Human T_{SCM} cell dynamics *in vivo* are compatible with long-lived immunological memory and stemness

Pedro Costa del Amo[1], Julio Lahoz Beneytez[1], Lies Boelen[1], Raya Ahmed[2], Kelly L. Miners[3], Yan Zhang[2], Laureline Roger[4], Rhiannon E. Jones[4], Silvia A. Fuertes Marraco[5], Daniel E. Speiser[5], Duncan M. Baird[4], David A. Price[3], Kristin Ladell[3], Derek Macallan[2,6], Becca Asquith[1]*

[1]Department of Medicine, Imperial College London, London, UK.
[2]Institute for Infection and Immunity, St. George's Hospital, University of London, London, UK.
[3]Division of Infection and Immunity, Cardiff University School of Medicine, Cardiff, UK.
[4]Division of Cancer and Genetics, Cardiff University School of Medicine, Cardiff, UK.
[5]Department of Oncology, Lausanne University Hospital, Lausanne, Switzerland.
[6]St George's University Hospitals NHS Foundation Trust, London, UK.

*b.asquith@imperial.ac.uk

Author Summary

The human immune system remembers previously encountered pathogens so, on meeting the same pathogen a second time, the response is quicker and more effective. This immune memory is the basis of all vaccinations. Immune memory persists for decades but how memory is maintained is unclear. It has been hypothesised that there is a dedicated population of cells called stem cell-like memory T (T_{SCM}) cells which have stem cell-like behaviour and are responsible for the persistence of memory. Here we show that a subset of T_{SCM} cells, in healthy humans in vivo, have the dynamic properties of self-renewal and clonal longevity necessary to maintain long-lived immune memory.

Abstract

Adaptive immunity relies on the generation and maintenance of memory T cells to provide protection against repeated antigen exposure. It has been hypothesized that a self-renewing population of T cells, named stem cell-like memory T (T_{SCM}) cells, are responsible for maintaining memory. However, it is not clear if the dynamics of T_{SCM} cells *in vivo* are compatible with this hypothesis. To address this issue, we investigated the dynamics of T_{SCM} cells under physiological conditions in humans *in vivo* using a multidisciplinary approach that combines mathematical modelling, stable isotope labelling, telomere length analysis, and cross-sectional data from vaccine recipients. We show that, unexpectedly, the average longevity of a T_{SCM} clone is very short (half-life<1 year, degree of self-renewal=430 days): far too short to constitute a stem cell population. However, we also find that the T_{SCM} population is comprised of at least two kinetically-distinct subpopulations which turnover at different rates. Whilst one subpopulation is rapidly replaced (half-life=5 months) and explains the rapid average turnover of the bulk T_{SCM} population; the half-life of the other T_{SCM} subpopulation is approximately 9 years, consistent with the longevity of the recall response. We also show that this latter population exhibited a high degree of self-renewal, with a cell residing without dying or differentiating for 15% of our lifetime. Finally, although small, the population was not subject to excessive stochasticity. We conclude that the majority of T_{SCM} cells are not stem cell-like but that there is a subpopulation of T_{SCM} cells whose dynamics are compatible with their putative role in the maintenance of T cell memory.

Introduction

The maintenance of long-lived T cell memory is one of the hallmarks of adaptive immunity [1, 2]. Multiple studies have shown that the recall response to a previously encountered antigen has a half-life of the order of decades [3, 4]. It has been hypothesised that this T cell memory is dynamically maintained by differentiation of a precursor stem cell-like memory population [5]. Alternative, non-exclusive, explanations include replacement by proliferation of differentiated memory T cells; or the existence of a putative subpopulation of long-lived memory T cells that has not yet been identified, either because such cells are very rare or because they reside primarily outside of the peripheral blood [6-9].

Central memory (T_{CM}) T cells ($CD45RA^{-}CCR7^{+}$ in humans) were previously thought to constitute the stem cell-like memory precursor population. Evidence supporting the "stemness" of T_{CM} cells includes their capacity to differentiate into effector memory T cells (T_{EM}) and T effector cells (T_{EFF}) [10, 11]. This hypothesis was further strengthened by cell fate tracking experiments in mice (using genetic barcoding and single cell transfer) showing that T_{CM} cells had the capacity to self-renew and that a single T_{CM} cell could reconstitute immune protection against an otherwise lethal pathogen [12, 13].

