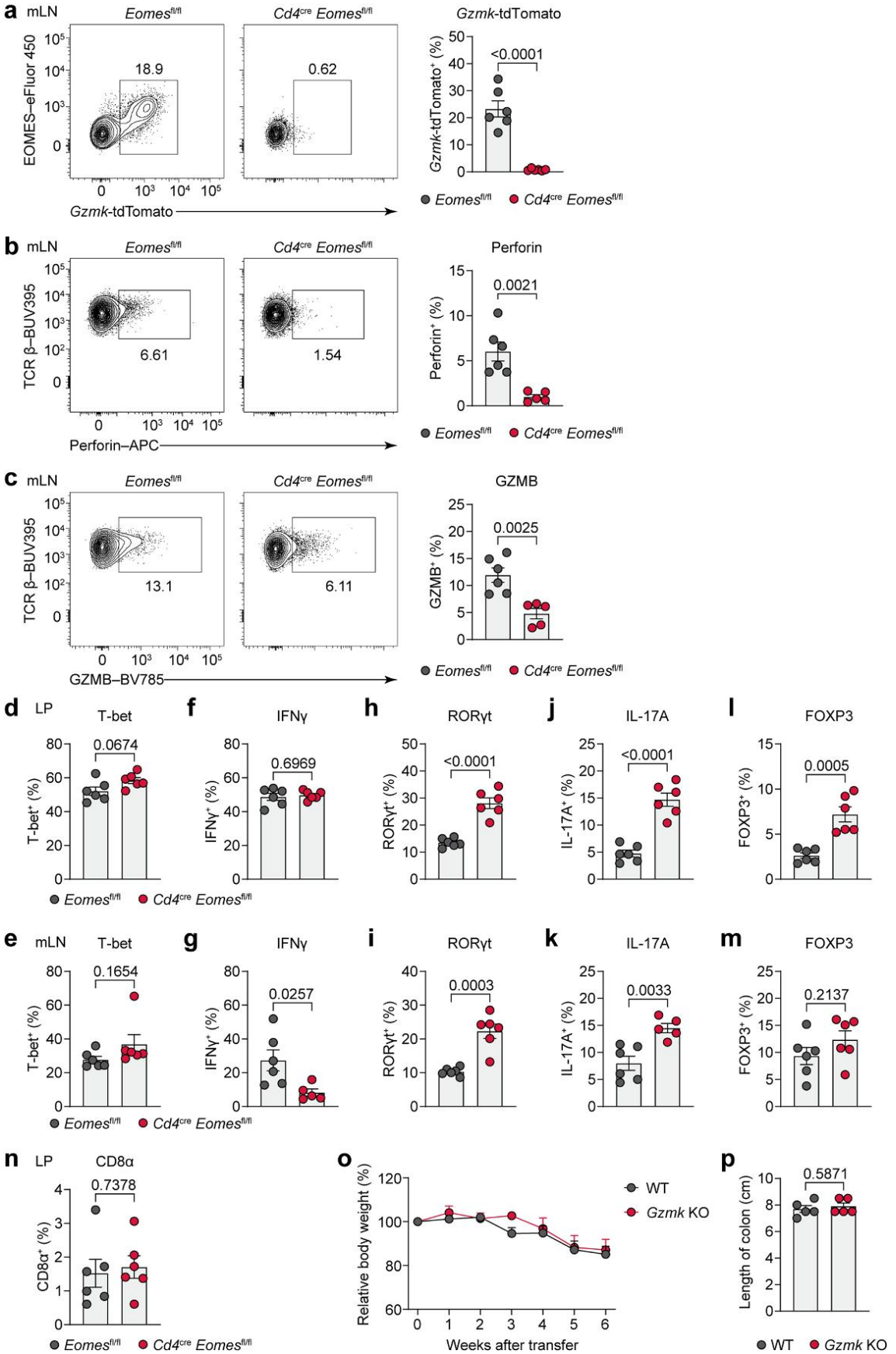
Extended Data Fig. 7 | EOMES drives the T_HK program and constrains alternative T_H cell fates. a, RT–qPCR analysis of *Gzmk*, *Prf1*, and *Gzmb* expression in control or *Eomes*-overexpressing CD4$^+$ T cells activated under non-polarizing conditions. Data normalized to *Actb* and presented relative to control ($n = 3$ per group; one undetected replicate in the *Gzmk* control group). **b**, Reanalysis of published RNA-seq data (GSE122889) from control or *Eomes*-overexpressing CD8$^+$ T cells. Expression levels of *Gzmk*, *Prf1*, and *Gzmb* shown as FPKM. **c**, EOMES and *Gzmk*-tdTomato expression in CD4$^+$ T cells under in vitro T_H1, T_H2, T_H17, T_{reg} polarizing conditions, and in vivo in adoptively transferred OT-II×*Gzmk*-tdTomato CD4$^+$ T cells from OVA-immunized recipient mice. **d**,**e**, Flow cytometry analysis of T-bet (**d**) and GATA3 (**e**) in control or *Eomes*-overexpressing CD4$^+$ T cells cultured under T_H1- and T_H2-polarizing conditions, respectively ($n = 3$). Left: representative plots. Right: quantification of frequency. **f**,**g**, Flow cytometry analysis of RORγt (**f**) and IL-17A (**g**) in control or *Eomes*-overexpressing CD4$^+$ T cells under T_H17-polarizing conditions ($n = 3$). Left: representative plots. Right: quantification of frequency. **h**, Flow cytometry analysis of FOXP3 in control or *Eomes*-overexpressing CD4$^+$ T cells under Treg-polarizing conditions ($n = 3$). Left: representative plots. Right: quantification of frequency. **i**,**j**, Flow cytometry analysis of PD-1 and CXCR5 (**i**) and BCL6 and CXCR5 (**j**) in donor-derived OT-II CD4$^+$ T cells 7 days after transfer into wild-type recipients and immunization with OVA protein ($n = 8$ per group). Left: representative plots. Right: quantification of frequency. **k**, Expression of EOMES and BLIMP1-EYFP in CD4$^+$ T cells isolated from the lamina propria (LP) and mesenteric lymph nodes (mLN) of colitis mice. Data are shown as mean ± s.e.m. and are representative of at least two independent experiments. Exact *P* values are shown in each graph. Statistical significance was assessed by two-tailed unpaired *t*-test.

