Single-cell transcriptomics of allo-reactive CD4^+ T cells over time reveals divergent fates during gut GVHD

Jessica A. Engel, … , Motoko Koyama, Ashraful Haque

*JCI Insight*. 2020. https://doi.org/10.1172/jci.insight.137990.

**Graphical abstract**

![Graphical abstract](image_url)

Find the latest version:

[https://jci.me/137990/pdf](https://jci.me/137990/pdf)
Single-cell transcriptomics of allo-reactive CD4+ T cells over time reveals divergent fates during gut GVHD.

Jessica A. Engel*1, Hyun Jae Lee*1, Cameron G. Williams*1, Rachel Kuns1, Stuart Olver1, Lianne I. M. Lansink1, Megan S. F. Soon1, Stacey B. Andersen2, Joseph E. Powell3,4, Valentine Svensson5, Sarah A. Teichmann6, Geoffrey R Hill1,7,8, Antiopi Varelias1, 9, Motoko Koyama7 and Ashraful Haque1,10.

Affiliations:

1QIMR Berghofer Medical Research Institute, Herston, Brisbane, Queensland 4006, Australia
2Institute for Molecular Bioscience, University of Queensland, St Lucia, Australia
3Garvan-Weizmann Centre for Cellular Genomics, Sydney, Australia.
4UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, Australia
5Flagship Labs 60 Inc, Cambridge, Mass, USA.
6Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, UK
7Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
8Division of Medical Oncology, University of Washington, Seattle, WA 98109, USA
9Faculty of Medicine, University of Queensland, St Lucia, Queensland, Australia
10Department of Microbiology and Immunology, University of Melbourne, located at The Peter Doherty Institute for Infection and Immunity, 792 Elizabeth Street, Parkville, Victoria 3000, Australia

*These authors contributed equally.

Corresponding author:

Ashraful Haque
Peter Doherty Institute for Infection and Immunity
792 Elizabeth Street
The University of Melbourne
Victoria 3000
Australia
+61 3 9035 5206
ashraful.haque@unimelb.edu.au

Conflict of interest statement: The authors declare that no conflicts of interests exist.
Abstract

Acute gastrointestinal Graft-versus-Host-Disease (GVHD) is a primary determinant of mortality after allogeneic hematopoietic stem-cell transplantation (alloSCT). It is mediated by alloreactive donor CD4⁺ T cells that differentiate into pathogenic subsets expressing IFNγ, IL-17A or GM-CSF, and is regulated by subsets expressing IL-10 and/or Foxp3. Developmental relationships between T-helper states during priming in mesenteric lymph nodes (mLN) and effector function in the GI tract remain undefined at genome-scale. We applied scRNA-seq and computational modelling to a mouse model of donor DC-mediated GVHD exacerbation, creating an atlas of putative CD4⁺ T-cell differentiation pathways in vivo. Computational trajectory inference suggested emergence of pathogenic and regulatory states along a single developmental trajectory in mLN. Importantly, we inferred an unexpected second trajectory, categorised by little proliferation or cytokine expression, reduced glycolysis, and high tcf7 expression. TCF1⁺ cells upregulated α4β7 prior to gut migration and failed to express cytokines therein. Nevertheless, they exhibited recall potential and plasticity following secondary transplantation, including cytokine or Foxp3 expression, but reduced TCF1. Thus, scRNA-seq suggested divergence of allo-reactive CD4⁺ T cells into quiescent and effector states during gut GVHD exacerbation by donor DC, reflecting putative heterogenous priming in vivo. These findings, the first at a single-cell level during GVHD over time, may assist in examination of T cell differentiation in patients undergoing alloSCT.
Allogeneic hematopoietic stem cell transplantation (alloSCT) is a curative therapy for a range of leukemias, due to the capacity of donor T cells within the transplant to kill tumour cells – known as the graft-versus-leukaemia (GVL) effect. Unfortunately, donor T cells can also damage non-cancerous host tissue, particularly the gastrointestinal (GI) tract, liver and skin, causing the serious condition, graft-versus-host-disease (GVHD). Acute GVHD in the GI tract remains the primary determinant of GVHD severity and risk of death (1). Thus, a primary goal for alloSCT is the prevention of acute gut GVHD while preserving GVL. We previously showed in pre-clinical models that donor CD4\(^+\) T cells are initially activated by recipient non-hematopoietic antigen-presenting cells (APC) within the gut, including epithelial cells that upregulate MHCII molecules (2, 3). Following this, donor-derived colonic dendritic cells (DC) also prime donor CD4\(^+\) T cells in mesenteric lymph nodes (mLN) and trigger T helper-cell differentiation, which serves to amplify and exacerbate GVHD (4). Using TCR transgenic T cells specific for a single allo-peptide, TEa cells (5), we revealed that donor CD4\(^+\) T cells within the same alloSCT recipient differentiate into multiple cellular states that express pro-inflammatory and pathogenic Th1/Th17-associated cytokines, including IFN\(\gamma\) and IL-17A, or the master transcription factor for induced regulatory T (iTreg) cells, Foxp3 (4). While cytokines such as IL-6 and IL-12 control differentiation of donor CD4\(^+\) T cells, fate-mapping studies based on the IL-17a promoter suggested complex dynamic relationships between apparent helper subsets (6). Thus, differentiation of allo-reactive donor CD4\(^+\) T cells is characterised by complexity, both in terms of dynamics and multiple cellular states adopted, neither of which have been explored at genome-scale.

Single-cell mRNA sequencing (scRNA-seq) enables unbiased genome-wide assessment of individual T cells without reliance upon pre-determined protein markers or genes. ScRNA-seq was previously employed to examine heterogeneity in CD4\(^+\) T cells isolated from IL-17a reporter mice undergoing experimental autoimmune encephalomyelitis (EAE) (7). Subsequently, scRNA-seq was used to examine CD4\(^+\) T cell differentiation during house dust mite-induced allergy (8), protozoan parasite infection (9), as well as CD8\(^+\) T cells in viral infections and cancer (10-12). Many of these studies were cross-sectional, offering insight into heterogeneity amongst clonal TCR transgenic cells at a single time-point. We previously examined CD4\(^+\) T cell transcriptomes over a range of time-points during experimental malaria, and employed computational approaches to re-construct the dynamics of Th1 versus Tfh differentiation. Using an approach based on Bayesian Gaussian Process Latent Variable Modelling (bGPLVM), we identified a bifurcation point between two trajectories, and revealed a role for T cell extrinsic factors in governing Th1/Tfh fate (9). More
recently, we employed scRNA-seq to reveal heterogeneity and tissue adaptation of thymic Tregs and colonic CD4$^+$ T cells during steady-state in mice and humans (13, 14). Here, we examined donor DC-mediated differentiation of allo-reactive donor CD4$^+$ T cells during exacerbation of acute gut GVHD using droplet-based scRNA-seq and computational modelling.
Results

Allo-reactive donor CD4⁺ T cells acquire heterogeneous pro-inflammatory, regulatory and uncharacterised states during the exacerbation phase of acute gut GVHD.

We previously established a pre-clinical model of acute GVHD exacerbation (4), in which late donor CD4⁺ T cell responses were targeted to a single, host-derived, allo-reactive peptide presented by donor MHCII molecules. In this model, TEa TCR transgenic CD4⁺ T cells (B6 background) exhibit specificity for a BALB/c-derived Ea peptide from the MHCII molecule, I-E^d, when presented by the donor (B6) MHCII molecule, I-A^b (Figure 1A). Here, we exposed BALB/c mice to total body irradiation and provided an MHC-mismatched B6 bone marrow transplant (containing donor T cells). As in our previous study (4), we opted not to model the GVL effect by instilling leukaemic cells, since tumour burdens would be relatively low during the exacerbation phase of GVHD. 12-days later, once donor allo-presenting APCs had developed (4), TEa cells were transferred. TEa cell priming occurred specifically within the mesenteric lymph nodes (mLN), and triggered differentiation, as expected by day 4, into subsets with a strong capacity to express pro-inflammatory cytokines IFNγ or IL-17A upon re-stimulation, or in rarer cases (~1%) express Foxp3 (Figure 1A & B). We also examined the frequency of TEa cells expressing IFNγ, IL-17A or Foxp3 directly ex vivo, without further stimulation (Figure 1C). Although some TEa cells expressed IFNγ, IL-17A or Foxp3 directly ex vivo (Figure 1C), the majority did not at this early time-point. Therefore, although T helper cell differentiation had occurred, use of three canonical T helper markers was insufficient for characterising the fate of most TEa cells in mLN.

Droplet-based single-cell RNA-seq and computational modelling reveals divergent fates in donor CD4⁺ T cells.

