



Transcriptome dynamics of CD4⁺ T cells during malaria maps gradual transit from effector to memory

Megan S. F. Soon^{1,12}, Hyun Jae Lee^{1,2,12}, Jessica A. Engel^{1,12}, Jasmin Straube¹, Bryce S. Thomas¹, Clara P. S. Pernold¹, Lachlan S. Clarke¹, Pawat Laohamonthonkul¹, Rohit N. Haldar¹, Cameron G. Williams^{1,2}, Lianne I. M. Lansink^{1,2}, Marcela L. Moreira², Michael Bramhall², Lambros T. Koufariotis¹, Scott Wood¹, Xi Chen^{3,4,5}, Kylie R. James^{4,5}, Tapio Lönnberg⁶, Steven W. Lane¹, Gabrielle T. Belz^{7,8,9}, Christian R. Engwerda¹, David S. Khoury¹⁰, Miles P. Davenport¹⁰, Valentine Svensson¹¹, Sarah A. Teichmann^{4,5,13} ✉ and Ashrafal Haque^{1,2,13} ✉

The dynamics of CD4⁺ T cell memory development remain to be examined at genome scale. In malaria-endemic regions, anti-malarial chemoprevention protects long after its cessation and associates with effects on CD4⁺ T cells. We applied single-cell RNA sequencing and computational modelling to track memory development during *Plasmodium* infection and treatment. In the absence of central memory precursors, two trajectories developed as T helper 1 (T_{H1}) and follicular helper T (T_{FH}) transcriptomes contracted and partially coalesced over three weeks. Progeny of single clones populated T_{H1} and T_{FH} trajectories, and fate-mapping suggested that there was minimal lineage plasticity. Relationships between T_{FH} and central memory were revealed, with antimalarials modulating these responses and boosting T_{H1} recall. Finally, single-cell epigenomics confirmed that heterogeneity among effectors was partially reset in memory. Thus, the effector-to-memory transition in CD4⁺ T cells is gradual during malaria and is modulated by antiparasitic drugs. Graphical user interfaces are presented for examining gene-expression dynamics and gene-gene correlations (http://haquelab.mdhs.unimelb.edu.au/cd4_memory/).

CD4⁺ TCRαβ⁺ T cells confer immunity to bacteria, viruses, fungi and protozoan parasites. They are repositories of immunological memory, facilitating secondary responses both faster than and of larger magnitude than primary infection. CD4⁺ T cell memory has been studied using major histocompatibility complex (MHC) tetramers and T cell antigen receptor (TCR) transgenic T cells, mostly in bacterial and viral infections^{1–5}, and less so in parasitic infection⁶. Previous studies have employed preselected markers, including fate-mapping of cytokine or chemokine genes^{1,2}, adoptive transfer of cell-sorted effectors^{3,4} and single-cell transfer and longitudinal sampling⁵. These indicate that a single naive CD4⁺ T cell generates memory clones that partly resemble their effector counterparts. This can be explained by a linear model wherein naive cells necessarily give rise to effector cells before transit to memory, or a branching model, wherein memory cells develop before or in parallel with effectors. Single-cell RNA sequencing (scRNA-seq) has recently revealed CD4⁺ central memory precursors during the first week of viral infection⁷, consistent with a branching model. In humans, longitudinal assessment after vaccination has indicated

that effectors gave rise directly to circulating memory cells^{8,9}. Given the class of infectious agent can influence the response of a CD4⁺ T cell clone¹⁰, it remains difficult to extrapolate from previous studies to CD4⁺ T cells in a parasitic infection.

Malaria remains a threat to human health, with 228 million cases and 405,000 deaths in 2018 (ref. ¹¹). Although non-sterilizing immunity can be achieved via natural infection, this takes multiple exposures over many months to achieve¹². In endemic regions, antimalarial chemoprevention can improve protection for at least one year after cessation of drug treatment^{13,14}. Mechanisms explaining this phenomenon remain unresolved, although qualitative changes in memory CD4⁺ T cells have been reported¹⁵. In mice, immune-checkpoint blockade reduced CD4⁺ T cell exhaustion and improved immunity¹⁶, suggesting that parasitic infection impairs, but does not delete, CD4-dependent cellular immunity. Defining molecular pathways that promote effective development of CD4⁺ T cell memory may reveal new strategies for improving immunity to malaria, either during natural exposure and chemoprevention or via vaccination.

¹QIMR Berghofer Medical Research Institute, Herston, Brisbane, Queensland, Australia. ²Department of Microbiology and Immunology, University of Melbourne, located at The Peter Doherty Institute for Infection and Immunity, Parkville, Victoria, Australia. ³Department of Biology, South University of Science and Technology of China, Shenzhen, China. ⁴Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK. ⁵Cavendish Laboratory, University of Cambridge, Cambridge, UK. ⁶Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland. ⁷The University of Queensland Diamantina Institute, The University of Queensland, Brisbane, Queensland, Australia. ⁸Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia. ⁹Department of Medical Biology, University of Melbourne, Parkville, Australia. ¹⁰Kirby Institute, University of New South Wales, Kensington, Sydney, New South Wales, Australia. ¹¹Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. ¹²These authors contributed equally: Megan S. F. Soon, Hyun Jae Lee, Jessica A. Engel. ¹³These authors jointly supervised this work: Sarah A. Teichmann, Ashrafal Haque. ✉e-mail: st9@sanger.ac.uk; ashrafal.haque@unimelb.edu.au

Although CD4⁺ T cells protect against *Plasmodium* parasites¹⁷, genome-scale characterization of the dynamics of memory development remains to be performed in malaria or other infection models. We previously employed TCR transgenic CD4⁺ T cells specific for a *Plasmodium* epitope (PbTII cells), in conjunction with scRNA-seq and computational modeling, to reveal mechanisms underlying T_{H1} and T_{FH} fate bifurcation during the first week of experimental malaria¹⁸. From hundreds of single-cell transcriptomes, T_{H1} and T_{FH} effectors emerged from a proliferating, intermediate state, the balance of which was externally influenced by monocytes or B cells. Here, we extend our previous work by studying thousands of parasite-specific CD4⁺ T cells over four weeks in the spleen using two complementary scRNA-seq platforms, single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq), computational modelling and in vivo validation. We define, at genome scale, the dynamics underlying CD4⁺ T cell memory development during experimental malaria and drug treatment, and provide user-friendly online resources to facilitate research in T cell biology.

Results

PbTIIs exhibit memory phenotypes boosted by anti-malarial drugs. We transferred naive eGFP⁺ *Plasmodium*-specific PbTIIs¹⁹ into C57BL/6J mice—seeding the spleen of each mouse with 100–500 cells—and infected them with blood-stage *Plasmodium chabaudi* chabaudi AS (PcAS) parasites¹⁷. Parasitemia peaked by day 7, followed by low-level parasitemia (Fig. 1a) and splenomegaly (Fig. 1b) for 2–3 months. We administered sodium artesunate or control saline at multiple time-points from day 7, referred to as intermittent antimalarial drug treatment (IAT).

PcAS infection triggers CD4⁺ T cell responses that largely comprise T_{H1} and T_{FH} cells^{20,21} (definitions for CD4⁺ T cell states are provided in Table 1). Splenic PbTIIs peaked by day 7, remaining detectable well beyond the first month, although IAT reduced these and splenomegaly (Fig. 1b,c and Supplementary Information Fig. 1a). PbTIIs expressed CD62L and CCR7 heterogeneously during the first month, which was modulated by IAT (Fig. 1d), suggesting central memory (T_{CM}) and effector memory (T_{EM}) PbTIIs had developed. Bifurcated expression of T_{H1}-associated C-X-C motif chemokine receptor 6 (CXCR6) and T_{FH}-associated CXCR5 was partially retained, and was enhanced by IAT at day 60. (Extended Data Fig. 1a). PbTIIs were located in splenic T cell, B cell and red pulp zones at day 7 (Extended Data Fig. 1b), and in GCs and B cell zones at later times (Extended Data Fig. 1c,d). Thus, PbTIIs entered GCs and persisted long enough to become memory cells.

We next determined whether PbTIIs exhibited recall responses. Direct ex vivo interferon- γ (IFN- γ) production progressively subsided in both groups by day 28, consistent with waning of primary T_{H1} responses (Fig. 1e). In contrast, in vitro T_{H1} recall was substantially enhanced on a per-cell basis by IAT (Fig. 1e), despite reductions in PbTII numbers (Fig. 1c), and those capable of expressing IFN- γ or interleukin-10 (IL-10) (Extended Data Fig. 2a,b). The potential for PbTIIs to mount T_{H1} and T_{FH} recall responses was examined by bulk ATAC-seq. Inaccessible chromatin in naive PbTIIs became accessible at day 7, and remained partially accessible at day 28 with or without IAT (Extended Data Fig. 2c). This was evident for specific T_{H1} or T_{FH}-associated genes *Ifng*, *Cxcr5*, T-box transcription factor 21 (*Tbx21*) and *Il21* (Extended Data Fig. 2d). To test recall in vivo, we performed homologous high-dose rechallenge, comparing day 28, antigen-experienced PbTIIs, with naive comparators transferred before rechallenge (Supplementary Fig. 1b). PbTIIs primed 28 days previously exhibited greater direct ex vivo IFN- γ production than that of naive PbTIIs (Fig. 1f), with reduced proliferative and early-activation markers (Extended Data Fig. 2e,f). Thus, PbTIIs had acquired memory phenotypes, with IAT boosting T_{H1} memory (Fig. 1f).

Central memory precursors are not essential for CD4⁺ T cell memory development. We next examined the mechanisms underlying memory development. Despite reporting only two fates in our previous study¹⁸, we hypothesized that rare memory precursors existed. To test this, we assessed thousands of splenic effector transcriptomes. PbTIIs at day 7 (10,251 cells following quality control) had bifurcated, exhibiting reported T_{H1} and T_{FH} gene signatures¹⁸, and a smaller group of proliferating cells expressing T_{H1} or T_{FH} features (Fig. 2a–c). Consistent with our previous study, the minor cluster was likely undergoing final cell division and fate bifurcation¹⁸. Most importantly, neither a recently published CD4⁺ T_{CM} precursor gene signature⁷ (Fig. 2d) nor CCR7 expression (Fig. 2e) supported emergence of CD4⁺ T_{CM} precursors at peak infection.

T_{H1} and T_{FH} effectors exhibit gradual transit towards memory.

We next hypothesized that memory developed directly from effectors. However, tracking this process using single-gene approaches was difficult since T_{H1} and T_{FH} phenotypic mixing is reported in *Plasmodium* infection²². scRNA-seq revealed that neither *Ifng*, *Tbx21* nor *Cxcr3* expression was confined to T_{H1} cells (Extended Data Fig. 3). Similarly, selectin p ligand (*Selplg*) (encoding the adhesion molecule PSGL1) and lymphocyte antigen 6 complex, locus C2 (*Ly6c2*) (encoding Ly6C) were expressed in both fates at peak (Extended Data Fig. 3). *Cxcr6* and *Cxcr5* were reasonably well-confined to their respective fates (Fig. 2b), although no single gene faithfully marked all T_{H1} or T_{FH} cells. Therefore, we applied scRNA-seq profiling over time to avoid pre-selected markers, to capture intermediate states and to examine dynamics at genome scale. As before, we transferred PbTIIs (Extended Data Fig. 4a), infected with PcAS, and administered IAT or saline from day 7 (Supplementary Information Fig. 1c). At various timepoints, splenic PbTIIs were sorted and processed via Smart-seq2 scRNA-seq (Extended Data Fig. 4b). From 4,548 wells, high-quality transcriptomes were obtained for 2,964 cells (Extended Data Fig. 4c). To study transcriptomes from naivety to memory, we integrated this Smart-seq2 dataset with our previous PbTII datasets¹⁸, resulting in 3,728 transcriptomes spanning 4 weeks of infection with and without IAT (Extended Data Fig. 4d). Principal-component analysis (PCA) and uniform manifold approximation and projection (UMAP) revealed, as before, 2 effector populations emerging by day 7, expressing *Cxcr5* or *Cxcr6* (Extended Data Fig. 5a). UMAP suggested that there was progressive transcriptomic change over time, with two trajectories being more apparent during IAT than during saline treatment (Extended Data Fig. 5a). Unsupervised trajectory inference using Slingshot²³ was unable to map trajectories that adhered to timepoint information (Extended Data Fig. 5a). Estimation of RNA velocity²⁴ indicated that, although change was rapid and trajectory inference was possible around fate bifurcation, subsequent timepoints exhibited slower rates of change, making inference of trajectory difficult (Extended Data Fig. 5b). In any case, our integrated scRNA-seq dataset supported the presence of intermediate transcriptomic states between effector and memory states, particularly under IAT, suggesting that there was a progressive effector-to-memory transition over several weeks.

Temporal mixture modelling of effector-to-memory transitions.

We next generated a probabilistic transcriptomic model to map gene-expression dynamics with or without IAT. We employed GPfates¹⁸, which involved Bayesian Gaussian process latent variable modeling (bGPLVM), calculation of pseudotime from resulting latent variables, and finally overlapping mixtures of Gaussian processes (OMGP), run on IAT and control saline groups separately—including data from day 0–7 in both datasets (Fig. 3a). Resulting models resembled incomplete circles in two dimensions, with T_{H1} and T_{FH} lineages coalescing and partially returning towards naivety

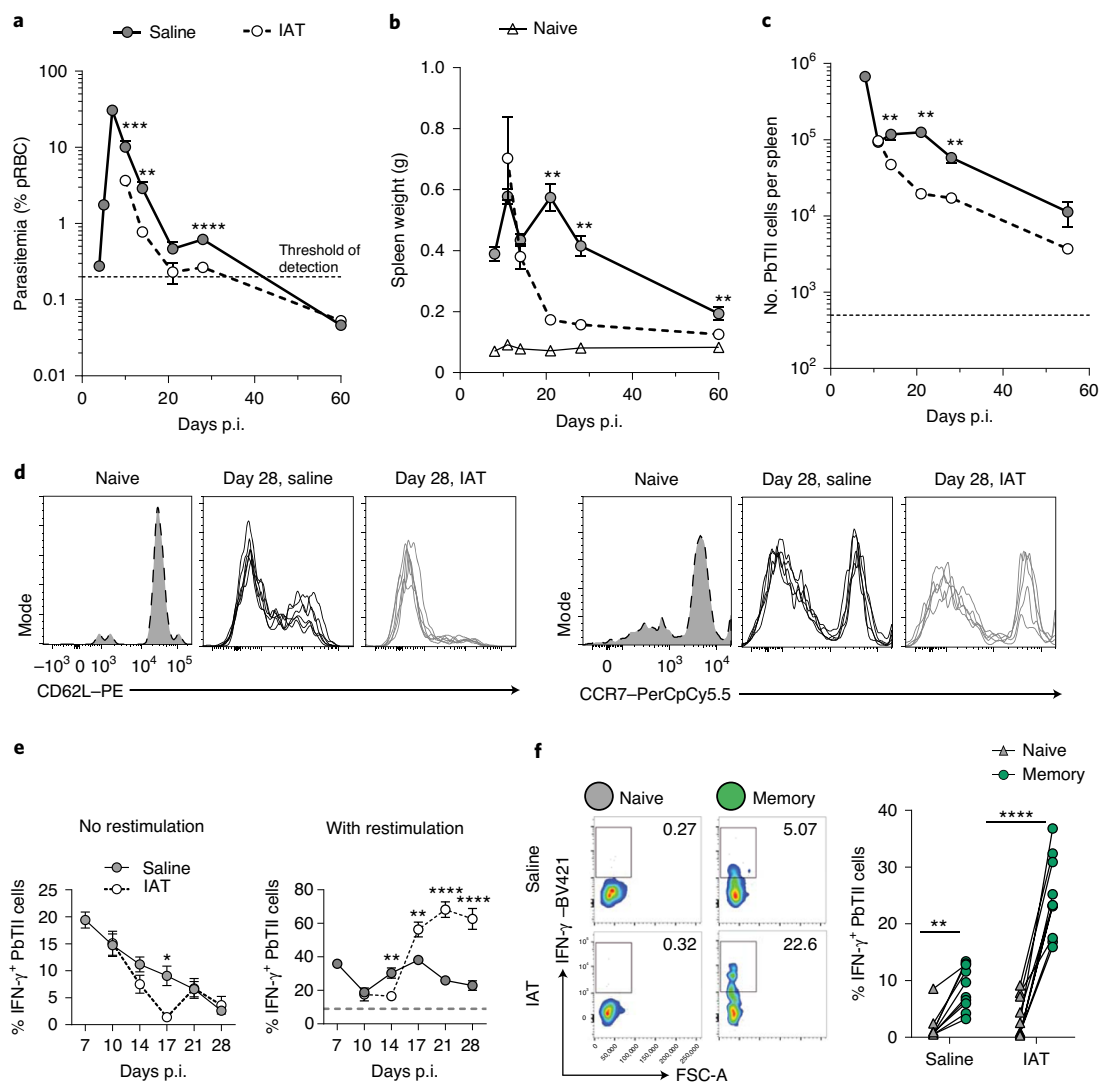


Fig. 1 | *Plasmodium*-specific TCR transgenic PbTII cells develop memory during infection. a, Parasitemia during *PcAS* infection in C57BL/6J mice in the presence or absence (saline-treated control group) of IAT. Data are pooled from 3 independent experiments ($n=5$ mice per group, per independent experiment). Statistical testing was performed using two-tailed Mann-Whitney test. Days p.i., days postinfection; pRBC, parasitized red blood cells. **b**, Spleen weight during *PcAS* infection in the presence or absence of IAT. Data are representative of 3 independent experiments ($n=5$ mice group, per individual timepoint). Statistical testing was performed using two-tailed Mann-Whitney test. **c**, Splenic PbTII cell numbers during *PcAS* infection in the presence or absence of IAT. Data are pooled from 2 independent experiments ($n=5$ mice per group, per individual timepoint for each independent experiment). Statistical testing was performed using two-tailed Mann-Whitney test. **d**, Histograms of flow-cytometric assessment for surface CD62L and CCR7 expression on PbTII cells. A representative histogram is presented for naive controls, and overlaid histograms ($n=5$ mice per group) are presented for day 28 (D28) saline and IAT groups. Data are representative of two independent experiments. **e**, Left, analysis of direct ex vivo IFN- γ production by PbTII cells without restimulation. Right, analysis of ex vivo IFN- γ production by PbTII cells after PMA and ionomycin restimulation *in vitro*. The dashed gray line represents the threshold of IFN- γ production by naive PbTII cells. Data are pooled from 2 independent experiments ($n=5$ mice per group, per individual timepoint for each independent experiment). Statistical analysis performed using two-tailed Mann-Whitney test. **f**, Representative fluorescence-activated cell sorting (FACS) plots showing direct ex vivo production of IFN- γ at 17 hours following rechallenge for naive (gray) or memory (green) PbTII cells in saline and IAT groups. The graph compares IFN- γ production between naive or memory PbTII cells in saline or IAT groups. Data are pooled from 2 independent experiments ($n=5$ mice per group, per independent experiment). Statistical analysis was performed using paired two-way analysis of variance with Tukey's multiple-comparisons test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. Data are presented as mean \pm s.e.m. (**a,b,c,e**). Statistical analysis was performed between saline and IAT groups for each timepoint (**a,b,c,e**).

(Fig. 3a). Our ability to assign effectors to either lineage diminished along pseudotime (Fig. 3b). This was more apparent for saline than for IAT (Fig. 3b). Also of note, the T_{H} lineage tracked closer to naive cells (Fig. 3a).

Gene numbers increased from $\sim 2,000$ at naivety to $\sim 5,000$ during clonal expansion, $\sim 4,000$ in effectors and $\sim 2,000$ by the end of pseudotime, a pattern mirrored by cell-cycling genes (Fig. 3c). IAT

did not alter the decline in detected genes over pseudotime, but did reduce expression of exhaustion-associated genes (Fig. 3d). To validate our GPfates models, we employed single-cell variational inference²⁵, which recapitulated observations from PCA and GPfates, including that *Cxcr5*- and *Cxcr6*-expressing effectors appeared to gradually change over three weeks (Extended Data Fig. 5c). Thus, computational modeling suggested that effectors gradually

Table 1 | T cell terminology

	Definition
Naive CD4 ⁺ T cell	A mature CD4 ⁺ T cell yet to encounter cognate antigen presented in the context of MHCII, expressing CCR7 and CD62L
Effector CD4 ⁺ T cell	A CD4 ⁺ T cell exhibiting helper functions such as cytokine production after being activated by cognate antigen and having progressed through clonal expansion
Memory CD4 ⁺ T cell	A CD4 ⁺ T cell previously activated and clonally expanded via cognate antigen, which although quiescent, retains a capacity to respond to a second round of activation with cognate antigen
T _{H1} cell	Effector CD4 ⁺ T cell expressing high levels of T-bet and IFN-γ, also marked by cell-surface expression of CXCR6
T _{FH} cell	Effector CD4 ⁺ T cell expressing high levels of BCL6 and CXCR5, as well as ICOS, with a capacity to leave T cell zones and move to B cell follicles in secondary lymphoid tissue
GC T _{FH} cell	Effector CD4 ⁺ T cell expressing high levels of BCL6, CXCR5 and PD1, with a capacity to interact with B cells in the germinal center
T _{R1} cell	Effector CD4 ⁺ T cell coexpressing IFN-γ and IL-10
T _{CM} cell	Antigen-experienced, quiescent T cell that expresses CCR7 (with or without CD62L)
T _{EM} cell	Antigen-experienced, quiescent T cell that rapidly expresses cytokine and/or specific T helper phenotypes during recall

transitioned to various quiescent cellular states that partially resembled naivety.

To facilitate interrogation of our GPfates models, we present an online graphical user interface (GUI) (Fig. 3e): http://haquelab.mdhs.unimelb.edu.au/cd4_memory/. The ‘Dynamic Model’ GUI enables visualization of expression dynamics for any gene in CD4⁺ T cells during persistent infection or IAT. We also present a gene-gene correlation function, allowing testing for coexpression of two genes in the same cell.

Transcriptome dynamics reveal immune signatures retained in memory. To discover genes associated with memory development, we categorized ~15,000 genes within our GPfates models according to expression dynamics along pseudotime (without consideration of T_{H1}–T_{FH} bifurcation). Using SpatialDE²⁶, seven distinct dynamics were identified (Fig. 4a,b). The shape of each was similar between IAT and saline groups, with minor differences, for example, in dynamic 1 (Fig. 4c). Gene Ontology (GO) analysis was conducted within each dynamic (Fig. 4b). Consistent with our previous study¹⁸, dynamic 6, associated with DNA replication, peaked before bifurcation and faded as effector phenotypes emerged. Dynamic 5 was more prolonged and was associated with energy metabolism, including ATPase and electron-carrier activity. Dynamic 7, featuring a late drop in gene expression, was overwhelmingly composed of ribosomal genes. Dynamic 1 was composed of relatively few genes (511 of 14,167 for IAT; 463 of 15,310 for saline) that peaked late and remained elevated at the end of pseudotime. GO analysis indicated that there were strong associations with immune processes, including cytokine signaling, TNF-receptor signaling, antigen binding and targeting by the immune-associated E3 ubiquitin ligase, TRAF6. Therefore, only 3–4% of genes exhibited a late peak and preserved expression, and were largely associated with immune processes.

Correlation and gene-network analysis of dynamic 1 revealed stronger positive correlations between genes under IAT than under

saline (Fig. 4d). Resulting networks suggested a linear ‘axis’ under IAT, and a single hub of weaker connections under saline (Fig. 4d,e). Several transcriptional regulators were distributed along the IAT axis with a graded T_{H1}–T_{FH} structure (Fig. 4d,e), with *Bcl6* and *Tox2* at one end, and T_{H1}-associated *Id2*, *Tbx21* and *Runx2* at the other (Fig. 4e). Foci for immune checkpoints (*Tigit*, *Lag3*, *Pdcd1*, *Ctla4*) and Tr1 (*Il10*, *Prdm1*, *Bhlhe40* (refs. 27,28)) were also noted. Thus, similar genes were detected in dynamic 1 for IAT and saline, but gene-gene correlations were stronger and T_{H1}, Tr1, exhaustion and T_{FH} groupings were clearer in IAT.

