Differential impact of self and environmental antigens on the ontogeny and maintenance of CD4$^+$ T cell memory

Thea Hogan[1¤], Maria Nowicka[2¤], Daniel Cownden[3], Claire Pearson[4], Andrew J. Yates[2‡*], Benedict Seddon[1‡*]

[1] Institute of Immunity and Transplantation, Division of Infection and Immunity, UCL, Royal Free Hospital, Rowland Hill Street, London NW3 2PF, United Kingdom
[2] Department of Pathology and Cell Biology, Columbia University Medical Center, 630 West 168th Street, New York, NY 10032, USA
[3] Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow G3 8TA, United Kingdom
[4] Kennedy Institute of Rheumatology, University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7FY, United Kingdom

¤ Contributed equally
‡ Contributed equally
*Address correspondence to either author; `benedict.seddon@ucl.ac.uk, andrew.yates@columbia.edu`

Abstract

Laboratory mice not exposed to overt infections nevertheless develop populations of circulating memory phenotype (MP) CD4$^+$ T cells, presumably in response to environmental, commensal or self-antigens. The relative importance and timing of the forces that generate these populations remain unclear. Here we combine mathematical models with data from studies tracking the generation of CD4$^+$ MP T cell subsets in mice of different ages, housed in facilities that differ in their 'dirtiness'. We infer that both central and effector CD4$^+$ MP T cell populations derive directly from naive CD4 T cells, and are heterogeneous in their rates of turnover. We also infer that early exposure to self and environmental antigens establishes persistent memory populations at levels determined largely, but not exclusively, by the dirtiness of the environment. After the first few weeks of life, however, these populations are continuously supplemented by new memory cells at rates that appear to be independent of environment, likely in response to self or ubiquitous commensal antigens.

Keywords: Memory T cells, Ontogeny, Mathematical modelling

Introduction

Conventional memory T cells are defined as lymphocytes which respond rapidly upon re-encounter with previously-encountered epitopes [1, 2]. In mice, memory T cells exhibit considerable heterogeneity in their function, circulation patterns, response to re-challenge, and capacities for proliferative self-renewal and survival [2–5]. This phenotypic heterogeneity is reflected in differential expression of various cell-surface molecules. In uninfected naive mice, there are at least two distinct populations of recirculating cells distinguished by their expression of the lymph node homing receptor L-selectin (CD62L); $CD44^{hi}$ $CD62L^{-}$ effector memory (T_{EM}) and $CD44^{hi}$ $CD62L^{+}$ central memory (T_{CM}) cells. During immune responses to active infection, there is an even more complex mix of effector and memory intermediates [6].

While it is clear that memory to infection resides amongst these $CD44^{hi}$ subsets, it is also evident that cells of this phenotype are also generated in naive mice in the absence of overt infection. Their functional significance is not fully understood, but there is evidence they can augment primary immune responses. MP $CD4^{+}$ T cells can facilitate rapid production of IFN-γ during the early inflammatory phase of the immune response to *Toxoplasma gondii* infection and enhance T_H1-type CD4+ T cell responses later in infection [7]. There is also evidence that MP $CD4^{+}$ cells are capable of making rapid cross-reactive responses during primary infections [8]. Given that such MP cells represent the majority of the memory compartment in specific pathogen-free (SPF) mice [7], a better understanding of how these cells are generated and maintained is crucial to better understand their function and impact upon conventional memory to defined challenges.

The precise nature of the forces driving the generation of these memory-phenotype (MP) populations remain unclear. Their development does appear to be dependent upon a TCR-mediated activation event. Evidence for this comes from observations that $CD28^{-/-}$ mice have greatly reduced numbers of MP populations [9], while mice lacking canonical NF-κB signalling, an obligate pathway in T cell activation, are completely devoid of such cells [10]. Whether the stimuli for these activation triggers derive from self or foreign recognition events is unclear. Their generation may reflect a stochastic process in which naive CD4 T cells occasionally encounter homeostatic stimuli that are above an activation threshold [11], and indeed MP-like cells are generated not only in lymphopenia but also constitutively under replete conditions throughout life [3, 7]. There is a positive correlation between the affinity of naive $CD4^{+}$ T cells to self-antigens and the potential for differentiating into MP cells [7], although self-reactivity is also positively correlated to reactivity to foreign antigens [12]. Stronger evidence supporting an autoreactive stimulus comes from the failure of broad-spectrum antibiotics to prevent conversion of naive cells to MP cells following adoptive transfer in vivo, and the observation that mice raised in germ free conditions contain similar numbers of MP cells as those in SPF conditions [7]. However, there is also a role for foreign environmental antigen in generating T cell memory compartments, since mice raised in more antigenically diverse environments, but still in the absence of overt infection, exhibit larger peripheral memory $CD8^{+}$ T cell pools than SPF mice [13]. When exactly such environmental stimuli impact upon memory compartment development, or how foreign and self reactivity combine to form the memory compartments, is unknown.

The differentiation pathways of these MP cells are also poorly understood. They may derive either directly from the naive T cell pool or through interconversion of other memory phenotypes. Amongst CD8 cells, there is evidence that some MP subpopulations are generated in the thymus [14]. In the

case of CD4 lineage cells, it has been suggested that MP $CD4^+$ T cells derive from peripheral naive phenotype cells in a thymus-independent fashion [7]. While both CD4 T_{EM} and CD4 T_{CM} are produced constitutively in adult mice [3], it remains unclear how their differentiation patterns relate to those of classically antigen-stimulated naive T cells.

In this study we aimed to characterise the development and maintenance of memory CD4 T cell subsets in adult mice to identify the nature and timing of the signals driving these dynamics in the absence of overt infection. To do this we quantified the homeostasis and ontogeny of memory CD4 T cells in identical strains of mice raised in two different animal facilities with distinct caging environments; those housed in individually ventilated cages (IVCs) and fed irradiated water, and those fed untreated tap water and housed in open cages, who might consequently be exposed to a greater variety of environmental antigens. We made use of an established temporal fate-mapping model in both cohorts, which allowed us to track the development of T cells under replete conditions [3, 15]. We also used data from germ-free (GF) mice to dissect further the relative contributions of self and environmental antigens to the generation and maintenance of MP cells.

Results

Modelling the ontogeny and homeostasis of constitutively-generated memory CD4 T cell subsets

We employed a system described previously [3, 15] to follow the flow of cells into memory subsets (Fig. 1A). Briefly, CD45.1$^+$ C57Bl6/SJL host mice of various ages were treated with optimised doses of the transplant conditioning drug busulfan to selectively deplete haematopoeitic stem cells (HSC). The HSC compartment was then reconstituted by bone marrow transplantation (BMT) with congenically labelled bone marrow from CD45.2$^+$ C57Bl6/J donors. The progeny of donor HSC were then followed as they developed in the thymus and percolated into the peripheral T cell pools, initially replete with host-derived cells. Total numbers of CD4$^+$ naive T cells and T_{CM} and T_{EM} cells in these busulfan chimeric mice were normal, in comparison to untreated WT controls (Fig. S1), confirming that that the busulfan treatment regime left peripheral compartments intact [3, 15].

We studied busulfan chimeric mice from two housing facilities that employed different levels of mouse containment. At the MRC National Institute for Medical Research (NIMR), mice were held in open cages and fed untreated tap water, while mice held at the UCL Comparative Biology Unit (UCL) were maintained in individually ventilated cages (IVCs) and fed irradiated water. Henceforth we refer to UCL sourced mice as 'clean' and NIMR sourced mice as 'dirty', in reference to the presumed difference in health status of the mice. In both cases, the same C57Bl6/SJL strain was analysed by the one researcher and enumerated using the same single CASY cell counter. In mice aged 10 weeks and older, the numbers of naive CD4 T cells in mice from clean and dirty environments were broadly similar (Fig. 1B, left panel). The total sizes (host+donor) of all circulating memory CD4 T cell subsets remained relatively stable over the time frame of analysis, but were already significantly larger in dirty mice (Fig. 1B, right panels) at age 10 weeks. Following BMT, donor-derived memory T cells accumulated in similar numbers in the two environments (Fig. 1C). The larger compartments of dirty mice were therefore associated with a reduced capacity of donor cells to replace the pre-existing host memory compartments (Fig. 1D).

To quantify the cellular processes underlying these kinetics, we first considered a simple mechanistic explanation shown schematically in Fig. 2A. In this model, each memory population (CD4 T_{CM} or T_{EM}) is fed at a constant *per capita* rate from a precursor population (source). We refer to this rate as the force of recruitment, φ. The total cellular flux into memory per day is then φ multiplied by the size of the source population, which in principle could be naive CD4 T cells, or the complementary memory population. We assume that memory cells are then lost at a constant net *per capita* rate λ which is the balance of loss and proliferative self-renewal. The inverse of this rate can be interpreted as the average lifetime of a clonal population of T cells – that is, it measures the persistence of of a lineage of cells and their progeny, and not the lifetime of individual memory cells.

We also considered an alternative model, illustrated in Fig. 2B, which was motivated by three observations. First, donor CD4 T_{CM} and T_{EM} in busulfan chimeras express Ki67, a marker of recent cell division, at higher levels than host cells for some time after BMT [3], although these levels eventually converge (data not shown). These observations suggest that the simplest model, in which host and donor memory cells behave identically at all times, may not fully capture memory T cell dynamics. Second, we previously found evidence, using BrdU labelling in WT mice, that both CD4 T_{EM} and T_{CM}