To examine donor CD4⁺ T cell differentiation without employing pre-selected markers, we opted for droplet-based scRNA-seq. TEa cells were transferred into BALB/c mice 13 days after irradiation, provision of alloSCT, and subsequent GVHD initiation. TEa cells were then recovered from mLN at days 1, 2, 3 & 4 post-transfer (Figure 2A and Figure S1A & B) – non-transferred control TEa cells were also examined, referred to as day 0 cells. Flow cytometric assessment of sorted TEa cells - pooled from mice at each time-point – revealed as expected, rapid, transient upregulation of CD69 and evidence of cell division by CFSE dilution by day 2 (Figure S1B). This was followed on days 3 and 4 by complete loss of CFSE, indicative of dramatic clonal expansion, as well as substantial upregulation of the gut homing integrin, α4β7 in many, but not all TEa cells. These data confirmed that TEa cell activation had occurred in mLN, and suggested emerging heterogeneity. TEa cells were therefore processed for droplet-
based scRNA-seq. After excluding poor-quality single-cell transcriptomes (Figure S2A & B), we advanced 22,854 high-quality TEa samples for further analysis. We noted a substantial increase in the average number of genes detected per cell from day 0 through day 1 and 2, which dropped by day 4 (Figure S2A). This was consistent with our previous study in which CD4+ T cells more than doubled the number of detected genes during clonal expansion (9).

Uniform Manifold Approximation and Projection (UMAP) (15) visualisation after Principal Component Analysis (PCA) suggested that day 0 and day 1 TEa cells, whilst relatively homogeneous within their own timepoints, were distinct from each other, and furthermore were distinct from cells assessed at days 2-4 (Figure S2C). In contrast, there was evidence both of transcriptomic overlap between cells from days 2, 3 and 4, as well as of transcriptomic progression from day 2 through to day 4 (Figure S2C). Therefore, we sought to model potential developmental trajectories within the scRNA-seq data. We employed non-linear probabilistic PCA, termed bGPLVM (9), which embedded the data in low-dimensional space and re-ordered transcriptomes independently of the time-point of capture. Previously we employed bGPLVM on a relatively small number of cells (<1000) (9). It was evident, here, that bGPLVM was scalable to ~25,000 cells. Running bGPLVM iteratively ten times on the dataset yielded similar learned embeddings, indicating the stability of the output (Figure S3). Variability between cells was clearly observable along three latent variables, with interpretable variation also evident in the first two latent variables, thus permitting modelling in two dimensions (Figure 2B).

We next sought to infer differentiation trajectories based on transcriptomic similarity between cells. We first employed Slingshot (16), a top-ranked trajectory inference tool (17) on the bGPLVM embedding with a start point specified in day 1, and with day 0 omitted due to transcriptomic distance from the rest of the data (Figure S4A & B). Slingshot suggested three potential trajectories in the bGPLVM space, two of which (trajectories I & II) progressed through days 2, 3 and 4, while trajectory III appeared to terminate in day 2 (Figure 2B). Consistent with flow cytometric data (Figure S1), Cd69 expression gradually reduced along trajectories I & II, while Itgb7, encoding subunit β7 of integrin α4β7, increased (Figure 2C & D). Most strikingly, trajectory II ceased expression of cell-cycling genes including Mki67 (encoding Ki-67)(Figure 2C & D) and exhibited lower expression of aerobic glycolysis genes compared to trajectory I.

We also employed Slingshot on UMAP of PCs and found four trajectories, two of which largely superimposed upon each other suggesting, as for bGPLVM, three main trajectories (Figure S5A). Two of these progressed through days 2-4, with one largely devoid of cellular proliferation, and the other exhibiting strong expression of Mki67 and other cycling genes.
To further test for trajectories in the data, we employed other highly-ranked computational approaches (17), including Monocle 2, Velocyto (18), and PAGA (19). In all cases (Figure S6), there appeared one main bifurcation event within days 2-3, leading to two trajectories. Therefore, taken together, our analysis suggested two apparent trajectories had emerged in TEa cells, which differed from each other in expression of genes including those related to cell cycling and aerobic glycolysis.

Pro-inflammatory and regulatory effector fates emerge within one trajectory, revealing a second, quiescent Tcf7-expressing fate.

We next assessed expression of the canonical markers, Ifng, Il17a and Foxp3 within our bGLPVM/Slingshot model (Figure 3A), and surprisingly, found them expressed only in trajectory I. In addition, we noted significant expression of Il17a and Il17f (Figure 3, Il17f not shown), in areas shared by trajectory III and I (Figure 3A). These observations pointed to early, partly transient, activation of the Il17a promoter at day 2, followed later by upregulation of Ifng and Foxp3. Most importantly, we saw little evidence of Th1, Th17 and iTReg cell types emerging from different developmental trajectories, either when examined using our previous Th gene modules or when relatively sparse Ifng, Il17a and Foxp3 expression levels were imputed using Adaptively-thresholded Low-Rank Approximation (ALRA) (20) (Figure S7). Instead, our analysis suggested pro-inflammatory and regulatory fates had emerged within trajectory I.

Interestingly, trajectory II lacked expression of Ifng, Il17a and Foxp3 (Figure 3A). Differential gene expression analysis between cells in trajectory II vs I revealed Tcf7 as a top transcription factor associated with trajectory II, as well as elevated expression of central memory-associated Ccr7 and Sell (encoding CD62L)(Figure 3B and Table 1 & S1). These observations were also seen in our UMAP/Slingshot model and in a second scRNA-seq experimental repeat (Figure S8 and Table 1). Thus, given an absence of cell cycle activity, lower expression of aerobic glycolysis genes, absence of pro-inflammatory or regulatory gene expression, and expression of memory or stem-like genes including Tcf7, we inferred that trajectory II contained quiescent TEa cells that had gone through a clonal burst in mLN, but had not acquired effector function.

We further examined genes expressed at a higher level in Tcf7hi TEa cells compared to effector counterparts (Table 1 and Table S1). We noted several genes reported to support T-cell quiescence, including Btg1, which mediates deadenylation and degradation of mRNA (21), Mxd4, a Myc-antagonist (22), Samhd1, which mediates degradation of dNTPs required for DNA synthesis, Laptm5, which has been reported to mediate lysosomal targeting and
degradation of components of the CD3 complex (23), and Ms4a4b, which may limit T cell-cycling (24). Similarly, Gimap3 encodes a GTP-binding protein linked with T-cell longevity (25), while Izumo1R, encoding Folate Receptor 4, was recently reported on long-lived memory Tfh cells (26). In addition, we noted genes reported to control T helper fate, including Klf2 (27, 28), and trafficking, such as Ccr7 and Rgs10 (29, 30). Thus, our transcriptomic data were broadly consistent with Tcf7hi TEa cells exhibiting a phenotype associated with central memory, quiescence and longevity.

To test predictions from our transcriptomic models, we assessed mLN TEa cells at day 4 post-transfer by flow cytometry. As expected (31), all TEa cells had upregulated the Th1-associated lineage transcription factor, T-bet (Figure 4A). We also observed a clear bifurcation in expression of the Tcf7-encoded protein, TCF1 (Figure 4A), with one population expressing higher levels of TCF1 and lower levels of T-bet than its counterpart (Figure 4A). We also noted that direct ex vivo expression of IFNγ or IL-17A was substantially reduced in TCF1hi cells relative to TCF1lo counterparts. Strikingly, Foxp3 expression was absent in TCF1hi cells relative to TCF1lo cells (Figure 4A), and in vitro re-stimulation did not recover expression of these molecules, (Figure 4B). Together these data were consistent with scRNA-seq prediction of the emergence of allo-reactive CD4+ T cells that were quiescent and marked by high expression of TCF1.

Tcf7hi allo-reactive CD4+ T cells change minimally during migration from mLN to gut.

Although TCF1hi TEa cells within the mLN failed to express Foxp3, or pro-inflammatory cytokines, IFNγ and IL-17A, scRNA-seq predicted their capacity to migrate to the gut due to expression of the integrin gene Itgb7 (Figure 2C). To test this, and to explore the developmental relationships between mLN and gut-migrating TEa cells, we conducted a third scRNA-seq experiment, examining TEa cells at day 5, both in mLN and in the intra-epithelial lymphocyte (IEL) fraction of the gut (Figure 5A).