The existence of T_{H1} and T_{FH} networks under IAT motivated examination of genes according to the strength of T_{H1}–T_{FH} bifurcation. All genes were scored for pseudotime correlation and bifurcation to either lineage (Fig. 4f). Some genes exhibited moderate pseudotime correlation and strong T_{H1} or T_{FH} assignment, including *Cxcr6*, *Ccr2*, *Ccr5*, *Tox2*, *Cxcr5* and *Bcl6*. Among strongly T_{H1}-bifurcating genes, some correlated strongly with pseudotime, including *Nkg7*, *S100a4*, *S100a6* and *Ccl5* (Fig. 4f). In contrast, only *Folr4* (encoding folate receptor 4) was T_{FH} bifurcating and highly correlated with pseudotime. Instead, many T_{FH}-associated genes correlated with early pseudotime, including *Tcf7*, highlighting their expression in naivety. Finally, some genes correlated with late pseudotime with little bifurcation, including transcription factors *Id2* and *Maf*, exhaustion-related genes *Tigit* and *Lag3* (in saline controls more so than under IAT) and *Cxcr3*.

To test predictions from our models (Extended Data Fig. 6a), we examined protein expression of various markers, including distribution among lineages. ID2 was strongly upregulated at day 7, but was not substantially retained at the protein level by day 28 (Extended Data Fig. 6b). Reciprocally related TCF1 (encoded by *Tcf7*), was heterogeneously downregulated in effectors, and partly recovered in some memory cells during IAT (Extended Data Fig. 6b). c-Maf was upregulated at day 7, and partially retained at day 28 with or without IAT (Extended Data Fig. 6c). CCL5 was absent from CXCR5⁺ effector and memory PbTIIs, and was associated instead with CXCR6⁺ memory PbTIIs (Extended Data Fig. 6d). CXCR3 was broadly expressed on CXCR5⁺ and CXCR6⁺ cells at day 28 under IAT (Extended Data Fig. 6e).

Id3, a transcription factor related to ID2, appeared to be confined to the T_{FH}-associated lineage (Extended Data Fig. 6a). We confirmed this using *Id3*^{GFP} PbTIIs, with GFP expression in CXCR5⁺ but not CXCR6⁺ PbTIIs at day 28 (Extended Data Fig. 6f). Thus, *Id3* and *Tcf7* were retained along the T_{FH} lineage. To confirm *Tcf7* promoted the T_{FH} lineage, we examined T_{H1}–T_{FH} fate in Cre^{ERT2}*Tcf*^{fl/fl} PbTIIs, compared with those from wild-type littermates (Extended Data Fig. 6g). TCF1-deficient PbTII effectors were defective in BCL6 and CXCR5 expression, and instead expressed more CXCR6, ID2 and T-bet (Extended Data Fig. 6h), confirming TCF1 promoted the T_{FH} lineage in our model.

Memory fate is determined during the first week of infection.

To examine memory fate in single naive PbTIIs, we employed the diverse endogenous TCRαβ sequences in Rag-sufficient PbTIIs as barcodes¹⁸. From 2,964 transcriptomes across all timepoints and conditions, where endogenous VDJ regions were reconstructed²⁹, we detected 201 families (Fig. 5a,b), defined as sharing the same endogenous sequences. Families ranged from two to five cells (Fig. 5b), and did not cross timepoints or conditions (Extended Data Fig. 7a), because individual mice were used for each, and PbTIIs in separate mice harbor unique endogenous barcodes. Cells from one family frequently occupied both T_{H1} and T_{FH} lineages (Fig. 5c), confirming single naive PbTIIs displayed heterogeneity in the effector and memory fates of their progeny.

We tested whether lineage choice was random within a family. We examined cells under IAT only, where binary T_{H1}–T_{FH} fates were more easily discerned than in saline controls (Fig. 5c). After

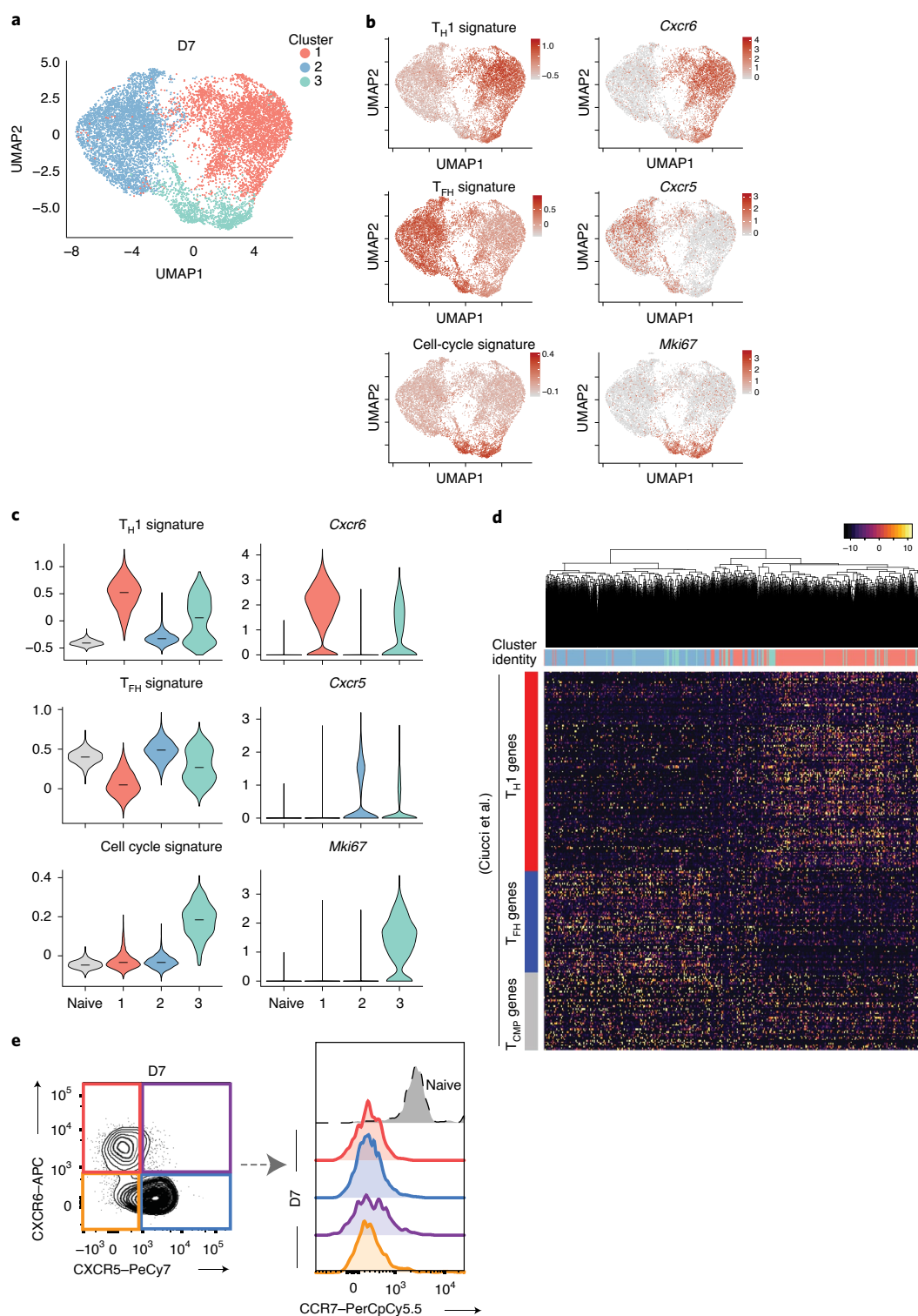


Fig. 2 | A lack of PbTII central memory precursors at peak infection. **a**, Droplet-based scRNA-seq analysis of 10,251 PbTII cells at peak effector stage (D7 postinfection) during *PcAS* infection using the 10x Genomics Chromium platform. Unsupervised clustering was performed on a UMAP representation of PbTII cells at D7 p.i., calculated using principal components (PC) 1–10 with 1,522 highly variable genes as input. **b**, UMAP visualization showing a T_{H1} , T_{FH} and cell-cycle signature score and genes (*Cxcr6*, *Cxcr5* and *Mki67*) associated with each signature on PbTII cells at D7 p.i. **c**, Violin plots of T_{H1} , T_{FH} and cell-cycle signature scores and genes (*Cxcr6*, *Cxcr5* and *Mki67*) associated with each signature on PbTII cells at D7 p.i. The expression value for naive (D0) PbTII cells is shown for each signature and gene for reference. Median expression for each cluster is denoted for gene signature scores as a horizontal line. **d**, Heatmap with unsupervised hierarchical clustering showing the expression of T_{H1} , T_{FH} and T central memory precursor (T_{CMP}) gene signatures as defined by Ciucci et al.⁷ for PbTII cells at D7 p.i. **e**, Representative FACS plot of PbTII cells at D7 p.i. showing the surface expression of CXCR6 and CXCR5. PbTII cells were put in subsets as follows: CXCR6⁺CXCR5⁻ (red), CXCR6⁺CXCR5⁺ (purple), CXCR6⁻CXCR5⁺ (blue) and CXCR6⁻CXCR5⁻ (orange). Representative histograms showing surface CCR7 expression for the different subsets of PbTII cells at D7 p.i. and naive PbTII cells (gray). Data are representative of 2 independent experiments ($n = 6$ mice per group, per independent experiment).

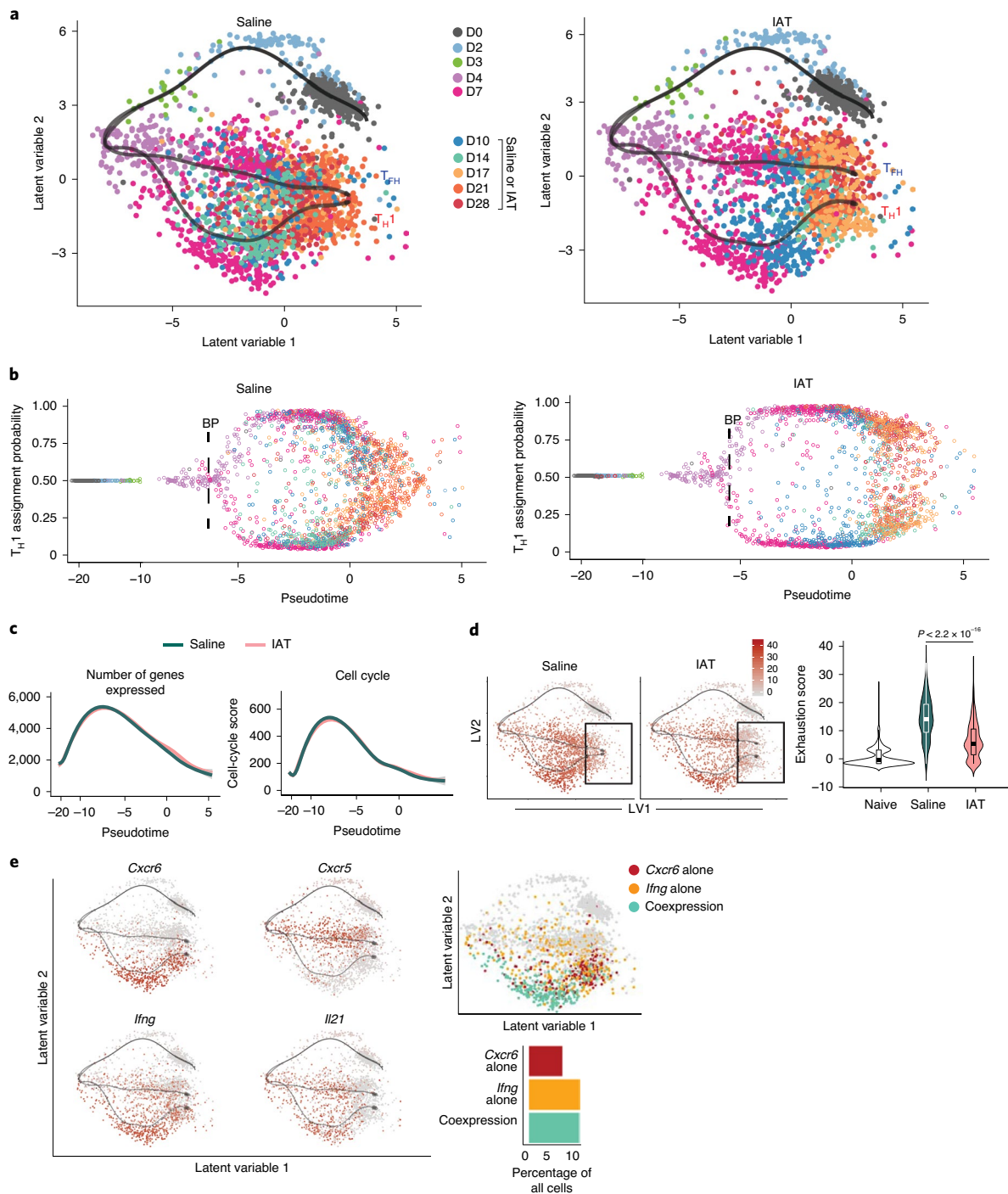


Fig. 3 | Temporal mixture modelling of scRNA-seq data maps effector-to-memory transit. **a**, Reconstruction of trajectories along pseudotime using GPfates[®]. bGPLVM analysis was performed on the full time series from the combined batch-corrected PbTII datasets (first batch: Smart-seq2 (96), D0, D4 and D7 p.i.; second batch: SMARTer C1, D0–D7 p.i.; third batch: Smart-seq2 (384), D0–D28 p.i.) and 2D bGPLVM representations were plotted for each treatment group (saline, left; IAT, right). D0, D2, D3, D4 and D7 p.i. cells were replicated for each representation. OMGP analysis was performed individually for saline and IAT. All Smart-seq2 datasets mentioned from here onwards comprise the batch-corrected full time series containing all three experimental batches, unless otherwise specified. Saline and IAT groups comprise shared timepoints (D0–D7) and treatment-unique timepoints (D10–D28, saline or IAT). **b**, T_H1-assignment probability assigned using OMGP along pseudotime for PbTII cells for saline (left) and IAT (right) groups. BP, T_H1–T_{FH} bifurcation point. **c**, Histogram showing the number of genes expressed (left) and cell-cycle score (right) along pseudotime for the saline and IAT groups. **d**, Left, degree of exhaustion visualized on a 2D representation of bGPLVM. Right, violin plots showing the exhaustion score in naive cells (D0) and cells positioned at late pseudotime (pseudotime > 0.9) for saline and IAT groups (shown in the black box). For the box plot, the center line indicates the median, box limits indicate the upper and lower quartiles and the whiskers indicate the maximum and minimum measures. Statistical analysis was performed using two-sided Wilcoxon rank-sum test between saline and IAT groups. LV, latent variable. **e**, Left, example gene-expression dynamics for T_H1- and T_{FH}-associated genes interrogated on the IAT ‘Dynamic’ dataset from our GUI: http://haquelab.mdhs.unimelb.edu.au/cd4_memory/. Right, example of the optional gene–gene correlation function.

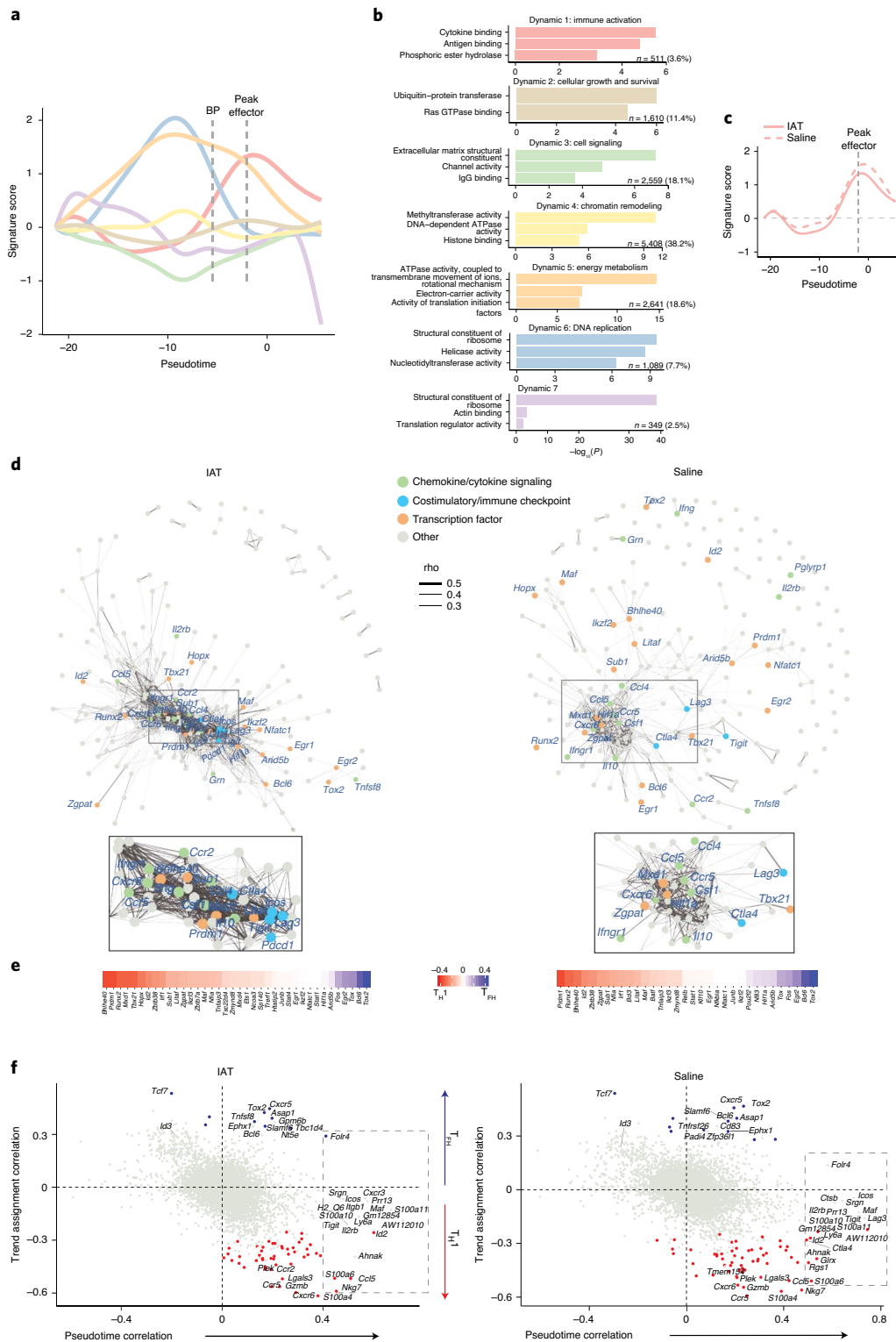


Fig. 4 | Transcriptome dynamics reveal gene signatures that were preserved during memory development. **a**, Genome-wide expression patterns of 14,167 significantly variable genes along pseudotime from the combined batch-corrected Smart-seq2 dataset, grouped according to individual dynamics and represented as signature scores. T_{H1} - T_{FH} BP and peak effector phase (median of D7 p.i. pseudotime values) are depicted to provide temporal context. **b**, Summary of GO enrichment analysis of molecular function associated with genes for each dynamic (dynamics 1–7) as in **a**. The number and proportion of genes for each dynamic are shown. **c**, Expression of signature scores of genes in dynamic 1 between IAT and saline groups along pseudotime. **d**, Coexpression network analysis of genes in dynamic 1 for IAT (left) and saline (right) groups. Edge weight corresponds to Spearman's rho values. **e**, Transcription factors in dynamic 1 for IAT and saline groups. Each transcription factor is scored for its T_{H1} - T_{FH} assignment probability. **f**, Gene-expression correlation with pseudotime (x axis) versus the correlation with T_{H1} trend assignment (y axis) is shown (Pearson correlation). Genes were assessed for the T_{H1} trend (T_{H1} - T_{FH} bifurcation) using GPfates (top 20 T_{H1} -associated, red; top 20 T_{FH} -associated, blue) for each treatment group. The top 20 pseudotime-correlated genes are enclosed within the gray dashed line window.

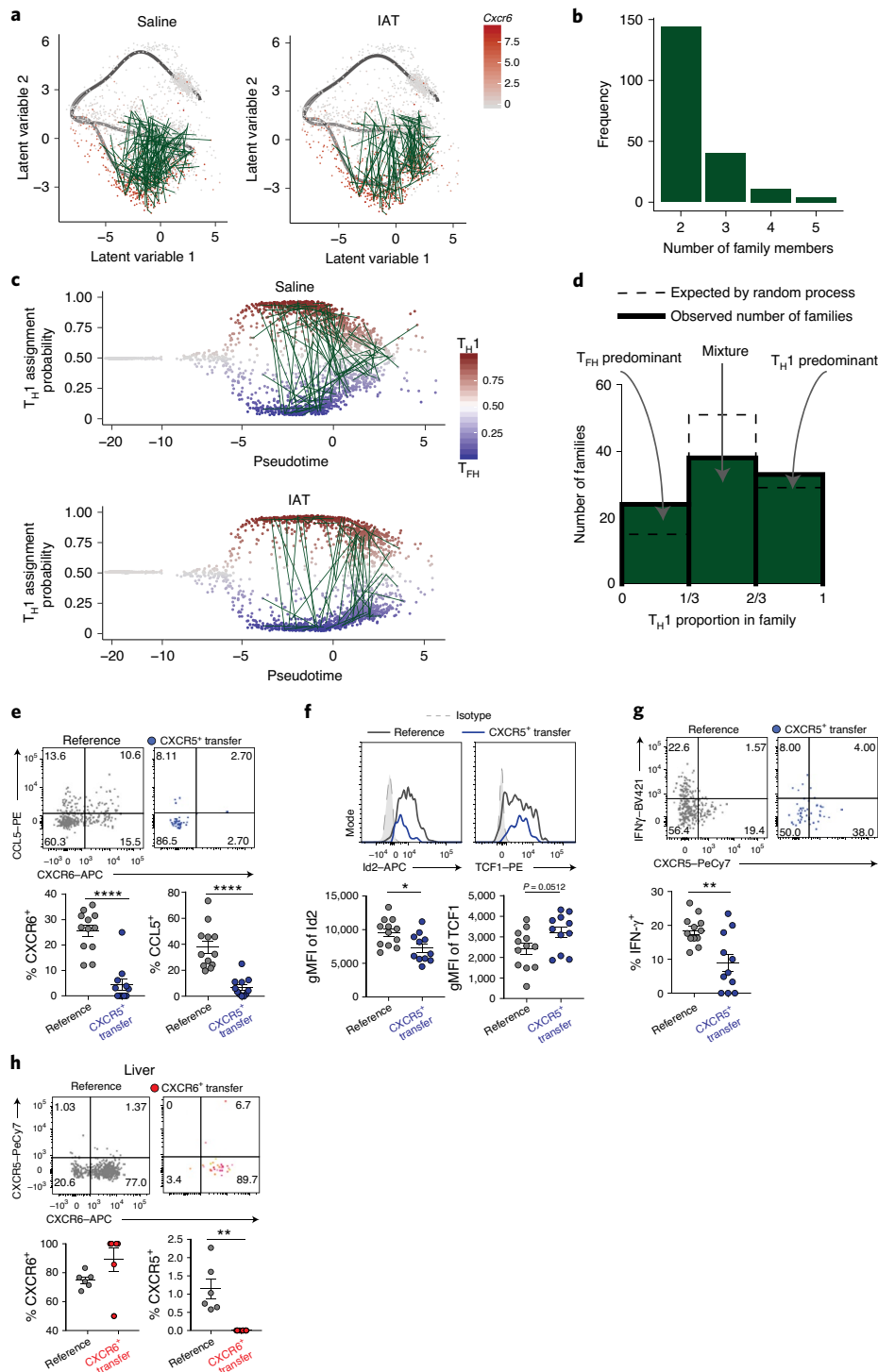


Fig. 5 | Memory fate of naive clones is set during the first week of infection. **a**, Visualization of families sharing endogenous TCR α and TCR β sequences on bGPLVM representation, with each family connected by green lines. Cells are colored by the level of *Cxcr6* expression. **b**, Bar graph showing the frequency of families for each family size. **c**, Visualization of families on a T_{H1} assignment probability along pseudotime. **d**, Bar graph showing the observed number of families (solid green bar) that are (from left to right): T_{FH} predominant, mixture of T_{FH} and T_{H1} , and T_{H1} predominant. The dotted line bars represent the expected number of families by random process as described in the Methods. **e–g**, Representative FACS plots and graphs showing the proportion or geometric mean fluorescence intensity (gMFI) of splenic PbTII cells at D28 p.i. expressing CXCR6 and CCL5 (**e**), Id2 and TCF1 (**f**) and IFN- γ (**g**) for the reference and CXCR5 $^{+}$ transfer groups. Data are pooled from 2 independent experiments (first experiment: $n = 6$ mice per group; second experiment: $n = 6$ mice in the reference group and $n = 5$ mice in the CXCR5 $^{+}$ transfer group). Staining for TCF1 and IFN γ was performed after stimulation with PMA and ionomycin. Statistical analysis was performed using the two-tailed Mann-Whitney test. **h**, FACS plots and graph showing the proportion of PbTII cells in the liver expressing CXCR6 or CXCR5 at D28 p.i. A representative plot is shown for the reference control, while pooled data are shown for the CXCR6 $^{+}$ transfer. Data are representative of 2 independent experiments ($n = 6$ mice in reference group and $n = 5$ mice in the CXCR6 $^{+}$ transfer group for each independent experiment). Statistical analysis was performed using two-tailed Mann-Whitney test. Data are presented as mean \pm s.e.m. (**e–h**). * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$.

excluding cells of indeterminate fate, we examined 97 families. We hypothesized cell fate was independent of family, and therefore distribution of fates within each family would follow a random, binomial distribution (Fig. 5d). Instead, sibling PbTIIIs exhibited greater tendencies towards the same fate than could be explained by random binomial process ($P < 0.0005$, Fig. 5d), the trend being evident also during examination of smaller families, albeit with reduced statistical significance owing to smaller sample sizes ($P = 0.054$ for 67 families of 2; $P = 0.052$ for 19 families of 3). Thus, T_{H1} – T_{FH} lineage fate was not entirely random within a family.