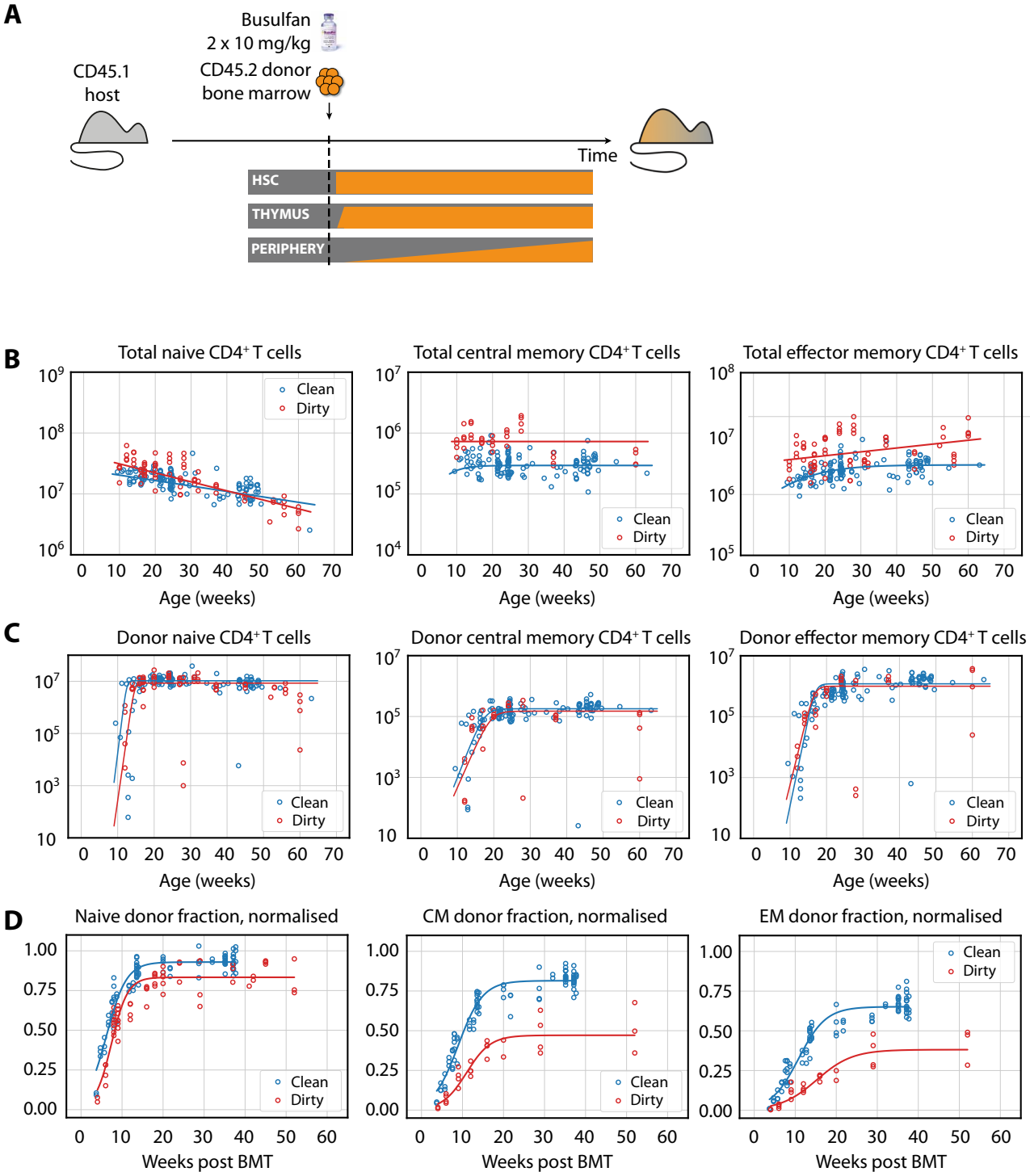

Figure 1 – **(A)** Generating bone marrow chimeras to map T cell homeostasis. **(B)** Comparing total numbers of $CD4^+$ naive, T_{CM} and T_{EM} cells in clean and dirty chimeric mice. **(C)** Numbers of donor-derived $CD4^+$ subsets recovered from spleen and lymph nodes of clean and dirty chimeric mice (BMT performed at age 5-26 weeks). **(D)** The donor fraction (chimerism) within $CD4^+$ T cell subsets, varying with time post bone marrow transplant (BMT), normalised to the chimerism in the double-positive thymocytes in each mouse. Lines in all panels represent empirical curves fitted to each population for visual guidance.

appear to be kinetically heterogeneous, with at least two subpopulations turning over and dividing at different rates [3]. Third, and consistent with this picture, the increases in donor chimerism that we observed in both T_{CM} and T_{EM} with time post-BMT were suggestive of a biphasic kinetic, with a relatively rapid accumulation of donor cells followed by a slower increase (Fig. 1D). All three observations are consistent with a mechanism in which cells newly recruited into memory first enter a subpopulation that divides and turns over rapidly, and transitions to a more quiescent population that turns over – and hence is replaced by donor cells – more slowly (Fig. 2B). We denote these rates λ_{fast} and λ_{slow}. Differences in Ki67 expression in donor and host cells could then be explained by an enrichment for 'new' memory (donor cells) in the faster subpopulation in the weeks immediately following BMT, and not by any intrinsic differences in the behaviour of donor and host cells.

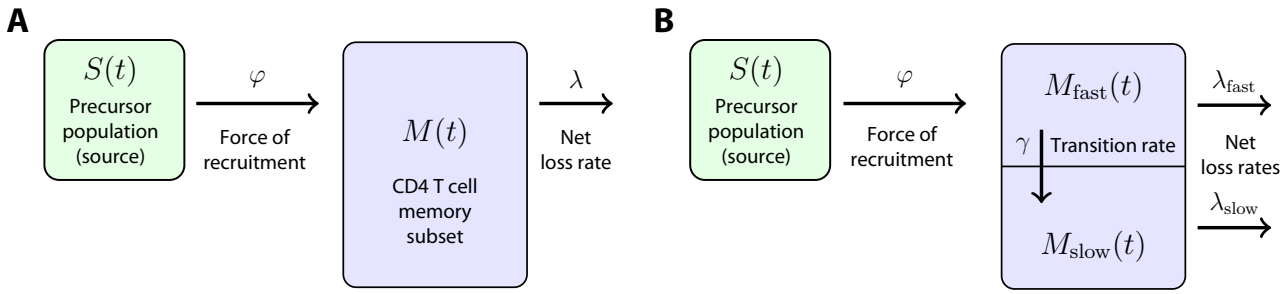

Figure 2 – Modelling the generation and maintenance of memory CD4 T cell subsets in adult mice. (**A**) New cells from a precursor (source) population of size $S(t)$ flow in to a memory subset $M(t)$ at total rate $\varphi S(t)$. The 'force of recruitment' φ is approximately the daily probability that any given cell from the source will be recruited into memory, multiplied by an expansion factor. This memory population may self-renew through division and may die, and is continually supplemented or replaced by cells recruited from the source. We assume that the net loss rate (loss minus division) is a constant, λ, such that $1/\lambda$ is the mean lifetime of a memory clone. (**B**) In the two-compartment model the force of recruitment φ is also assumed to be constant. New cells are recruited into a population $M_{\text{fast}}(t)$ that has a high net turnover rate λ_{fast} and so this population replaced by donor cells relatively quickly. These cells transition into a slower subset $M_{\text{slow}}(t)$ at constant rate γ. This subset may also self-renew through division and die, such that the population turns over at net rate $\lambda_{\text{slow}} < \lambda_{\text{fast}}$.

Naive CD4 T cells are likely the direct precursors of both CD4 T_{CM} and T_{EM} in the absence of infection, and both CD4 memory populations are heterogeneous in their turnover

We compared the ability of the two mechanisms, encoded as mathematical models (see Methods), to describe the replacement kinetics of memory subsets in the dirty and clean environments. The kinetic of donor chimerism in CD4 T_{EM} clearly lagged that of CD4 T_{CM} (Fig. 1D), ruling out T_{EM} as a potential source for T_{CM}. We therefore considered only naive CD4 T cells as the source for T_{CM}, but considered both naive and T_{CM} cells as potential sources of T_{EM}.

For each environment and target cell population (T_{EM}/T_{CM}), we considered each putative source (naive or T_{CM}) and description of the kinetics of homeostasis (the simple model or the two compartment model). We fitted each combination of source and model simultaneously to the timecourses of the total cell numbers and the proportion of donor cells within the subset. We then calculated a measure of the combined probability (joint likelihood) that the replacement kinetics in both dirty and clean environments derive from that combination of source and model. We then used these joint likelihoods to assess our relative confidence in each combination of source and model, using the Akaike Information Criterion (see Methods). We found extremely strong support for the two-compartment model of memory and with T_{EM} deriving directly from naive CD4 T cells and not from T_{CM} (Table 1; fits shown in Fig. 3; fits using the simple model shown in Fig. S2). The latter result is consistent with conclusions we derived previously from analysis of a smaller set of data from busulfan chimeras [3].

	CD4 central memory	**CD4 effector memory**	
	Source population	Source population	
Model	CD4 naive	CD4 naive	CD4 central memory
Simple	**107** (0)	**42** (0)	**93** (0)
Two-compartment	**0** (1)	**0** (1)	**112** (0)

Table 1 – Measures of relative support (using differences in the Akaike information criterion, AIC - AIC_{min}; see Methods) for models in which CD4 T_{CM} derive directly from naive CD4 T cells, and T_{EM} derive either from naive T cells or T_{CM}. AIC differences are shown in bold, with zero indicating the model with strongest support and large AIC differences representing reduced support, with differences of 10 or more generally considered highly significant. Figures in parentheses are an alternative measure, Akaike model weights, with values close to 1 indicating extremely strong support, and values close to zero indicating very weak support.

Constitutive generation of circulating CD4 T_{EM} and T_{CM} cells in adult mice occurs at constant rates irrespective of environment

The kinetics of accumulation of donor-derived memory cells were visually indistinguishable in the two environments from age 10 weeks onwards (Fig. 1C), suggesting similar net rates of recruitment and loss. Consistent with this, we found no significant differences between clean and dirty mice in the forces of recruitment (φ) into circulating CD4 T cell memory subsets between the ages of 11-64 weeks (Fig. 4A and Table 2). Therefore, the rates at which new CD4 T_{CM} and T_{EM} are generated both appear to remain in constant proportion to the size of the naive CD4 T cell pool throughout adulthood and into old age. We infer that antigenic stimuli common to both environments drive the continuous and steady generation of new effector and central memory CD4 T cells in adult mice.