TEa cells were readily recovered from lamina propria (LP) and IEL fractions from the gut. However, given the longer protocol required to isolate cells from LP versus IEL, and the potential to interfere with transcriptome fidelity, we confined scRNA-seq analysis to IEL TEa cells. To integrate day 5 data with our previous mLN dataset, we repeated assessments of days 0 and 4 mLN. Finally, to control for possible technical variation in IEL induced by the isolation protocol, we treated day 5 mLN cells with and without the IEL-isolation pre-digestion buffer. After scRNA-seq, and quality control as before (Figure S9A), we assessed 21,632 cells across all samples.
Firstly, we noted no effect of the IEL isolation buffer on transcriptomes (Figure S9B), indicating that direct comparison of cells from mLN and IEL was possible. Next, unsupervised clustering of day 5 IEL cells, revealed four main clusters (0, 1, 2 & 3), and minor cluster 4, with \textit{Ifng}, \textit{Il17a} and \textit{Foxp3} upregulated across a broad area in Clusters 0, 1 & 3 but not 2 (Figure S9C). The frequency of TEa cells expressing pro-inflammatory cytokine genes was substantially elevated compared to mLN, with ~15% and ~40% of TEa cells in Clusters 0, 1 & 3 expressing \textit{Il17a} or \textit{Ifng} respectively (Figure S9C). Pronounced expression also revealed that patterns for \textit{Ifng} and \textit{Il17a} expression were not identical in IEL TEa cells. While \textit{Ifng} was relatively uniform in expression across Clusters 0, 1, & 3, \textit{IL17A} was focussed in specific areas. However, a clear distinction between Th1 and Th17 could not be drawn transcriptomically, consistent with previous data that allo-reactive CD4$^+$ T cells can co-express IFN\textgamma and IL-17A at protein level. Interestingly, \textit{Foxp3} was not tightly confined to a particular area in IEL cells, suggesting that iTregs were transcriptomically varied, and could not be partitioned into a specific lineage separate from those expressing pro-inflammatory cytokines. Together, these data reveal that even after priming in mLN and migration to the gut, allo-reactive effector CD4$^+$ T cells remain transcriptomically similar to each other, regardless of their pro-inflammatory or regulatory phenotype.

Unsupervised clustering of IEL TEa cells also revealed Cluster 2, which completely lacked expression of \textit{Ifng}, \textit{Il17a} and \textit{Foxp3}, but was elevated for \textit{Tcf7}, \textit{Ccr7} and \textit{Cd27} expression relative to other clusters (Figure S9C). Cluster 4 contained a small population of cells that could not be reliably analysed (Figure S9D). Thus, our assessment of TEa cells in the gut revealed the presence of both effector and quiescent cell states, mirroring those seen in mLN.

To determine molecular relationships between TEa cells in mLN and gut, we integrated our datasets across all time-points and organs using single-cell Variational Inference (scVI) (32), which accounted for possible batch effects across independent experiments, and provided a temporal atlas of differentiation for allo-reactive CD4$^+$ T cells (Figure S10A & B). Output from scVI was first visualised via 3-dimensional UMAP. Day 0 and day 4 mLN TEa cells, regardless of experiment, occupied the same space as their time-point counterparts, suggesting that any technical variation from different experiments, sequencing platforms, and protocols had been removed. Next, we noted that day 0 and day 1 cells existed in discrete transcriptional states, separate from each other and from day 2-5 cells (Figure S10A & B). This suggested TEa cells had undergone rapid and uniform change upon initial exposure to allo-antigen in mLN, with potential intermediate cellular states not captured by scRNA-seq assessment at a single time-point 24 hours after transfer. Similarly, we noted few
transcriptomic intermediates between day 1 and day 2, suggesting again further uniform, rapid
change during the second 24-hour period. Differential gene expression analysis between
consecutive days revealed gene families associated firstly with ribosomal processes, then
cellular division were upregulated during the first 48-hour period of allo-antigen exposure
(Table S2), consistent with initiation of clonal expansion.

Next, we noted from day 2-5 in mLN, a substantial, gradual increase in heterogeneity,
with modest effector molecule expression, such as Ifng, Il17 and Foxp3 expression confined
to one space, with quiescent Tcf7hi cells occupying a separate space (Figure 5C & D).
Importantly, from our integrated model we inferred further transcriptomic change as effector
cells migrated from mLN to gut, including for example, upregulation of Csf2 (encoding GM-
CSF), Tr1-associated immune-suppressive Il10, and increased expression of Ifng, Il17a and
Foxp3 (Figure S10C). This suggested differentiation into IFNg+ or IL-17A+ pro-inflammatory
effectors, Foxp3+ iTreg cells or IL-10+ Tr1 cells, though initiated in secondary lymphoid tissue
appeared to continue during and after migration to the gut. In contrast, we noted substantial
transcriptomic overlap between Tcf7hi cells in mLN versus IEL. Moreover, once Tcf7hi cells had
emerged in mLN by day 3, their phenotype altered very little either over the following 2 days,
or during migration to the gut (Figure 5D). These data suggest that Tcf7hi cells remained
transcriptomically stable across different tissues, while effector cells, including those
expressing Ifng, Il17a, or Foxp3, underwent progressive transcriptomic change over this
period.

Finally, using our scVI model, we sought to further study how quiescent, allo-reactive
CD4+ T cells might develop in mLN during acute gut GVHD. Trajectory inference, for example
using Slingshot on bGPLVM or UMAP embedding, assumes developmental changes in cells
are gradual enough to capture intermediate states. However, while emergence of effector cells
appeared gradual in the mLN and during migration to the gut, this was less clear for
emergence of Tcf7hi cells. Transcriptomic intermediate states were apparent between Tcf7hi
and Tcf7lo cells by days 4-5 (Figure 5D), consistent with possible linear transitions from effector
to memory. However, at day 2, prior to effector differentiation and clonal expansion, small
numbers of cells appeared in the quiescent, low cell-cycling Tcf7hi population (Figure 5D),
suggesting an alternative mechanism of development unlinked to effector differentiation. Thus,
scRNA-seq analysis suggested that some quiescent, Tcf7hi allo-reactive CD4+ T cells could
have emerged rapidly within 48 hours of allo-presentation in mLN, with developmental
intermediates being difficult to capture in our experimental design. In summary, our integrated
atlas of allo-reactive CD4+ T cell differentiation

(https://camerongw.github.io/Engel_Lee_Williams_Supplementary_File.html), revealed the
emergence of pro-inflammatory, regulatory, and quiescent cell states within secondary lymphoid tissue, which emerged rapidly and evolved to differing degrees during migration to the gut.

**Clonally-expanded, TCF1^{hi} CD4^{+} T cells can mount secondary effector responses in vivo**

Finally, we sought to determine the functional potential of quiescent TCF1^{hi} TEa cells. We tested the hypothesis that TCF1^{hi} TEa cells could give rise to a secondary effector response and re-generate themselves within the gut. Firstly, we noted these cells expressed less of the canonical co-inhibitory markers *Pdcd1* (encoding PD-1), *Havcr2* (encoding Tim-3), and *Tigit* but not *Lag3*, compared with effector TEa cells, consistent with the idea that TCF1^{hi} TEa cells might be responsive to re-activation in the gut (Figure S11A). Secondly, as for naïve cells, *Tcf7^{hi}* cells expressed higher levels of both *Il6st* and *Il6ra* than effectors, which are required for classical IL-6 signalling that promotes CD4^{+} T cell responses in acute gut GVHD (Figure S11B). These data suggested that *Tcf7^{hi}* cells might retain the capacity to respond to allo-antigen.

In the absence of a *Tcf7* reporter system, we designed a cell-sorting strategy to enrich for *Tcf7^{hi}* TEa cells and remove effector counter-parts. We noted in our bGPLVM/Slingshot scRNA-seq model, that trajectory II cells expressed much lower levels of *IL2ra* (encoding CD25) and the Th1-associated chemokine receptor gene, *Cxcr6* compared with trajectory I effector cells (Figure 6A). Consistent with this, flow cytometric assessment of day 4 mLN revealed that IFNγ/IL-17A production and Foxp3 expression was confined to CD25^{−} TCF1^{lo} TEa cells (Figure 6B & C). Given that CD25^{−} TEa cells uniformly expressed higher levels of intracellular TCF1 compared to CD25^{+} counterparts (Figure 6D), we concluded that cell-sorting clonally expanded cells based on the lack of CD25/CXCR6 expression and loss of CellTrace™ dye, provided a feasible alternative to employing a *Tcf7*-reporter (Figure 6E).

We sorted clonally-expanded (CFSE^{lo}) CD25^{−} CXCR6^{−} TEa cells from mLN at day 4 post-transfer, and transferred these or naïve TEa comparator cells into irradiated BALB/c alloSCT recipients (Figure 7A & Figure 6E). Although our experiment was designed to directly compare the abilities of antigen-experienced TCF1^{hi} TEa and naïve TEa cells to mount effector responses and regenerate themselves, we also transferred CD25^{+} and/or CXCR6^{+} counterparts into a third cohort of alloSCT recipients - due to their lower prevalence, this was performed with 50% fewer cells than the other two groups. When assessed 4 days later, recipients of TCF1^{+} TEa cells harboured as many cells as those receiving naïve TEa cells (Figure 7B). In contrast, recipients of TCF1^{−} TEa cells, albeit transferred with 50% fewer cells
harboured ~10% of the original input, compared to 40-60% for naïve and TCF1\textsuperscript{hi} cells. While naïve TEa cells diverged to give rise to both TCF1\textsuperscript{hi} and TCF1\textsuperscript{lo} cells, TCF1\textsuperscript{hi} TEa cells were less able to re-generate the TCF1\textsuperscript{hi} phenotype in a secondary transplant, and TCF1\textsuperscript{lo} TEa cells were almost incapable of doing so (Figure 7C). TCF1\textsuperscript{hi}-derived TEa cells were capable of expressing IFN\(\gamma\), IL-17A or Foxp3 directly ex vivo and after re-stimulation (Figure 7D & E), with IL-17A expression increased relative to primary responses by naïve cells. Together, these data suggested that antigen-experienced TCF1\textsuperscript{+} CD4\textsuperscript{+} T cells, though quiescent after clonal expansion in a primary response, could mount secondary effector responses, but regenerated the TCF1\textsuperscript{hi} phenotype poorly. Thus, our data are consistent with the priming of quiescent, gut migratory, allo-reactive CD4\textsuperscript{+} T cells, which nevertheless retain effector potential.
Discussion