We next hypothesized that splenic effectors remained on their trajectories with no lineage plasticity. To test this, we focused on IAT, where trajectories were better preserved over time. Since *Cxcr5* and *Cxcr6* were the top bifurcating effector genes (Fig. 4e), we sorted $CXCR5^+CXCR6^-$ (T_{FH}) and $CXCR6^+CXCR5^-$ (T_{H1}) PbTIIIs at day 7, transferred them separately into infection-matched mice and administered IAT until day 28; the reference group of control mice harbored PbTIIIs that were unperturbed after transfer, infection and IAT (Supplementary Fig. 1d). At day 28, PbTIIIs derived from $CXCR5^+CXCR6^-$ effectors expressed neither T_{H1} -associated *CXCR6* or *CCL5* (Fig. 5e), high levels of *Id2* (Fig. 5f) nor reduced expression of *TCF1* relative to those in reference controls (Fig. 5f). They also exhibited reduced in vitro IFN- γ recall relative to that in controls (Fig. 5g). PbTIIIs derived from $CXCR6^+CXCR5^-$ effectors were undetectable in spleens, but were found expressing *CXCR6*, not *CXCR5*, in the liver (Fig. 5h). Fate-mapping performed in saline-treated mice elicited similar results, suggesting that IAT had not restricted lineage plasticity (Extended Data Fig. 7b,c). Thus, single naive clones produce effector progeny that progress along T_{H1} and T_{FH} lineages with only a modest clone-specific preference for either; having done so, memory fate is set because little plasticity exists between splenic trajectories.

T_{CM} and Tr1 emerge within T_{FH} and T_{H1} lineages. *CCR7* and *CD62L* mark naive and T_{CM} cells. *Sell* (encoding *CD62L*), after downregulation at priming, was re-expressed along both lineages (Extended Data Fig. 8a). *Ccr7* was retained during clonal expansion and by only T_{FH} lineage cells (Extended Data Fig. 8a,b), supporting the idea that T_{CM} cells develop along the T_{FH} lineage. We hypothesized that T_{CM} cells in lymph nodes could not display a T_{H1} phenotype. Consistent with this, fewer PbTIIIs in inguinal lymph nodes were T_{H1} -like than in spleen (Extended Data Fig. 8c). We next searched for GC T_{FH} cells along the T_{FH} lineage. *Pdcd1* (encoding *PD1*) and *Ccr7* were poorly coexpressed within the T_{FH} lineage (Extended Data Fig. 8d,e), suggesting that if *Pdcd1*⁺*Cxcr5*⁺ GC T_{FH} cells and *Ccr7*⁺ T_{CM} emerged within the T_{FH} lineage, the GPfates model had insufficient power to distinguish them.

Finally, *Ifng* was not confined to the T_{H1} lineage, but was expressed more so in that lineage than in the T_{FH} lineage (Extended Data Fig. 8f,g). In contrast, *Il10* expression was largely confined to the T_{H1} lineage: 34% of T_{H1} lineage cells coexpressed *Ifng* and *Il10* during persistent infection, consistent with progressive Tr1 development from T_{H1} effectors, that was limited by IAT (Extended Data Fig. 8f,g). Thus, on the basis of GPfates modeling, we propose that T_{FH} effectors gave rise to T_{CM} cells, GC T_{FH} cells and potentially T_{FH} memory, while T_{H1} cells gave rise to Tr1 cells and T_{H1} -phenotype T_{EM} cells.

IAT limits exhaustion in both lineages and boosts GC T_{FH} cells. We next determined broader effects of IAT, examining ‘late pseudotime’ in GPfates models to focus on exhaustion and memory. Differential-gene-expression analysis between IAT and saline within either lineage revealed that IAT altered the expression of 1,110 genes (false-discovery rate (FDR) < 0.05) in the T_{H1} lineage (Supplementary Table 1) compared with 252 (FDR < 0.05) for the T_{FH} lineage (Supplementary Table 2). GO analysis revealed IAT increased expression of immune-associated genes in the T_{H1} lineage,

with fewer immune effects on the T_{FH} lineage (Fig. 6a). There were 175 immune genes upregulated in the T_{H1} lineage by IAT, 17 of which were also downregulated in the T_{FH} lineage, including *Il7r*, *Cd96* and *Cd40lg* (Fig. 6a). Fifty-four immune genes were downregulated by IAT in the T_{H1} lineage, 20 of which were also downregulated in the T_{FH} lineage, including *Ctla4*, *Lag3*, *Tigit*, *Icos*, *Il21*, *Cd3d* and *Cd3g*. These data suggest that IAT influenced both lineages, but more so T_{H1} .

Power to detect differentially expressed genes between groups of single cells is influenced by heterogeneity. We reasoned that increased heterogeneity in the T_{FH} lineage had obscured IAT effects, and that distinguishing GC T_{FH} from T_{CM} cells required analysis of more cells. We also questioned whether persisting infection had merged T_{FH} and T_{H1} lineages into a single state. We assessed PbTIIIs from IAT and control groups via droplet-based scRNA-seq at day 28 (3,227 and 3,380 cells, respectively, following quality control), performing unsupervised clustering on the combined datasets (Fig. 6b), available for interrogation as the ‘Memory’ GUI at http://haquelab.mdhs.unimelb.edu.au/cd4_memory/. During persisting infection, PbTIIIs exhibited two main phenotypes (Fig. 6b): T_{FH} -like cluster 1, and T_{H1} -like cluster 3 (Fig. 6c,d). Minor cluster 5 was proliferative based on cell-cycling genes, including marker of proliferation *Ki-67* (*Mki67*) (Fig. 6c,d). Finally, during persisting infection, rare PbTIIIs occupied distal clusters 2, 4, 6 and 7, as well as minor cluster 8 (characterized by type I IFN genes). Separated clusters 1 and 3 suggested that persisting infection had not merged T_{H1} and T_{FH} lineages into a single state. In addition to these, rarer phenotypes had also been generated during persisting infection.

Considering IAT effects, T_{H1} -like clusters 3 and 4 (Fig. 6c,d) were mainly from saline and IAT groups, respectively (Fig. 6b). They were distinct from each other, with cluster 4 being smaller in size, suggesting that IAT had uniformly altered the T_{H1} lineage and reduced heterogeneity. T_{FH} -like clusters 1 and 2 were less distinct, remaining partly adjacent to each other. Most strikingly, in contrast to T_{H1} cells, IAT increased heterogeneity among non- T_{H1} clusters 2, 6 and 7 (Fig. 6b). Nevertheless, IAT reduced expression of exhaustion-related genes in both lineages, including *Lag3* (Fig. 6e and Supplementary Tables 3 and 4), abrogated *Il10* expression in the T_{H1} lineage (Fig. 6e) and boosted *Il7r* expression, all of which are consistent with the GPfates models. Predicted effects of IAT on *Lag-3*, *TIGIT* and *IL-7R* were confirmed at the protein level (Fig. 6f). Collectively, our results revealed that IAT had reduced exhaustion in both lineages, limited Tr1 development, condensed T_{H1} memory phenotypes and increased heterogeneity within the T_{FH} lineage.

To search for GC T_{FH} and T_{CM} cells, we employed small curated gene lists: *Cxcr5*, *Bcl6*, *Il21*, *Pdcd1* and *Icos* for GC Tfh, and *Ccr7* and *Sell* for T_{CM} . Cluster 6 exhibited the strongest GC T_{FH} signature, and cluster 7 the strongest T_{CM} signature (Fig. 6b,c,g). Cluster 6 also appeared to be moderately exhausted, likely owing to *Pdcd1* and *Cd160*, markers common to exhaustion and GC biology (Fig. 6e,g). Clusters 6 and 7 were smaller than the adjacent T_{FH} -like cluster 2, consistent with the GPfates model having missed phenotypes present at $< 5\%$. Importantly, the proportion of T_{FH} -like cells with T_{CM} or GC T_{FH} signatures increased by $\sim 80\%$ under IAT (Fig. 6h), confirmed for GC T_{FH} by flow cytometry (Fig. 6i). Finally, we identified genes differentially expressed between GC T_{FH} and T_{CM} cells (Supplementary Table 5), and those elevated in GC T_{FH} compared with other fates at day 28 (Supplementary Table 6). In addition to T_{FH} -associated genes, *Pdcd1*, *Tox2*, *Cxcr5*, *Id3*, *Tcf7*, *Tox* and *Bcl6*, we noted those previously unstudied, including *Rgs10* and *Ppp1r14b* (Supplementary Table 6). Thus, IAT had improved $CD4^+$ T cell responses by preventing exhaustion and boosting T_{CM} and GC T_{FH} cells.

scATAC-seq reveals effector heterogeneity and partial resetting of chromatin in memory. We hypothesized that lineage convergence and partial reversion to naivety was an artefact of transcriptomics.

We examined chromatin using scATAC-seq in naive (day 0), proliferating (day 4), effector (day 7) and memory (day 32 with IAT or control saline) PbTII cells (Fig. 7a), as well as sorted T_{FH} lineage (CXCR5^{hi}) and T_{H1} lineage (CXCR6^{hi}) effectors (Fig. 7a). Across all cells, we identified 59,197 accessible peaks. Using latent semantic indexing (LSI)³⁰ and PCA, PC2 separated proliferating and effector cells from naive and memory cells, while PC3 separated T_{H1} from T_{FH} effectors (Fig. 7a). Chromatin accessibility was substantially more heterogeneous among effectors than among their naive counterparts. PCA and UMAP (Fig. 7a,b) revealed epigenomes at day 32 that were more homogenous than effectors, and clustered close to naive cells. Targeted analysis of transcription-factor-binding motifs suggested differential and inverse patterns of accessibility for *Id2* and *Tcf7* in T_{H1} effectors compared with in T_{FH} effectors, which was partially reset to naive levels by day 32 (Fig. 7c). *Tbx21* motifs were enriched, and *Bcl6* motifs depleted during the first week, which was partially lost by day 32 (Fig. 7c). Thus, phenotypic convergence of effectors during transit to memory was evident by epigenomics.

Nevertheless, epigenomic differences remained between naive and day 32 cells. Reactome pathway analysis revealed enrichment for peaks associated with tyrosine-kinase and CD28 signaling at day 32 that was absent in naive cells (Fig. 7d). Thus, specific changes in chromatin had been retained in memory. We sought to infer transcriptional regulators whose access to chromatin varied between naivety and memory³¹. PCA of motif bias-corrected deviation revealed that day 32 motif variability was similar, although distinct from, that of naive cells (Fig. 7e). Binding motifs were variable for *Fos*, *Jun*, *Runx*, *Rel*, *Maf*, *Tbx* and *Irf* families (Fig. 7e and Supplementary Table 7), with motif enrichment for *Irf8*, *JunD* and *MafF*, and loss for *Tcf7*, *Ctcf*, *Etv6*, *Nfatc3*, and *Runx1* in memory compared to naivety (Fig. 7f and Extended Data Fig. 9). Thus, epigenomic and transcriptomic heterogeneity induced during effector differentiation was partially, but incompletely, reset during transit to memory (Extended Data Fig. 10).

Discussion

Here, we examined the dynamics of CD4⁺ T cell memory development using a model of malaria and treatment with antimalarials. We mapped transcriptomic change as CD4⁺ T cells gradually transitioned from effector to memory over four weeks, and in doing so inferred genomic relationships between T_{FH} , T_{H1} , Tr1, GC T_{FH} , T_{CM} , T_{EM} , naive, proliferating and exhausted cells (terms defined in Table 1). We revealed that CD4⁺ T cell fate toward either T_{H1} -effector memory or various T_{FH} and T_{CM} fates was determined during the first week of infection, whose phenotypes were altered by antimalarial drugs.

Currently, the only available malaria vaccine, RTS,S/AS01, generates short-lived immunity that correlates with antibodies and CD4⁺ T cell-derived IL-2 and TNF^{32,33}. Since T_{H1} cells also provide protection³⁴, one vaccine strategy might be dual promotion of antibodies and T_{H1} memory. Our study highlights that these mechanisms can be elicited simultaneously among T cells of the same specificity during infection, although development of one likely influences and interferes with the other^{22,35}. One possible solution might be to temporally segregate T_{H1} and T_{FH} differentiation, by priming and boosting with different antigens, and employing adjuvants that block and then promote T_{H1} immunity.

One conclusion from our work is that memory fate is determined during the first week of infection. Therefore, diverting CD4⁺ T cells along a particular lineage may be possible at early, but not late, stages of infection or vaccination. Nevertheless, consistent with clinical studies^{13,14}, our data suggest that reducing parasite load profoundly affects cellular phenotypes and functional potential. We speculate that anti-malarial drug treatment in endemic regions minimizes deleterious effects of chronic infection on naturally acquired or vaccine-mediated immunity.

In contrast to LCMV infection⁷, CCR7⁺ T_{CM} precursors were absent during *Plasmodium* infection. One interpretation might be that all effectors expressed equal T_{CM} potential, although this appears unlikely given the restriction of *Ccr7* along one lineage, and that fate-mapping suggested minimal crossover between trajectories. The discrepancy between studies may be due to differences in viral versus parasitic infection¹⁰, the specific transgenic TCRs used or specifics of infectious dose or *Plasmodium* species employed. Nevertheless, our data argue that T_{CM} precursors are not essential for CD4⁺ T cell memory development.

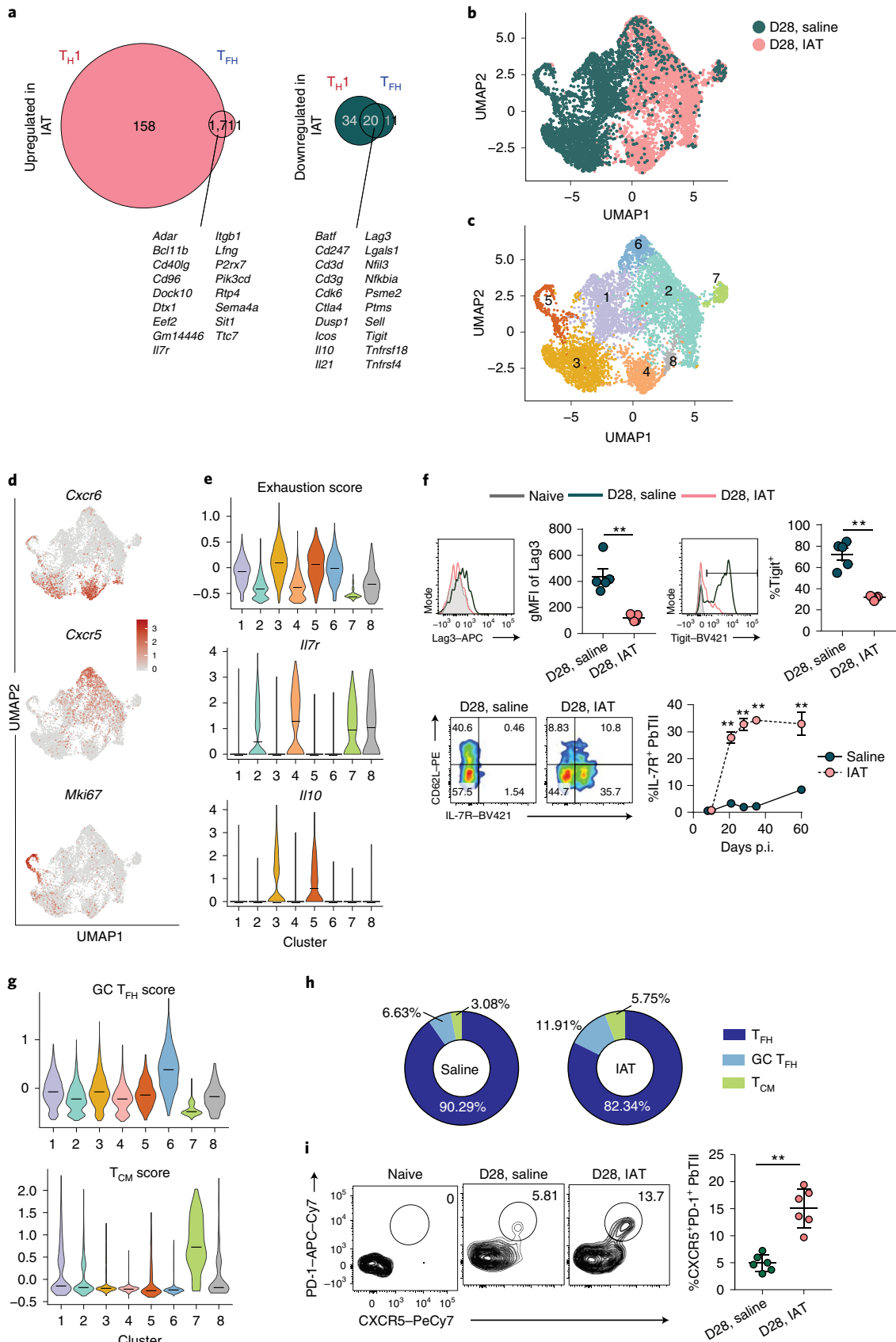
Clonal analysis confirmed that diversity in fate is common among clones during parasitic infection¹⁸. Mathematical modeling revealed a modest predisposition to one fate among sibling cells. This suggested heterogeneity in responses by individual clones, despite transcriptomic and epigenomic homogeneity in naivety, and predictable frequencies of each fate at a population level. Our observations are consistent with those from bacterial infection³⁶. Given heterogeneity in microanatomical location, and since B cells and myeloid cells influence effector fate¹⁸, we hypothesize that cell–cell interactions are the primary drivers of differences between clones.

A caveat of our dynamic modeling was its inability to distinguish GC T_{FH} from T_{CM} cells within the T_{FH} lineage. A focused analysis at a late time-point revealed that, although the bulk of T_{FH} lineage cells exhibited a heterogeneous T_{FH} phenotype, rarer GC T_{FH} and T_{CM} subpopulations were present. GC T_{FH} and T_{CM} cells differentially expressed many genes, including those previously reported such as *Klf2* (ref. 37) and *Satb1* (ref. 38). Our analysis suggested a continuum between T_{FH} and GC T_{FH} cells, with T_{CM} cells as a separate,

Fig. 6 | IAT boosts development of GC T_{FH} and memory CD4⁺ T cells during malaria. **a**, Venn diagrams showing the number of differentially expressed genes (genes with FDR < 0.05) involved in immune-system process (GO:0002376) performed on the Smart-seq2 (384) dataset only. An overlapping list of genes between genes upregulated in the T_{H1} and T_{FH} branches of late cells (pseudotime > 0.9) during IAT (pink, left), or upregulated in the T_{H1} and T_{FH} branch of late cells in the saline group (green, right). **b**, UMAP representation of PbTII cells isolated at D28 p.i. from saline or IAT groups, analyzed using the droplet-based 10x Genomics Chromium platform. UMAP was calculated from the first 10 principle components using 1,394 highly variable genes. Data are available as 'Memory' on our web-based GUI: https://haquelab.mdhs.unimelb.edu.au/cd4_memory/. **c**, Unsupervised clustering of UMAP representation in **b**. **d**, Visualization of *Cxcr6*, *Cxcr5* and *Mki67* on UMAP representation in **b**. **e**, Violin plots showing the exhaustion score, *Ilf7r* expression and *Ilf10* expression for all clusters shown in UMAP representation in **b**. Median expression for each cluster is denoted for each gene and score. **f**, Top, representative FACS plots and graph showing surface expression of LAG3 and TIGIT on PbTII cells at D28 p.i. Bottom, representative FACS plots and graph showing the kinetics of IL7R expression over time. Data are representative of 2 independent experiments ($n = 5$ mice per group, per individual timepoint for each independent experiment) and are presented as mean \pm s.e.m. Statistical analysis was performed between saline and IAT groups for each timepoint individually using two-tailed Mann–Whitney test. **g**, Violin plots show GC T_{FH} and T_{CM} score for all clusters shown in UMAP representation in **b**. Median expression for each cluster is denoted for each score. **h**, Pie charts show the proportion of GC T_{FH} and T_{CM} cells within the ' T_{FH} ' population for PbTII cells at D28 p.i. with or without IAT as described in **b** and **g**. **i**, Representative FACS plots and graph showing proportion of PbTIIs as GC T_{FH} cells (CXCR5⁺PD-1⁺) at D28 p.i. in the presence or absence of IAT. Data are representative of 3 independent experiments ($n = 6$ mice per group, per independent experiment) and are presented as mean \pm s.e.m. Statistical analysis was performed using two-tailed Mann–Whitney test. ** $P < 0.01$.

adjacent population. We propose that cells progress along the T_{FH} lineage, with small proportions differentiating further into GC T_{FH} or T_{CM} cells.