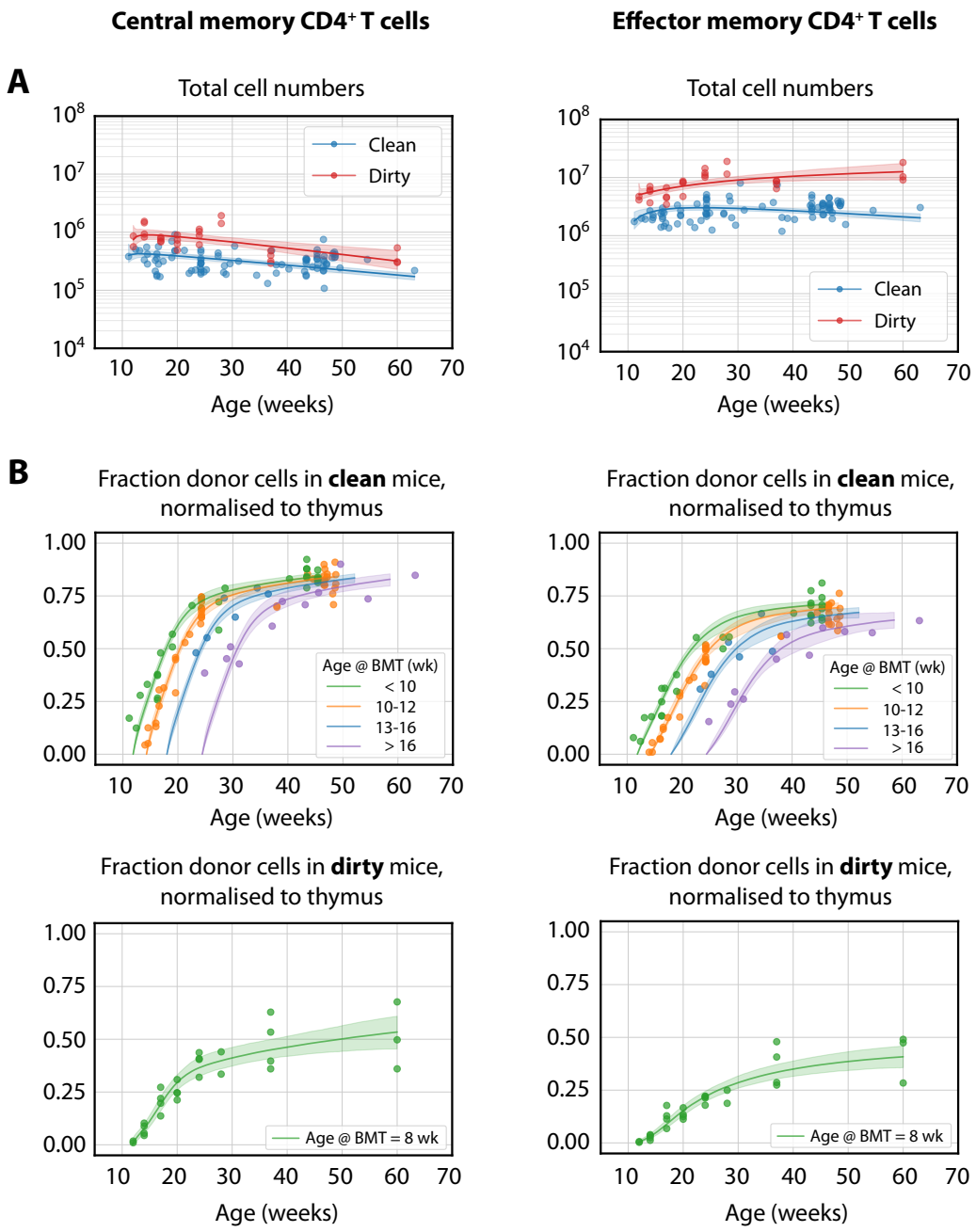

Figure 3 – Timecourses and best-fitting models of **(A)** total (donor+host) numbers of memory cells and **(B)** chimerism, modelled from 4 weeks post bone marrow transplant (BMT). The upper panels show fits for four groups of clean mice who differed in their age at BMT, but whose trajectories are described with a single set of parameters. The lower panels show the fitted trajectories of CD4 T_{CM} and T_{EM} chimerism in mice in the dirty environment, all of which underwent BMT at a similar age.

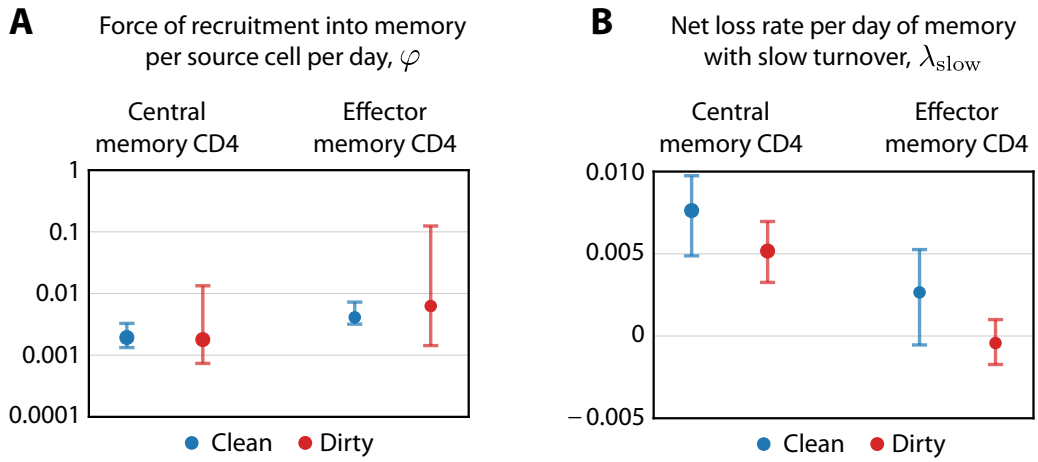

Figure 4 – Key parameters describing the constitutive production of memory CD4 T cells in adult mice. (**A**) Estimates of the baseline *per capita* rate of recruitment from the source (φ, cell^{-1} day^{-1}). Vertical bars represent bootstrapped 95% confidence intervals. (**B**) The net rate of loss of slow memory (loss minus self-renewal, λ_{slow} (day^{-1}))

Reduced chimerism in memory derives from both slow turnover and waning influx from the naive pool

In both environments, the donor chimerism in memory reached substantially lower levels than in the naive precursor pool (Fig. 1D). The rate at which donor cells replace memory is positively correlated with the rate of influx from the precursor population and the rate of memory turnover (see Methods). It can also be shown that if influx declines faster than the rate of turnover, donor cells will never be able to establish chimerism at the level of the source – in effect, the source 'dries up' more quickly than the memory pool can be replenished by immigrants (see Methods). We see signs of this effect in these mice. Due to thymic involution, naive CD4 T cell numbers decay exponentially in both clean and dirty adult mice (Fig. 1B) with half lives of 228 days (95% CI 227-231 days) and 143 (142-144) days respectively. These timescales are comparable to the half lives of T_{CM} clones with slow turnover in both clean and dirty mice and T_{EM} clones in clean mice (Table 2). Further, we infer that T_{EM} clones in dirty mice may persist indefinitely ($\lambda_{\text{slow}} \simeq 0$). Therefore, the limited chimerism we observed in all memory subsets derives from both waning influx and the persistence of existing memory clones.

Fast and slow T_{CM} and T_{EM} subpopulations are comparable in size but the transition to slow memory is inefficient

Fitting the models to the data gives us a quantitative picture of the process governing entry into the MP pools in both environments. We inferred that both CD4 T_{CM} and T_{EM} clones in the fast subpopulations have mean lifetimes of only a few days, a timescale that is largely determined by their net loss rate λ_{fast} and not by the transition to the slow population, which occurs very slowly (Fig. 4B and C, Table 2). Differentiation into the slower subsets is therefore relatively inefficient, with only ∼3-4% of T_{CM} and ∼20% of T_{EM} surviving to become the more persistent slow memory cells. These efficiencies appeared to be largely independent of environment, although with considerable uncertainty for T_{EM} in dirty mice. Despite the disparity in lifetimes of the two subpopulations, the substantial constitutive influx

Parameter	Quantity	CD4 T_{CM} Clean	CD4 T_{CM} Dirty	CD4 T_{EM} Clean	CD4 T_{EM} Dirty
Force of recruitment ($d^{-1} \times 10^{-3}$)	φ	1.7 (1.3, 3.3)	1.8 (0.74, 11)	3.8 (3.2, 7.3)	7.5 (1.4, 120)
Daily cell influx at age $t^* = 20$ wk ($\times 10^4$)	$\varphi S(t^*)$	2.8 (2.2, 5.4)	4.0 (1.6, 25)	6.3 (5.3, 12)	17 (3.2, 270)
Net loss rate of fast subset (d^{-1})	λ_{fast}	0.10 (0.072, 0.20)	0.13 (0.036, 1.1)	0.039 (0.019, 0.089)	0.52 (0.0020, 9.5)
Net loss rate of slow subset ($d^{-1} \times 10^{-3}$)	λ_{slow}	7.3 (4.9, 9.8)	5.3 (3.3, 7.0)	3.0 (-0.53, 5.3)	-0.33 (-1.7, 0.92)
Half-life of clones in fast subset (d)	$\ln(2)/\lambda_{\text{fast}}$	6.7 (3.4, 9.7)	5.3 (0.65, 19)	18 (7.9, 37)	1.3 (0.06, 160)
Half-life of clones in slow subset (d)	$\ln(2)/\lambda_{\text{slow}}$	96 (71, 140)	130 (100, 210)	230 (— , 3100)	—
Proportion of fast transitioning to slow	$\gamma/(\lambda_{\text{fast}} + \gamma)$	0.03 (0.01, 0.40)	0.04 (0.01, 0.09)	0.16 (0.00, 0.19)	0.19 (0.01, 1.0)
Proportion in slow memory at t^* =20 wk	$M_{\text{slow}}(t^*)/M(t^*)$	0.29 (0.18, 0.36)	0.64 (0.48, 0.75)	0.48 (0.19, 0.58)	0.96 (0.67, 0.99)

Table 2 – Estimates of parameters governing CD4 T_{CM} and T_{EM} homeostasis in adult mice. Where point estimates or confidence limits on the loss rate λ_{fast} were negative (indicating net growth), decay half-lives are not defined and are indicated with a dash (—).

into fast memory means that we estimate that the two pools are broadly comparable in size, again consistent with our previous analysis of BrdU labelling of CD4 T_{CM} and T_{EM} in WT mice [3], although with a notable enrichment for T_{EM} with slow turnover in dirty mice (Table 2).

Larger memory populations in dirty mice derive from early antigen exposure

Given the similarity of the rates of generation of memory in adult mice in different environments, the larger, stable memory populations in dirty mice (Fig. 3A) must derive from the early establishment of greater numbers of both T_{CM} and T_{EM}, and/or lower loss rates. While we saw a trend for greater persistence of memory in dirty environments (Fig. 4D), and even a suggestion of a low level of inflation of CD4 T_{EM} in dirty mice, these differences in net loss rates were not significant. Therefore, the larger memory population sizes in dirty mice must derive predominantly from the establishment of a higher 'set point' early in life. This difference is then sustained well into adulthood by the very slow loss, or even net growth, of these early memory populations.

To examine the forces contributing to the establishment of the different memory set points, we used the model parameters to predict the kinetics of ontogeny of the memory subsets. We drew on a timecourse of naive CD4 T cell numbers in WT mice aged between 5 days and 15 weeks, kept in the clean facility (Fig. 5A). We assumed that numbers of naive CD4 T cells early in life were similar in the clean and dirty environments, and that the forces of recruitment from the naive CD4 T cell pool into T_{CM} and T_{EM} and their loss rates were the same as in adulthood. We then used the timecourse of naive T cell numbers to predict the kinetics of accumulation of memory from age 5d in both environments, starting both from the mean numbers of CD4 T_{CM} and T_{EM} recovered from 5d old WT mice in the clean facility. These trajectories predicted the numbers of CD4 T_{CM} and T_{EM} at age 11 weeks in clean chimeric mice remarkably well (Fig. 5B). Even more strikingly, the trajectories closely tracked the observed growth in numbers of CD4 T_{CM} and T_{EM} recovered from clean WT mice aged 5d-15wk (Fig. 5B, grey/black data points), which spanned two orders of magnitude. However, the predicted trajectories in dirty mice underestimated CD4 T_{CM} numbers at age 11 weeks by a factor of roughly 5, and T_{EM} numbers by a factor of 2 (Fig. 5C). These levels were insensitive to the presumed size of the memory population in dirty mice at age 5d, because they are rapidly outnumbered by the influx of new memory cells from the exponentially expanding naive CD4 T cell pool.