Although alloSCT is an established curative therapy for a range of hematological malignancies, a major limitation is acute GVHD, where alloreactive naive donor T cells differentiate into pro-inflammatory effectors that damage the GI tract, liver and skin (1). Cytokines such as IL-17A, IFNγ and GM-CSF produced by allo-reactive Th1 and Th17 cells in the GI tract promote disease (33), while IL-10 produced by Tr1 and iTreg cells can protect (31). An important goal in alloSCT is to preserve GVL effects while reducing GVHD. Key to this endeavour is a consideration of the spatial and temporal differences between GVL and GVHD. While GVL exerts beneficial effects in primary and secondary lymphoid organs, acute and lethal GVHD often occurs in the GI tract. By understanding CD4+ T cell differentiation in the gut after alloSCT, we may define new strategies to block pathogenic cellular states and encourage protective ones. Although CD4+ T cell differentiation has been explored at genome-scale in infection, auto-immune and allergy models (7-9), extrapolating to alloSCT remains challenging. For example, in the alloSCT setting, alloantigen is ubiquitous and constant, while pathogen-derived antigen may be more dynamic or transient. Secondly, alloSCT often features profound lymphopenia unlike other models. Hence, we specifically examined transcriptome dynamics of T helper cell differentiation in mLN and the GI tract of mice after alloSCT.

By sampling transcriptomes from thousands of alloreactive CD4+ T cells of a single specificity across lymphoid and nonlymphoid gut-associated tissue, we detected cellular states expressing canonical Th1/Th17 cytokine genes, Ifng and Il17a, and the regulatory genes, Foxp3 and Il10 – at frequencies similar to that observed by flow cytometry. Importantly, however, unbiased clustering and trajectory inference tools suggested substantial similarity between the transcriptomes of these effector subsets, particularly in lymphoid tissue but also in the gut. Subtle differences became more evident amongst gut trafficked TEa cells that had stopped proliferating, with Ifng expressed more uniformly than either Il17a or Foxp3. Given that TEa cells can upregulate T-bet and IFNγ but not IL-17A or Foxp3 in the complete absence of MHCII-presentation (2, 4), our data are broadly consistent with a Th1-like state being a default program in the gut, that may be countered by alloantigen presentation via MHCII towards iTreg or Th17-like states. Nevertheless, our main inference from transcriptome dynamics was that pro-inflammatory states were not readily distinguished from immune-suppressive iTreg states. One question, however is whether our inability to separate proinflammatory and regulatory states was due to reliance on droplet-based scRNAseq, without supplementing this with high-dimensional protein assessment via flow or mass cytometry, single-cell epigenomic assessment via scATACseq, or scRNAseq at higher
sequencing depths. Future experiments using high-dimensional flow cytometry will determine whether post-transcriptional regulation plays any role in distinguishing emerging effector states within secondary lymphoid tissue.

Our unbiased, single-cell genomic approach suggested an unexpected, apparent trajectory characterised by TCF1$^{\text{hi}}$ expression, rapid shutdown of cellular proliferation, a lack of pro-inflammatory or immune-regulatory gene expression, an ability to migrate to the gut, and a capacity to mount a secondary recall response (Figure 8). In addition, many genes upregulated in these cells have been associated with T-cell quiescence or longevity, including Btg1, Samhd1, Mxd4, Laptm5, Gimap3, and Izumo1r (21-23, 25, 26). Based on these observations we infer TCF1$^{\text{hi}}$ TEa cells to be generally quiescent, memory or stem-like cells that emerged very rapidly during alloSCT. Transcriptomic modelling suggested that Tcf7$^{\text{hi}}$ cells could under certain circumstances arise from the cytokine-expressing effector lineage at day 3 or 4, which would be consistent with a linear model in which effector cells give rise directly to memory-like cells (34, 35). However, we also noted rare instances of Tcf7$^{\text{hi}}$ cells emerging at day 2 post-transfer, as clonal expansion was beginning and effector differentiation had yet to occur. We did not detect transcriptomic intermediates between this distal state and more naïve cells, either because such states do not exist, or because our study was not designed to detect such rare, transient events. Given that trajectory inference from scRNA-seq data tends to rely upon, indeed assume, gradual transcriptomic change, it is likely that very rapid state changes cannot be mapped using this approach. Our data does not resolve the extent to which TCF1$^{\text{hi}}$ CD4$^{+}$ T cells emerge via gradual linear effector-memory transition versus a more rapid, branching process via asymmetric cell division, which was recently reported for similar cells during a respiratory virus infection model (36), and for CD8$^{+}$ T-cells during LCMV infection (11). We speculate, given the presence of rare “pioneer”-like Tcf7$^{\text{hi}}$ cells as well as apparent transcriptomic intermediates, that both these developmental pathways may operate during alloSCT. New research tools are required to quantify the relative use of these mechanisms in vivo.

Antigen-experienced TCF1$^{\text{hi}}$ cells have been reported frequently in recent CD8$^{+}$ T cell studies (37-39), where the expression of TCF1 has been associated with long-term persistence of CD8$^{+}$ T cells, either as memory or stem-like cells. Moreover, TCF1$^{\text{hi}}$ CD8$^{+}$ T cells are thought to represent the pool of cells that responds well to immune checkpoint blockade during cancer treatment (40-42). Thus, emerging dogma from CD8$^{+}$ T-cell studies is that antigen-experienced TCF1$^{\text{hi}}$ T cells are functionally relevant. In addition, earlier reports suggested that some Th17 cells, defined by expression of IL17A, might also express high levels of Tcf7, which was associated with their persistence (43). Our data reveal that allo-
reactive CD4+ T cells can also adopt a TCF1hi state during GVHD exacerbation. Moreover, these cells may be functionally relevant at later times during GVHD. Thus, our data are broadly consistent with CD8+ T cell studies in suggesting that clonally expanded, antigen-experienced T cells can adopt a TCF1hi state that may persist and function at later timepoints. Therefore, under certain circumstances, primary activation of small numbers of antigen-specific CD4+ T-cells could give rise to a larger pool of highly plastic counterparts, with obvious implications for the magnitude and quality of secondary immune responses to the same antigen.

One question to consider from our transcriptomic modelling is how separate states, loosely referred to as “effector” or “quiescent”, emerge amongst clonal T cells during alloSCT. Heterogeneity amongst clonal TEa cells could have been induced via various non-mutually exclusive mechanisms, including asymmetric cell division (36), differential APC engagement and differential access to early local cytokine signals. Given the stark difference in CD25 expression between trajectories, it appears feasible that IL-2-signalling promotes effector function at the expense of quiescence. A recent study revealed that a reciprocal relationship in production and receipt of IL-2 controlled fate bifurcation in CD4+ T cells (44). Hence, we speculate that a similar mechanism might be acting during alloSCT. In our model, we previously reported that colonic CD103+ dendritic cells played a crucial role in amplifying acute GVHD (4). It is possible that naïve CD4+ T cells that failed to access these APC may have been programmed towards the quiescent, TCF1hi state. As part of this scenario, given that our model is characterised by profound lymphopenia, it is possible that homeostatic proliferation, perhaps via IL-7 signalling, might have partly contributed to the proliferation and stabilisation of a quiescent cellular state. However, given that donor allo-antigen presentation via MHCII is important for supporting clonal expansion in this model, exposure to diverse donor APC may contribute to heterogeneity in CD4+ T cell differentiation.

A crucial element of our study design was its focus on CD4+ T cells of a single-specificity. Given that TCR sequence can influence effector fate (45), we intentionally reduced the complexity of our system to increase the likelihood of computationally modelling Th1/Th17/iTreg effector differentiation. While numerous scRNAseq-based studies, including our own, have mapped cellular change in TCR transgenic T cells over time (7-12), an ongoing challenge has been to replicate this approach with highly diverse polyclonal T cell populations. We expect that increases in phenotypic diversity and heterogeneity in response kinetics will present additional computational challenges. Nevertheless, efforts to progress longitudinal genomic studies of T-cell differentiation from TCR transgenic systems to polyclonal T cells are warranted, in particular to facilitate analogous studies in humans.
We envisage a model for development of Th1, Th17, Tr1, or iTreg-like states during alloSCT controlled by specific micro-anatomical, T cell extrinsic factors, such as access to MHCII-presentation on different types of APC, or exposure to cytokines including IL-6 and IL-12. Given the apparent continued maturation of pro-inflammatory TEa cells as they migrated from mLN to the gut, inferred from our scVI-based atlas, this raises the question of whether peripheral tissue signals in the gut, such as local cytokine-signalling or unique cell-cell interactions contribute to this process. We speculate that spatial transcriptomic assessment of gut-located TEa cells will shed light on this matter. One further implication of our data is that conversion of emerging pro-inflammatory CD4⁺ T cells into protective, non-pathogenic iTreg cells in the gut may be feasible, since developmental pathways are similar between the two. Conversely, our data also support those iTreg-based therapies with appropriate mitigation for the effects of reversion to pro-inflammatory states in vivo (46).