Tcf7 promotes T_{FH} cell differentiation, but its role in $CD4^+$ T cell memory is unclear. We noted *Tcf7* expression in $CD4^+$ T cells along the T_{FH} lineage—indeed, TCF1 supported this trajectory in our



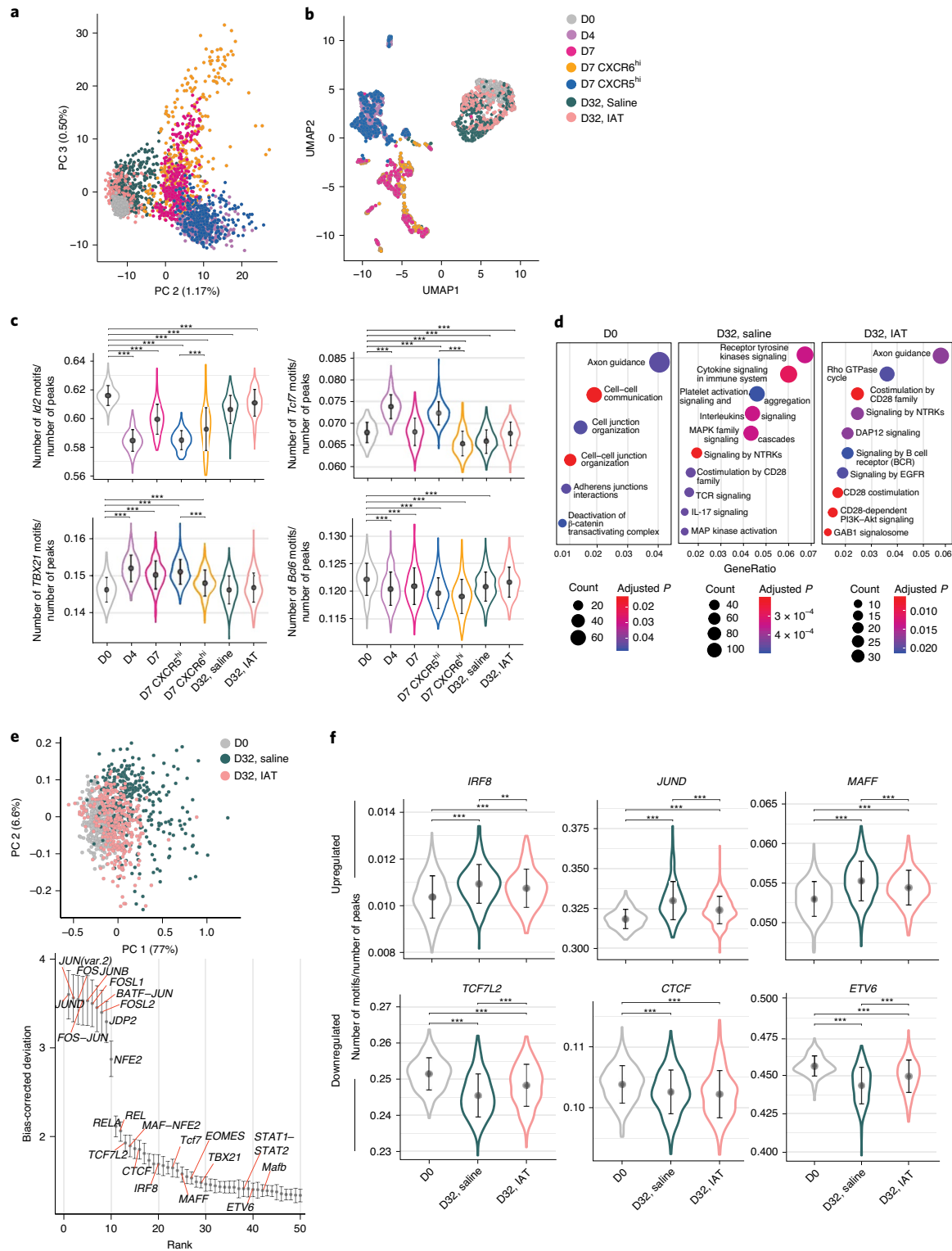


Fig. 7 | scATAC-seq reveals partial epigenomic resetting and homogenization during memory development. a, PCA of peaks after application of latent semantic indexing analysis³⁰. **b**, UMAP representation of the epigenomic landscape of the epigenomic landscape of PbTII cells from PC 2-10. **c**, Changes in scATAC-seq peaks associated with different T_H1- or T_{FH}-associated transcription factors (*Id2*, *Tcf7*, *Tbx21*, *Bcl6*) for different timepoints. Error bars represent mean \pm s.d. Statistical test was performed using two-sided Wilcoxon rank-sum test. **d**, Reactome analysis of the top 10 enriched pathways for different timepoints. The P value was adjusted for multiple testing using Benjamini-Hochberg correction. **e**, Top, PCA analysis of 500 variable motifs (combined database of human and mice motifs from Jaspar database) calculated using chromVAR analysis for comparison between cells from naive (D0), saline and IAT groups at D32 p.i. Bottom, top 50 motifs explaining variability between cells from naive (D0), saline and IAT groups at D32 p.i. **f**, Changes in scATAC-seq peaks associated with the top variable transcription factors detected from chromVAR analysis as described in **e**. Error bars represent mean \pm s.d. Statistical test was performed using two-sided Wilcoxon rank-sum test. ***P* < 0.01, ****P* < 0.001.

model. *Tcf7* was also preserved during persisting infection, suggesting that exhaustion was not an impediment to its potential function. This is reminiscent of CD8⁺ T cell reports^{39–42}, wherein TCF1 supported stem-like memory cells in humans, and precursor exhausted cells were reinvigorated via immune checkpoint blockade. These observations suggest a generalized model in which TCF1 is retained in one lineage after bifurcation, supporting their subsequent recall responses.

An important feature of the T_{H1} and T_{FH} lineages was their gradual, partial coalescence and incomplete return to naivety during memory. Nevertheless, changes in transcription-factor-binding motifs were preserved among individual memory cells—for example, *Maf* and *Irf8* motifs. Several studies have examined epigenomic and transcriptomic changes in CD8⁺ T cells during memory development⁴³, or in humans undergoing vaccination or cancer treatment⁹. These support a model of CD8⁺ T cell memory characterized by partial reversion to naive-like states, consistent with that presented here for CD4⁺ T cells.

Drug treatment boosted T_{H1} memory cells and GC T_{FH} formation in our mouse model, perhaps consistent with what happens in humans in malaria-endemic regions, where immunity may be improved by antimalarial drugs. A caveat of our model is that immunity induced by a single infection is so robust that improvements to parasite control cannot be assessed functionally⁴⁴. Nevertheless, we previously associated T_{FH} boosting with improved primary immunity⁴⁵. We speculate that exhaustion, impaired antigen presentation⁴⁶ and IL-10 immune regulation⁴⁷ impair GC T_{FH} responses, and are alleviated by removing parasites with antimalarial drugs. In endemic areas, children are infected multiple times each season. The effect of multiple infections and drug treatment on CD4⁺ T cells remains to be tested.

Thus, transition from effector to memory and GC T_{FH} cell states occurs gradually in CD4⁺ T cells during experimental malaria. The extended timeframe over which transcriptomic change occurred, and the effect of antimalarial drugs, suggests that there are opportunities for manipulating memory development after primary immune responses have been triggered. We identify potential targets for manipulating CD4⁺ T cell memory, with potential clinical utility in malaria-vaccine development or in prevention of ongoing, persisting *Plasmodium* infections. Given pivotal roles for CD4⁺ T cells in bacterial, fungal and viral infections, our datasets may be relevant across a range of human infectious diseases.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41590-020-0800-8>.

Received: 12 February 2020; Accepted: 28 August 2020;

Published online: 12 October 2020

References

- Harrington, L. E., Janowski, K. M., Oliver, J. R., Zajac, A. J. & Weaver, C. T. Memory CD4 T cells emerge from effector T-cell progenitors. *Nature* **452**, 356–360 (2008).
- Pepper, M., Pagan, A. J., Igyarto, B. Z., Taylor, J. J. & Jenkins, M. K. Opposing signals from the Bcl6 transcription factor and the interleukin-2 receptor generate T helper 1 central and effector memory cells. *Immunity* **35**, 583–595 (2011).
- Hale, J. S. et al. Distinct memory CD4⁺ T cells with commitment to T follicular helper- and T helper 1-cell lineages are generated after acute viral infection. *Immunity* **38**, 805–817 (2013).
- Marshall, H. D. et al. Differential expression of Ly6C and T-bet distinguish effector and memory T_{H1} CD4⁺ cell properties during viral infection. *Immunity* **35**, 633–646 (2011).
- Tabo, N. J. et al. Most microbe-specific naive CD4⁺ T cells produce memory cells during infection. *Science* **351**, 511–514 (2016).
- Opat, M. M. et al. Early effector cells survive the contraction phase in malaria infection and generate both central and effector memory T cells. *J. Immunol.* **194**, 5346–5354 (2015).
- Ciucci, T. et al. The emergence and functional fitness of memory CD4⁺ T cells require the transcription factor Thpok. *Immunity* **50**, 91–105.e4 (2019).
- Waickman, A. T. et al. Dissecting the heterogeneity of DENV vaccine-elicited cellular immunity using single-cell RNA sequencing and metabolic profiling. *Nat. Commun.* **10**, 3666 (2019).
- Akondy, R. S. et al. Origin and differentiation of human memory CD8 T cells after vaccination. *Nature* **552**, 362–367 (2017).
- Arroyo, E. N. & Pepper, M. B cells are sufficient to prime the dominant CD4⁺ T_{FH} response to *Plasmodium* infection. *J. Exp. Med.* **217**, e20190849 (2020).
- WHO Global Malaria Programme: World Malaria Report 2019 (World Health Organization, 2019).
- Langhorne, J., Ndungu, F. M., Sponaas, A. M. & Marsh, K. Immunity to malaria: more questions than answers. *Nat. Immunol.* **9**, 725–732 (2008).
- Muhindo, M. K. et al. Intermittent preventive treatment with dihydroartemisinin-piperazine and risk of malaria following cessation in young Ugandan children: a double-blind, randomised, controlled trial. *Lancet Infect. Dis.* **19**, 962–972 (2019).
- Schellenberg, D. et al. Intermittent preventive antimalarial treatment for Tanzanian infants: follow-up to age 2 years of a randomised, placebo-controlled trial. *Lancet* **365**, 1481–1483 (2005).
- Jaganathan, P. et al. Effective antimalarial chemoprevention in childhood enhances the quality of CD4⁺ T cells and limits their production of immunoregulatory interleukin 10. *J. Infect. Dis.* **214**, 329–338 (2016).
- Butler, N. S. et al. Therapeutic blockade of PD-L1 and LAG-3 rapidly clears established blood-stage *Plasmodium* infection. *Nat. Immunol.* **13**, 188–195 (2011).
- Suss, G., Eichmann, K., Kury, E., Linke, A. & Langhorne, J. Roles of CD4- and CD8-bearing T lymphocytes in the immune response to the erythrocytic stages of *Plasmodium chabaudi*. *Infect. Immun.* **56**, 3081–3088 (1988).
- Lonnberg, T. et al. Single-cell RNA-seq and computational analysis using temporal mixture modelling resolves T_{H1}/T_{FH} fate bifurcation in malaria. *Sci. Immunol.* **2**, eaal2192 (2017).
- Fernandez-Ruiz, D. et al. Development of a novel CD4⁺ TCR transgenic line that reveals a dominant role for CD8⁺ dendritic cells and CD40 signaling in the generation of helper and CTL responses to blood-stage malaria. *J. Immunol.* **199**, 4165–4179 (2017).
- Meding, S. J., Cheng, S. C., Simon-Haarhaus, B. & Langhorne, J. Role of gamma interferon during infection with *Plasmodium chabaudi chabaudi*. *Infect. Immun.* **58**, 3671–3678 (1990).
- Perez-Maziah, D. et al. Follicular helper T cells are essential for the elimination of *Plasmodium* infection. *EBioMedicine* **24**, 216–230 (2017).
- Obeng-Adjei, N. et al. Circulating T_{H1}-cell-type T_{FH} cells that exhibit impaired B cell help are preferentially activated during acute malaria in children. *Cell Rep.* **13**, 425–439 (2015).
- Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
- La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
- Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
- Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nat. Methods* **15**, 343–346 (2018).
- Huynh, J. P. et al. Bhlhe40 is an essential repressor of IL-10 during *Mycobacterium tuberculosis* infection. *J. Exp. Med.* **215**, 1823–1838 (2018).
- Yu, F. et al. The transcription factor Bhlhe40 is a switch of inflammatory versus antiinflammatory T_{H1} cell fate determination. *J. Exp. Med.* **215**, 1813–1821 (2018).
- Stubbington, M. J. T. et al. T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* **13**, 329–332 (2016).
- Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309–1324.e18 (2018).
- Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
- White, M. T. et al. The relationship between RTS,S vaccine-induced antibodies, CD4⁺ T cell responses and protection against *Plasmodium falciparum* infection. *PLoS ONE* **8**, e61395 (2013).
- Olotu, A. et al. Circumsporozoite-specific T cell responses in children vaccinated with RTS,S/AS01E and protection against *P. falciparum* clinical malaria. *PLoS ONE* **6**, e25786 (2011).
- Pinzon-Charry, A. et al. Low doses of killed parasite in CpG elicit vigorous CD4⁺ T cell responses against blood-stage malaria in mice. *J. Clin. Invest.* **120**, 2967–2978 (2010).

35. Ryg-Cornejo, V. et al. Severe malaria infections impair germinal center responses by inhibiting T follicular helper cell differentiation. *Cell Rep.* **14**, 68–81 (2016).
36. Tubo, N. J. et al. Single naive CD4⁺ T cells from a diverse repertoire produce different effector cell types during infection. *Cell* **153**, 785–796 (2013).
37. Lee, J. Y. et al. The transcription factor KLF2 restrains CD4⁺ T follicular helper cell differentiation. *Immunity* **42**, 252–264 (2015).
38. Yasuda, K. et al. Satb1 regulates the effector program of encephalitogenic tissue T_H17 cells in chronic inflammation. *Nat. Commun.* **10**, 549 (2019).
39. Siddiqui, I. et al. Intratumoral Tcf1⁺PD-1⁺CD8⁺ T cells with stem-like properties promote tumor control in response to vaccination and checkpoint blockade immunotherapy. *Immunity* **50**, 195–211.e10 (2019).
40. Chen, Z. et al. TCF-1-centered transcriptional network drives an effector versus exhausted CD8 T cell-fate decision. *Immunity* **51**, 840–855.e5 (2019).
41. Nish, S. A. et al. CD4⁺ T cell effector commitment coupled to self-renewal by asymmetric cell divisions. *J. Exp. Med.* **214**, 39–47 (2017).
42. Lin, W. W. et al. CD8⁺ T lymphocyte self-renewal during effector cell determination. *Cell Rep.* **17**, 1773–1782 (2016).
43. Scott-Browne, J. P. et al. Dynamic changes in chromatin accessibility occur in CD8⁺ T cells responding to viral infection. *Immunity* **45**, 1327–1340 (2016).
44. Akter, J. et al. Plasmodium-specific antibodies block in vivo parasite growth without clearing infected red blood cells. *PLoS Pathog.* **15**, e1007599 (2019).
45. James, K. R. et al. IFN regulatory factor 3 balances T_H1 and T follicular helper immunity during nonlethal blood-stage plasmodium infection. *J. Immunol.* **200**, 1443–1456 (2018).
46. Haque, A. et al. Type I IFN signaling in CD8⁺ DCs impairs T_H1-dependent malaria immunity. *J. Clin. Invest.* **124**, 2483–2496 (2014).
47. Couper, K. N. et al. IL-10 from CD4⁺CD25⁺Foxp3⁺CD127⁺ adaptive regulatory T cells modulates parasite clearance and pathology during malaria infection. *PLoS Pathog.* **4**, e1000004 (2008).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Experimental mice, adoptive transfer and infections. C57BL/6J and SJL.Ptprca mice were purchased from Animal Resources Centre (Canning Vale), PbTII¹⁹, nzEGFP and Cre^{ROSA26Ert2} mice were bred in-house. nzEGFP mice express enhanced green fluorescent protein (eGFP) from the ubiquitously expressed CMV early enhancer/chicken beta actin promoter. *Id3^{GFP}* mice and *Tcf7*-floxed (*Tcf7^{fl/fl}*) were crossed with PbTII mice. Only *Id3^{GFP/+}* mice were used as reporters. All mice were female, aged 8–12 weeks, and maintained under specific-pathogen-free conditions within the animal facility at QIMR Berghofer Medical Research Institute. Mice were housed in exhaust-ventilated cages (Opti-mice) with ≤6 mice per cage, and room temperature was maintained between 19°C and 22°C, with humidity between 55% and 65% and with 12-h/12-h dark/light cycle with a 15-min sunrise and sunset. All animal procedures and protocols were approved (A1503-601M) and monitored by the QIMR Berghofer Medical Research Institute Animal Ethics Committee.

Spleens were collected and homogenized through a 100-µm cell strainer to create a single-cell suspension. Red blood cells (RBCs) were lysed using RBC Lysing Buffer Hybri-Max (Sigma-Aldrich) and CD4⁺ T PbTII cells were enriched using CD4 microbeads (Miltenyi Biotec). Cells (1 × 10⁴ per mouse) were transferred to each mouse via lateral tail-vein intravenous (i.v.) injection.

PcAS parasites were used after thawing frozen, infected blood stabilites and performing a single in vivo passage in C57BL/6J mice. PcAS-infected RBCs were collected from passage mice by cardiac puncture and mice infected (1 × 10⁵ infected RBCs per mouse) via lateral tail vein i.v. injection.

Intermittent antimalarial drug treatment (IAT). Sodium artesunate (Guilin Pharmaceutical or sourced from J. Mohrle) was prepared according to the manufacturer's protocol by diluting in 0.9% saline (Baxter) to a final concentration of 5 mg ml⁻¹. Mice were treated via intraperitoneal (i.p.) injection with sodium artesunate (1 mg per mouse), or vehicle control saline, twice daily from day 7 to day 9 p.i., once daily from day 10 to day 16 p.i. and then twice weekly until experimental endpoint. Mice treated with sodium artesunate were also administered pyrimethamine (0.07 g l⁻¹; Sigma-Aldrich) in drinking water for the duration of the treatment.

Tamoxifen preparation and administration. Tamoxifen (Sigma Aldrich) was dissolved in 10% ethanol in corn oil (Sigma Aldrich) at a concentration of 50 mg ml⁻¹. This was sonicated for 30 min–1 h, until all tamoxifen particles had dissolved. Mice were administered with a single 5-mg dose of tamoxifen via i.p. injection.

Parasitemia assessment. Parasitemia assessment was carried out as previously described⁴⁵. Briefly, a single drop of blood was collected via tail bleed and diluted in 250 µl of RPMI medium containing 5 U ml⁻¹ heparin sulphate. Diluted blood was stained with Syto84 (5 µM; Life Technologies) and Hoechst 33342 (10 µg ml⁻¹; Sigma-Aldrich) for 30 min in the dark at room temperature (RT). Staining was quenched with ×10 volume of ice-cold RPMI medium, and samples were immediately acquired by flow cytometry.

Flow cytometry. Spleens and lymph nodes were collected and homogenized through a 100-µm cell strainer to create a single cell suspension, and RBCs were lysed using RBC Lysing Buffer Hybri-Max (Sigma-Aldrich). Cells were assessed for viability using a LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (1:200 dilution, catalog no. 423102, Biogend), according to the manufacturer's protocol, unless otherwise specified. Prior to antibody staining, Fc receptors were blocked using antibodies against CD16 and CD32. Cells were incubated with surface marker antibodies for 20 min at 4°C. Staining for PerCpCy5.5 CCR7 surface antibody (1:10 dilution, clone: 4B12, catalog no. 120116, Biogend) was performed at 37°C for 1 h as per the manufacturer's recommendations, after Fc receptor blocking. Other surface marker antibodies used were: APC-CD4 (1:200 dilution, clone: RM4-5, catalog no. 100516, Biogend), PE-CD4 (1:200 dilution, clone: RM4-5, catalog no. 100512, Biogend), BV421-CCRβ (1:200 dilution, clone: H57-597, catalog no. 109230, Biogend), PerCpCy5.5-TCRβ (1:200 dilution, clone: H57-597, catalog no. 109228, Biogend), PerCpCy5.5-Vβ12 (1:200 dilution, clone: MR11-1, catalog no. 46-5798-80, eBioscience), AF700-CD44 (1:200 dilution, clone: IM7, catalog no. 103026, Biogend), PE-CD62L (1:200 dilution, clone: MEL-14, catalog no. 104408, Biogend), APC-CXCR6 (1:200 dilution, clone: SA051D1, catalog no. 151106, Biogend), biotinylated CXCR5 (1:50 dilution, clone: 2G8, catalog no. 551960, BD Biosciences), PE-Ki67 (1:200 dilution, clone: SolA15, catalog no. 12-5698-82, eBioscience), BV605-CD69 (1:200 dilution, clone: H1.2F3, catalog no. 104530, Biogend), BV421-IL-7R (1:200 dilution, clone: A7R34, catalog no. 135924, Biogend), PE-CXCR3 (1:200 dilution, clone: CXCR3-173, catalog no. 126506, Biogend), APC-Lag3 (1:100 dilution, clone: C9B7W, catalog no. 125210, Biogend), BV421-TIGIT (1:100 dilution, clone: 1G9, catalog no. 565270, BD Biosciences), AF700-CD45.1 (1:200 dilution, clone: A20, catalog no. 110724, Biogend). Other staining dyes or fluorophores include streptavidin-PeCy7 (1:200 dilution, catalog no. 405206, Biogend) and propidium iodide (1:400 dilution, catalog no. 12-9855-41, Sigma-Aldrich).

To assess cytokine production, cells were incubated with brefeldin-A (10 mg ml⁻¹) with or without ionomycin (500 ng ml⁻¹) and PMA (25 ng ml⁻¹)

at 37°C for 3 h. Intracellular staining for cytokines, transcription factors and chemokines was then performed using the eBioscience Foxp3/Transcription Factor Staining Buffer Set. Staining with intracellular antibodies was conducted at 4°C for 1 h. Where TCF1 was stained after stimulation with PMA and ionomycin for 3 h is explicitly stated in figure legends. All other transcription factors and chemokine molecules were stained without prior stimulation. Intracellular marker antibodies used were: APC-Id2 (1:100 dilution, clone: ILCID2, catalog no. 17-9475-82, eBioscience), PE-TCF1 (1:100 dilution, clone: C63D9, catalog no. 2203S, Cell Signaling Technology), PE-CCL5 (1:100 dilution, clone: 2F9, catalog no. 149104, Biogend), PE-c-Maf (1:100 dilution, clone: sym0F1, catalog no. 12-9855-41, Thermo Fischer), PE-IL-10 (1:100 dilution, clone: JES5-16E3, catalog no. 505008, Biogend), BV421-IFN-γ (1:200 dilution, clone: XMG1.2, catalog no. 505830, Biogend), APC-T-bet (1:50 dilution, clone: 4B10, catalog no. 17-5825-82, Thermo Fischer); PE-BCL6 (1:10 dilution, clone: K112-91, catalog no. 561522, BD Biosciences). Samples were acquired on a LSRII Fortessa analyzer (BD Biosciences) and subsequently analyzed using FlowJo software (Treestar).

Livers were collected in 1% (vol/vol) FCS/PBS and homogenized through a 200-µm metal sieve. Liver suspension was resuspended in 33% (vol/vol) percoll/PBS before centrifugation at 1,700 r.p.m. for 12 min at room temperature. After removal of supernatant containing unwanted cells and debris, liver leukocyte pellet was processed similarly as splenic and lymph-node tissue described above.

PbTII memory recall during in vivo rechallenge. Mice harboring previously in vivo primed eGFP⁺ PbTIIs received CD4-microbead enriched, non-eGFP⁺ congenically-marked (CD45.1⁺) PbTII cells from naive donors (1 × 10⁵ cells per mouse) 1 day prior to homologous high-dose rechallenge with 1 × 10⁷ PcAS-infected RBCs.