We conclude that in the clean facility, the force of recruitment into memory is essentially constant from birth up to over a year in age and derives from self antigens and ubiquitous environmental or commensal antigens. An additional force early in life, presumably from increased exposure to environmental antigens, is required to account for the increased memory T cell numbers in adulthood in the dirtier facility.

Analysis of germ free mice confirms roles for both autoreactive and commensal stimuli in the establishment of memory compartments

We found that the rate of constitutive recruitment into memory in adult mice was insensitive to variations in environmental commensals, and that these antigens must exert their biggest influence on the establishment of MP cells in neonates and young mice. However, earlier studies reported that numbers of memory cells in the spleens of SPF and germ-free (GF) mice are similar and argue that self-recognition is therefore the sole driver of MP cell generation in early life [7]. To reconcile these apparent differences, and dissect the contributions of self and foreign antigens to the establishment of MP T cells in young mice, we compared the size and behaviour of memory CD4 T cell subsets in C57Bl6/J and/or C57Bl6/SJL mice housed in a range of environments. In addition to the clean (UCL) and dirty (NIMR) mice analysed above, we enumerated cells from GF and SPF mice obtained from the Kennedy Institute (KI) in Oxford. Consistent with these earlier studies, substantial numbers of both T_{CM} and T_{EM} MP cells were recovered from GF mice aged between 40-200d, confirming that the generation of MP cells does not depend exclusively on commensal-derived foreign antigens (Fig. 6A). However, the memory CD4 T cell compartments of GF mice, enumerated from spleen and lymph nodes combined, were significantly smaller than in the SPF mice in all facilities. Clean mice from KI and UCL had similar-sized memory compartments, and in turn both were substantially smaller than those in mice from the dirty facility (NIMR) (Fig. 6A). Indeed the dirty mice played host to around five times the number of MP cells found in GF mice, indicating that antigens from commensal organisms are a substantial driver of MP cell expansion.

We also analysed the proliferative activity of MP cells in mice from the different environments, by measuring the expression of Ki67. Division of MP cells is dependent on TCR [16] and CD28 costimulation signaling [7]. A substantial fraction of MP cells in GF mice expressed Ki67 (Fig. 6B), indicating cell cycle activity within the previous 3-4 days [3]. This proportion was broadly similar to that in mice from both clean and dirty environments, indicating that the level of proliferation of MPs in adult mice was relatively insensitive to environmentally-derived antigenic stimuli.

Finally, we wanted to make quantitative comparisons of the rates of memory generation in the different environments, in order to gain further insights into the magnitude of the forces responsible. We began with the parameters estimated for the clean (UCL) adult mice, which, as shown above, also accurately predicted their accumulation of memory from age 5d onward (Fig. 5B). We then estimated the fold changes in the force of recruitment φ needed to seed memory at the levels observed in the dirty (NIMR) mice and the GF mice (Fig. 7). For CD4 T_{CM}, GF mice needed about 0.4 times the force of recruitment estimated for clean mice, and dirty mice needed a force approximately 2.5 times larger. The figures for CD4 T_{EM} were similar (0.3 and 2, respectively). These estimates assumed that memory cell numbers in all facilities were identical at age 5d and equal to those observed in WT clean mice. Although there was uncertainty around the validity of this assumption, we found that allowing these initial numbers to vary had very little effect on the correction factors or the levels of memory attained at week 11 (Fig. S3). This

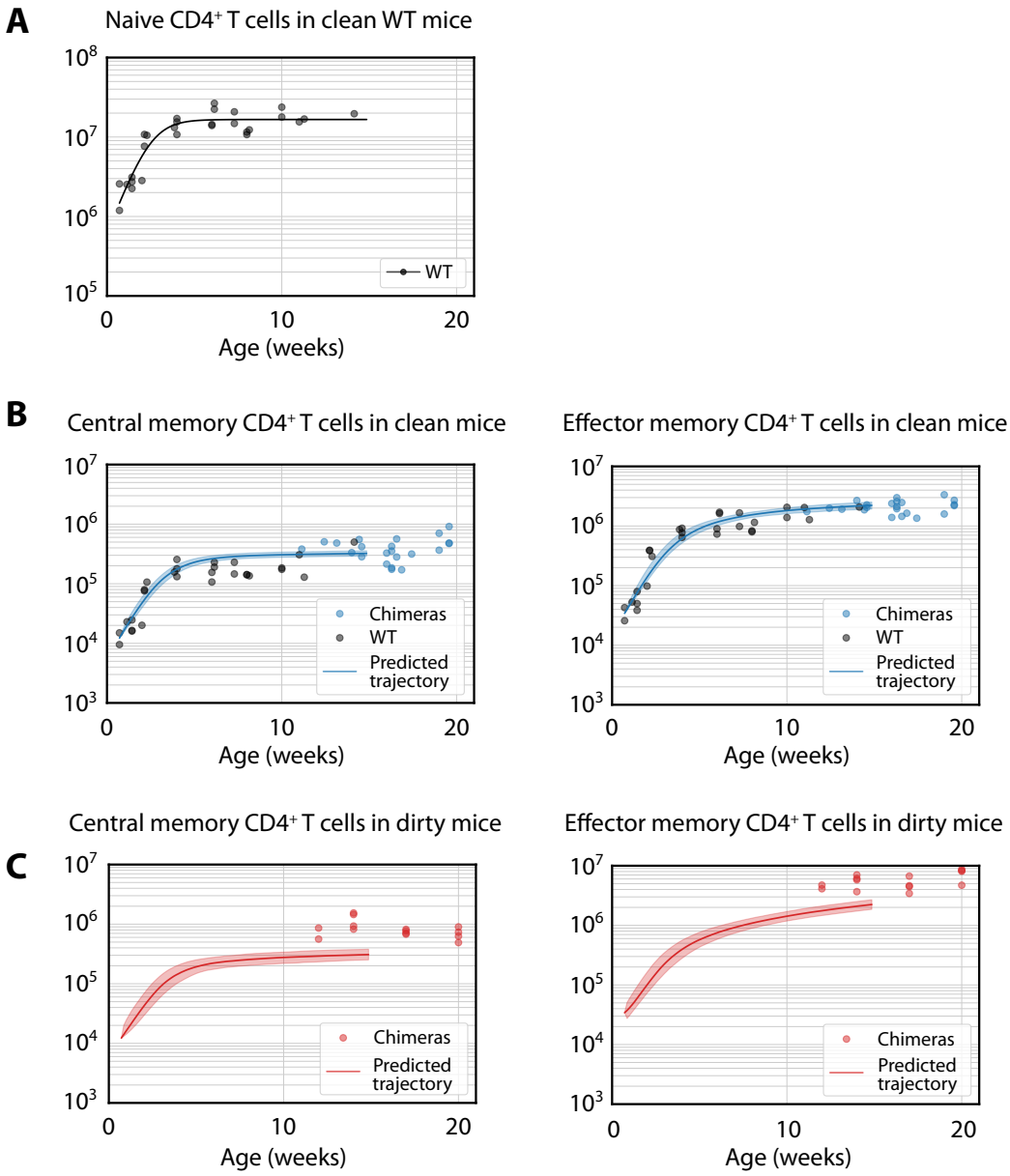

Figure 5 – Predicting the kinetics of ontogeny of CD4 T_{CM} and T_{EM} in clean and dirty environments. (A) The timecourse of numbers of naive CD4 T cells recovered from the spleen and lymph nodes of WT mice housed in the clean facility at UCL, aged 5 days to 15 weeks. We fitted an empirical function $S(t)$ to these data (black line) as the basis for predicting the generation of memory cells from birth onwards. **(B)** Predicting the development of memory cell numbers in clean mice (blue line) using the estimated daily rate of influx into memory, $\varphi S(t)$, from birth, and the loss rates λ_{fast} and λ_{slow} estimated in adults. Shaded regions are indicate the spread of predictions over the 95% confidence intervals of the parameters. Overlaid – but not used for fitting – are the numbers of CD4 T_{CM} and T_{EM} recovered from the spleen and lymph nodes of WT control mice in the same facility (black points). Blue points denote cell numbers recovered from the busulfan chimeras, which were used to estimate the parameters governing the prediction curve. **(C)** Predicted accumulation of memory cells in dirty chimeric mice, using the same method.

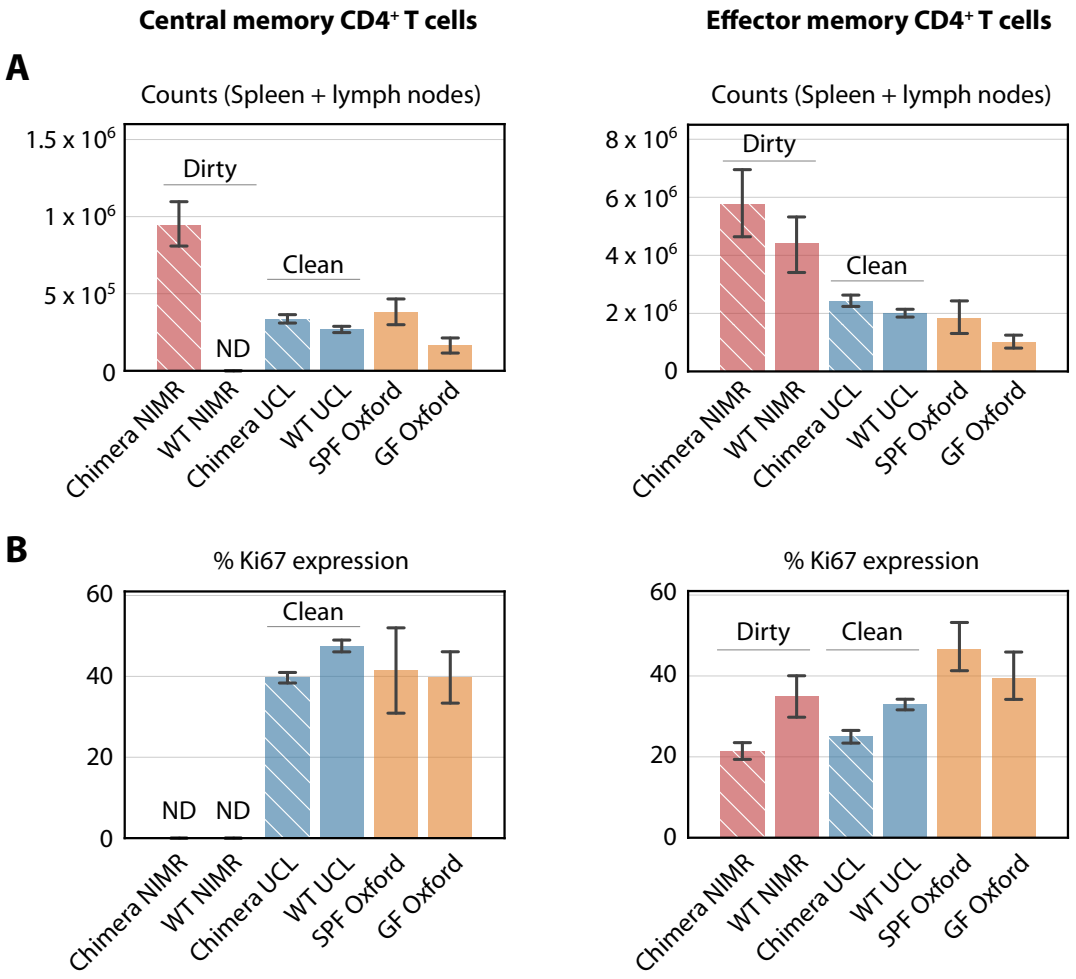

Figure 6 – Comparing numbers and proliferative activity of CD4 T_{CM} and T_{EM} in mice in different SPF and GF facilities. Cross-hatched bars denote busulfan chimeras, solid bars WT mice. NIMR and UCL are the 'dirty' and 'clean' facilities used for the bulk of the analysis presented here. 'ND' denotes data not available.