https://doi.org/10.1038/s41590-026-02479-6



Extended Data Fig. 8 | EOMES directly binds to regulatory elements of T_HK-associated genes. a, Genomic annotation of EOMES-bound regions identified by CUT&Tag in *Eomes*-overexpressing $CD4^+$ T cells. Pie chart shows distribution across promoters, introns, distal intergenic regions, and other features. **b**, Transcription factor motif enrichment analysis in EOMES-bound peaks. Shown are fold enrichment (target/background) and−Log(p value) (natural log) from HOMER (binomial test). Dashed gray lines: $P = 0.05$ and fold enrichment = 1. **c**, Reanalysis of published ChIP-seq data (GSE122888) showing EOMES occupancy at the *Gzmk* and *Prf1* loci in control or *Eomes*-overexpressing $CD8^+$ T cells. Tracks display EOMES ChIP-seq signal and control signal (mm10).

https://doi.org/10.1038/s41590-026-02479-6

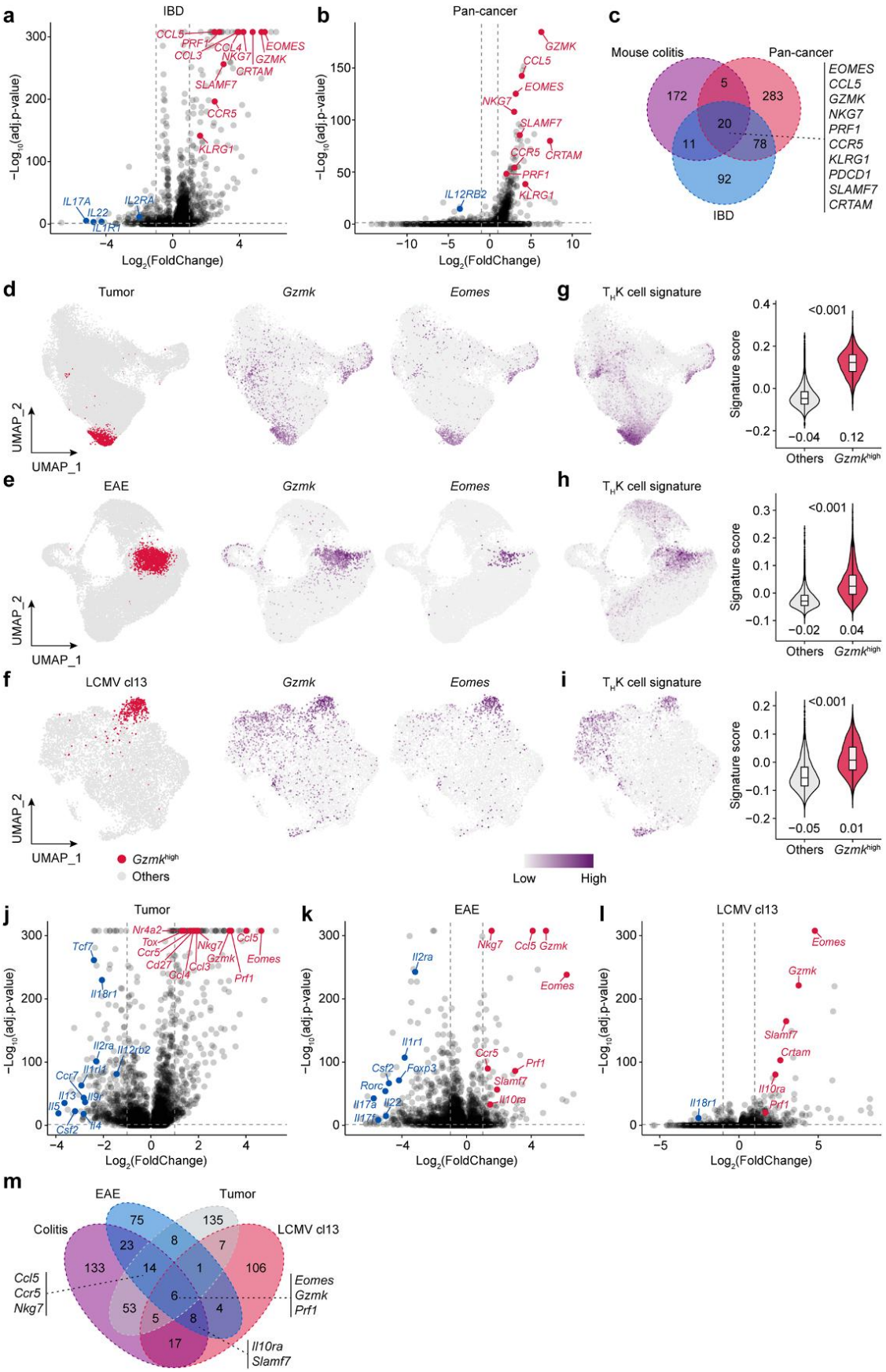


Extended Data Fig. 9 | See next page for caption.

https://doi.org/10.1038/s41590-026-02479-6