Immunofluorescence microscopy. Spleens were fixed with 2% paraformaldehyde for 2–4 h at RT in the dark followed by dehydration with 30% sucrose (Chem-supply) overnight at RT in the dark. Spleens were snap-frozen in Tissue-Tek Optimal Cutting Temperature embedding medium (Sakura Finetek) on dry ice and stored at -80°C. Spleens were sectioned at 10–30 µm on polysine slides such that consecutive sectioning is avoided. Sections were allowed to dry overnight. The dried sections were rehydrated for 15–20 min before fixation in 4% paraformaldehyde for 15–20 min at RT in dark. Slides were washed 3 times for 5 min in washing buffer (0.01% Tween20 in PBS) before permeabilization with 0.1% Triton X-100 in washing buffer for 10–15 min. After washing, slides were incubated with Medical Background Sniper (Biocare) for 30 minutes. Slides were rinsed for 2 min and endogenous biotin was blocked using an Avidin/Biotin Blocking kit (Vector Laboratories), according to manufacturer's protocol. Tissue sections were then stained with rabbit anti-GFP (1:500 dilution, catalog no. ab6556, Abcam), rat anti-mouse CD3-AF594 (1:200 dilution, clone: 1742, catalog no. 100240, Biogend), rat anti-mouse IgD-AF647 (1:50 dilution, clone: 11–26.c.2a, catalog no. 405708, Biogend) and biotinylated peanut agglutinin (1:500 dilution, catalog no. B-1075-5 PNA; Vector Laboratories, Inc.) for 1–2 h at RT in the dark. Secondary antibody staining for PNA and GFP was performed using streptavidin-AF555 (1:300 dilution, catalog no. S21381, Thermo Fisher Scientific) and donkey anti-rabbit AF488 (1:300 dilution, catalog no. R37118, Thermo Fisher Scientific), respectively, for 1–2 h at RT in the dark. Tissue sections were incubated with DAPI for 10–15 min to counterstain nuclei, and slides were mounted in Dako Mounting Media (Agilent Technologies). Image acquisition was performed using an Aperio FL slide scanner or a Zeiss 780-NLO point scanning confocal microscope at ×20, ×40 and ×63 objective.

Cell detection and quantification were performed using the spot-detection function in Imaris image analysis software (Bitplane), with thresholds <10 µm. All objects were manually inspected for accuracy before data were plotted and analyzed in GraphPad Prism. Colocalization analysis was performed using Imaris colocalization functions. B cell follicle was defined as IgD-positive region, and GC was defined as IgD-negative and PNA-positive region within B cell follicle.

Bulk Fast-ATAC sequencing and analysis. Fast-ATAC sequencing was performed as previously described⁴⁸. Briefly, 5,000 viable PbTII cells were sorted by flow cytometry and pelleted by centrifugation. Supernatant was removed and 50 µl of transposase mixture (25 µl TD buffer (Illumina), 2.5 µl of TDE1 (Illumina), 0.5 µl of 1% digitonin (Promega), 22 µl nuclease-free water) was added to the cells. Cells were then incubated at 37°C for 30 min at 300 r.p.m. in an Eppendorf ThermoMixer. Transposed DNA was amplified and purified using a Qiagen MinElute PCR Purification kit, according to the manufacturer's protocol. Purified DNA was eluted in 20 µl Buffer EB (Qiagen) and sequenced using paired-end sequencing on a NextSeq 550 instrument (Illumina).

Raw ATAC-seq reads were mapped to mouse genome MGScv37 (mm9) using BWA-MEM⁴⁹ (version 0.7.15). Unmapped reads, reads mapping to unassigned contigs and mate unmapped reads were removed, as well as mitochondrial genes and PCR duplicates. The resulting .bam files were first converted to bedGraph using the bedtools genomecov command from the BEDTools suite⁵⁰ (version 2–2.29.0) and then converted to bigwig format using the bedGraphToBigWig program from University of California, Santa Cruz (UCSC) Genome Browser (<https://genome.ucsc.edu/>). Read counts were normalized to the number of

uniquely mapped reads per million. Peak calling was performed using MACS2 (version 2.1) with the following parameters: `--nomodel`, `--shift 37`, `--extsize 73`, `--pvalue 1e-5`. The bigWig tracks and narrowPeak files were visualized in a custom UCSC Genome Browser track hosting the mouse mm9 reference genome. Peaks overlapping with the mm9 blacklist region generated by UCSC were removed, and overlapping peaks between replicate samples were identified using the `findOverlaps` command from the `GenomicRanges` (version 1.38) package and used for downstream analysis. Peaks were annotated using the R package `ChIPseeker`⁵¹ (version 1.22.1) with transcription start site (TSS) reaching from $-3,000$ to $3,000$ to the UCSC mm9 gene model and the `org.Mm.eg.db` R package.

Single-cell capture and sequencing. *Droplet-based scRNA-seq.* Recipient mice ($n = 6$ mice per group per timepoint) received 1×10^4 sorted naive (CD62L⁺CD44⁻) eGFP-expressing PbTII cells from a single donor 24 h prior to infection with *PcAS*. Spleens from 5 infected, recipient mice were pooled, and viable PbTII cells were sorted at D7 and D28 (saline and IAT groups) into cold 1% BSA/PBS buffer. PbTIIs were counted, and $\sim 8,000$ cells were loaded per channel onto a Chromium controller (10x Genomics) for generation of gel bead-in-emulsions. Sequencing libraries were prepared using Single Cell 3' Reagent Kits v3 (10x Genomics) and then converted using the MGIEasy Universal Library Conversion Kit (BGI) before sequencing on a MGISEQ-2000 instrument (BGI).

Smart-seq2 scRNA-seq. We prepared 384-well lo-bind plates with 0.5 μ l of Triton-X lysis buffer, 0.25 μ l of 10 μ M olido-dT30-VN, 0.25 μ l dNTP mix (25 mM each) and ERCC controls (final dilution of 1:64 million) per well and stored at -20°C until use. Plates were thawed on ice before sorting single PbTII cells isolated from 1 selected donor mouse at each timepoint ($n = 6$ mice per group per timepoint) into each well. Single-cell lysates were sealed and spun at 100g for 1 min, then immediately frozen on dry ice and stored at -80°C . Reverse transcription and PCR were performed following the Smart-Seq2 protocol with the following modifications: 1 μ l of reverse transcription mix and 5.7 μ l PCR Master mix. After amplification, complementary DNA was subjected to quality control using 1 μ l of amplified cDNA on an Agilent 2100 BioAnalyzer and Agilent High Sensitivity DNA kits (Agilent Technologies). Next, 5 μ l of cDNA per cell was cleaned using Agencourt AMPure XP beads (Beckman Coulter) at a 1 \times ratio on a Hamilton STAR liquid handler (Hamilton Robotics). cDNA was quantified using Biotium AccuClear High Sensitivity DNA quantification reagent and normalized to 1 ng μ l⁻¹ in a total volume of 500 nl before library preparation using Nextera XT DNA Sample Preparation Kit (Illumina). Then, 125 nl of in-house index adapters (Integrated DNA Technologies) similar to Illumina N7 and N5 indices were added to the tagmentation reaction before adding 1.5 μ l of KAPA HiFi DNA polymerase (KAPA Biosystems) and performing 12 cycles of PCR according to the manufacturer's instructions. After PCR, all samples were pooled into 384-plex pools using a 24 \times 16 dual indexing approach, and the pool was cleaned using Agencourt AMPure XP beads at a 0.6 \times ratio. Library pools were eluted in buffer EB and quality controlled using an Agilent 2100 BioAnalyzer and Agilent High Sensitivity DNA kits before the concentration was adjusted to 4 nM. Each library was sequenced on 1 lane of Illumina HiSeq 2000 version 4 chemistry (paired-end 75-base-pair reads).

Processing of scRNA-seq data. *Smart-seq2 data.* Raw reads were mapped to a mouse genomic index (Ensembl version 70) using STAR aligner⁵² version 2.5.2 and transcripts per millions (TPM) were estimated by RSEM⁵³ version 1.2.30. ERCC RNA spikes were included in the index, but removed for down-stream analysis.

10x Genomics data. For the BGI FASTQ files to be made compatible with 'cellranger count' pipeline from Cell Ranger version 3.0.2 (10x Genomics), file names and FASTQ headers were reformatted using code from https://github.com/IMB-Computational-Genomics-Lab/BGIvsIllumina_scRNASeq⁵⁴. Data were processed using 10x mouse genome 1.2.0 release as a reference.

Quality control of scRNA-seq data. *10x Genomics data.* Cells were filtered to remove those expressing fewer than 500 genes and more than 6,000 genes, and those with more than 35% mitochondrial content. Only genes expressed in three or more cells were considered. Regression of the number of unique molecular identifiers was performed for each analysis individually.

Smart-seq2 data. Cells were filtered to remove those with fewer than 100,000 reads mapping to the mouse genome, those expressing fewer than 1,000 genes and more than 5,000 genes and those with more than 35% mitochondrial content. Only genes expressed at one or more TPM in three or more cells were considered, unless otherwise specified.

Batch-effect correction. Two PbTII datasets (Smart-seq2 (96) and SMARTer C1) from our previous study underwent quality control as described previously¹⁸ and were combined with Smart-seq2 (384) dataset from our current study. Genes expressed at less than one TPM in fewer than three cells were globally removed. The 'removeBatchEffect' function from Limma⁵⁵ was then used to remove the batch effect. All three datasets shown have been regressed as per described unless otherwise specified.

Dimensionality reduction. PCA dimensionality reduction (prcomp) was performed on Smart-seq2 dataset with the genes expressed at 100 or more TPM in 15 or more cells as input. On the combined, batch-effect corrected dataset, dimensionality reduction was performed with BGPLVM using GPfates as previously described¹⁸.

To generate UMAP of the combined dataset, the 'removeBatchEffect' function from Limma was used to remove the batch effect from the dataset, which contained highly variable genes selected using 'trendVar' and 'decomposeVar' from Scran (v1.6.9)⁵⁶ (FDR < 0.05, bio > 0.5), followed by PCA dimensionality reduction and usage of 'RunUMAP' from Seurat⁵⁷ package (v2.3.4 unless otherwise specified).

Alternative integration of the 3 datasets was performed using single-cell variational inference (version 0.3.0). The expected-counts matrix obtained from RSEM was used as an input. All parameters were kept at default except: up to 4,000 genes were considered as an input, 30 latent variables were considered, 2 hidden layers were used for encoder and decoder neural networks and 100 epochs and a 0.01 learning rate were used to train the model. The computed latent variables were used as an input to generate UMAP using 'RunUMAP' from Seurat.

To perform PCA dimensionality reduction on the 10x Genomics dataset, we identified highly variable genes using the 'FindVariableGenes' function from the Seurat package and used these as an input. For each subset of data used for dimensionality reduction, highly variable genes were computed individually and used as an input. The number of highly variable genes used in each analysis is noted in the figure legends. The PCA output was then used as an input to generate UMAP using 'RunUMAP' from Seurat.

Gene-signature scoring. T_{H1} and T_{FH} signature scores were calculated using the top 50 T_{H1} - and T_{FH} -bifurcating genes derived from Lonnberg et al.¹⁸. Cell-cycle signature score was calculated using 226 cell-cycle genes derived from Cyclebase⁵⁸. Eight coinhibitory receptors (*Ctla4*, *Pdcd1*, *Lag3*, *Havcr2*, *Btla4*, *Cd160*, *2b4*, *Tigit*) established for their role in T cell exhaustion⁵⁹ were used for exhaustion-phenotype scoring.

Each cell on the 10x Genomics dataset was scored for various signatures defined in each figure legend using the 'AddModuleScore' function from Seurat. To score cells in the Smart-seq2 dataset, a summative expression of all genes defined within each signature was calculated for each cell.

Pseudotime and trajectory inference. Pseudotime was inferred on the combined, batch-effect-corrected dataset based on latent variable (LV) 1 coordinates from bGPLVM dimensionality reduction. Firstly, DBSCAN clustering was performed on D0–D3 cells and D4–28 cells separately, using LV1 and LV2 coordinates as inputs. From the D0–D3 group, only cells with LV2 coordinates lower than the lowest LV2 coordinate from the non-outlier cells were considered outliers, and from the D4–D28 group, only cells with LV2 coordinates higher than the highest LV2 coordinate from the non-outlier cells were considered outliers. Outliers from each group were grouped together with the cells from the opposing groups. LV1 coordinates of the first group (D0–D3 cells and outliers from D4–D28 cells) were flipped along the lowest LV1 coordinate from all pf the dataset and concatenated with the LV1 coordinates of the second group (D4–D28 cells and outliers from D0–D3 cells) to infer the pseudotime coordinates.

Overlapping mixtures of Gaussian processes (OMGP). Lineages were traced along the pseudotime with OMGP using GPfates¹⁸, using LV2 coordinates to model as a function of pseudotime. Trajectory inferences were performed separately for saline and IAT groups, with the same set of parameters used (two trends assumed, global variance = 0.5, per-trend variance = 3 and per-trend lengthscale = 2). Here, the saline-treated group consists of all cells from D0–D7 and saline-treated cells from D10 onwards. The IAT group consists of all cells from D0–D7 and IAT-administered cells from D10 onwards.

Slingshot. Trajectories were inferred through UMAP cell embeddings using Slingshot v0.99.12 (ref. ²³). Cluster information from unsupervised clustering was used as an input. Slingshot analysis was performed separately for saline and IAT groups. A semisupervised approach was taken whereby clusters with a high proportion of D0 were specified as starting points.

RNA velocity. RNA-velocity analysis was performed using Velocyto version 0.17.16 (ref. ²⁴). Analysis was performed separately for saline and IAT groups and only timepoints from D4 onwards were included in the analysis. All parameters were kept at default except: minimum number of spliced molecules = 40, minimum number of cells expressing spliced molecules of a gene = 30, 1,000 gs to rank, and 20 PCs and 150 neighbors considered for knn imputation. Calculated velocity was projected onto pre-computed BGPLVM embeddings.

Functional dynamics analysis. Automatic expression histology (AEH) from spatialDE²⁶ was applied on the combined, batch-effect-corrected dataset. Only significantly variable genes along pseudotime were considered (FDR < 0.05). AEH was applied separately for saline and IAT samples, with the same set of parameters (number of patterns (c) = 7, lengthscale = 9). c specifies the number of groups of genes, with each group displaying distinct expression dynamic along pseudotime.

Genes in Dynamic 1 ($n=511$) were identified together in a small number of groups even when $c=70$, indicating that gene groupings were robust to variation in parameter choice for c .

Gene-signature scores of each dynamic were computed using the 'AddModuleScore' from Seurat. Correlations between expression profiles of genes in dynamic 1 were quantified by calculating Spearman's rho. Only correlations between genes higher than $\rho > 0.3$ were considered as input for generation of a coexpression network. The network plot was generated using ggraph v2.0.3 with layout = 'fr' parameter.

Differential-gene-expression analysis. As an input for differential-gene-expression analysis (DGEA), expected counts from Smartseq2 data were estimated by RSEM version 1.2.30 and then normalized and log-transformed using 'computeSumFactors' and 'normalize' functions from scran. Pairwise Welch t -test was performed using 'pairwiseTTests' function from scran to identify genes differentially expressed between two groups of cells. Comparisons were done as follows: (1) cells at late pseudotime from IAT group on the T_{H1} arm versus cells at late pseudotime from saline group on the T_{H1} arm, (2) cells at late pseudotime from IAT group on the T_{FH} arm versus cells at late pseudotime from saline group on the T_{FH} arm (Fig. 4e). Genes with a FDR below 0.05 were considered to be differentially expressed. Cells with a pseudotime value higher than 0.9 were considered to be late pseudotime.

To generate the differential gene signature from comparison of different clusters within D28 data from the 10x Genomics dataset, cluster-specific genes were defined using 'FindAllMarkers' from Seurat by comparing between clusters as specified in each analysis.

GO term enrichment analysis. GO terms were obtained from the 'org.Mm.eg.db' Bioconductor annotation package⁶⁰. Fisher's exact test was used to identify significantly over-represented GO terms. Input gene lists were differentially expressed genes from the comparisons described above, with up- and downregulated genes considered separately. The background gene set was reduced to those used as an input for DGEA. GO term enrichment analysis was conducted for functional-dynamics analysis using genes from each dynamic as input and all significantly variable genes along pseudotime as the background gene set. Enrichment analysis for biological process and molecular function terms was performed using the 'goana' function in edgeR⁶¹.

Testing the distribution of cell fates among clonotypes. TCR sequences for each cell from Smart-seq2 data were reconstructed using TraCeR as previously described²⁹. Cells were classified as belonging to the T_{H1} group if their T_{H1} assignment probability (TAP) was greater than the median TAP + 0.05, and cells were classified as belonging to the T_{FH} group if their TAP was less than the median TAP - 0.05. Analysis to test whether the distribution of cell fate among families was consistent with cells acquiring fates independently of the family to which they belong was focused on only those cells from the 97 families found in IAT group, excluding cells where fate was indeterminate. To test the null hypothesis that cell fate is randomly distributed across cells, independent of their family (that is ancestry), a bootstrapping approach was used. First, we constructed a likelihood function to describe the likelihood of observing the a given number of T_{H1} cells, Th_i , in a particular family, i , of size n_i , assuming the process is binomial with some probability P_h . The overall likelihood of observing the distribution of cell fates across all families is then given by:

$$L = \prod_{i=1}^N Bi(Th_i, n_i, P_h)$$

where N is the number of families. Note that the P_h that maximizes the likelihood will be given by the proportion of all cells that have a T_{H1} fate (demonstrated by evaluating the derivative of when log of the likelihood function and solving for when derivative is zero). In this dataset the proportion of cells that were T_{H1} was $P_h = 0.575$. Using the observed numbers of Th_i and n_i for each family, L in equation 1 was computed and was called the likelihood of the observed distribution (L_0).

To determine whether the likelihood of the observed distribution of cell fates by family is consistent with a random process, cells were randomly permuted among the different families 100,000 times. Each permutation, p , was performed in such a way as to preserve the size of families, but redistribute the cells across the families. After permuting the cells, the likelihood, L_p , was recalculated for each permutation p . This process produced a distribution of the likelihood associated with cell fates being randomly and independently distributed across families. The likelihood of the observed distribution of cell fates across families, L_0 , was compared with the set of likelihoods, L_p , associated with each random permutation of cells across the fates, and the likelihood of finding the observed distribution of cell fates by family by random chance was less than approximately 0.05%, that is $p = 0.0005$. Thus, the null hypothesis that cell fate is independent of family was rejected.

The histogram of the observed versus expected distribution of families with a T_{H1} predominance, a T_{FH} predominance or a mixture of fates (Fig. 6d), was determined by first drawing a histogram of the observed proportions of T_{H1} cells

in each family (noting that families of size 3 or 6 that had a proportion of 1/3 or 2/3 were included in the middle bin in the histogram). To determine the expected distribution of families, the above described permutation of cells between families was performed and the histogram redrawn each time. The expected distribution of families was created by taking the mean of the bin counts across these 100,000 random permutations.

Plate-based scATAC-seq. Plate-based scATAC-seq was performed as previously described⁶². Briefly, 17,000–50,000 viable PbTII cells were sorted using flow cytometry and pelleted by centrifugation. Supernatant was removed, and cells were resuspended in 50 μ l tagmentation mix (33 mM Tris-acetate, pH 7.8, 66 mM potassium acetate, 10 mM magnesium acetate, 16% dimethylformamide (DMF), 0.01% digitonin and 5 μ l of Tn5 from the Nextera kit from Illumina, cat. no. FC-121-1030). Cell/tagmentation mixture was then incubated at 37 °C, 800 r.p.m. on an Eppendorf thermomixer for 30 min. Tagmentation reaction was stopped by the addition of equal volume (50 μ l) of stop buffer (10 mM Tris-HCl, pH 8.0, 20 mM EDTA, pH 8.0) followed by incubation in ice for 10 min. Then, 150 μ l of DPBS/0.5% BSA was added to the mixture and nuclei suspension was then transferred to a polypropylene FACS tube. 0.0006% DAPI (in-house) was added for staining the nuclei prior to single-nuclei sort into 384-well plate (Eppendorf, cat. no. 0030128508) containing 2 μ l of 2 \times lysis buffer (100 mM Tris-HCl, pH 8.0, 100 mM NaCl, 40 μ g ml⁻¹ Proteinase K (New England Biolabs, P8107S), 0.4% SDS. Single-nuclei lysates were sealed and spun at 100g for 1 min, then were immediately frozen on dry ice and stored at -80 °C. Tn5 release and proteinase K digestion was then performed on a PCR machine (MJ Research Peltier Thermal Cycler) at 65 °C for 15 min followed by the addition of 2 μ l of 10% Tween to quench the SDS. Five microliters of NEBNext High-Fidelity 2 \times PCR Master Mix (New England Biolabs, cat. no. M0541L) was then added to a preprepared 384-well plate (Eppendorf, cat. no. 0030128508) containing 1 μ l of 10 μ M i5 and i7 indexing primer mix (5 μ M each) (Integrated DNA Technologies). The indexing primers were arranged in a combinatorial manner to allow multiplexing of up to 1,536 (4 \times 384-well plates) sorted nuclei into one sequencing pool. The plate containing the PCR Master Mix and indexing primers was stamped onto the single nuclei, mixed briefly, sealed and spun at 100g for 1 min. Amplification was performed on a PCR machine (MJ Research Peltier Thermal Cycler) with 72 °C for 10 min, 98 °C for 5 min and 20 cycles of 98 °C for 10 s, 63 °C for 30 s and 72 °C for 20 s. The 384-well plate containing the PCR product was inverted into a VBLOK200 reservoir (Clickbio cat. no. CBVBLOK200-1) and spun at 200g for 1 min. The combined reactions were then transferred to a 5-ml Eppendorf tube, (Eppendorf, cat. no. 0030119401) and the DNA was purified and size-selected with an AMPure XP workflow (Beckman Coulter, cat. no. A63880). The purified pool was quantified on an Agilent Bioanalyser and combined with 3 additional pools, each containing 384 nuclei, to produce a sequencing library containing 1,536 single nuclei, and this was sequenced on 1 lane of an Illumina HiSeq 4000 instrument.

Single-cell ATAC sequencing and processing. Raw ATAC-seq reads were mapped to mouse genome MGScv37 (mm9) using BWA-MEM (version 0.7.15). For each single cell, .bam reads with mapping quality below 30 and duplicated reads marked with Picard (version 2.19.0) were removed. For each time point, we aggregated the single-cell .bam files and marked and removed duplicated reads again. Peak calling was performed using MACS2 (version 2.1) with parameters --q 0.01 --nomodel --shift -100 --extsize 200 for each aggregated time point. The union of aggregated time point peaks ($q < 0.01$) was generated using bedops⁶³. To generate the full read-count matrix over the union of peaks for each single cell we used bedtools (version 2-2.29.0) coverageBed⁶⁰.

Quality-control metrics. For each single cell, the numbers of total reads sequenced, uniquely mapped reads, duplicated reads and mitochondrial reads were calculated. In addition, we calculated the fraction of reads in peaks and the fraction of open chromatin as previously described⁶².

Filtering cells and peaks. To identify poor-quality single-cell sequencing for each time point, we marked cells with log₂-transformed uniquely mapped reads counts, mitochondrial content and fraction of open chromatin were twice as high as the median absolute deviation. If a cell deviates twice from any of the three metrics, it was removed prior to analysis. Peaks were removed if they overlapped with the mm9 blacklist region generated by UCSC genome browser. Peaks mapped to regions present in less than 30 cells across all cells were removed.

Dimensionality reduction. The filtered coverage matrix was binarized and used as input for latent semantic analysis (lsa) followed by principle component analysis (PCA) using R package lsa (<https://CRAN.R-project.org/package=lsa>) (version 3.6.3). UMAP was generated using 'RunUMAP' from Seurat (version 3.0).