analysis suggested that approximately two-fold changes in the force of memory generation observed in UCL mice were sufficient to account for the reduced compartments in GF mice and the expanded memory compartments observed in dirty mice from NIMR.

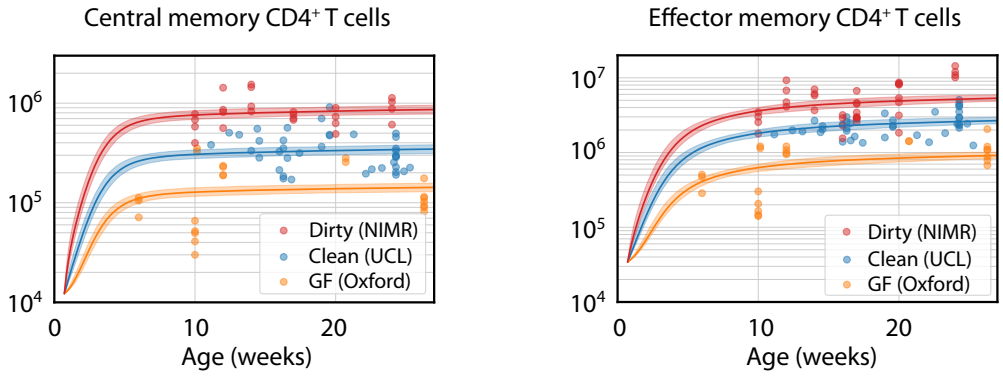

Figure 7 – Modelling the ontogeny of CD4 memory populations at different facilities. As in Fig. 5B, we took the parameters estimated for clean (UCL) adult mice and used them together with naive CD4 T cell numbers to predict the kinetic of accumulation of CD4 T_{CM} and T_{EM} from age 5d onwards (blue curve). Then, using these parameters, we estimated the correction to the force of recruitment φ needed to generated the observed memory cell numbers in the GF and dirty (NIMR) mice (orange and red curves). To do this we fitted the trajectories, with the correction factor as the single unknown, to available data from these facilities from mice up to 21 weeks of age. Curves show the projection as far as 27 weeks.

Discussion

In this study, we compared mice housed in facilities with distinct antigenic burdens to investigate the nature and timing of the forces that establish and maintain MP CD4 T cell compartments. We examined (i) the generation/establishment of the compartment in the young, (ii) tonic recruitment from the naive pool in adults, and (iii) the kinetic substructure and maintenance of these compartments throughout life. Our data suggest that self recognition contributes to all these processes, but that the contribution of reactivity to commensals is largely restricted to the neonatal period.

Our analysis of GF mice, which lack exposure to commensal organisms, confirmed earlier work showing that generation of MP cells is not driven exclusively by foreign commensals [7] and suggesting that self-recognition is instead the key driver. While GF mice lack exposure to live organisms, they are not entirely free of environmental antigens and it is possible that proteins in bedding material and diet could represent foreign antigenic stimuli. However, the argument for self-recognition is made through the correlation of the degree of MP cell conversion and steady state proliferation with affinity for self-MHC, as indicated by CD5 expression. We also found that cell division amongst memory cells, as reported by Ki67, was substantial and largely independent of the level of commensals. We could not determine the extents to which Ki67 expression derived from the homeostatic proliferation of existing cells or and influx of newly generated MP cells. However, the high levels of Ki67 in GF mice suggest that the processes of generation of memory cells are normal in the clean environment so are not driven by commensal organisms. More direct evidence came from our fate mapping of T cells in clean and

dirty hosts, revealing that both tonic generation of memory and their cell division were insensitive to environment, suggesting that self-recognition represents a major driver of both recruitment and proliferative renewal in adult memory compartments.

While the recruitment and dynamics of MP cells in adults appeared to be insensitive to commensal exposure, these antigens had a substantial impact upon the size of the memory compartment generated during ontogeny. CD4 T_{CM} and T_{EM} numbers in mice raised in dirty environments were 3-5 times greater than those in either GF or cleaner IVC or SPF facilities, and this difference in memory sizes reflected similar differences in the forces of recruitment driving influx into memory during ontogeny. The estimated five fold difference in this force between GF and 'dirty' mice prompts the simplistic interpretation that the drive from self recognition only accounts for $\sim$20% of the memory compartment in dirty mice, but might be the dominant force in clean mice. The broad range of relatively stable memory compartment sizes we observed in different hosts also supports the view that there is not a strict homeostatic limit on the size of memory compartment [17]. This flexibility is also evident from the rates of tonic memory generation in adults, which appear to be insensitive to the size of the memory compartment they are entering. As such, we saw no evidence that larger compartments resisted entry of new memory cells, which might be expected if there were significant competition between memory cells for homeostatic niches.

While we identified substantial differences in the memory compartments of 'clean' and 'dirty' mice, it was less clear whether the reduced commensal load in clean mice represented a significant force driving memory development. CD4 memory compartments of 'clean' mice were modestly larger than those of GF mice. However, we also found that the dynamics of memory generation in clean adult mice, which were largely insensitive to environment and likely driven by self-antigen recognition, also predicted with remarkable accuracy both the expansion of memory during ontogeny and subsequent maintenance in adults. This suggests any additional force contributing to memory development during ontogeny in clean mice by commensals maybe negligible next to those environment independent stimuli. The differences between clean and GF mice may, in part, be due to the smaller lymph nodes that develop in GF mice. Bacterial stimulation of DCs is required for their migration into LN and they are required for normal LN development [18–20]. Therefore, when comparing clean and GF mice, it is not straightforward to separate the indirect influence of commensals on lymphoid development from any direct influences upon memory generation. The apparent absence of additional forces driving memory ontogeny in clean mice is also surprising because previous studies suggest that lymphopenia in neonates drives memory formation [21]. This was demonstrated by transferring naive cells from adults into neonates, so it is possible that such naive cells do not represent the activity of naive T cells in neonates, that are almost exclusively RTE. What is clear from our analyses is that application of a constant force to the dynamics of naive T cell numbers throughout life is sufficient to explain the expansion and maintenance of memory compartments and no additional force, either from lymphopenia or environmental stimulus is required during the neonatal period.

Our analysis supports previous reports of heterogeneity in the rates of division and turnover of CD4 T_{CM} and T_{EM} [3, 22]. Using our observations of Ki67 within donor memory cells in busulfan chimeras, which begin higher than those in host cells but converge over time, and our estimates of the net loss rate λ_{fast}, we infer that naive CD4 T cells entering the T_{CM} and T_{EM} pools divide and die rapidly. These cells then transition slowly and inefficiently to more quiescent states. Our analysis also suggests that both T_{CM} and T_{EM} are fed independently from naive CD4 T cells, rather than through a linear naive-T_{CM}-T_{EM} path. It is possible however that both modes operate, with the linear pathway perhaps contributing

to a lesser degree. It may also be that cells transitioning to T_{EM} reside as T_{CM} too transiently for us to distinguish the two models.

We observed that environment not only impacted memory but also naive T cells. Their numbers naturally decline with age, but this decline occurred almost twice as fast in dirty mice as in genetically identical mice housed in cleaner facilities (Supplementary Fig. S4). The dynamics of replacement of host cells by donor cells in the busulfan chimeras were similar in the two cohorts, indicating that the lifespans of naive CD4 T cells were similar. Instead, more rapid loss of naive T cells derived from a more rapid reduction in thymic output in dirty mice, as measured by the size of thymocyte subsets (data not shown). Therefore, it appears that a further consequence of life in more antigenically diverse environments is more rapid involution of the thymus. Whether this effect is due to the stress or inflammation associated with an increased commensal burden, or is a consequence of expanded memory populations, is unclear. Whichever the case, our data clearly indicate that environmental factors can impact the maintenance of naive compartments. We emphasise that we control for this difference in naive T cell dynamics when studying memory generation, by measuring the force of recruitment per naive T cell, which is independent of total numbers of naive T cells, rather than the absolute influx.

The post-thymic age of naive T cells appears to have a significant impact upon their homeostatic fitness, such that older cells are more likely to be retained in the repertoire [23–25]. It may be that a similar effect operates in memory compartments, consistent with our conclusion that the subpopulations that we infer to be the end-points of T_{EM} and T_{CM} differentiation have very slow net rates of loss. Any such age effects will generate a richer substructure within memory, with clones established earlier in life better able to persist over time. It is possible, therefore, that memory established by commensals soon after birth may endure longer than memory generated in response to self antigens later in life.

Acknowledgements

This work was supported by the NIH (R01 AI093870) and the Medical Research Council (MC-PC-13055). We would like to thank Fiona Powrie and the Oxford Centre for Microbiome Studies for germ free mice.

Materials and methods

Generating busulfan chimeric mice

Mice were treated with optimised low doses of busulfan to deplete HSC but leave peripheral T cell subsets intact. HSC were reconstituted with congenically-labelled, T-cell depleted bone marrow to generate stable chimeras (Fig. 1A). Details of the protocols are given in refs. 26 and 27.

Mice

Busulfan chimeric mice and wildtype control mice were housed in conventional animal facilities, either at the National Institute for Medical Research, London, UK (NIMR); or at the UCL Royal Free Campus, London, UK (UCL). At NIMR, mice were housed in open cages and drank tap water. At UCL, mice were housed in individually ventilated cages and drank irradiated water. Germ Free and SPF mice were housed at the Oxford Centre for Microbiome Studies, Oxford, UK.