Extended Data Fig. 9 | *Eomes* ablation alters CD4$^+$ T cell differentiation in vivo. a–**m**, Flow cytometric analysis of transcription factors and effector molecules in donor-derived CD4$^+$ T cells from colonic lamina propria (LP) and mesenteric lymph nodes (mLN) ($n = 5$ or 6). Representative plots showing EOMES and *Gzmk*-tdTomato expression, with quantification of the frequency of *Gzmk*-tdTomato$^+$ cells in mLN (**a**). Representative plots (left) and frequencies (right) of perforin$^+$ (**b**) and GZMB$^+$ (**c**) cells in mLN. **d**–**m**, Frequencies of T-bet$^+$ (**d**,**e**), IFNγ$^+$ (**f**,**g**), RORγt$^+$ (**h**,**i**), IL-17A$^+$ (**j**,**k**), and FOXP3$^+$ (**l**,**m**) cells in LP (upper), mLN (lower). **n**, Frequencies of CD8α$^+$ cells in CD4$^+$ T cells in LP. **o**,**p**, Disease severity in *Rag1*$^{-/-}$ recipients of naive CD4$^+$ T cells from *Gzmk*$^{-/-}$ ($n = 5$) or control ($n = 5$) donors. Body weight change (**o**); colon length at endpoint (**p**). Data are shown as mean ± s.e.m. and are representative of three independent experiments. Each data point represents an individual mouse. Exact *P* values are shown in each graph. Statistical significance was assessed by two-tailed unpaired *t*-test (**a**–**n**,**p**) and two-way ANOVA with Bonferroni's post hoc test (**o**).