Motif in peaks fraction and motif variability. To assess the role of transcription factor (TF) motif accessibility during memory formation we utilized R packages Biostrings⁶⁴ (version 2.54.0), TFBSTools⁶⁵ (version 1.24.0) and transcription factor motif database JASPAR2016 (ref. ⁶⁶) (version 1.14.0). For each TF motif of interest, we annotated each peak if the underlying sequence contained the motif with a

minimum matching score of 80%. Then for each cell, we counted the number of peaks with the TF motif. We then corrected the number of peaks with a motif by the number of total peaks of that cell. In addition, we assessed TF motif variability across all time points or selected time points of interest using chromVar³¹ (version 1.8) with default parameters. In total, we assessed 514 TF motifs from mouse and human from the JASPAR2016 TF motif database.

Peak annotation and enrichment analysis. Peaks were annotated with R package ChIPseeker (version 1.5.1) with transcription start site (TSS) reaching from -3000 to 3000 to the UCSC mm9 gene model and the 'org.Mm.eg.db' Bioconductor annotation package. Genes with TSS closest to peaks that were identified solely during peak calling of the aggregate of D0, D32 saline or D32 IAT were assessed for pathway enrichment using R package ReactomePA⁶⁷ (version 1.30.0).

Statistical analyses. Statistical analyses for all flow-cytometry-based protein-expression data were performed using Prism 7 (version 7.0c, GraphPad software). P values are shown as * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, unless exact P values are indicated.

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

Data availability

All relevant data from this study, including raw flow-cytometry data, are available from the corresponding author upon request. Raw single-cell RNA sequencing data from our previous publication have been deposited in the ArrayExpress under accession number [E-MTAB-4388](#). Raw scRNA-seq data, bulk ATAC-seq data and scATAC-seq data generated from the current manuscript have been deposited in the ArrayExpress under accession numbers [E-MTAB-9317](#) (10x Genomics scRNAseq data), [E-MTAB-9393](#) (bulk ATAC-seq data), [E-MTAB-9403](#) (Smart-seq2 data), and [E-MTAB-9402](#) (scATAC-seq data). JASPAR 2016 database (version 1.14.0) was used for transcription factor motif analyses.

References

- Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
- Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://arxiv.org/abs/1303.3997> (2013).
- Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
- Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
- Senabouth, A. et al. Comparative performance of the BGI and Illumina sequencing technology for single-cell RNA-sequencing. *NAR Genom. Bioinform.* **2**, lqaa034 (2020).
- Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
- Lun, A. T., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75 (2016).
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
- Santos, A., Wernersson, R. & Jensen, L. J. Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. *Nucleic Acids Res.* **43**, D1140–D1144 (2015).
- Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nat. Rev. Immunol.* **15**, 486–499 (2015).
- Carlson, M. org.Mm.eg.db: Genome wide annotation for Mouse. R package v.3.8.2 (2019).
- Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
- Chen, X., Miragaia, R. J., Natarajan, K. N. & Teichmann, S. A. A rapid and robust method for single cell chromatin accessibility profiling. *Nat. Commun.* **9**, 5345 (2018).
- Neph, S. et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
- Pages, H., Aboyoun, P., Gentleman, R., DebRoy, S. Biostrings: efficient manipulation of biological strings. R package v.2.52.0 (2019).
- Tan, G. & Lenhard, B. TFBSTools: an R/Bioconductor package for transcription factor binding site analysis. *Bioinformatics* **32**, 1555–1556 (2016).
- Tan, G. JASPAR2016: data package for JASPAR 2016. R package v.1.12.0 (2019).
- Yu, G. & He, Q. Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479 (2016).

Acknowledgements

This study was funded by the Australian National Health & Medical Research Council: Project Grant 1126399 (awarded to A.H. and S.A.T.). Smart-seq2 work was funded by ERC Consolidator grant ThDEFINE (awarded to S.A.T.). We are grateful to J. Moehrle at Medicines for Malaria Venture for providing sodium artesunate. We acknowledge the expertise and assistance of several QIMR Berghofer Medical Research Institute Core Staff: the Flow Cytometry and Microscopy Core, particularly M. Rist and T. Hong Nguyen, for single-cell sorting; the Histology Core, including C. Winterford; the Animal Facility, including all technicians involved with animal husbandry; the Next-Generation-Sequencing Core, including P. Collins for assistance with droplet-based scRNA-seq and Illumina sequencing. We acknowledge the Single-Cell Genomics Core Facility and sequencing pipeline at the Wellcome Sanger Institute for plate-based Smart-seq2 and scATAC-seq processing.

Author contributions

A.H. conceived the study with T.L., K.R.J. and S.A.T. A.H. led efforts to acquire funding and managed the project in consultation with S.A.T. M.S.F.S., H.J.L. and J.A.E. ran the study equally, managing different aspects of cellular immunology, single-cell genomics and computational biology—first authorship order is determined by workload. B.S.T., C.P.S.P., L.S.C., P.L., R.N.H., X.C., K.R.J. and L.I.M.L. carried out experimentation in consultation with M.S.F.S., H.J.L., J.A.E., S.A.T. and A.H. G.T.B., C.R.E. and S.W.L. assisted in experimental design and data analysis and interpretation. M.S.F.S., H.J.L., J.A.E., J.S., C.G.W., M.L.M., M.B., L.T.K., S.W. and D.S.K. conducted analysis in consultation with A.H., S.A.T., V.S. and M.P.D. M.S.F.S., H.J.L., J.A.E. and J.S. interpreted results in consultation with A.H., S.A.T. and K.R.J. A.H., M.S.F.S., H.J.L. and J.A.E. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

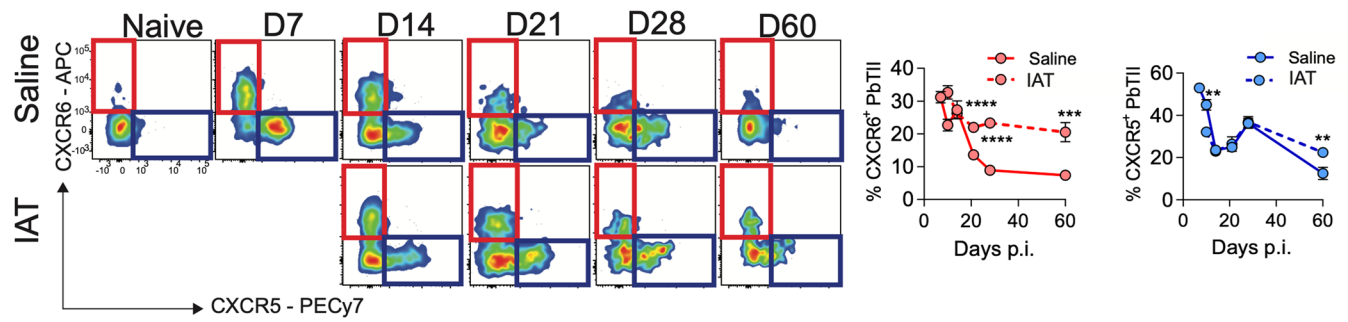
Extended data is available for this paper at <https://doi.org/10.1038/s41590-020-0800-8>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41590-020-0800-8>.

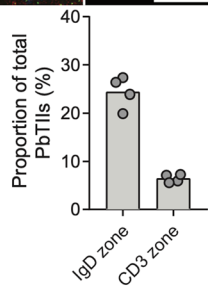
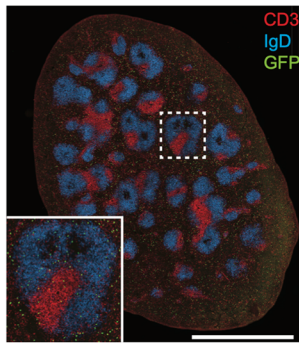
Correspondence and requests for materials should be addressed to S.A.T. or A.H.

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

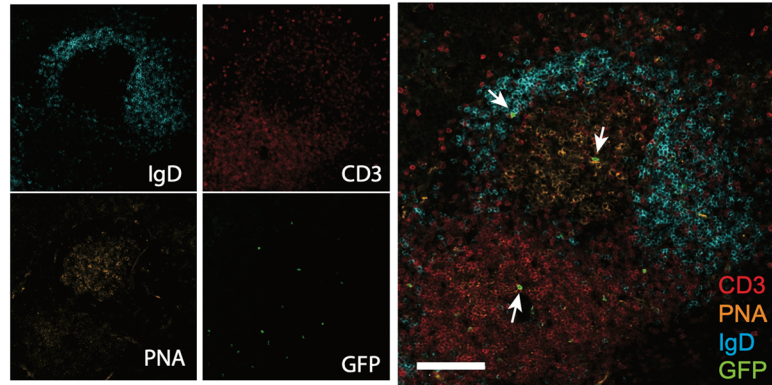
Reprints and permissions information is available at www.nature.com/reprints.



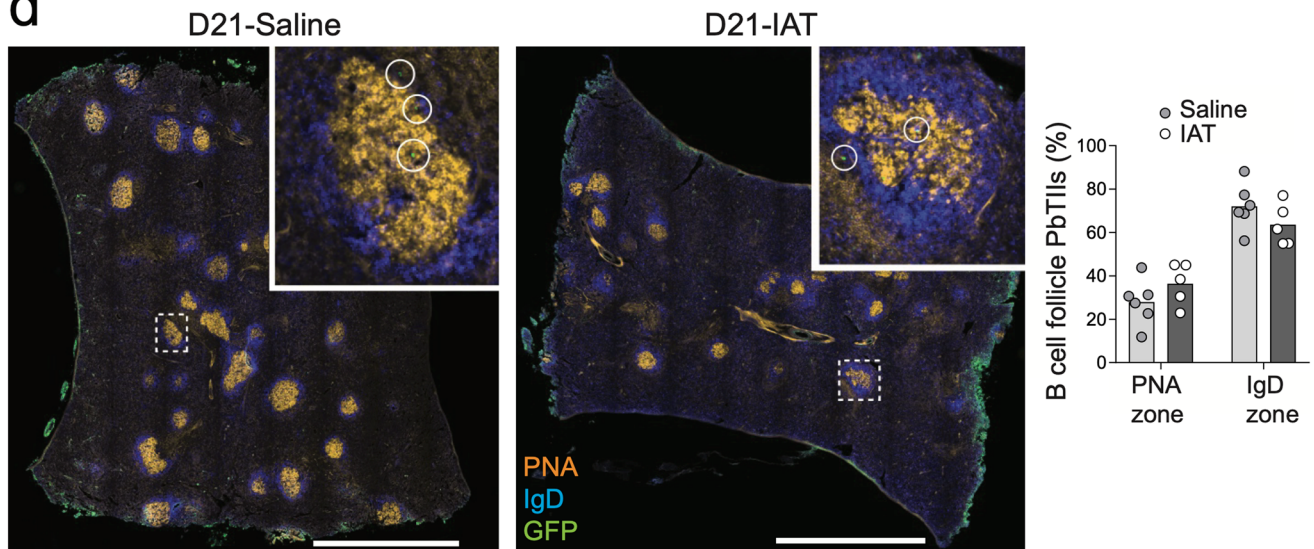
b



c

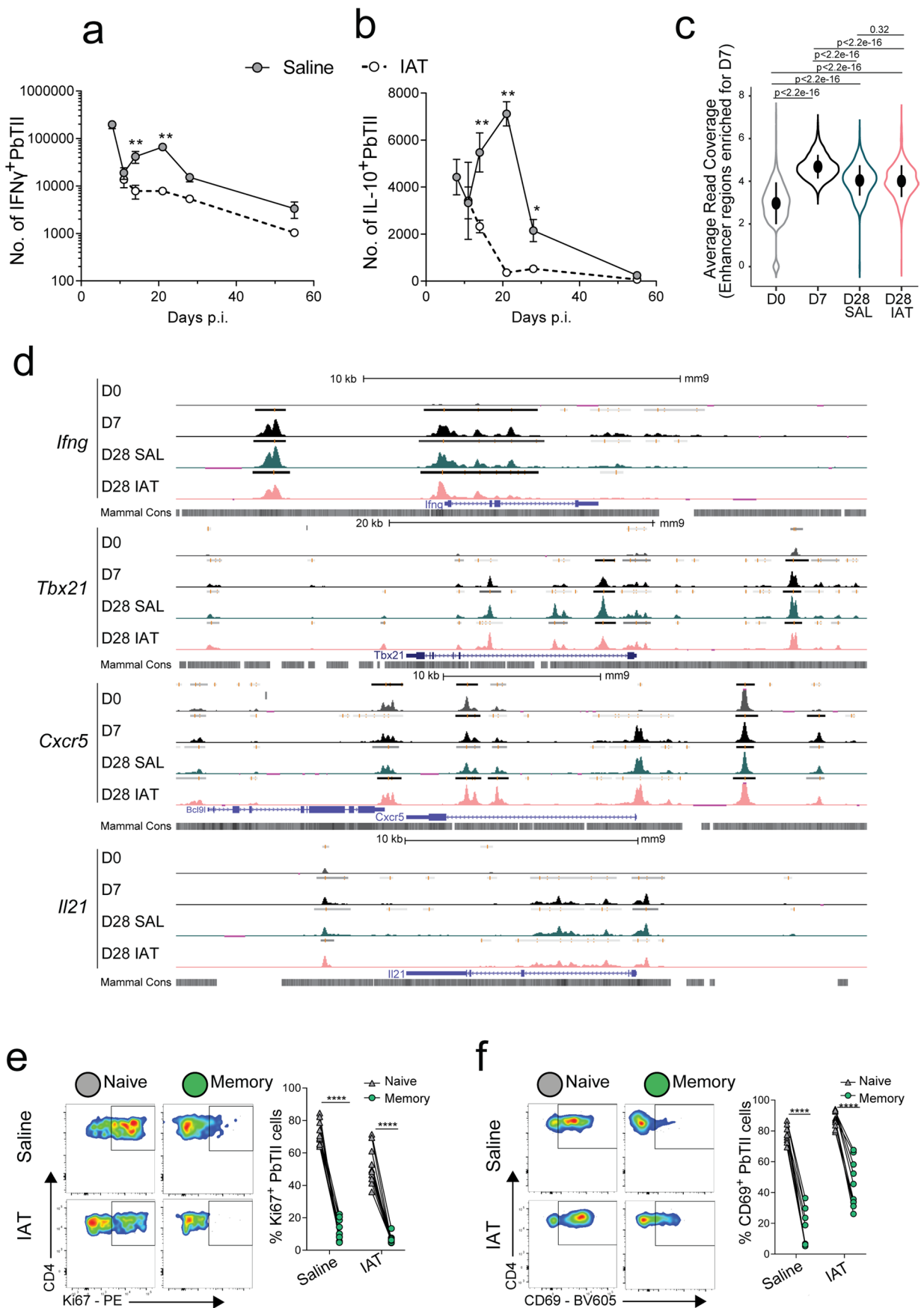


d



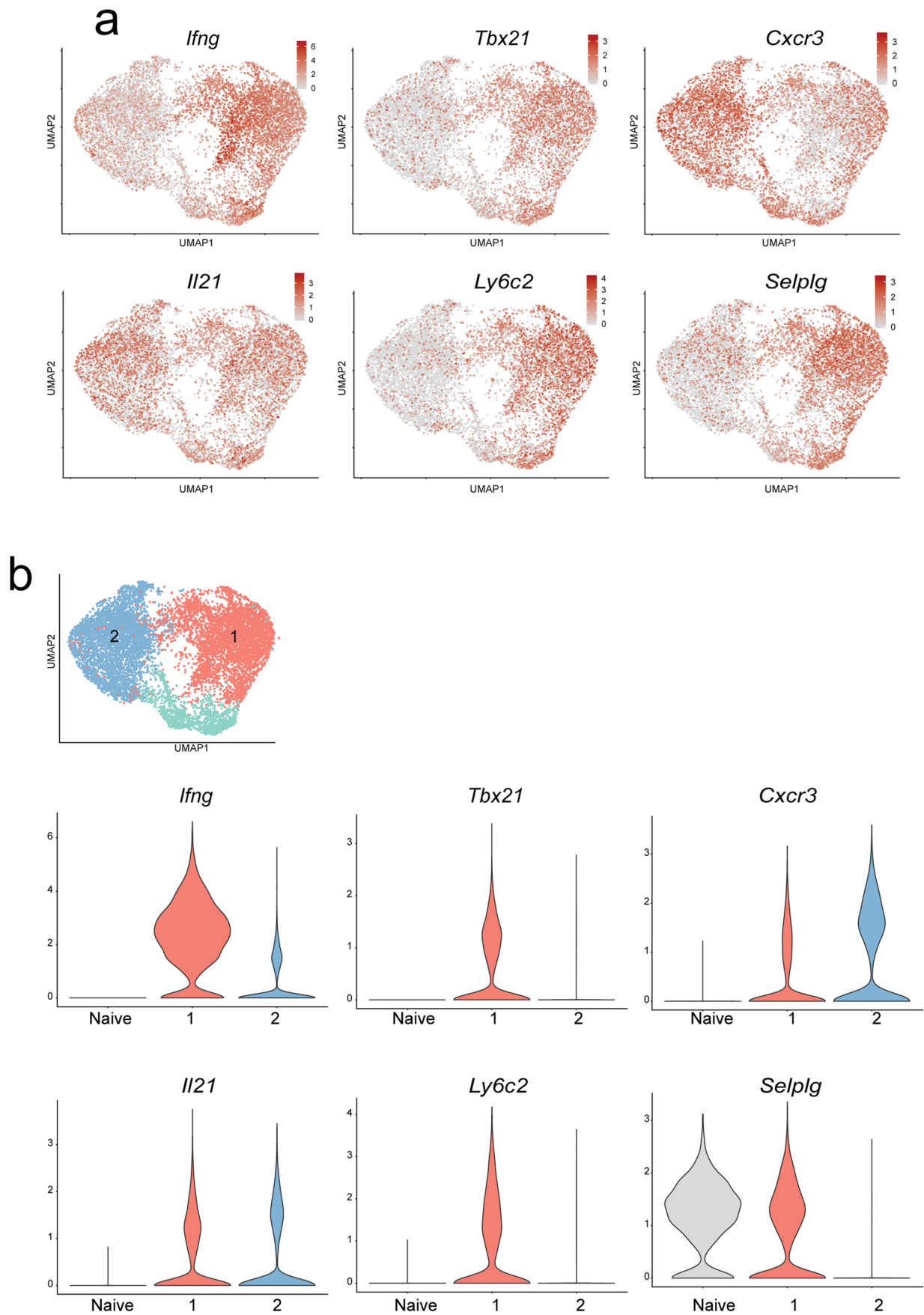
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | PbTII cells localise to T cell zones, B cell follicles and GCs during infection. **a**, Representative FACS plots for surface CXCR6 and CXCR5 expression on PbTII cells at various timepoints during infection. Graphs showing proportion of CXCR6⁺ or CXCR5⁺ PbTII cells over time. Data are pooled from 3 independent experiments (n=5 mice per group, per individual timepoint for each independent experiment) and presented as mean +/- SEM. Statistical analysis was performed between saline and IAT groups for each individual timepoint using two-tailed Mann-Whitney test. p-values where indicated are **p< 0.01, ***p<0.001, ****p<0.0001. **b**, eGFP⁺ PbTII cells were distributed throughout splenic B cell follicles and T cell zones at D7 p.i. (n=4). Graph showing proportion of PbTII cells localised within B cell (IgD) or T cell (CD3) zones. Scale bar, 2mm. **c**, Standard microanatomical structures observed in saline-treated mice at D21 p.i. White arrows indicate eGFP⁺ PbTII cells. Scale bar, 100µm. **d**, Localisation of PbTII cells within GCs and IgD⁺ naive B cell follicle regions of mice at D21 p.i. in the presence or absence of IAT. Graph showing proportion of PbTII cells localised within GCs or follicular regions of saline- (n=6) or IAT- (n=5) treated mice. Scale bar, 2mm. Images were acquired on an Aperio FL slide scanner (**b**, **d**) or a Zeiss 780-NLO point scanning confocal microscope (**c**).



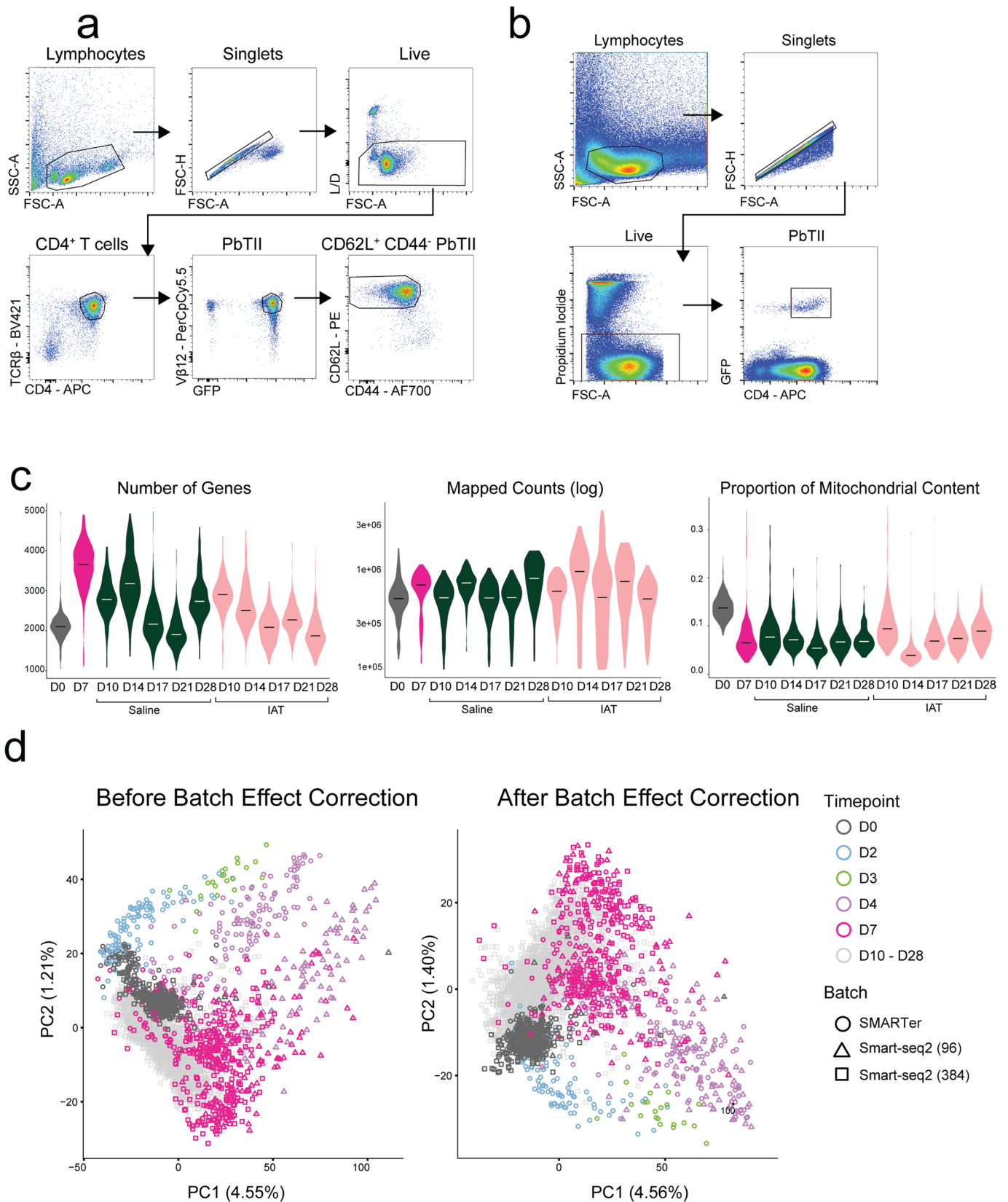
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | PbTII cells exhibit functional memory characteristics. **a, b**, Number of **(a)** IFN γ -expressing or **(b)** IL-10-expressing splenic PbTII cells after restimulation with PMA and ionomycin for 3 hours *in vitro* over time, in the presence or absence of IAT. Data are presented as mean \pm SEM. Data are representative of 2 independent experiments (n=5/ group, per individual timepoint for each experiment). Statistical test performed using two-tailed Mann-Whitney test. **c**, Average number of reads for ATAC-seq peaks within enhancer regions enriched for D7 p.i. sample. Bulk ATAC-seq experiments were performed as 2 independent experiments. Data are derived using overlapping peaks shared between biological replicates from each experiment for individual timepoints. Error bars represent mean \pm SD. Statistical test performed using two-sided Wilcoxon rank-sum test. **d**, Mean ATAC-seq peak coverage at *Irfng*, *Tbx21*, *Cxcr5* and *Il21* gene loci with a scale of 0-45 for all tracks. Boxes represent peaks called using MACS2. Data is shown for a representative experiment out of 2 independent biological repeats. **e**, Representative FACS plots and graph showing proliferative marker Ki67 expression for memory (green) or naive (grey) PbTII cells at 17 hours post-rechallenge. Data are pooled from 2 independent experiments (n= 5 mice per group, per independent experiment). Statistical test performed using paired two-way ANOVA with Tukey's multiple comparison test. **f**, Representative FACS plots and graph showing expression of early activation marker, CD69 for memory (green) or naive (grey) PbTII cells at 17 hours post-rechallenge. Data are pooled from 2 independent experiments (n= 5/ group, per independent experiment). Statistical test performed using paired two-way ANOVA with Tukey's multiple comparison test. p-values are indicated where *p< 0.05, **p< 0.01, ****p<0.0001. Statistical analysis was performed between saline- and IAT-treated groups for each individual timepoint (**a, b**).