Flow cytometry

Single cell suspensions were prepared from the thymus, spleen and lymph nodes of busulfan chimeric mice, wildtype control mice, or germ free mice. Cells were stained with the following monoclonal antibodies and cell dyes: CD45.1 FITC, CD45.2 FITC, CD45.2 AlexaFluor700, TCR-β APC, CD4 PerCP-eFluor710, CD44 APC-eFluor780, CD25 PE, CD25 eFluor450, CD25 PE-Cy7, CD62L eFluor450, NK1.1 PE-Cy7 (all eBioscience), CD45.1 BV650, CD45.2 PE-Dazzle, TCR-β PerCP-Cy5.5 CD4 BV711, CD44 BV785, CD25 BV650 (all Biolegend), CD62L BUV737 (BD Biosciences), LIVE/DEAD nearIR and LIVE/DEAD blue viability dyes. For Ki67 staining, cells were fixed using the eBioscience Foxp3/ Transcription Factor Staining Buffer Set and stained with either anti-mouse Ki67 FITC or PE (both eBioscience). Cells were acquired on a BD LSR-II or a BD LSR-Fortessa flow cytometer and analysed with Flowjo software (Treestar). Conventional CD4 cells were identified as live TCR-β+ CD4+ CD25- NK1.1-, and then CD44 and CD62L were used to identify EM (CD44+CD62L-) and CM (CD44+CD62L+) subsets.

Modelling the fluxes between naive, central memory and effector memory CD4 subsets

The simple homogeneous model. The simplest mechanism we invoked to describe the kinetics of constitutive generation and renewal of the CD4 T_{CM} and T_{CM} is illustrated in Fig. 2A and was formulated as follows. We assume that cells flow into a given memory subset of total size $M(t)$ from a precursor population $S(t)$ at total rate $\varphi S(t)$, where t is the age of the animal. The rate constant φ is the 'force of recruitment', a compound parameter which is the product of the *per capita* rate of recruitment of cells from the source population per day multiplied by number representing any net expansion that occurs during recruitment. Memory is also lost at net *per capita* rate λ. This rate is the balance of loss through death and/or differentiation, and self-renewal through division. We place no constraints on

the net rate of growth or decay of each memory subsets, so λ may be positive or negative.

$$\frac{dM(t)}{dt} = \varphi S(t) - \lambda M(t). \tag{1}$$

We assume host and donor cells obey the same kinetics, so that

$$\frac{dM_{\text{host}}(t)}{dt} = \varphi S_{\text{host}}(t) - \lambda M_{\text{host}}(t)$$
$$\frac{dM_{\text{donor}}(t)}{dt} = \varphi S_{\text{donor}}(t) - \lambda M_{\text{donor}}(t),$$

where the total population size $M(t) = M_{\text{donor}}(t) + M_{\text{host}}(t)$. We can then write an equation for the dynamics of donor chimerism. We work with the normalised chimerism of the memory population, $\chi_{\text{M, norm}}(t)$. This is defined to be the fraction of cells in memory that are donor-derived divided by the (stable) fraction of early double-positive (DP1) thymocytes that are donor-derived, measured in the same mouse;

$$\chi_{\text{DP}} = \frac{\text{DP1}_{\text{donor}}}{\text{DP1}_{\text{host+donor}}}, \tag{2}$$

$$\chi_M(t) = \frac{M_{\text{donor}}(t)}{M(t)}, \quad \chi_{\text{M, norm}}(t) = \frac{\chi_M(t)}{\chi_{\text{DP}}}. \tag{3}$$

This quantity evolves over time according to

$$\begin{aligned}\frac{d}{dt}\chi_{\text{M, norm}}(t) &= \frac{1}{\chi_{\text{DP}}}\frac{d}{dt}\left(\frac{M_{\text{donor}}(t)}{M(t)}\right) \\ &= \frac{1}{\chi_{\text{DP}}M(t)}\left(\varphi S_{\text{donor}}(t) - \lambda M_{\text{donor}}(t) - \frac{M_{\text{donor}}(t)}{M(t)}\frac{dM(t)}{dt}\right) \\ &= \frac{\varphi S(t)}{M(t)}\left(\chi_{\text{source, norm}} - \chi_{\text{M, norm}}\right),\end{aligned} \tag{4}$$

where

$$\chi_{\text{source, norm}}(t) = \frac{1}{\chi_{\text{DP}}}\frac{S_{\text{donor}}(t)}{S(t)}. \tag{5}$$

By normalising the chimerism of both the source and the memory populations to that in the thymus, we removed the mouse-specific values of χ_{DP} and χ_M from the calculation. This process allowed us to account for any variation between animals in the degree of chimerism achieved with busulfan treatment and BMT.

The timecourses of naive CD4 T cell numbers in mice from age $t_0 = 11$ weeks in both facilities were well described with a simple model of exponential decay, $S(t) = S(t_0)e^{-\nu t}$ (Fig. S4). With this form we can write a solution for $M(t)$;

$$\begin{aligned}M(t) &= M(t_0)e^{\lambda(t_0-t)} + \int_{\tau=t_0}^{t} \varphi S(\tau)e^{-\lambda(t-\tau)}d\tau \\ &= M(t_0)e^{\lambda(t_0-t)} + \varphi S(t_0)e^{-\lambda t}\int_{\tau=t_0}^{t} e^{-(\nu-\lambda)\tau}d\tau \\ &= M(t_0)e^{\lambda(t_0-t)} + \frac{\varphi S(t_0)}{\lambda-\nu}(e^{-\nu t} - e^{\lambda(t_0-t)-\nu t_0}).\end{aligned} \tag{6}$$

Using this form for $M(t)$ in eqn. (4), the kinetics of normalised chimerism in memory can be written as

$$\begin{aligned}\frac{d}{dt}\chi_{\text{M, norm}}(t) &= \frac{\varphi S(t_0)e^{-\nu t}}{M(t_0)e^{\lambda(t_0-t)} + \frac{\varphi S(t_0)}{\lambda-\nu}\left(e^{-\nu t} - e^{\lambda(t_0-t)-\nu t_0}\right)} \times \left(\chi_{\text{source, norm}}(t) - \chi_{\text{M, norm}}(t)\right) \\ &= \frac{1}{e^{(\nu-\lambda)t}\left(\frac{e^{-(\nu-\lambda)t_0}}{\nu-\lambda} + \frac{M(t_0)e^{\lambda t_0}}{\varphi S(t_0)}\right) - \frac{1}{\nu-\lambda}} \times \left(\chi_{\text{source, norm}}(t) - \chi_{\text{M, norm}}(t)\right). \qquad (7)\end{aligned}$$

The rate of increase in donor chimerism in memory then depends on the force of recruitment φ, the dynamics of the source $S(t_0)e^{-\nu t}$, the pool size $M(t_0)$ and the net turnover rate λ. Note that Eq. (7) predicts that given sufficiently long, and if the rate of decline of naive T cells is slower than the rate of loss of memory ($\nu < \lambda$), the chimerism in memory will stabilise at the chimerism of naive CD4 T cells.

The simple model is characterised by the three unknowns $M(t_0)$, λ and φ, where t_0 was the earliest timepoint considered, age 11 weeks. To estimate these parameters we solved Eqs. (1) and (7) and fitted them simultaneously to the timecourses of total memory cell numbers and normalised chimerism, from age 11 weeks onwards, using the reduced likelihood approach described in ref. 28. We used empirical functions to describe the time-varying sizes and donor chimerism of naive CD4 T cells, $S(t)$ and $\chi_{\text{source, norm}}(t)$ (Fig. S4). The initial value of the normalised chimerism in the memory population was set to zero at 25 days post-BMT. Memory cell numbers were log transformed and chimerism values were arcsine-square-root transformed to normalise the distributions of residuals. We estimated 95% confidence intervals on parameters by bootstrapping residuals 1000 times, and taking the 2.5% and 97.5% quantiles of the resulting distributions of their estimates.

The two-compartment model. The second mechanism (Fig. 2B) describes the kinetics of constitutive renewal of CD4 T_{CM} and T_{EM} assuming that both comprise subpopulations with distinct rates of loss,

$$M(t) = M_{\text{fast}}(t) + M_{\text{slow}}(t), \qquad (8)$$

where t is the mouse age. We assume that cells flow only into one subset from the precursor population $S(t)$ at total rate $\varphi S(t)$. Note that when estimating the rates of loss of these subsets, we did not constrain their net rates of loss; rather, the model fits indicated that the upstream population assumed to be fed by naive T cells has a greater net rate of loss than the downstream subset, which motivates the description naive $\rightarrow$ fast $\rightarrow$ slow. This pathway is consistent with the empirical evidence for heterogeneity in memory in busulfan chimeras. As before, the force of recruitment φ is the product of the *per capita* rate of recruitment of cells from the source population per day multiplied by number representing any net expansion that occurs during recruitment. We assume that cells then transition to the slow population at *per capita* rate γ. The clonal turnover rates λ_{fast} and λ_{slow} are the balance of the rates of loss and growth in each subpopulation.

$$\frac{dM_{\text{fast}}(t)}{dt} = \varphi S(t) - \gamma M_{\text{fast}}(t) - \lambda_{\text{fast}} M_{\text{fast}}(t) \qquad (9)$$

$$\frac{dM_{\text{slow}}(t)}{dt} = \gamma M_{\text{fast}}(t) - \lambda_{\text{slow}} M_{\text{slow}}(t). \qquad (10)$$

Assuming that host and donor cells obey the same kinetics, such that eqns. 9 and 10 hold separately for both populations, we obtain the following equations for the dynamics of donor chimerism in the fast and slow subsets, each normalised to the chimerism of DP1 thymocytes (see Eq. (2));

$$\frac{d}{dt}\chi_{\text{fast, norm}}(t) = \frac{\varphi S(t)}{M_{\text{fast}}(t)}(\chi_{\text{source, norm}}(t) - \chi_{\text{fast, norm}}(t)) \tag{11}$$

$$\frac{d}{dt}\chi_{\text{slow, norm}}(t) = \frac{\gamma M_{\text{fast}}(t)}{M_{\text{slow}}(t)}(\chi_{\text{fast, norm}}(t) - \chi_{\text{slow, norm}}(t)), \tag{12}$$

where

$$\chi_{\text{fast, norm}}(t) = \frac{1}{\chi_{\text{DP}}}\frac{M^{\text{donor}}_{\text{fast}}(t)}{M_{\text{fast}}(t)}, \quad \chi_{\text{slow, norm}}(t) = \frac{1}{\chi_{\text{DP}}}\frac{M^{\text{donor}}_{\text{slow}}(t)}{M_{\text{slow}}(t)}. \tag{13}$$

The normalised chimerism in the fast and slow populations combined is

$$\chi_{\text{M, norm}}(t) = \chi_{\text{fast, norm}}(t)\frac{M_{\text{fast}}(t)}{M(t)} + \chi_{\text{slow, norm}}(t)\frac{M_{\text{slow}}(t)}{M(t)}. \tag{14}$$

Eqn. (11) predicts that the normalised chimerism in the fast subset will eventually stabilise at the level of naive CD4 T cells, if their rate of decline ν is slower than the total rate of loss of fast memory ($\nu < \gamma + \lambda_{\text{fast}}$). Similarly, eqn. (12) predicts that the normalised chimerism in the slow subset will stabilise at the level of the upstream fast subset (and eventually, that of naive CD4 T cells) if the rate of decline of naive CD4 T cells ν is slower than the the net rates of loss of both subsets ($\nu < \gamma + \lambda_{\text{fast}}$ and $\nu < \lambda_{\text{slow}}$).