However, the concept of T_{CM} as the stem cell population has been challenged by the identification of "stem cell-like" memory T cells (T_{SCM}), in mice [14], non-human primates [15], and humans [16], which have enhanced stem cell-like properties compared to T_{CM} cells. In humans, like naïve cells, T_{SCM} cells are $CD45RA^{+}CD45RO^{-}$ and they express high levels of CD27, CD28, IL-7Rα, CD62L and CCR7. Unlike naïve cells, T_{SCM} cells are clonally expanded and express the memory markers CD95 and CD122 [1, 16]. T_{SCM} cells exhibit enhanced proliferative capacity compared to T_{CM} cells, the potential to differentiate into all other classically defined T cell memory subsets (including T_{CM}) as well as the ability to retain their phenotype following proliferation both *in vitro* and in mice *in vivo* [1, 14-16]. In the light of these observations, it has been suggested that T_{SCM} cells are the main stem cell memory population and play a key role in maintaining long-term memory *in vivo* [15-18].

There are three basic prerequisites for T cell memory stemness: multipotency, self-renewal and clonal longevity. In this study we focus on the related dynamic properties of self-renewal and clonal longevity. Self-renewal of human T_{SCM} cells has been demonstrated *in vitro* [19], but it remains a concern that the local micro-environment, which may crucially affect the degree of self-renewal, will be different *in vivo* and *in vitro*. However, proving self-renewal of human T_{SCM} cells *in vivo* has so far not been possible because of ethical and technical limitations. The second property we investigate is clonal longevity. Long-lived T cell memory requires that memory T cells clonotypes expressing the same T cell receptor (TCR) persist for several decades *in vivo*. For example, influenza immunity has been shown to last for several decades [20], and small pox vaccine-induced T cell memory has a half-life of 8-15 years [3, 4]. For T_{SCM} cells to constitute a potential precursor population for T cell memory *in vivo*, the survival of T_{SCM} clones needs to be consistent with those estimates. A number of studies

suggest that T_{SCM} clones can survive for several years. Biasco et al. [17] observed that genetically engineered T_{SCM} cells could persist for many years in patients suffering from severe combined immunodeficiency disease. Fuertes Marraco et al. [21] identified a Yellow Fever Virus (YFV)-specific T_{SCM} population up to 25 years after vaccination. Finally, in leukaemia patients who had undergone hematopoietic stem cell transplantation, Oliveira et al. [22] reported that gene-modified T_{SCM} cells could be detected in the circulation up to 14 years after treatment. These studies support the concept of T_{SCM} longevity, albeit in scenarios of lymphocyte depletion or profound $CD8^+$ T cell expansion. However, it has been shown that the dynamics of post-transplant haematopoiesis in mice differs significantly from normal, unperturbed haematopoiesis [23-25] and so it cannot be assumed that these transplantation studies in humans necessarily recapitulate the healthy human system. In short, T_{SCM} longevity has not been quantified in normal, unperturbed homeostasis in humans and the related question of the ability of T_{SCM} to self-renew has not been addressed in any human *in vivo* system.

In order to investigate human T_{SCM} cells in homeostasis, we previously performed stable isotope labelling of healthy volunteers and analysed label uptake in $CD4^+$ and $CD8^+$, naïve and T_{SCM} T cells. We found that the T_{SCM} population was rapidly turning over (median 0.02 per day, inter-quantile range 0.016-0.037 per day, half-life<1 year), and concluded that the T_{SCM} population is dynamically maintained [26]. However, in this previous study only labelling data were modelled and so it was not possible to address the central question of the "stemness" of the T_{SCM} pool. First, to constitute a stem cell population it is not enough to have a stably maintained population of cells; stemness requires long term clonal persistence [18]. That is, whilst the size of the T_{SCM} population as a whole may be stably maintained, the lifespan of any given antigen-specific precursor population could be short; such limited lifespans would be difficult to reconcile with the hypothesis that T_{SCM} cells are the repository of T cell memory. Second, the high turnover rates obtained in this labelling study [26] do not necessarily indicate that the majority of the T_{SCM} population is replaced by the self-renewal of the T_{SCM} pool; frequent naïve cell differentiation could also be responsible. Indeed, given the very large size of the naïve pool compared to the T_{SCM} pool [19], a relatively low proportion of proliferating naïve cells would be sufficient to replace lost T_{SCM} cells. In this scenario T_{SCM} cells would simply represent transit cells on the differentiation pathway from naïve to effector rather than self-renewing stem cells.