We and others have previously studied Th17 biology via fate-mapping using IL-17a promoter-driven, Cre-mediated fluorescent tagging of cells (6, 47, 48). This binary approach is powerful, but does not differentiate between cells that might transiently express Il17a compared with those exhibiting prolonged expression. In our scVI transcriptomic model, we noted early transient expression of Il17a and Il17f at day 2, which disappeared only to reappear in some cells at day 5 in the gut. This suggests that CD4⁺ T cells can transiently express IL17a in mLN without becoming bona fide Th17 cells. Interestingly, a recent report employed an Il17a-Cre fate-mapping approach in the murine EAE model, and suggested two states existed amongst cells that had expressed Il17a at some point. One state was a quiescent, stem-like population that expressed TCF1 and CD27 (47). Hence, we propose that studies employing fate-mapping approaches should be interpreted with possible transient expression taken into account.

Our experiments revealed that clonally expanded TCF1 hi CD4⁺ T cells, which had exhibited no effector function and had shut down cellular proliferation, were capable of expressing pro-inflammatory cytokines or Foxp3 at a later date, as shown during secondary transplantation. This raises the hypothesis that after alloSCT, the gut is populated with quiescent alloreactive T cells that could influence disease outcome. Further experiments are required to examine the longer term persistence and functional relevance of TCF1 hi CD4⁺ T cells during GVHD. As a note of caution, however, the above hypothesis is founded on an exacerbation model, in which donor DC emerge after the initiation phase of GVHD and amplify symptoms. Therefore, whether the differentiation trajectories inferred in this study occur in allo-reactive CD4⁺ T cells during GVHD initiation remains to be tested. Moreover, the possible influence of TCF1 hi CD4⁺ T cells on GVL warrants exploration, given the presence of these T
cells in secondary lymphoid tissue and the pivotal role of alloSCT in mediating GVL. Future experiments should employ pre-clinical models and clinical samples to examine the possible relevance of our findings to the initiation phase of acute gut GVHD. Nevertheless, we view clonally expanded TCF1^{hi} CD4^{+} T cells as an opportunity to lodge immune-suppressive, possibly tissue-resident CD4^{+} T cells within the gut that could contribute to disease prevention after alloSCT. However, the potential for these to produce pathogenic cytokines would clearly need to be addressed. In summary, our examination of transcriptome dynamics during GVHD exacerbation not only highlights possible developmental relationships between effector CD4^{+} T cells, but also suggests the existence of a quiescent, memory-like state that exhibits functional potential in vivo. It will be important next to interrogate our experimentally-derived observations in preclinical models of acute GVHD initiation, and in patient samples after alloSCT.
Methods

Mice.

C57BL/6J and BALB/c mice were purchased from the Animal Resources Centre (Canning Vale, Australia), transgenic TEa (Vα2^+, Vβ6^+, CD45.1^+, CD90.1^+) mice were bred in-house (4, 5). All mice were female and aged between 6-12 weeks of age and were maintained under specific pathogen-free conditions within the animal facility at QIMR Berghofer Medical Research Institute.

Bone marrow transplantation.

Balb/c mice were transplanted as previously described (4). Briefly, BALB/c mice received 900 cGy total body irradiation (TBI; ^137^Cs source at 84 cGy/min) on day -13. On day -12, Balb/c mice were transplanted with 10 x 10^6 bone marrow cells and 2 x 10^5 FACs purified T cells from C57BL/6J donor mice. On day 0, recipient BALB/c mice were injected intravenously with 1-2 x 10^6 FACS sorted TEa T cells (Vβ6^+Vα2^+CD45.1^+). Cell Trace™ CFSE Cell Proliferation (Life Technologies) and Violet Proliferation Dye 450 (VPD450; BD Biosciences) staining were performed according to manufacturer’s protocol. For secondary transfer of TEa cells, CD25^-CXCR6^-TEa cells or CD25^-CXCR6^-TEa cells were FACS isolated on D4 from the mLN and then injected intravenously into BALB/c mice that had received total body irradiation 13 days prior and transplanted with bone marrow cells and purified T cells from C57BL/6J mice 12 days prior. For reasons of cell availability post-sort, 0.8 x 10^6 CD25^-CXCR6^- or 0.4 x10^6 CD25^-CXCR6^- TEa cells were transferred into respective groups of mice, and 0.8 x 10^6 naïve TEa cells transferred as a reference control.

Cell isolation from small intestine and colon.

Intraepithelial lymphocytes were isolated from the small intestine and colon of mice using the Lamina Propria Dissociation Kit, Mouse (Miltenyi Biotec), according to the manufacturer’s protocol.

Flow cytometry.

Cells were assessed for viability by staining with 7AAD (Sigma-Aldrich) or using a LIVE/DEAD™ Fixable Aqua Dead Cell Stain Kit (Life Technologies), according to manufacturer's protocol. Prior to staining, cells were incubated with antibodies against CD16
and CD32 (2.4G2) to block Fc receptors. For surface staining, cells were incubated with various combinations of the following antibodies for 20 minutes at 4°C: CD4 (RM4-5) – PerCPCy5.5, CD4 (RM4-5) – PE Dazzle 594, Va2 (B20.1) – APC Cy7, CD45.1 (A20) – PeCy7, CD69 (H1.2F3) – PB, α4β7 (KATK32) – PE, CD25 (PC61) – PECy7, CXCR6 (SA051D1) – APC, Armenian hamster IgG isotype control – PB, rat IgG2a isotype control – PE, rat IgG1k isotype control – PECy7, rat IgG2bk isotype control – APC (all Biolegend). For assessment of intracellular cytokine production and transcription factor expression, cells were incubated with or without ionomycin (500 ng/ml) and PMA (50 ng/ml) for 4 hours at 37°C, brefeldin-A (5 µg/ml) was added to cells after 1 hour of incubation. Intracellular staining was then performed using the eBioscience™ Foxp3/Transcription Factor Staining Buffer Set with the following antibodies: IFNγ (XMG1.2) – BV421, IL-17A (TC11-18H10.1) – BV605, rat IgG1k isotype control – BV421, rat IgG1k isotype control – BV605, rat IgG2a isotype – AF700, mouse IgG1k isotype control – APC from Biolegend; Tcf1 (C63D9) – PE, rabbit IgG isotype control – PE from Cell Signaling Technology; Tbet (4B10) – APC, Foxp3 (FJK-16S) – AF700 from eBioscience. Samples were acquired on a LSRII Fortessa Analyser (BD Biosciences) and data analysed using FlowJo software (Treestar).

Single-cell RNA capture and sequencing.

Three independent experiments were performed for sequencing: termed mGVHD1, mGVHD2, and mGVHD3. TEa cells were isolated by flow cytometry into a 1% BSA/PBS buffer. Approximately 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 v2 (mGVHD1 and mGVHD2) or v3.0 (mGVHD3) (10x Genomics) and either sequenced on an Illumina NextSeq550 (mGVHD1 and mGVHD2) or converted using the MGIEasy Universal Library Conversion Kit (BGI) before sequencing on a MGISEQ-2000 instrument (BGI; mGVHD3).

Data Availability

The full raw single-cell RNA sequencing data from this study have been submitted to the ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-9125.

Processing of scRNA-sequencing data.
FASTQ files were processed using “cellranger count” pipeline from Cell Ranger version 2.1.0 and 3.0.2 (10x Genomics) with 10x mouse genome 1.2.0 release as a reference. For BGI FASTQ files (mGVHD3), it was made compatible with Cell Ranger by reformatting file names and FASTQ headers using code from https://github.com/IMB-Computational-Genomics-Lab/BGIvsIllumina_scRNASeq (49).

Quality control of scRNA-seq data.

Cells outside the thresholds of 200-6000 expressed genes and up to 15% mitochondrial content were removed. Further filtering was performed after unsupervised clustering of cells, where clusters of cells with low Cd3 and Cd4 expression were removed. Cells that expressed Cd8a were globally removed from downstream analysis. Only genes expressed in 3 or more cells were considered.

Data transformation.

scRNA-seq data was normalised using “NormalizeData” function from Seurat v2.3.4 (50) where UMI counts for each gene from each cell were divided by the total UMI counts from that cell, multiplied by the scale factor of 10,000 and natural log-transformed. Total UMI content and mitochondrial content per cell were considered unwanted sources of variation and were removed by individual linear regression. Final residuals were then scaled to have mean feature expression of 0 and variation of 1 across cells.

Feature selection and dimensionality reduction.