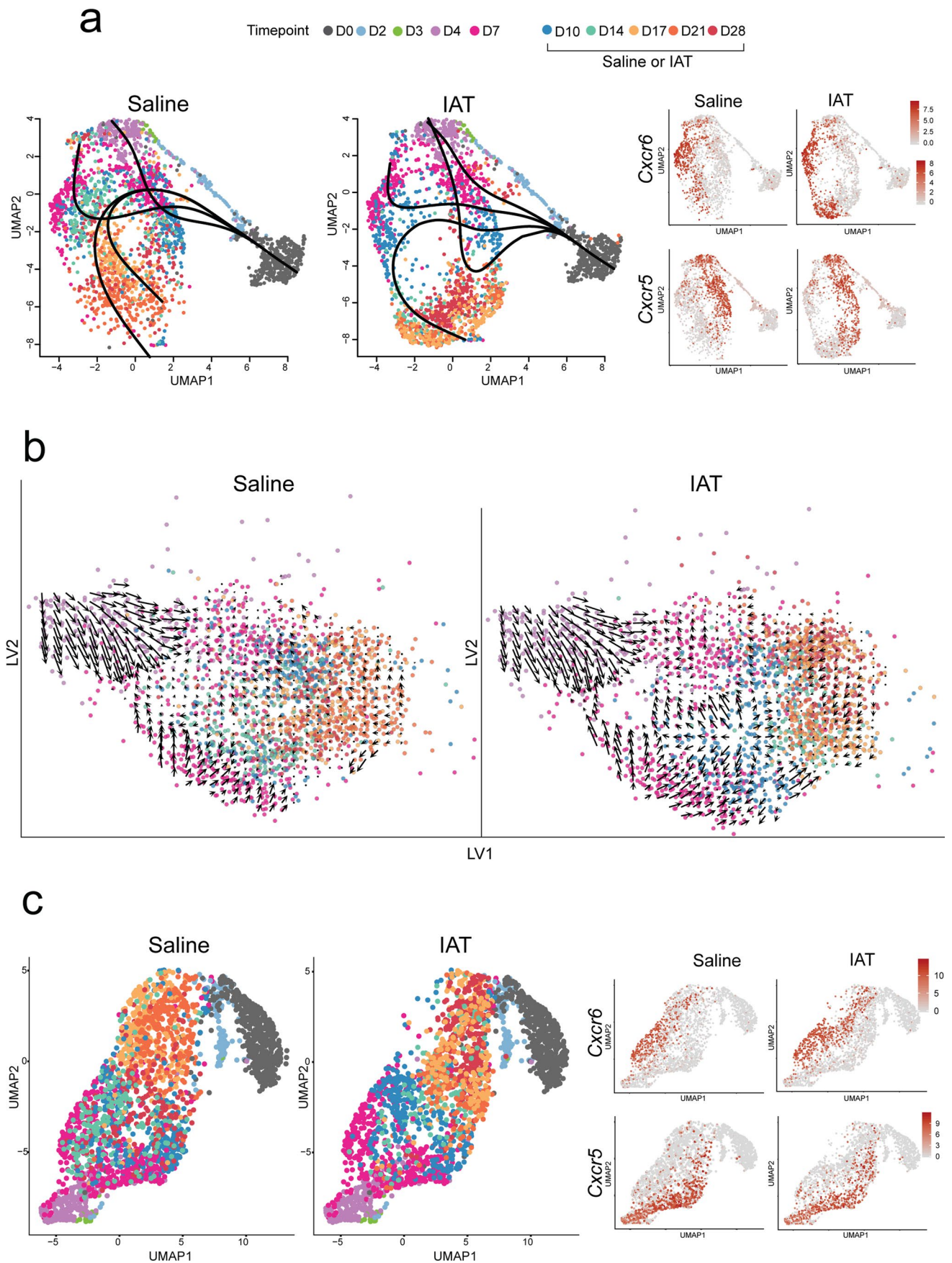
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | PbTII cells at peak effector stage display mixed expression of canonical T_H1 and T_{FH} markers. **a**, Visualisation of the expression of common markers used for T_H1 and T_{FH} lineage tracing experiments on UMAP representation of PbTII cells at D7 p.i. assessed using the droplet-based 10x Genomics platform. **b**, (*top*) UMAP representation as in (**a**) showing the individual clusters identified by unsupervised clustering analysis. (*bottom*) Violin plots showing the expression of genes described in (**a**) within each cluster. The expression value for naive (DO) PbTII cells is shown for each gene for reference.



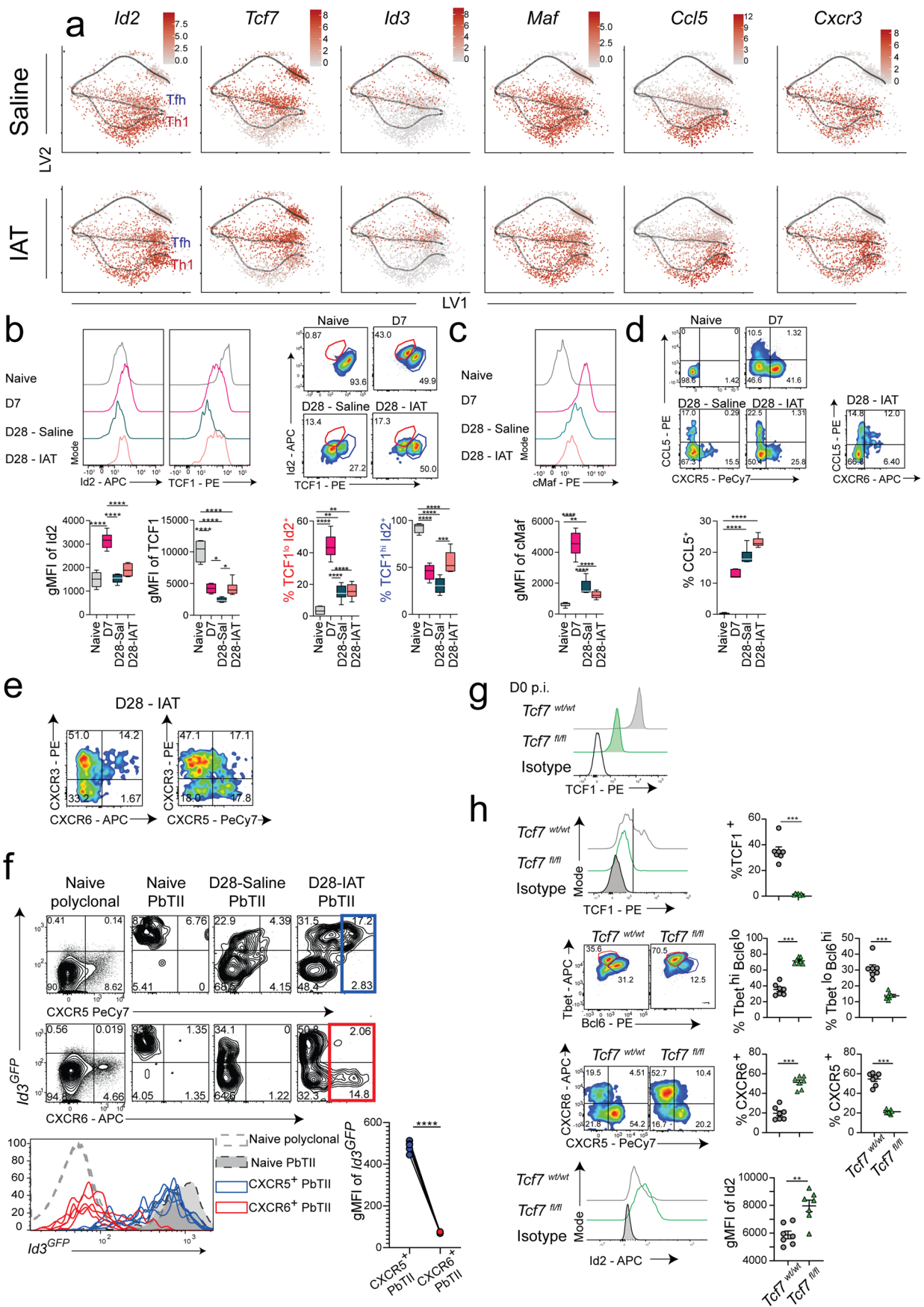
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Quality control checks for scRNA-seq assessment of memory PbTII cell differentiation. **a**, FACS gating strategy for isolation of naive (CD62L⁺ CD44⁻) donor PbTII cells for transfer into recipients, 24 hours prior to infection. A subset of these cells was also used for scRNA-seq assessment for PbTII responses at D0 p.i. Cells were then either sorted as single-cells onto 384-well plates for Smart-seq2 assessment or sorted as single-cells using the 10X Genomics platform. **b**, FACS gating strategy for isolation of PbTII cells from D7 p.i. onwards for scRNA-seq assessment for either Smart-seq2 or 10X Genomics assessment. **c**, Distribution of PbTII cells from the Smart-Seq2 dataset after filtering for number of genes ($1000 < n_{\text{Gene}} < 5000$), mapped counts ($> 100,000$) and percentage of mitochondrial content (< 0.35). **d**, The current Smart-seq2 (384) PbTII dataset of D0, D7, D10*, D14*, D17*, D21* and D28* p.i. cells were combined with our previous datasets⁸ containing PbTII cells isolated from D0 to D7 p.i. (SMARTer batch: D0, D2, D3, D4, D7 p.i.; Smart-seq2 (96) batch: D0, D4, D7 p.i.) PCA plots showing the entire time series, with shapes denoting the different experimental batches, (*left*) before and (*right*) after batch effect correction as described in methods. (*) = samples were isolated either from saline or IAT groups.



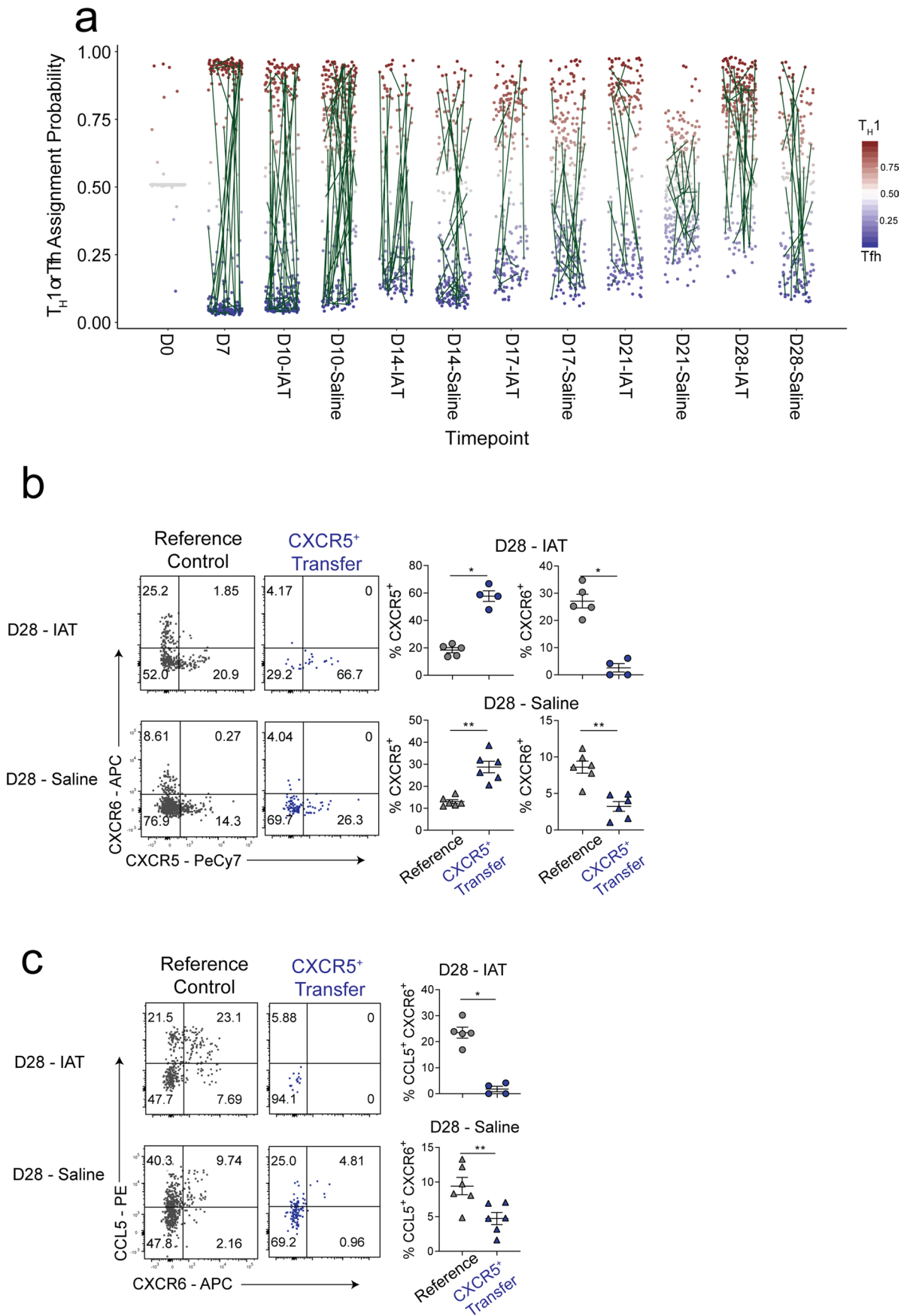
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Multiple computational approaches for trajectory inference of PbTII scRNA-seq data. **a**, (*left*) UMAP representation of batch-corrected Smart-seq2 PbTII dataset superimposed with trajectory inferences calculated using Slingshot. (*right*) Visualising *Cxcr6* and *Cxcr5* expression on UMAP representations as described previously. **b**, Grid-view of RNA velocities for each cell from the Smart-seq2 PbTII dataset (only D4-D28 p.i.) visualised on 2D bGPLVM representations. **c**, (*left*) Integration of the three PbTII datasets (Smart-seq2(96/ 384) and SMARTer) using scVI represented on a UMAP plot. (*right*) Visualising *Cxcr6* and *Cxcr5* on a UMAP representation of the scVI-integrated dataset.



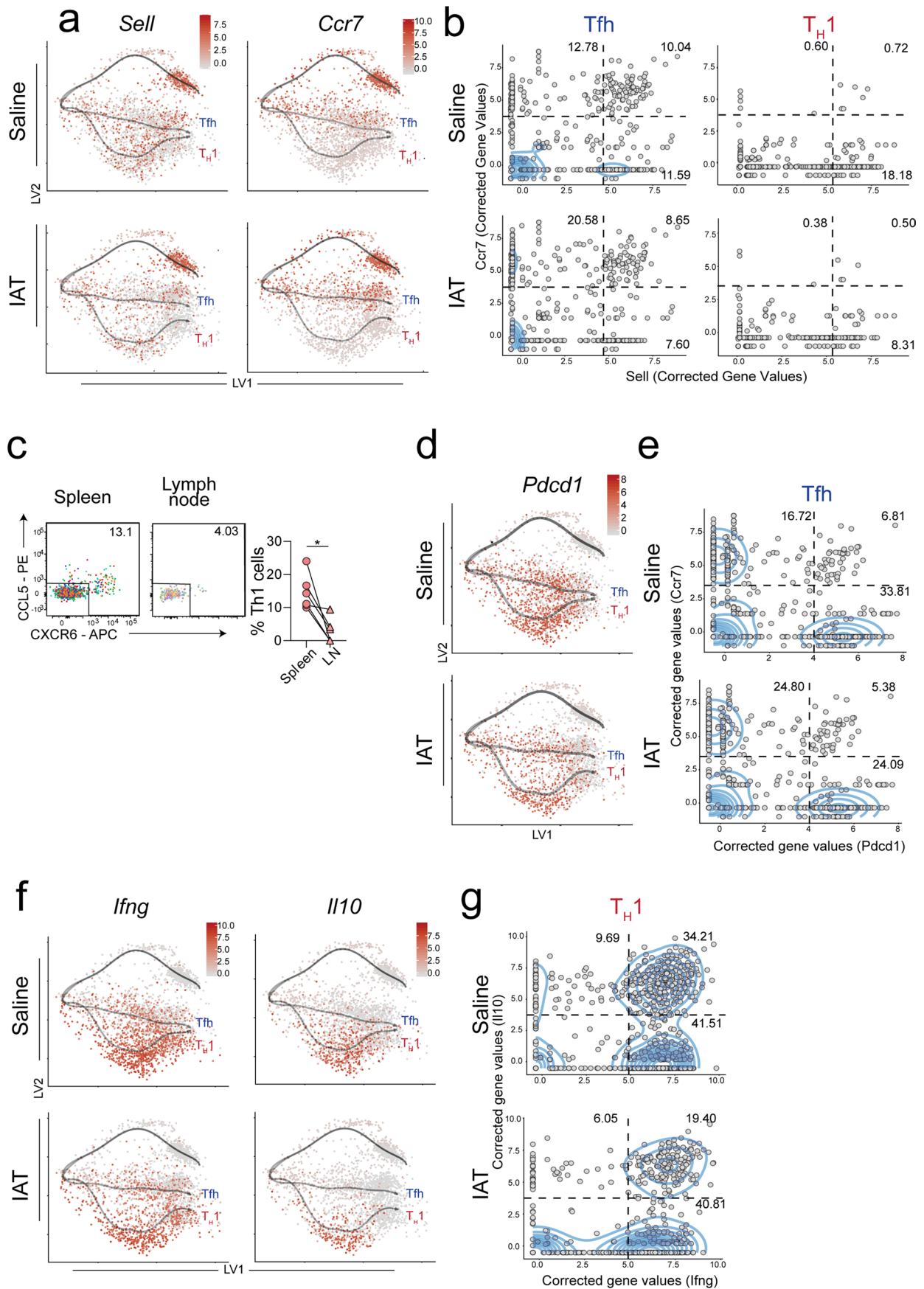
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Gene expression and protein validation of multiple memory-associated genes in PbTII cells. **a**, Visualisation of the expression of various memory-associated genes identified using GPfates modelling on 2D bGPLVM representations. LV, latent variable. **b–d**, Representative FACS plots and histograms showing protein level expression of various memory-associated markers over time, in the absence or presence of IAT. Graphs showing kinetics of Id2, TCF1, cMaf and CCL5 protein expression over time. Data are representative of 3 independent experiments (n=3 for naive and n=6 for day 7 and day 28 saline- or IAT-treated groups, per individual timepoint). For the box plot shown, the centre line indicates the median, box limits indicate the upper and lower quartiles, and the whiskers indicate the maximum and minimum measures. Statistical analysis was performed using one-way ANOVA between all groups. TCF1 staining was performed after stimulation with PMA/ Ionomycin. **e**, Representative FACS plot showing co-expression of CXCR3 with either CXCR6 or CXCR5 on PbTII cells at D28 p.i. during IAT. **f**, Representative FACS plots, overlaid histograms and graph showing the relationship between promoter activity of *Id3* from *Id3^{GFP}* reporter mice with either CXCR6 or CXCR5. Data are representative of 2 independent experiments (n= 6 mice per group, per independent experiment). Statistical analysis performed using two-tailed Wilcoxon matched-pairs signed-rank test. **g**, Expression of TCF1 in PbTII cells from individual PbTII^{cre-ERT2} *Tcf7^{wt/wt}* and PbTII^{cre-ERT2} *Tcf7^{fl/fl}* donor mice on the day of transfer, 7 days post-tamoxifen treatment. **h**, Representative FACS plots and graphs showing expression of various T_H1 and Tfh-associated markers at D7 p.i. in *Tcf7^{wt/wt}* and *Tcf7^{fl/fl}* PbTIIIs. Experiment performed once (n=7 mice per group). Data are presented as mean±/− SEM. Statistical test performed using two-tailed Mann-Whitney test. p-values are indicated where *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.



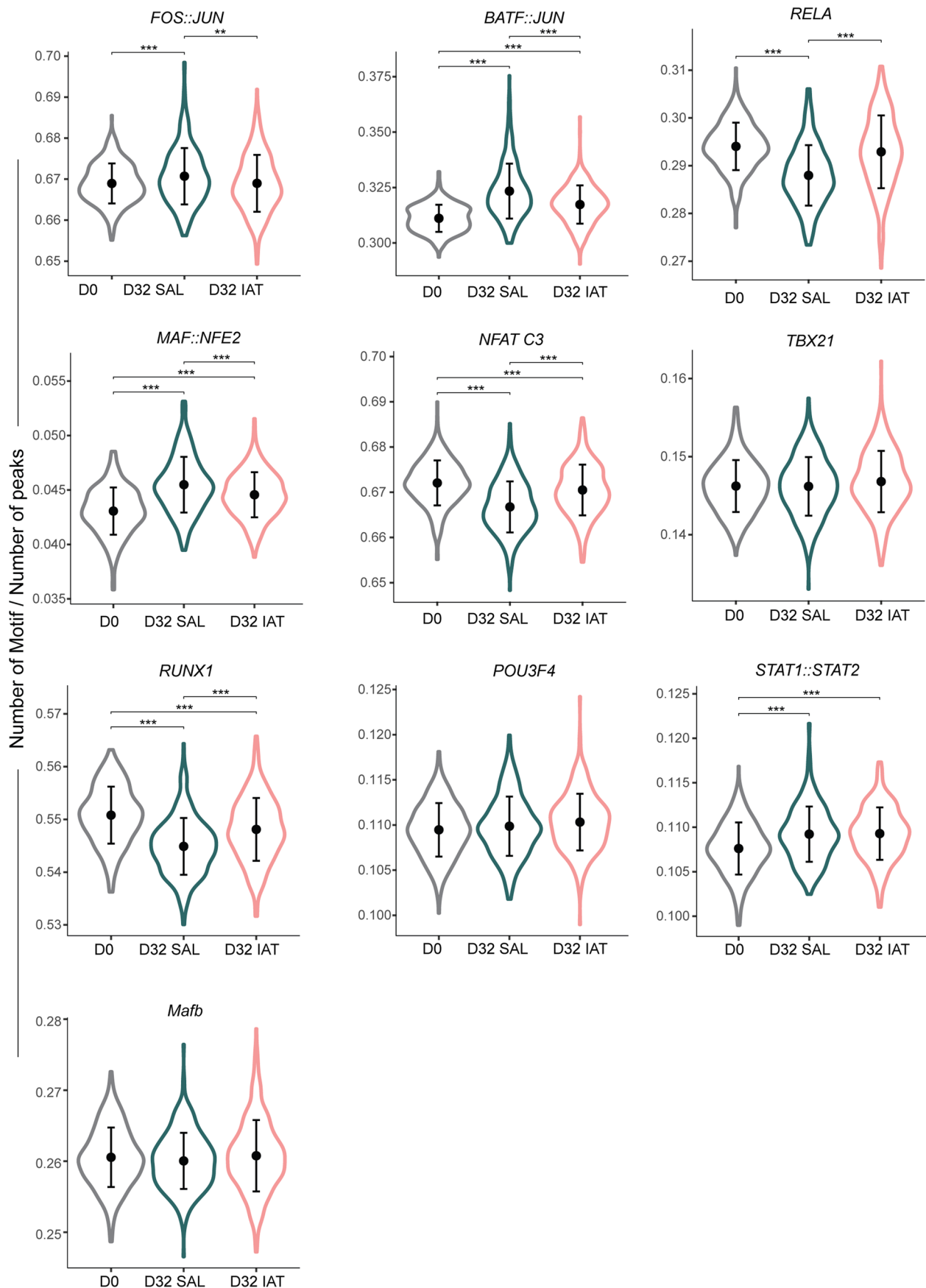
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Clonal analysis of memory fate and lineage plasticity testing with and without IAT. **a**, Families sharing endogenous TCR α and TCR β sequences displaying no crossover between timepoints or treatments. **b**, Lineage tracing of effector to memory transition of PbTII_s with and without the presence of IAT. Representative FACS plots and graphs showing expression of CXCR5 and CXCR6 in splenic PbTII_s at D28 p.i. Experiment performed once (n=5 for Reference and n=4 for CXCR5⁺ Transfer in the IAT group, and n=6 for Reference and n=6 for CXCR5⁺ Transfer in the Saline group). Statistical test performed using two-tailed Mann-Whitney test. **c**, Representative FACS plots and graphs showing co-expression of two T_H1-markers in splenic PbTII_s at D28 p.i. Experiment performed once (n=5 for Reference and n=4 for CXCR5⁺ Transfer in the IAT group, and n=6 for Reference and n=6 for CXCR5⁺ Transfer in the Saline group). Statistical test performed using two-tailed Mann-Whitney test. Data are presented as mean \pm SEM (b, c). p-values are indicated where * p<0.05, ** p<0.01.