https://doi.org/10.1038/s41590-026-02479-6



Extended Data Fig. 10 | See next page for caption.

https://doi.org/10.1038/s41590-026-02479-6

Extended Data Fig. 10 | Analysis of the T_HK transcriptional program across species and diseases. a,**b**, Differential gene expression analysis comparing human *GZMK*^{high} CD4$^+$ T cells to other CD4$^+$ subsets in published scRNA-seq datasets from inflammatory bowel disease (IBD) patients (**a**; data from Fig. 1a) and pan-cancer samples (**b**; data from Extended Data Fig. 1). Differential expression was assessed using the Wilcoxon rank-sum test (two-tailed) implemented in the Seurat package. *P* values were adjusted using the Bonferroni correction. Red dots: upregulated in *GZMK*^{high} CD4$^+$ T cells; blue dots: upregulated in other CD4$^+$ T cells. Horizontal dashed line: adjusted $P = 0.05$; vertical dashed lines: $\log_2$(fold change) = ±1. **c**, Overlap between human orthologs of upregulated genes from bulk RNA-seq of *Gzmk*-tdTomato$^+$ versus *Gzmk*-tdTomato$^-$ CD4$^+$ T cells in murine colitis (Fig. 4a) and upregulated genes in human *GZMK*^{high} subsets (**a**,**b**). **d**–**f**, UMAP visualization of CD4$^+$ T cells from murine disease models: Hepa1-6 tumor-infiltrating lymphocytes (GSE285225; **d**), CNS-infiltrating CD4$^+$ T cells in EAE (GSE156196; **e**), and splenic CD4$^+$ T cells during chronic LCMV cl13 infection (GSE201730; **f**). *Gzmk*^{high} subsets are highlighted; *Gzmk* and *Eomes* expression are shown. **g**–**i**, Gene set enrichment scores of the 'T_HK cell signature' in *Gzmk*^{high} versus other CD4$^+$ T cell subsets in tumor (**g**; *Gzmk*^{high}, $n = 2{,}943$; others, $n = 42{,}419$), EAE (**h**; *Gzmk*^{high}, $n = 1{,}882$; others, $n = 18{,}003$), and LCMV cl13 (**i**; *Gzmk*^{high}, $n = 341$; others, $n = 7{,}031$) models. Violin plots with embedded box plots show median (center line) and 25th–75th percentiles (box bounds). Numbers above: adjusted *P* values (Wilcoxon rank-sum test, two-tailed, *Gzmk*^{high} versus others); below: mean scores. **j**–**l**, Volcano plots of differentially expressed genes in *Gzmk*^{high} CD4$^+$ T cells versus other CD4$^+$ T cell subsets in tumor (**j**), EAE (**k**), and LCMV cl13 (**l**) models. Differential expression was assessed using the Wilcoxon rank-sum test (two-tailed) implemented in the Seurat package. *P* values were adjusted using the Bonferroni correction. Red dots: upregulated in *Gzmk*^{high} CD4$^+$ T cells; blue dots: upregulated in other CD4$^+$ T cells. Horizontal dashed line: adjusted $P = 0.05$; vertical dashed lines: $\log_2$(fold change) = ±1. **m**, Venn diagram showing intersection of significantly upregulated genes across all four murine disease models. All upregulated genes in this figure were defined using the thresholds: adjusted $P < 0.05$ and $\log_2$(fold change) > 1.