Highly variable genes (HVGs) were identified using “FindVariableGene” function from Seurat with default parameters. For each subset of data used for dimensionality reduction, HVGs were computed individually and used as an input, unless otherwise specified. Number of HVGs used in each analysis is noted in the figure legends.

Principal component analysis (PCA) dimensionality reduction was performed using “RunPCA” function from Seurat and the computed PCs were used to generate uniform manifold approximation and projection (UMAP) of scRNA-seq data using “RunUMAP” function from Seurat. Number of PCs used in each analysis is noted in the figure legends.
Bayesian Gaussian Processes Latent Variable Modelling (bGPLVM) dimensionality reduction was performed using GPfates v1.0.0 (9) where datasets containing all genes after initial gene filtering step were used as input. Up to five latent variables were considered.

Integrated dimensionality reduction of mGVHD2 and mGVHD3 datasets was performed using scVI v0.4.1 (32). Each experiment was identified as separate batch. All parameters were kept at default except: up to 30 latent variables were considered, 2 hidden layers were used for encoder and decoder neural networks and up to 100 epochs were used to train the model. The computed latent variables were used as an input to generate UMAP using “RunUMAP” from Seurat, or using the umap-learn v0.3.5 python package (15).

Unsupervised clustering.

“FindClusters” function from Seurat was used to perform unsupervised clustering of cells. The resolution parameter and the number of PCs or variables used in each analysis are noted in the figure legends.

Trajectory inference.

Slingshot

Trajectories were inferred through the mGVHD2 dataset using Slingshot v0.99.12 (16). Slingshot requires clusters of cells and embeddings for those cells as input. bGPLVM latent variables 1 and 2 were used, and unbiased clustering based on these variables was also performed. A semi-supervised approach was taken whereby a cluster with high proportion of day 1 cells were specified as a starting point (Figure S4A).

Monocle

Trajectories were inferred through the mGVHD2 dataset using Monocle v2.8.0 (51). PCA dimensionality reduction was performed and the first 10 PCs were used as an input for unsupervised clustering using “plot_pc_variance_explained” and “clusterCells” functions, where number of clusters was specified (n=6). Differential gene expression analysis was performed between clusters using “differentialGeneTest” function and the list of significant genes (qval < 0.01) was used as an input to order the cells using “orderCells” function.
Trajectories were inferred through mGVHD2 dataset using SCANPY v1.4.4, which includes the PAGA trajectory inference algorithm (19). PCA dimensionality reduction was performed using the "scanpy.tl.pca" function with ARPACK SVD solver to aid computation. A neighbourhood graph was computed via "scanpy.pp.neighbors" with the first 10 PCs and the size of local neighbourhood specified as 30. Unsupervised clustering of cells was performed using "scanpy.tl.louvain" function with resolution 1.0. Finally, coarse-grained connectivity structures connecting the computed clusters of cells was mapped using "scanpy.tl.paga" with default parameters.

RNA velocity analysis was performed using scVelo version 0.1.16 (52). Only days 2-4 cells from mGVHD2 dataset were included in the analysis. All parameters were kept at default, except: 3,000 genes, 20 PCs, and 30 neighbours considered for RNA velocity estimation. Calculated velocity was projected onto pre-computed bGPLVM embeddings.

The “AddModuleScore” function from Seurat was used to calculate gene signatures. The cell cycle score was calculated using 226 cell cycle genes derived from Cyclebase (53), the aerobic glycolysis score used 41 genes associated with the Gene Ontology (GO) I.D. GO:0006096 and the oxidative phosphorylation score used 30 genes associated with I.D. GO:0006119.

Gene expression inference from missing data, or imputation, was performed using ALRA (initial release) (https://github.com/KlugerLab/ALRA/)(20) and MAGIC v1.5.0 (54).

Differential gene expression analysis (DGEA) was performed using “FindMarkers” function from Seurat at default parameters. Comparisons were done as follows: (1) trajectory I cells and trajectory II cells, (2) day 0 and day 1 cells, and (3) day 1 and day 2 cells. (1) was
performed using mGVHD2 dataset as input, and (2) and (3) were performed using the integrated mGVHD2 and mGVHD3 datasets as input. Genes from (2) and (3) with Bonferroni adjusted P-value below 0.01 and average log fold change greater than 0.5 were considered as input for gene ontology (GO) term enrichment analysis. GO terms were obtained from “org.Mm.eg.db” Bioconductor annotation package 99 (55). Fisher’s exact test was performed using “goana” function from edgeR (56) to identify over-represented GO terms. Input genes for DGEA was used as the background gene set.

Other packages.

scRNA-seq data was primarily visualised with the ggplot2 v3.2.1 R package (57). Additional functions were provided by cowplot v0.9.2, in particular via Seurat’s inbuilt plotting features (50). 3D projections were created by the scatterplot3d v0.3-41 R package (58). Seurat v3.0.1 was used for conversion of data to loom format.

Study approval.

All animal procedures and protocols were approved (A0412-617M; P832) and monitored by the QIMR Berghofer Medical Research Institute Animal Ethics Committee.

Author Contributions

JAE designed research studies, conducted wet-lab experiments, acquired data and analysed data. HJL and CGW designed research studies and performed bioinformatics analyses of data. RK, SO, LIML, MSFS, SBA, AV, MK, AH designed research studies and conducted experiments. VS provided input to the bGPLVM computational modelling of the data. JEP, SAT, GRH, AV and MK provided valuable expertise to the project. SAT, MK and AH conceived the project and interpreted results. AH, JAE, HJL and CGW wrote and edited the manuscript and prepared figures. All authors provided input and approved the manuscript. JAE, HJL and CGW are co-first authors of the manuscript, the order of authors was determined by work load.

Acknowledgments
This work was supported by research grants from the Australian National Health and Medical Research Council: AH, SAT and MK were supported by Project and Ideas Grants (1126399 and 1180951). JEP is supported by an Investigator Grant (1175781). We would like to thank staff from the QIMR Berghofer flow cytometry facility for their expertise and assistance with experiments, and staff from QIMR Berghofer animal facility for animal care and husbandry. We acknowledge Paul Collins at the QIMR Berghofer sequencing facility, and Scott Wood and his team at the QIMR Berghofer High Performance Computing facility for support with scRNA-seq data management. We acknowledge helpful discussions with Dr. Kate Gartlan (QIMR Berghofer) on matters relating to T cell differentiation during alloSCT. We thank Dr. Kylie R. James (Wellcome Sanger Institute) for reading the manuscript and providing helpful comment on issues relating to CD4^+ T-cell differentiation.
1. Koyama M, and Hill GR. The primacy of gastrointestinal tract antigen-presenting cells in lethal graft-versus-host disease. Blood. 2019;134(24):2139-48.
2. Koyama M, et al. Recipient nonhematopoietic antigen-presenting cells are sufficient to induce lethal acute graft-versus-host disease. Nat Med. 2011;18(1):135-42.
3. Koyama M, et al. MHC Class II Antigen Presentation by the Intestinal Epithelium Initiates Graft-versus-Host Disease and Is Influenced by the Microbiota. Immunity. 2019;51(5):885-98 e7.
4. Koyama M, et al. Donor colonic CD103+ dendritic cells determine the severity of acute graft-versus-host disease. J Exp Med. 2015;212(8):1303-21.
5. Grubin CE, Kovats S, deRoos P, and Rudensky AY. Deficient positive selection of CD4 T cells in mice displaying altered repertoires of MHC class II-bound self-peptides. Immunity. 1997;7(2):197-208.
6. Gaublomme JT, et al. Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity. Cell. 2015;163(6):1400-12.
7. Tibbitt CA, et al. Single-Cell RNA Sequencing of the T Helper Cell Response to House Dust Mites Defines a Distinct Gene Expression Signature in Airway Th2 Cells. Immunity. 2019;51(1):169-84 e5.
8. Lonnberg T, et al. Distinct microbial and immune niches of the human colon. Nat Immunol. 2020.
9. Miragaia RJ, et al. Single-Cell Transcriptomics of Regulatory T Cells Reveals Trajectories of Tissue Adaptation. Immunity. 2019;50(2):493-504 e7.
10. La Manno G, et al. RNA velocity of single cells. Nature. 2018;560(7719):494-8.
11. Saelens W, Cannoodt R, Todorov H, and Saeys Y. A comparison of single-cell trajectory inference methods. Nat Biotechnol. 2018;36(3):375-81.
Kawai Y, Ouchida R, Yamasaki S, Dragone L, Tsubata T, and Wang JY. LAPTM5 promotes lysosomal degradation of intracellular CD3zeta but not of cell surface CD3zeta. *Immunol Cell Biol.* 2014;92(6):527-34.

Xu H, et al. MS4a4B, a CD20 homologue in T cells, inhibits T cell propagation by modulation of cell cycle. *PLoS One.* 2010;5(11):e13780.

Yano K, et al. Gimap3 and Gimap5 cooperate to maintain T-cell numbers in the mouse. *Eur J Immunol.* 2014;44(2):561-72.

Kunzli M, et al. Long-lived T follicular helper cells retain plasticity and help sustain humoral immunity. *Sci Immunol.* 2020;5(45).