The two compartment model is characterised by six unknowns; the sizes of the fast and slow subsets at time t_0, $M_{\text{fast}}(t_0)$, $M_{\text{slow}}(t_0)$; the turnover rates λ_{fast} and λ_{slow}; the transition rate γ; and the force of recruitment φ. The initial values of the normalised chimerism in both fast and slow subsets were set to zero at the age 25 days after BMT. To estimate these parameters we fitted the solutions of eqns. 8, 9, 10 and 14 simultaneously to the timecourses of total memory cell numbers and normalised chimerism, from age 11 weeks onwards, using the reduced likelihood approach described in ref. 28 and empirical descriptions of the size and chimerism of the naive CD4 T cell pool (Fig. S4). Memory cell numbers were log transformed and chimerism values were arcsine-square-root transformed to normalise the distributions of residuals. As before, we estimated 95% confidence intervals on parameters by bootstrapping residuals 1000 times, and taking the 2.5% and 97.5% quantiles of the resulting distributions of their estimates.

Measuring the relative support for models using the Akaike Information Criterion

Differences in the Akaike Information Criterion (AIC) can be used to generate degrees of support (or weights) for competing models [29, 30]. Consider a set of models $M_i, = 1 \ldots n$ and assume M_1 is the most strongly supported by the data (with the lowest AIC value). The difference in AIC between each model and the strongest model is

$$\Delta_i = -2\log(\mathcal{L}_i/\mathcal{L}_1) + 2(K_i - K_1), \tag{15}$$

where $\mathcal{L}_i$ is the likelihood of the data under model i and K_i is number of parameters estimated in maximising the likelihood. Differences of 10 or more between two models are considered highly significant, with most support for the model with the lowest AIC value.

An alternative measure of support is the Akaike weight w_i, which is the probability of model i being the one from the set of candidates that best captures the information in the data;

$$w_i = \frac{\exp(-\Delta_i/2)}{\sum_{j=1}^{n} \exp(-\Delta_j/2)}. \tag{16}$$

The weights w_i represent our relative belief in model i among the set of n candidates. The quantities in equations 15 and 16 are the measures shown in Table 1.

References

1. Tania S Gourley, E John Wherry, David Masopust, and Rafi Ahmed. Generation and maintenance of immunological memory. In *Seminars in Immunology*, volume 16(5), pages 323–333. Elsevier, 2004.

2. Susan M Kaech and E John Wherry. Heterogeneity and cell-fate decisions in effector and memory CD8+ T cell differentiation during viral infection. *Immunity*, 27(3):393–405, 2007.

3. Graeme Gossel, Thea Hogan, Daniel Cownden, Benedict Seddon, and Andrew J Yates. Memory CD4 T cell subsets are kinetically heterogeneous and replenished from naive T cells at high levels. *Elife*, 6:e23013, 2017.

4. Stephen C Jameson and David Masopust. Diversity in T cell memory: an embarrassment of riches. *Immunity*, 31(6):859–871, 2009.

5. Donna L Farber. T cell memory: heterogeneity and mechanisms. *Clinical Immunology*, 95(3):173–181, 2000.

6. Stephen C Jameson and David Masopust. Understanding Subset Diversity in T Cell Memory. *Immunity*, 48(2):214–226, 02 2018.

7. Takeshi Kawabe, Dragana Jankovic, Shuko Kawabe, Yuefeng Huang, Ping-Hsien Lee, Hidehiro Yamane, Jinfang Zhu, Alan Sher, Ronald N Germain, and William E Paul. Memory-phenotype CD4+ T cells spontaneously generated under steady state conditions exert innate Th1-like effector function. *Science Immunology*, 2(12), 2017.

8. Booki Min and William E Paul. Endogenous proliferation: burst-like CD4 T cell proliferation in lymphopenic settings. In *Seminars in Immunology*, volume 17(3), pages 201–207. Elsevier, 2005.

9. Motoko Kotani, Kazuya Hirata, Shuhei Ogawa, Katsuyoshi Habiro, Yasuo Ishida, Seiichi Tanuma, Reiko Horai, Yoichiro Iwakura, Hidehiro Kishimoto, and Ryo Abe. CD28-dependent differentiation into the effector/memory phenotype is essential for induction of arthritis in interleukin-1 receptor antagonist-deficient mice. *Arthritis & Rheumatism*, 54(2):473–481, February 2006.

10. Louise V Webb, Alessandro Barbarulo, Jelle Huysentruyt, Tom Vanden Berghe, Nozomi Takahashi, Steven Ley, Peter Vandenabeele, and Benedict Seddon. Survival of Single Positive Thymocytes Depends upon Developmental Control of RIPK1 Kinase Signaling by the IKK Complex Independent of NF-κB. *Immunity*, 50(2):348–361.e4, Feb 2019.

11. Jonathan Sprent and Charles D Surh. Normal T cell homeostasis: the conversion of naive cells into memory-phenotype cells. *Nature Immunology*, 12(6):478, 2011.

12. Judith N Mandl, João P Monteiro, Nienke Vrisekoop, and Ronald N Germain. T cell-positive selection uses self-ligand binding strength to optimize repertoire recognition of foreign antigens. *Immunity*, 38(2):263–274, 2013.

13. Lalit K Beura, Sara E Hamilton, Kevin Bi, Jason M Schenkel, Oludare A Odumade, Kerry A Casey, Emily A Thompson, Kathryn A Fraser, Pamela C Rosato, Ali Filali-Mouhim, et al. Normalizing the environment recapitulates adult human immune traits in laboratory mice. *Nature*, 532(7600):512, 2016.

14. You Jeong Lee, Stephen C Jameson, and Kristin A Hogquist. Alternative memory in the CD8 T cell lineage. *Trends in Immunology*, 32(2):50–56, 2011.

15. Thea Hogan, Graeme Gossel, Andrew J Yates, and Benedict Seddon. Temporal fate mapping reveals age-linked heterogeneity in naive T lymphocytes in mice. *Proceedings of the National Academy of Sciences*, 112(50):E6917–E6926, 2015.

16. Benedict Seddon, Peter Tomlinson, and Rose Zamoyska. Interleukin 7 and T cell receptor signals regulate homeostasis of CD4 memory cells. *Nature Immunology*, 4(7):680–686, July 2003.

17. Vaiva Vezys, Andrew Yates, Kerry A Casey, Gibson Lanier, Rafi Ahmed, Rustom Antia, and David Masopust. Memory CD8 T-cell compartment grows in size with immunological experience. *Nature*, 457(7226):196–9, 2009.

18. Meike Wendland, Stefanie Willenzon, Jessica Kocks, Ana Clara Davalos-Misslitz, Swantje I Hammerschmidt, Kathrin Schumann, Elisabeth Kremmer, Michael Sixt, Angelika Hoffmeyer, Oliver Pabst, and Reinhold Förster. Lymph node T cell homeostasis relies on steady state homing of dendritic cells. *Immunity*, 35(6):945–957, December 2011.

19. Christine Moussion and Jean-Philippe Girard. Dendritic cells control lymphocyte entry to lymph nodes through high endothelial venules. *Nature*, 479(7374):542–546, November 2011.

20. Zongde Zhang, Jianjian Li, Wencheng Zheng, Guang Zhao, Hong Zhang, Xiaofei Wang, Yaqian Guo, Chuan Qin, and Yan Shi. Peripheral Lymphoid Volume Expansion and Maintenance Are Controlled by Gut Microbiota via RALDH+ Dendritic Cells. *Immunity*, 44(2):330–342, February 2016.

21. Booki Min, Rebecca McHugh, Gregory D Sempowski, Crystal Mackall, Gilles Foucras, and William E Paul. Neonates support lymphopenia-induced proliferation. *Immunity*, 18(1):131–140, January 2003.

22. Souheil-Antoine Younes, George Punkosdy, Stephane Caucheteux, Tao Chen, Zvi Grossman, and William E Paul. Memory phenotype CD4 T cells undergoing rapid, nonburst-like, cytokine-driven proliferation can be distinguished from antigen-experienced memory cells. *PLoS biology*, 9(10):e1001171, 2011.

23. Hirotake Tsukamoto, Karen Clise-Dwyer, Gail E Huston, Debra K Duso, Amanda L Buck, Lawrence L Johnson, Laura Haynes, and Susan L Swain. Age-associated increase in lifespan of naive CD4 T cells contributes to T-cell homeostasis but facilitates development of functional defects. *Proc Natl Acad Sci U S A*, 106(43):18333–8, Oct 2009.

24. Sanket Rane, Thea Hogan, Benedict Seddon, and Andrew J Yates. Age is not just a number: Naive T cells increase their ability to persist in the circulation over time. *PLoS Biol*, 16(4):e2003949, Apr 2018.

25. Arnold Reynaldi, Norah L Smith, Timothy E Schlub, Cybelle Tabilas, Vanessa Venturi, Brian D Rudd, and Miles P Davenport. Fate mapping reveals the age structure of the peripheral T cell compartment. *Proc Natl Acad Sci U S A*, Feb 2019.

26. T Hogan, AJ Yates, and B Seddon. Generation of Busulfan chimeric mice for the analysis of T cell population dynamics. *Bio-protocol*, 4(24), 2017.

27. T Hogan, AJ Yates, and B Seddon. Analysing temporal dynamics of T cell division *in vivo* using Ki67 and BrdU co-labelling by flow cytometry. *Bio-protocol*, 7(24), 2017.

28. Thea Hogan, Graeme Gossel, Andrew J Yates, and Benedict Seddon. Temporal fate mapping reveals age-linked heterogeneity in naive T lymphocytes in mice. *Proc Natl Acad Sci U S A*, 112(50):E6917–26, Dec 2015.