Here we investigate whether the dynamics of T_{SCM} cells in healthy humans are consistent with their putative role as memory stem cells. Specifically, we investigate both the capacity of T_{SCM} cells to self-renew and the longevity of T_{SCM} clones. It is challenging to address these questions in humans and they cannot be answered using stable isotope labelling alone since different scenarios (e.g. "all new T_{SCM} cells come from naïve cell differentiation" versus "all new T_{SCM} cells come from T_{SCM} proliferation") can give rise to very similar levels of label in the T_{SCM} population. To enable us to deconvolute these possibilities we performed telomere length analysis and utilised cross-sectional T_{SCM} cell data from YFV-vaccine recipients. Deterministic

and stochastic mechanistic mathematical modelling were then used to analyse all three datasets. This novel approach allows us to investigate human T_{SCM} cell dynamics *in vivo* and to address questions previously only investigated in animal models.

Results

To investigate the dynamics of T_{SCM} cells in humans *in vivo* we analysed experimental data that we obtained in a 7 week stable isotope labelling study of 5 healthy individuals [26], in which label incorporation into $CD4^+$ and $CD8^+$ naïve ($CD45RA^{bright}CD27^{bright}CCR7^+CD95^-$) and T_{SCM} ($CD45RA^{bright}CD27^{bright}CCR7^+CD95^+$) T cells was measured at multiple time points (Fig 1). In addition, we performed single telomere length analysis of each cell population in the same individuals. We then constructed an ordinary differential equation-based mathematical model to describe both stable isotope (heavy water) labelling and telomere length, in which we assumed a linear differentiation pathway from naïve to T_{SCM} cells (Fig 2, Methods [27]).

Fig 1. Cell surface phenotype of naïve and T_{SCM} populations

A) Gating strategy used to sort $CD4^+$ and $CD8^+$ naïve and T_{SCM} cells for isotope label and telomere length analysis: the top panels show the consecutive gating to detect $CD8^+$ or $CD4^+$ T cells; the bottom panels show the further gating to detect Naïve or T_{SCM} cells within $CD8^+$ or $CD4^+$ populations **B)** Expression of CD45R0, CD28, CD127 and CD45RA on $CD8^+$ $CCR7^+$ $CD95^-$ naïve cells (blue cloud) and $CD8^+$ $CCR7^+$ $CD95^+$ T_{SCM} cells (red cloud) compared with bulk $CD8^+$ T cells (black cloud). **C)** as for B but depicting $CD4^+$ naïve and $CD4^+$ T_{SCM} compared to bulk $CD4^+$ T cells.

Fig 2. Model to describe the naïve and T_{SCM} populations

A) Schematic representation of the model for naïve (T_N) and T_{SCM} populations. **B)** Schematic representation of the model for telomere length data for naïve and T_{SCM} populations when $C=k$ (inactive telomerase). T_{Ni} (or T_{SCMi}) represent the number of T_N (or T_{SCM}) cells which have divided i times; p_n, p_s, d_n, and d_s are the proliferation and disappearance rates of T_N and T_{SCM} populations; Δ is the fraction of naïve cells recruited per day, and k is the number of divisions that occur during clonal expansion.

Heterogeneity in the T_{SCM} population.

Estimates of the rate of T_{SCM} renewal and T_{SCM} clonal lifespan will depend upon the kinetic structure of the T_{SCM} pool. We therefore first asked whether there was evidence for kinetic heterogeneity (i.e. existence of subpopulations with differing kinetics) within the T_{SCM} pool by comparing the quality of fit of a homogenous and heterogeneous version of the mathematical model. In the homogeneous version of the model we constrain the input rate (proliferation rate + rate of new entrants due to differentiation from naïve cells) of the whole T_{SCM} population to be equal to the disappearance rate of labelled T_{SCM} cells; this condition will be met for a kinetically homogenous population of constant size. In the heterogeneous version of the model this constraint was relaxed to allow for the possibility of kinetic heterogeneity