Lee JY, et al. The transcription factor KLF2 restrains CD4(+) T follicular helper cell differentiation. *Immunity.* 2015;42(2):252-64.

Weber JP, et al. ICOS maintains the T follicular helper cell phenotype by downregulating Kruppel-like factor 2. *J Exp Med.* 2015;212(2):217-33.

Garcia-Bernal D, Dios-Esponera A, Sotillo-Mallo E, Garcia-Verdugo R, Arellano-Sanchez N, and Teixido J. RGS10 restricts upregulation by chemokines of T cell adhesion mediated by alpha4beta1 and alphaLbeta2 integrins. *J Immunol.* 2011;187(3):1264-72.

Lee JK, Kannarkat GT, Chung J, Joon Lee H, Graham KL, and Tansey MG. RGS10 deficiency ameliorates the severity of disease in experimental autoimmune encephalomyelitis. *J Neuroinflammation.* 2016;13:24.

Zhang P, et al. Eomesoderm promotes the development of type 1 regulatory T (TR1) cells. *Sci Immunol.* 2017;2(10).

Lopez R, Regier J, Cole MB, Jordan MI, and Yosef N. Deep generative modeling for single-cell transcriptomics. *Nat Methods.* 2018;15(12):1053-8.

Gartlan KH, et al. Donor T-cell-derived GM-CSF drives alloantigen presentation by dendritic cells in the gastrointestinal tract. *Blood Adv.* 2019;3(19):2859-65.

Pepper M, Pagan AJ, Igyarto BZ, Taylor JJ, and Jenkins MK. Opposing signals from the Bcl6 transcription factor and the interleukin-2 receptor generate T helper 1 central and effector memory cells. *Immunity.* 2011;35(4):583-95.

Tubo NJ, Fife BT, Pagan AJ, Kotov DI, Goldberg MF, and Jenkins MK. Most microbe-specific naive CD4(+) T cells produce memory cells during infection. *Science.* 2016;351(6272):511-4.

Nish SA, et al. CD4(+) T cell effector commitment coupled to self-renewal by asymmetric cell divisions. *J Exp Med.* 2017;214(1):39-47.

Utzschneider DT, et al. T Cell Factor 1-Expressing Memory-like CD8(+) T Cells Sustain the Immune Response to Chronic Viral Infections. *Immunity.* 2016;45(2):415-27.

Wu T, et al. The TCF1-Bcl6 axis counteracts type I interferon to repress exhaustion and maintain T cell stemness. *Sci Immunol.* 2016;1(6).

Yao C, et al. Single-cell RNA-seq reveals TOX as a key regulator of CD8(+) T cell persistence in chronic infection. *Nat Immunol.* 2019;20(7):890-901.

Im SJ, et al. Defining CD8(+) T cells that provide the proliferative burst after PD-1 therapy. *Nature.* 2016;537(7620):417-21.

Jansen CS, et al. An intra-tumoral niche maintains and differentiates stem-like CD8 T cells. *Nature.* 2019;576(7787):465-70.

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.* 2019;50(1):195-211 e10.

Muranski P, et al. Th17 cells are long lived and retain a stem cell-like molecular signature. *Immunity.* 2011;35(6):972-85.

DiToro D, et al. Differential IL-2 expression defines developmental fates of follicular versus nonfollicular helper T cells. *Science.* 2018;361(6407).

Tubo NJ, et al. Single naive CD4(+) T cells from a diverse repertoire produce different effector cell types during infection. *Cell.* 2013;153(4):785-96.
Di Stasi A, et al. Inducible apoptosis as a safety switch for adoptive cell therapy. *N Engl J Med.* 2011;365(18):1673-83.

Karmaus PWF, et al. Metabolic heterogeneity underlies reciprocal fates of TH17 cell stemness and plasticity. *Nature.* 2019;565(7737):101-5.

Varelias A, et al. Lung parenchyma-derived IL-6 promotes IL-17A-dependent acute lung injury after allogeneic stem cell transplantation. *Blood.* 2015;125(15):2435-44.

Senabouth A, et al. Comparative performance of the BGI and Illumina sequencing technology for single-cell RNA-sequencing. *BioRxiv.* 2019.

Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* 2018;36(5):411-20.

Trapnell C, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol.* 2014;32(4):381-6.

Bergen V, Lange M, Peidli S, Wolf FA, and Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. *BioRxiv.* 2019.

Santos A, Wernersson R, and Jensen LJ. Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. *Nucleic Acids Res.* 2015;43(Database issue):D1140-4.

van Dijk D, et al. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell.* 2018;174(3):716-29 e27.

Carlson M. *org.Mm.eg.db: Genome wide annotation for Mouse.* R package version 3.8.2. 2019.

Robinson MD, McCarthy DJ, and Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139-40.

Wickham, and Hadley. *ggplot2: Elegant Graphics for Data Analysis.* 2016.

Ligges U, and Mächler M. scatterplot3d- An R Package for Visualizing Multivariate Data. *Journal of Statistical Software.* 2003;8(11).
Figure 1. Allo-reactive donor CD4+ T cells acquire pro-inflammatory and regulatory states during acute GVHD exacerbation.

(A) Schematic for model of acute gut-mediated GVHD exacerbation where donor CD4+ T cells respond to host allogeneic peptide presented within donor MHC class II, which in turn drives CD4+ T cell expansion in the mesenteric lymph node (mLN).

(B & C) Representative FACS plots for IFNγ and IL-17A production and Foxp3 expression on T Ea cells from the mLN at day 4 post-transfer, with PMA/ionomycin restimulation (B) or directly ex vivo (C). Graphs of percentage IFNγ+, IL-17A+ or Foxp3+ T Ea cells. Data shown are combined from three replicate experiments (n = 10 mice) and are represented as box-and-whisker plots, with bounds from 25th to 75th percentile, median line and whiskers ranging from minimum to maximum values. Statistical analysis was performed between the isotype control (ISO) and respective cytokine samples using a Mann-Whitney test. *p<0.05, **p≤0.01, ***p≤0.001, ****p≤0.0001.
Figure 2. Droplet-based scRNA-seq and computational modelling of TEa T cells over time suggests diverging fates in mLN.

(A) Schematic of scRNA-seq experiment used to study donor CD4+ T cell differentiation. TEa cells were recovered from the mLN at various timepoints and cells pooled from multiple mice (n = number of mice pooled; at day 4 samples from 3 mice ran separately) for droplet-based scRNA-seq using the 10x Chromium.

(B) bGPLVM visualisation of TEa cells from day 1 to day 4 overlaid with the developmental trajectories identified by Slingshot (Trajectory I, II and III).

(C) Visualisation of Cd69, Mki67 and Itgb7 expression, or the cell cycle, aerobic glycolysis, and oxidative phosphorylation gene signature scores, in bGPLVM overlaid with trajectories.

(D) Violin plots showing expression of genes or gene signature scores as described in (C) for each trajectory. Statistical analysis was performed using a Wilcoxon-rank sum test between trajectories. *p<0.05, **p≤0.01, ***p≤0.001, ****p≤0.0001.
Figure 3. *Ifng*, *Il17a* and *Foxp3* expressing fates emerge along one trajectory, identifying a second *Tcf7* hi trajectory.

(A) bGPLVM visualisation of TEa cells expressing *Ifng*, *Il17a* and *Foxp3* with Slingshot trajectories overlaid. Bar graphs show the proportion of cells within Trajectory I or II that express *Ifng*, *Il17a* or *Foxp3*.

(B) bGPLVM visualisation of TEa cells expressing *Tcf7*, *Ccr7*, *Cd27* and *Sell* with Slingshot trajectories overlaid. Violin plots show the level of expression for each gene in Trajectory I and II. Statistical analysis was performed using a Wilcoxon-rank sum test between trajectories.

*p<0.05, **p≤0.01, ***p≤0.001, ****p≤0.0001.*
**Figure 4.** Clonally expanded TEa T cells acquire a TCF1<sup>hi</sup> non-effector state in mLN.

(A & B) Representative FACS plots showing the expression of Tbet and Tcf1 in TEa cells at day 4 post-transfer in the mLN directly ex vivo (A) or after restimulation with PMA/ionomycin (B). Graphs show the geometric mean fluorescence intensity of the Tcf1<sup>lo</sup> (purple) or Tcf1<sup>hi</sup> (turquoise) population. The expression of IFNγ, IL-17A and Foxp3 for the Tcf1<sup>lo</sup> (purple) or Tcf1<sup>hi</sup> (turquoise) population is also shown. Graphs show the percentage of IFNγ<sup>+</sup>, IL-17A<sup>+</sup> or Foxp3<sup>+</sup> TEa cells for the Tcf1<sup>lo</sup> or Tcf1<sup>hi</sup> population. Data shown are combined from two independent experiments (n = 3 mice / experiment). Statistical analysis was performed using a paired t-test. *p<0.05, **p≤0.01, ***p≤0.001, ****p≤0.0001.
**Figure 5.** An integrated transcriptomic atlas of TEa T cells over time and across organs.