nature portfolio

Corresponding author(s): Chen Dong

Last updated by author(s): Feb 21, 2026

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	BD FACSDIVA for flow cytometry; DNBSEQ-T7 platform for RNA-seq data; Illumina NovaSeq X Plus platform for ATAC-seq and CUT&Tag data.
Data analysis	Flow cytometry data were analyzed with FlowJo software version 10.9.0. Statistical analyses were performed using GraphPad Prism (v10.1.2) and R (v4.4.3). scRNA-seq analysis: Seurat (v5.2.1), Harmony (v1.2.3), Nebulosa (v1.16.0) Bulk RNA-seq analysis: fastp (v0.23.4), HISAT2 (v2.2.1), samtools (v1.18), featureCounts (v2.0.6), DESeq2 (v1.46.0), GSEABase (v1.68.0), deepTools (v3.5.4) ATAC-seq analysis: fastp (v0.24.0), Bowtie2 (v2.5.4), Picard (v3.1.0), samtools (v1.21), bedtools (v2.31.1), MACS2 (v2.2.9.1), deepTools (v3.5.6), DiffBind (v3.16.0), HOMER (v5.1) CUT&Tag analysis: fastp (v0.24.0), Bowtie2 (v2.5.4), Picard (v3.1.0), samtools (v1.21), bedtools (v2.31.1), MACS2 (v2.2.9.1), deepTools (v3.5.6), ChIPseeker (v1.42.1), TxDb.Mmusculus.UCSC.mm39.knownGene (v3.20.0), HOMER (v5.1)

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The integrated CD4⁺ T cell scRNA-seq dataset for IBD patients was obtained from Figshare (https://doi.org/10.6084/m9.figshare.22316641). The pan-cancer integrated CD4⁺ T cell scRNA-seq dataset was obtained from Zenodo (https://doi.org/10.5281/zenodo.5461803). The scRNA-seq data of mouse colonic CD4⁺ T cells were obtained from datasets CRA016814 and GSE235664. Additional mouse scRNA-seq datasets include: tumor-infiltrating conventional CD4⁺ T cells from Hepa1-6–bearing mice (GSE285225), CNS-infiltrating CD4⁺ T cells from EAE mice (GSE156196), and GP66⁺ splenic CD4⁺ T cells from mice with LCMV cl13 infection (GSE201730). The raw sequence data reported in this paper have been deposited in National Genomics Data Center with accession code PRJCA048888. The processed RNA-seq, ATAC-seq and CUT&Tag data of this study are available at Figshare (https://doi.org/10.6084/m9.figshare.30477716).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	No human participant involved
Reporting on race, ethnicity, or other socially relevant groupings	No human participant involved
Population characteristics	No human participant involved
Recruitment	No human participant involved
Ethics oversight	No human participant involved

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications. Sample sizes were stated in the figure legends and include at least three biological replicates per group.
Data exclusions	No data were excluded from the analysis.
Replication	Experiments were performed independently at least two times as described in detail in the Methods section and all attempts were successful.
Randomization	The mice were randomly assigned to different groups to avoid artificial conditions.
Blinding	Data collection and analysis were not performed blind to the conditions of the experiments.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used

The following antibodies were used at 1:400 dilutions unless otherwise noted: BUV395 Rat Anti-Mouse CD4 (BD, Cat. No. 563790, clone GK1.5, AB_2738426); Spark PLUS UV395 anti-mouse TCR β chain Antibody (Biolegend, Cat. No. 109264, clone H57-597, AB_3662168); RB705 Rat Anti-Mouse CD279 (PD-1) (BD, Cat. No. 570566, clone 29F.1A12, AB_3685850); EOMES Monoclonal Antibody, eFluor 450 (eBioscience, Cat. No. 48-4875-82, clone Dan11mag, AB_2574062); T-bet Monoclonal Antibody, PE-Cyanine7 (eBioscience, Cat. No. 25-5825-82, clone eBio4B10 (4B10), AB_11042699); RB705 Mouse Anti-T-bet (BD, Cat. No. 570294, clone O4-46, AB_3685647); Gata-3 Monoclonal Antibody, PerCP-eFluor 710 (eBioscience, Cat. No. 46-9966-42, clone TWAJ, AB_10804487); RB705 Mouse Anti-Mouse RORγt (BD, Cat. No. 570259, clone Q31-378, AB_3685620); PE-CF594 Mouse Anti-Bcl-6 (BD, Cat. No. 562401, clone K112-91, AB_11152084, 1:200); FOXP3 Monoclonal Antibody, FITC (eBioscience, Cat. No. 11-5773-82, clone FJK-16s, AB_465243); FOXP3 Monoclonal Antibody, PerCP-eFluor 710 (eBioscience, Cat. No. 46-5773-82, clone FJK-16s, AB_2811810); APC anti-mouse Perforin Antibody (Biolegend, Cat. No. 154304, clone S16009A, AB_2721463); Brilliant Violet 785 anti-human/mouse Granzyme B Recombinant Antibody (Biolegend, Cat. No. 396438, clone QA18A02, AB_3106140); PE/Dazzle 594 anti-human/mouse Granzyme B Recombinant Antibody (Biolegend, Cat. No. 372216, clone QA16A02, AB_2728383); Alexa Fluor 700 Rat Anti-Mouse IFN-γ (BD, Cat. No. 557998, clone XMG1.2, AB_396979); Biotin Rat Anti-Mouse CD185 (CXCR5) (BD, Cat. No. 551960, clone 2G8, AB_394301, 1:50); PE/Cyanine7 Streptavidin (Biolegend, Cat. No. 405206, 1:200).