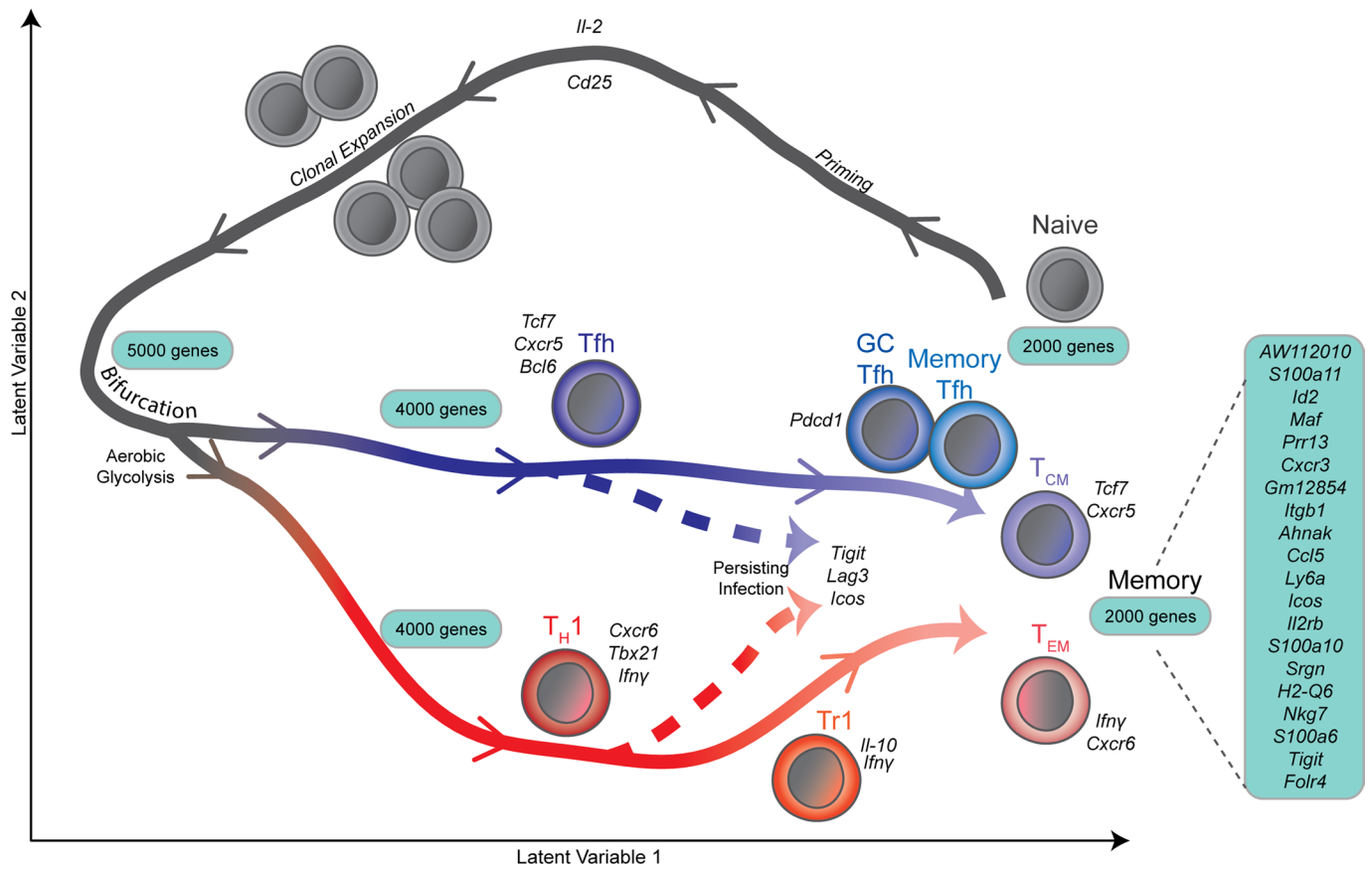


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Assessing relationships of T_H1 and Tfh-lineages with T_{CM}, GC Tfh and Tr1 subsets in the GPfates model. **a**, Visualisation of *Sell* and *Ccr7* expression on 2D bGPLVM representations overlaid with OMGP trajectories representing T_H1 or Tfh branch. **b**, Co-expression of *Sell* and *Ccr7* along pseudotime, for all cells annotated for the Tfh (left) or T_H1 (right) branch as outlined in methods. Kernel density estimation failed to estimate a population of *Sell*⁺ *Ccr7*⁺ cells, hence a threshold was estimated where cells express both *Sell* ≥ 4.7 and *Ccr7* ≥ 3.75 corrected gene values. **c**, Representative FACS plots and graph showing expression of T_H1-markers (CCL5 and CXCR6) on PbTIIIs at D28 p.i. isolated from the spleen and inguinal lymph nodes of the IAT group. Experiment performed once (n=6 mice). Statistical test performed using Wilcoxon matched-pairs signed-rank test. p-value where indicated is *p<0.05. **d**, Visualisation of *Pdcd1* expression on 2D bGPLVM representations overlaid with OMGP trajectories representing T_H1 or Tfh branch. **e**, Co-expression of *Pdcd1* and *Ccr7* along pseudotime, for all cells annotated for the Tfh branch. Kernel density estimation failed to estimate a dense population of *Pdcd1*⁺ *Ccr7*⁺ cells, hence a threshold was estimated where cells express both *Pdcd1* ≥ 4.0 and *Ccr7* ≥ 3.50 corrected gene values. **f**, Visualisation of *Il10* and *Ifng* expression on 2D bGPLVM representations overlaid with OMGP trajectories representing T_H1 or Tfh branch. **g**, Co-expression of *Il10* and *Ifng* along pseudotime for all cells annotated for T_H1. Tr1 cells are annotated as those cells expressing *Il10* and *Ifng* above the threshold drawn (cells expressing both *Il10* ≥ 4.9 and *Ifng* ≥ 3.75 corrected gene values), according to kernel density estimation.



Extended Data Fig. 9 | Changes in availability of transcription factor binding motifs in memory PbTII cells via scATAC-seq. Changes in scATAC-seq peaks for cells from naive (D0), saline and IAT groups at D32 p.i. associated with different transcription factors. Error bars represent mean \pm SD. Statistical test was performed using two-sided Wilcoxon rank-sum test. p-values are indicated where *** p <0.001.



Extended Data Fig. 10 | A conceptual view for development of memory CD4⁺ T cells during malaria. A broad overview of transcriptome dynamics accompanying transition of CD4⁺ T cells from naivety to memory. CD4⁺ T cells branch into effector T_{H1} and T_{fh} after undergoing clonal expansion, and exhibit partial retention of effector phenotypes as they transition to memory. T_{H1} effector cells give rise to Tr1 cells and T_{H1} phenotype T_{EM} memory cells. T_{fh} effector cells give rise to T_{CM} cells, GC T_{fh} cells, or memory T_{fh} cells. Genes correlating strongly with memory development are summarised in the box (right). Numbers in boxes denote the average number of genes detected by high-resolution scRNA-seq at different stages of CD4⁺ T cell differentiation.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see 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 Software used include: BD FACS DIVA software (version 8.0.1) was used for flow-sorting of cells and other relevant flow cytometry acquisition. Detailed parameters of methods are mentioned in the relevant sections in the Methods.

Data analysis Softwares used include:

Analysis of single cell RNAseq data was performed with 10x Genomics Cell Ranger (version 3.0.2), STAR (version 2.5.2), RSEM (version 1.2.30), Seurat (version 2.3.4), scVI (version 0.3.0), DBSCAN (version 0.22.1), GPFates (<https://github.com/Teichlab/GPFates>), Slingshot (version 0.99.12), Velocity (version 0.17.16), edgeR (version 3.10), TraCeR (<https://github.com/teichlab/tracer>).

Analysis of all ATAC-seq data was performed with BWA-MEM (version 0.7.15), BEDTools (version 2-2.29.0), MACS2 (version 2.1), GenomicRanges (version 1.38), ChIPseeker (version 1.22.1), Picard (version 2.19.0), Isa (version 3.6.3), Biostrings (version 2.54.0), TFBSTools (version 1.24.0), chromVAR (version 1.8), ChIPseeker (version 1.5.1), ReactomePA (version 1.30.0).

FlowJo (version 10.4) was used for analyzing flow cytometry data. Graph pad prism (version 7.0c) were used for analyses of flow cytometry data.

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 Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

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

All relevant data from this study including raw flow cytometry data are available from the corresponding author upon request. Raw single-cell RNA-sequencing data from our previous publication have been deposited in the ArrayExpress under accession number E-MTAB-4388. Raw scRNA-seq data, bulk ATAC-seq data, and scATAC-seq data generated from the current manuscript have been deposited in the ArrayExpress under accession numbers: E-MTAB-9317 (10x Genomics scRNAseq data); E-MTAB-9393 (bulk ATAC-seq data); E-MTAB-940 (Smart-seq2 data); and E-MTAB-9402 (scATAC-seq data). Other public databases used include ENSEMBL (www.ensembl.org) and JASPAR 2016 database (version 1.14.0; <http://jaspar2016.genereg.net/>).

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	In experiments where two groups of mice were compared at a single timepoint, for example, sodium artesunate drug treatment (IAT) versus control saline, we opted for a Type I error rate (alpha) of 0.05, and a power of 0.8. For comparison of two groups exhibiting a signal to noise ratio of 2 or greater (ie difference in means between two groups is twice as great at their standard deviations) 5-6 mice/group was adequate. Experimental repeats were employed in most situations unless specified otherwise to reduce Type I errors to 0.01. For droplet-based scRNAseq experiments, we estimated numbers of cells needed based on https://satijalab.org/howmanycells , developed by Rahul Satija and his research group at the New York Genome Center. Assuming 5 rare populations, the smallest present at 2% of all cells, and with a desired cell number of 100 per state, analysis of ~6000 cells permitted detection of these states with a probability of 0.95.
Data exclusions	No mice were excluded based on pre-established criteria. No exclusion was applied to the uploaded raw data in ArrayExpress. For the final count matrix, we excluded cells based on pre-established criteria for single-cells: we excluded low quality samples and contaminating cells (i.e. - cells with low number of detected genes and high mitochondria content) - exclusion criteria for each case are comprehensively detailed in the relevant Methods section.
Replication	All experiments other than scRNAseq were performed with at least 2 independent experiments (sometimes 3), unless otherwise noted in the figure legend of specific figures (Extended Data Figure 7b-c and Extended Data Figure 8c). All replicates performed showed similar findings. Exact numbers of independent experiments are detailed in each Figure legend.
Randomization	Infected mice were treated from day 7 with either IAT or control saline. At Day 0, infected mice were randomly assigned to the drug treated or the control saline group. Subsequent parasitemia measurements were not used to re-assign or manipulate groups prior to drug treatment.
Blinding	Drug-treated, control saline and naive groups of mice were not assessed in a blinded manner because a specific hypothesis was not being tested regarding the effect of anti-malarial drugs on T cell biology.

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
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies used

APC CD4 (1:200 dilution, clone: RM4-5, catalog no. 100516, Biolegend), PE CD4 (1: 200 dilution, clone: RM4-5, catalog no. 100512, Biolegend), BV421 TCR β (1:200 dilution, clone: H57-597, catalog no. 109230, Biolegend), PerCpCy5.5 TCR β (1: 200 dilution, clone: H57-597, catalog no. 109228, Biolegend), PerCpCy5.5 V β 12 (1:200 dilution, clone: MR11-1, catalog no. 46-5798-80, eBioscience), AF700 CD44 (1: 200 dilution, clone: IM7, catalog no. 03026, Biolegend, PE CD62L (1: 200 dilution, clone: MEL-14, catalog no. 104408, Biolegend), APC CXCR6 (1: 200 dilution, clone: SA051D1, catalog no. 151106, Biolegend), Biotinylated CXCR5 (1: 50 dilution, clone: 2G8, catalog no. 551960, BD Biosciences), PE Ki67 (1:200 dilution, clone: SoIA15, catalog no. 12-5698-82, eBioscience), BV605 CD69 (1: 200 dilution, clone: H1.2F3, catalog no. 104530, Biolegend), BV421 IL-7R (1: 200 dilution, clone: A7R34, catalog no. 135924, Biolegend), PE CXCR3 (1:200 dilution, clone: CXCR3-173, catalog no. 126506, Biolegend), APC Lag3 (1: 100 dilution, clone: C9B7W, catalog no. 125210, Biolegend), BV421 Tigit (1: 100 dilution, clone: 1G9, catalog no. 565270, BD Biosciences), AF700 CD45.1 (1:200 dilution, clone: A20, catalog no. 110724, Biolegend), PerCpCy5.5 CCR7 (clone 4B12, catalog no. 120116, Biolegend), APC Id2 (1: 100 dilution, clone: ILCID2, catalog no. 17-9475-82, eBioscience), PE TCF1 (1: 100 dilution, clone: C63D9, catalog no. 14456S, Cell Signalling Technology), PE CCL5 (1: 100 dilution, clone: 2F9, catalog no. 149104, Biolegend), PE c-Maf (1: 100 dilution, clone: sym0F1, catalog no. 12-9855-41, Thermo Fischer), PE IL-10 (1: 100 dilution, clone: JES5-16E3, catalog no. 505008, Biolegend), BV421 IFN γ (1: 200 dilution, clone: XMG1.2, catalog no. 505830, Biolegend), APC Tbet (1:50 dilution, clone: 4B10, catalog no. 17-5825-82, Thermo Fischer); PE Bcl6 (1: 10 dilution, clone: K112-91, catalog no. 561522, BD Biosciences); rabbit anti-GFP (1:500 dilution, catalog no. ab6556, Abcam), rat anti-mouse CD3-AF594 (1:200 dilution, clone: 1742, catalog no. 100240, Biolegend), rat anti-mouse IgD-AF647 (1:50 dilution, clone: 11-26c.2a, catalog no. 405708, Biolegend), biotinylated peanut agglutinin (1:500 dilution, catalog no. B-1075-5 PNA; Vector Laboratories, Inc.), streptavidin-AF555 (1:300 dilution, catalog no. S21381, Thermo Fisher Scientific), donkey anti-rabbit AF488 (1:300 dilution, catalog no. R37118, Thermo Fisher Scientific)

Validation

All antibodies used above have been validated by the manufacturer. The catalog number and research resource identifiers (RRID) are associated with records containing validation data and specific citations in which these antibodies were used.

APC CD4: validated in C57BL/6 mouse splenocytes (RRID: AB_312718). FACS plot shown on website.
 PE CD4: validated in C57BL/6 mouse splenocytes (RRID: AB_312715). FACS plot shown on website.
 BV421 TCR β : validated in C57BL/6 mouse splenocytes (RRID: AB_2562562). FACS plot shown on website.
 PerCpCy5.5 TCR β : validated in C57BL/6 mouse splenocytes (RRID: AB_1575173). FACS plot shown on website.
 PerCpCy5.5 V β 12: validated in BALB/c and C57BL/6 mouse lymph node cells (RRID: AB_10852552). FACS plot shown on website.
 AF700 CD44: validated in C57BL/6 mouse splenocytes (RRID: AB_493713). FACS plot shown on website.
 PE CD62L: validated in C57BL/6 mouse splenocytes (RRID: AB_313095). FACS plot shown on website.
 APC CXCR6: validated in C57BL/6 mouse splenocytes (RRID: AB_2572143). FACS plot shown on website.
 Biotinylated CXCR5: validated in mouse splenocytes (RRID: AB_394301). FACS plot shown on website.
 PE Ki67: validated in C57BL/6 mouse splenocytes (RRID: AB_11150954). FACS plot shown on website.
 BV605 CD69: validated in C57BL/6 mouse splenocytes (RRID: AB_2563062). FACS plot shown on website.
 BV421 IL-7R: validated in C57BL/6 mouse splenocytes (RRID: AB_11218800). FACS plot shown on website.
 PE CXCR3: validated in C57BL/6 mouse splenocytes (RRID: AB_1027650). FACS plot shown on website.
 APC Lag3: validated in C57BL/6 mouse splenocytes (RRID: AB_10639727). FACS plot shown on website.
 BV421 Tigit: validated in mouse splenic leucocytes (RRID: AB_2688007). FACS plot shown on website.
 AF700 CD45.1: validated in SJL mouse splenocytes (RRID: AB_493733). FACS plot shown on website.
 PerCpCy5.5 CCR7: validated in C57BL/6 mouse splenocytes (RRID: AB_2291144). FACS plot shown on website.
 APC Id2: validated in C57BL/6 mouse splenocytes (RRID: AB_2735051). FACS plot shown on website.
 PE TCF1: validated in Jurkat cells (RRID: AB_2199302). FACS plot shown on website. Used in mouse T cells in other publications such as: Utzschneider et al. (2016), T cell factor-1 expressing memory-like CD8+ T cells sustain the immune response to chronic viral infections, *Immunity*.
 PE CCL5: validated in C57BL/6 mouse splenocytes (RRID: AB_2564406). FACS plot shown on website.
 PE c-Maf: validated in Th17-polarized mouse splenocytes (RRID: AB_2572746). FACS plot shown on website.
 PE IL-10: validated in PMA/Ionomycin-stimulated Th2-polarized BALB/c mouse splenocytes (RRID: AB_315362). FACS plot shown on website.
 BV421 IFN γ : validated in C57BL/6 mouse splenocytes (RRID: AB_2563105). FACS plot shown on website.
 APC Tbet: validated in human peripheral blood lymphocytes (RRID: AB_2744712). FACS plot shown on website. Used in mouse T cells other publications such as: Zimmermann et al. (2016), T-bet expression by Th cells promotes type 1 inflammation but is dispensable for colitis, *Mucosal Immunology*.
 PE Bcl6: validated in both human and mouse B lymphocytes. (RRID: AB_10717126). FACS plot shown on website.
 rat anti-mouse CD3-AF594: validated in frozen C57BL/6 mouse spleen. (RRID: AB_2563427).
 rat anti-mouse IgD-AF647: validated in C57BL/6 mouse splenocytes. (RRID: AB_893528).

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

C57BL/6J and SJL.Ptprca mice were purchased from Animal Resources Centre (Canning Vale). All other transgenic mice used are of the C57BL/6J background and bred and maintained in-house. The specific information for all of these mice are described in the Methods section under "Experimental mice". Only females of 8-12 weeks were used for all experimental purposes. Mice were housed in exhaust ventilated cages (Opti-mice) \leq 6 per cage, where room temperature was maintained between 19°C to 22°C, with humidity range between 55% to 65% and with 12-12 hour dark/light cycle with a 15 minute sunrise and sunset.

Wild animals

Study did not involve wild animals.

Field-collected samples

Study did not involve samples collected from the field.

Ethics oversight

The study was approved by QIMR Berghofer Animal Ethics Committee (A1503-601M), in accordance to the guideline provided by the National Health and Medical Research Council of Australia.

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

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Spleens and lymph nodes were collected and homogenised through a 100 µm cell strainer to create a single cell suspension and RBCs were lysed using RBC Lysing Buffer Hybri-Max™ (Sigma-Aldrich). Cells were assessed for viability using a LIVE/DEAD™ Fixable Aqua Dead Cell Stain Kit (Life Technologies), according to the manufacturer's protocol, unless otherwise specified. Prior to antibody staining, Fc receptors were blocked using antibodies against CD16 and CD32. Cells were incubated with surface marker antibodies for 20 minutes at 4°C. To assess cytokine production and transcription factor expression, cells were incubated with brefeldin-A (10 mg/ml) with or without ionomycin (500 ng/ml) and PMA (25 ng/ml) at 37°C for 3 hours. Intracellular staining for cytokines, transcription factors and chemokines was then performed using the eBioscience™ Foxp3/Transcription Factor Staining Buffer Set. Staining with intracellular antibodies was done at 4°C for 1 hour. Where TCF1 was stained after stimulation with PMA and ionomycin for 3 hours, this was explicitly stated in the specific figure legend. All other transcription factors and chemokine molecules were stained without prior stimulation. Staining for CCR7 surface antibody was performed at 37°C for 1 hour as per manufacturer's recommendation, after Fc receptor blocking.

Livers were collected in 1% (vol/vol) fetal calf serum / phosphate-buffered saline (PBS) and homogenised through a 200 µm metal sieve. Liver suspension was resuspended in 33% (vol/vol) Percoll/ PBS before centrifugation at 1700 rpm for 12 minutes at room temperature. After removal of supernatant containing unwanted cells and debris, liver leukocyte pellet was processed similarly as splenic and lymph node tissue described above.

Instrument

Samples were acquired on a LSRII Fortessa analyser (BD Biosciences). Sorting was carried out on FACS Aria III (BD Biosciences).

Software

BD FACS DIVA software (version 8.0.1). was used to acquire data and FlowJo (version 10.4) was used for analyzing flow-sorting data.

Cell population abundance

Transgenic CD4+ PbTII T cells were found to be in a range of 0.1% - 2.0% of all single, live CD4+ TCR+ cells from a population of enriched CD4+ T cells. This range varies due to the kinetics of expansion and contraction of CD4+ T cells during infection, characterised in Fig 1. These values are according to analysis of the FCS files recorded at the time of sorting or from FACS analysis only. Purity of sorted CD4+ PbTII cells was always determined post-sort to be >99.9% of the sorted population.

Gating strategy

PbTII cells were identified as single, live (zombie aqua low), CD4+, TCR+ and GFP+ cells. All other markers of interest were gated against an isotype control where applicable or a naive infection control.

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