(A) Schematic of scRNA-seq experiment used to study the developmental relationships between TEa cells from the mLN and the gut (IEL).

(B) Two perspectives of 3-Dimensional UMAP visualisation produced from 30 scVI latent variables representing all samples from day 2 - day 5. Only day 2, day 5 gut, and day 5 mLN samples are shown for clarity.

(C) *Ifng*, *Il17a*, and *Foxp3* expression in the same representation as (B).

(D) *Tcf7* expression in the same representation as (B) and (C).
Figure 6. Defining a cell-sorting strategy for TCF1^{hi} TEa T cells.

(A) bGPLVM visualisation of TEa cells from the mLN from day 1 – day 4 expressing Il2ra and Cxcr6, with Slingshot trajectories overlaid.

(B) Representative flow cytometry plots showing the expression of CD25, IFNγ, IL-17A and Foxp3 on Tcf1^{lo} or Tcf1^{hi} TEa cells from the mLN at day 4 post-transfer, directly ex vivo or after restimulation with PMA/ionomycin.

(C) Graphs show the percentage of CD25^{+} or CD25^{-} TEa cells that express IFNγ, IL-17A and Foxp3. Data shown are combined from two independent experiments showing similar results (n = 9 mice). Statistical analysis was performed using a Paired t test. *p<0.05, **p≤0.01, ***p≤0.001, ****p≤0.0001.

(D) Representative flow cytometry plot showing the expression of CD25 and TCF1 on TEa cells from the mLN at day 4 post-transfer.

(E) Flow cytometry gating strategy used to isolate CD25^{+}CXCR6^{-} and CD25^{+}CXCR6^{+} TEa cells from the mLN at day 4 post-transfer. The corresponding purity checks and CTV expression for each population are shown.
Figure 7. TCF1\textsuperscript{hi} T\textsubscript{Ea} T cells can mount secondary effector responses in vivo.

(A) Schematic of secondary T\textsubscript{Ea} transfer experiment. On day 4 of primary transplant, CD25\textsuperscript{-} CXCR6\textsuperscript{-} or CD25\textsuperscript{+} and/or CXCR6\textsuperscript{+} (referred to as CD25\textsuperscript{+}CXCR6\textsuperscript{+}) T\textsubscript{Ea} were FACS isolated from mLN. 0.8 x 10\textsuperscript{6} CD25\textsuperscript{-}CXCR6\textsuperscript{-} and 0.4 x 10\textsuperscript{6} CD25\textsuperscript{+}CXCR6\textsuperscript{+} T\textsubscript{Ea} (or control 0.8 x 10\textsuperscript{6} Naïve T\textsubscript{Ea}) were transferred into recipient BALB/c mice that had received total body irradiation 13 days prior and a bone marrow transplant 12 days prior. FACS assessment of mLN T\textsubscript{Ea} cells was performed on day 4 or day 5 post-secondary transfer.

(B) Absolute numbers of T\textsubscript{Ea} cells in the mLN at day 4/5 post-secondary transfer. Arrows along y-axis denote the number of T\textsubscript{Ea} cells that were transferred per mouse for each group on day 0 (0.8 x 10\textsuperscript{6} Naïve T\textsubscript{Ea} (grey) and CD25\textsuperscript{-}CXCR6\textsuperscript{-} T\textsubscript{Ea} (turquoise), 0.4 x 10\textsuperscript{6} CD25\textsuperscript{+}CXCR6\textsuperscript{+} T\textsubscript{Ea} (purple)).

(C) Representative flow cytometry plots showing the expression of Tbet and Tcf1 on T\textsubscript{Ea} cells from the mLN at day 4/5 post-secondary transfer. Graph shows % of T\textsubscript{Ea} cells that are Tcf1\textsuperscript{lo} or Tcf1\textsuperscript{hi} in each group.

(D & E) Representative FACS plots showing IFN\textgamma, IL-17A and Foxp3 expression on T\textsubscript{Ea} cells after secondary transfer on day 4/5 from the mLN directly ex vivo (D) or post-restimulation with PMA and ionomycin (E). Graphs show the percentage of IFN\textgamma\textsuperscript{+}, IL-17A\textsuperscript{+} and Foxp3\textsuperscript{+} T\textsubscript{Ea} for each group. Data shown are combined from two independent experiments showing similar results (B-D; Naïve n = 12, CD25\textsuperscript{-}CXCR6\textsuperscript{-} n = 7, CD25\textsuperscript{+}CXCR6\textsuperscript{+} n = 6) or from one experiment only (E; Naïve n = 6, CD25\textsuperscript{-}CXCR6\textsuperscript{-} n = 4). Data are represented as median (B) or mean ± SEM (C - E). Statistical analysis was performed using a Mann-Whitney test (B,D, E) or a Paired t test (C), *p<0.05, **p≤0.01, ***p≤0.001, ****p≤0.0001.
Allo-reactive CD4+ T cells (TEa) activated by donor antigen presenting cells differentiate along two developmental trajectories in the mLN. Trajectory I cells express high levels of cell cycling and aerobic glycolysis genes and encompass pathogenic subsets that produce IFNγ and IL-17A and regulatory subsets expressing FOXP3. Trajectory II cells are largely quiescent, expressing TCF1 and upregulating stemness-associated genes including Ccr7, Sell, Cd27 and Klf2. Cells following both trajectories upregulate expression of α4β7, allowing trafficking to the gut (IEL). Trajectory I cells continue to evolve transcriptomically during migration to the gut, by further upregulating expression of pro-inflammatory and regulatory genes. In contrast, Trajectory II cells migrate to the gut, yet remain transcriptomically unchanged. Nevertheless, TCF1hi cells are able to mount a secondary effector response.
**Table 1.** Top 50 genes upregulated in Cluster 0 in Trajectory II compared to Cluster 3 in Trajectory I - genes marked with an asterisk (*) were detected within the Top 50 genes in an analogous comparison in a second, independent scRNAseq experiment.

| Gene       | avg_logFC | p_val_adj |
|------------|-----------|-----------|
| Malat1*    | 2.07      | 0         |
| Btg1*      | 2.04      | 0         |
| Ifi27l1a*  | 1.79      | 0         |
| Rgs10*     | 1.62      | 2.45E-292 |
| Izumo1r*   | 1.59      | 4.81E-126 |
| Igfbp4*    | 1.54      | 8.47E-65  |
| Tcf7*      | 1.52      | 1.02E-187 |
| Ypel3*     | 1.42      | 1.97E-138 |
| Npc2       | 1.29      | 6.60E-210 |
| Shisa5     | 1.24      | 0         |
| Cox7a2l*   | 1.19      | 2.52E-282 |
| Ms4a6b     | 1.17      | 2.83E-176 |
| Ms4a4b     | 1.14      | 1.33E-136 |
| Ccr7*      | 1.11      | 3.58E-66  |
| Gm26740    | 1.10      | 2.49E-74  |
| Use1*      | 1.04      | 1.20E-98  |
| Limd2      | 1.03      | 1.75E-264 |
| Gm8369*    | 1.01      | 3.42E-54  |
| Laptm5     | 1.01      | 2.34E-139 |
| Id3        | 1.01      | 5.28E-25  |
| Gimap6     | 1.01      | 2.92E-98  |
| Gbp2*      | 1.00      | 3.15E-13  |
| Stat1*     | 0.99      | 3.06E-67  |
| Saraf      | 0.98      | 2.09E-79  |
| Evl        | 0.97      | 7.43E-52  |
| Sesn3*     | 0.94      | 6.70E-52  |
| Pold4      | 0.92      | 6.78E-40  |
| Cd52       | 0.91      | 4.55E-204 |
| Gltsr2     | 0.90      | 3.75E-136 |
| Gimap3     | 0.90      | 3.02E-63  |
| Mxd4*      | 0.89      | 4.91E-49  |
| Fyb*       | 0.89      | 1.53E-79  |
| Tspan32    | 0.89      | 1.58E-35  |
| Slamf6*    | 0.88      | 1.30E-55  |
| Ltb        | 0.88      | 3.57E-108 |
| Pfdn5      | 0.88      | 4.07E-203 |
| Samhd1     | 0.88      | 1.03E-29  |
| Cd3d       | 0.86      | 1.82E-216 |
| Cd27       | 0.85      | 2.21E-37  |
| Hmha1      | 0.85      | 1.01E-32  |
| Nsg2*      | 0.84      | 4.70E-48  |
| Eva1b      | 0.83      | 1.14E-30  |
| Satb1*     | 0.82      | 6.48E-64  |
| Klf2       | 0.82      | 2.78E-55  |
| Pnrc1*     | 0.82      | 6.74E-27  |
| Asap1*     | 0.81      | 5.99E-30  |
| Sepp1      | 0.81      | 5.30E-24  |
| Ankrd12*   | 0.80      | 1.24E-19  |
| Cd3g       | 0.80      | 1.84E-169 |
| Ndrg3      | 0.79      | 4.91E-25  |