Validation

All antibodies are from commercial sources and have been validated by the manufacturers. Validation data are available on the manufacturer website.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	293T cells were from ATCC; P815 cells from Sunncell.
Authentication	No specific authentication was performed.
Mycoplasma contamination	Cell lines were not checked for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	In this study, we employed 6- to 12-week old male and female mice for the experiment. C57BL/6J (Cat. No. 000664), CD45.1 (Cat. No. 002014), Rag1−/− (Cat. No. 002216), Tbx21−/− mice (Cat. No. 004648), Stat6−/− (Cat. No. 005977), Cd4cre (Cat. No. 022071), Vav1cre (Cat. No. 008610), BLIMP1-EYFP (Cat. No. 008828) and OVA323–339-specific T cell receptor transgenic OT-II (Cat. No. 004194) mice were originally from the Jackson Laboratory. Gzmk−/− mice (Cat. No. T029683) were originally from GemPharmatech. Runx3fl/fl mice were provided by Dr. Yingli Sun (Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing) and crossed with Vav1cre mice to generate Vav1cre Runx3fl/fl mice. Gzmk-2A-CreERT2-2A-tdTomato mice (referred to as Gzmk-tdTomato, Cat. No. NM-KI-241127) were purchased from Shanghai Model Organisms Center. Eomesfl/fl, Rorcfl/fl and Bcl6fl/fl mice were crossed with Cd4cre mice to generate T cell–specific knockout lines. All animals were housed under specific pathogen-free (SPF) conditions at the animal facility of Westlake University and handled in accordance with the guidelines of the Institutional Animal Care and Use Committee of Westlake University.
Wild animals	No wild animals used.
Reporting on sex	Both male and female mice were used in experiment.
Field-collected samples	No field-collected samples used.
Ethics oversight	All the animal experiments were performed according to the guidelines of the Institutional Animal Care and Use Committee of Westlake University.

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

Plants

Seed stocks	No plants were used in this study.
Novel plant genotypes	No plants were used in this study.
Authentication	No plants were used in this study.

Flow Cytometry

Plots

Confirm that:

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

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

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

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

Methodology

Sample preparation	Before surface staining, cells were stained with Fixable Viability Dye eF506 (eBioscience, Cat. No. 65-0866-18) and blocked with CD16/CD32 antibodies (BD, Cat. No. 553142, clone 2.4G2, AB_394656) to prevent nonspecific binding. Cells were first stained with various surface markers, then fixed and permeabilized using the FOXP3/Transcription Factor Staining Buffer Set (eBioscience, Cat. No. 00-5523-00) for intracellular staining. If fluorescent protein signals were to be detected, cells were fixed with 2% paraformaldehyde for 20 minutes prior to permeabilization to preserve fluorescence integrity. For intracellular cytokine detection, cells were re-stimulated with phorbol 12-myristate 13-acetate (PMA, 50 ng/ml, Sigma-Aldrich, Cat. No. P8139) and ionomycin (500 ng/ml, Sigma-Aldrich, Cat. No. I0634) in the presence of GolgiStop (BD, Cat. No. 554724) for 4.5 hours.
Instrument	BD LSRFortessa; BD FACSAria Fusion.
Software	Flow cytometry data were collected using FACSDiva software and analyzed with FlowJo software (version 10.9.0).
Cell population abundance	Post-sort fractions were >98% pure, as confirmed by re-analysis on the flow cytometer.
Gating strategy	Initial gating strategy: Lymphocytes were selected based on FSC-A vs SSC-A. Doublets were excluded using FSC-W vs FSC-H and SSC-W vs SSC-H. Live CD4+ T cells were gated as Fixable Viability Dye–, CD4+, TCRβ+. Subsequent gating steps are described in the figures and figure legends.

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