29. H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

30. K. P. Burnham and D. R. Anderson. *Model Selection and Multimodel Inference*. Springer, second edition, 2002.

Supporting Information

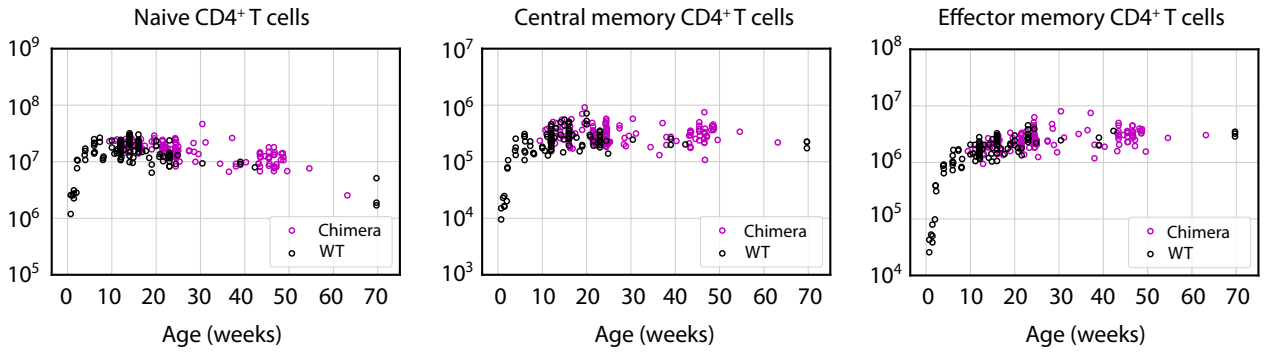

Figure S1 – Busulfan chimeric mice exhibit normal numbers of CD4 naive, T_{CM} and T_{EM} cells.

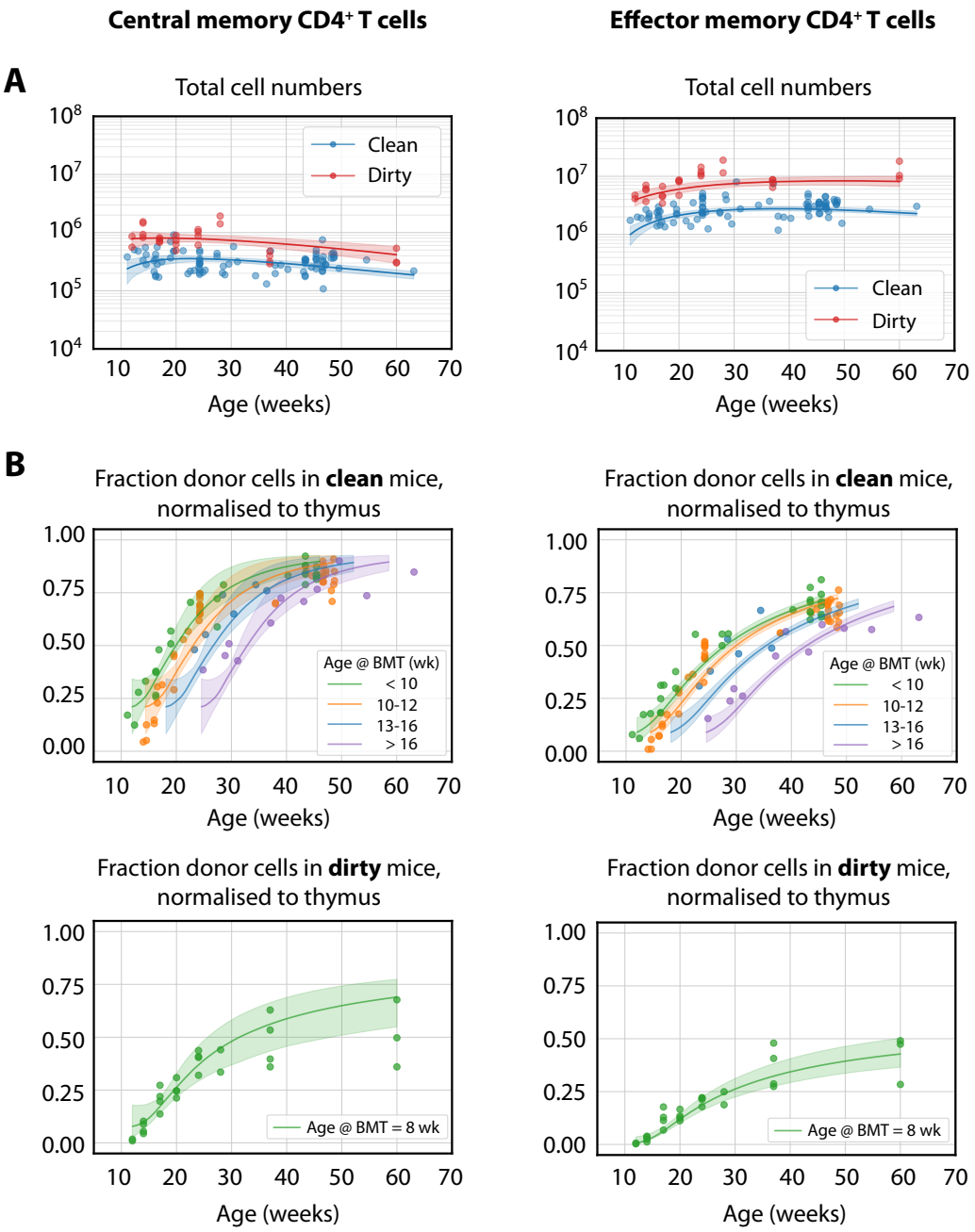

Figure S2 – Best fits using the single compartment model. (A) Total numbers of memory cells (host + donor) and **(B)** chimerism, modelled from age 11 weeks. The upper panels show fits for four groups of clean mice who differed in their age at BMT, but whose trajectories are described with a single set of parameters. The lower panels show the fitted trajectories of T_{CM} and T_{EM} chimerism for mice in the dirty environment, all of which underwent BMT at a similar age.

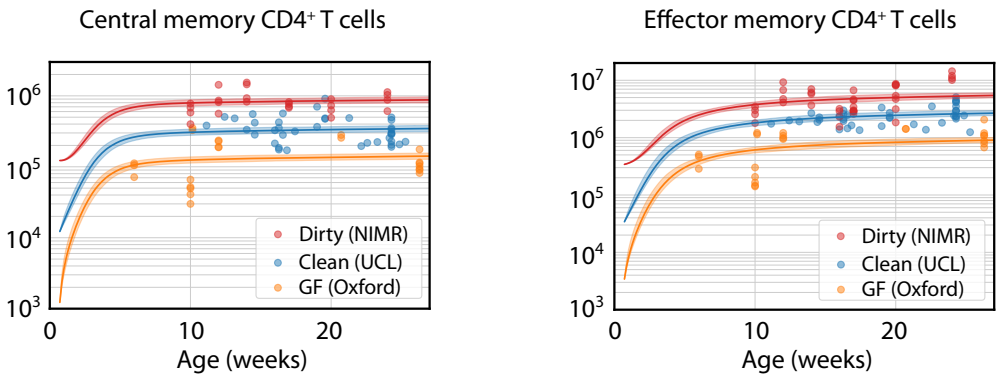

Figure S3 – Modelling of the ontogeny of CD4 T_{CM} and T_{EM} is insensitive to initial conditions. As in Fig. 7, we took the parameters estimated for clean (UCL) adult mice and used them together with naive CD4 T cell numbers to predict the kinetics of accumulation of CD4 T_{CM} and T_{EM} from age 5d onwards. Then, using these parameters, we estimated the corrections to the force of recruitment φ needed to generate the observed memory cell numbers in the GF and dirty (NIMR) mice, by fitting the trajectories to memory cell numbers up to 21 weeks. Curves show the projections as far as 27 weeks. The initial cell numbers in the dirty (NIMR) and GF (Oxford) environments were set to 10 and 0.1 times the UCL baseline numbers, respectively. Estimates of the correction factors were indistinguishable from those derived assuming the UCL baseline numbers in all three facilities, showing that our conclusions regarding the force of recruitment into memory CD4 T cell subsets in different environments are robust to uncertainty in the sizes of these subsets just after birth.

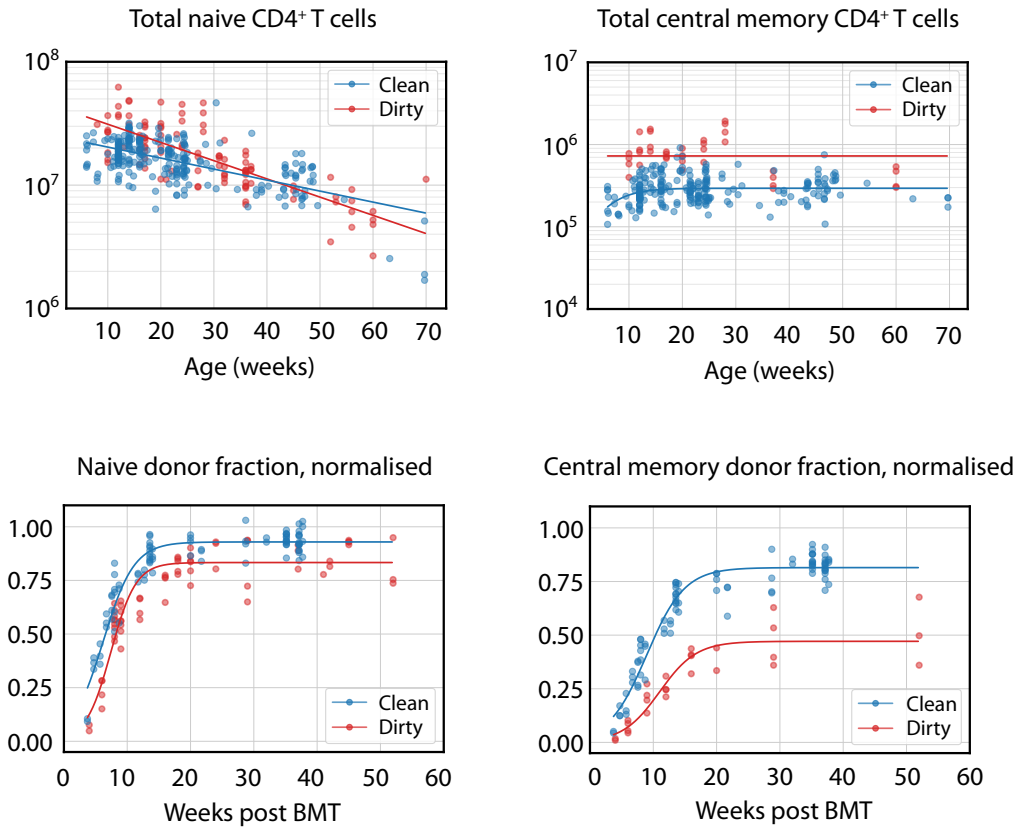

Figure S4 – Empirical descriptions of the size and chimerism of the putative source populations for T_{CM} and T_{EM}.