in the T_{SCM} pool (Methods, [28]). We used this implicit description of kinetic heterogeneity rather than an explicit description of the subpopulations because it requires fewer parameters (two compared with three for the explicit model [28-30]) and furthermore does not suffer from the parameter identifiability issue inherent in the explicit kinetic heterogeneity model which arises due to the very strong correlation between the proliferation rate and size of a subpopulation [31]. A total of 9 datasets were included in this analysis, representing $CD4^+$ and $CD8^+$ T cells (naïve and T_{SCM}) from 5 individuals (one $CD8^+$ T_{SCM} cell dataset from one subject was not available). We found that, in 7 out of the 9 cases, constraining the T_{SCM} population to be homogeneous resulted in a substantially worse description of the data (Fig 3); and there was strong evidence to reject the assumption of homogeneity $P=5.8\text{x}10^{-7}$, $P=4.1\text{x}10^{-6}$ (median of p-values calculated using Fisher's F-test for nested models between the homogeneous and heterogeneous models for $CD4^+$ and $CD8^+$ T_{SCM} resp.), indicating considerable support for the heterogeneous description of the T_{SCM} pool in both $CD4^+$ and $CD8^+$ T cell populations (S1 Table). In contrast, there was no evidence to reject the null hypothesis of homogeneity in the naïve cell pool ($P=0.6$, $P=0.5$, for $CD4^+$ and $CD8^+$ T_N resp; median of p-values calculated using Fisher's F-test for nested models between the homogeneous and heterogeneous models; S1 Table).

Fig 3. Label incorporation and telomere length in $CD4^+$ and $CD8^+$ naïve and T_{SCM} T cells.

A) Experimental labelling data (black dots) and best fit of model to the data assuming kinetic homogeneity (blue solid line) and kinetic heterogeneity (red dashed line) of the $CD4^+$ naïve (T_N) pool (left column) and the $CD4^+$ T_{SCM} pool (right column). **B)** Experimental measurement (red symbol) and best fit of the homogeneous (homog) and heterogeneous (heterog) models (black symbols) to the average telomere length differences (Θ) between $CD4^+$ T_N and $CD4^+$ T_{SCM} cells. Labelling data and telomere length data were fitted simultaneously. We found strong evidence to reject the null hypothesis of homogeneity in the $CD4^+$ T_{SCM} population (median $P=5.8\text{x}10^{-7}$, pooled $P=3.5\text{x}10^{-23}$) but not in the $CD4^+$ T_N population (median $P=0.6$, pooled $P=0.9$). **C)** and **D)** as for A and B but for $CD8^+$ cells rather than $CD4^+$ cells. Again, for $CD8^+$ cells as for $CD4^+$, we found strong evidence to reject the null hypothesis of homogeneity in the T_{SCM} population (median $P=4.1\text{x}10^{-6}$, pooled $P=6.1\text{x}10^{-15}$) but not in the T_N population (median $P=0.5$, pooled $P=0.7$). Experimental data depicted in this figure can be found in S1 Data).

Magnitude of clonal expansion.

The size of a newly generated T_{SCM} clone will be an important determinant of clonal longevity as this determines not just the initial magnitude of a new clone but also the rate at which an existing clone is displaced by new entrants bearing different TCRs. Unfortunately, the size of the clonal expansion accompanying the differentiation of naïve to T_{SCM} cells (k in the model. Fig 2, Methods) was not identifiable. Different fitting runs to the same data set (with different

initial conditions or different random seeds) gave different estimates of the clonal expansion parameter *k*. Consistent with this, we found that if *k* was fixed to different constant values in the range 0 to 20 then, with the exception of one individual where the sum of squares increases dramatically for k above 15, the sum of squares remained constant in every case for all values of *k* (Fig 4A and Fig 4E). Henceforth, we systematically repeat all analyses for multiple values of k in the range 0 to 20 to ensure that results are robust despite uncertainty in the clonal expansion parameter. Values of k above 20 were not considered biologically plausible [13, 32, 33].

Fig 4. Estimates of T_{SCM} cell parameters from the implicit kinetic heterogeneity model

A-H) Estimates of T_{SCM} cell parameters as a function of *k* in the $CD4^+$ (top row: A - D) and $CD8^+$ (bottom row: E - H) lineages. Isotope labelling and telomere length data were fitted simultaneously, fixing the clonal expansion size *k* to values between 0 and 20 and leaving the remaining parameters free. **A**) and **E**) show the variation in the sum of squared residuals (ssr) with *k* in $CD4^+$ and $CD8^+$ T cells respectively; **B**) and **F**) show the variation in the fraction of newly generated T_{SCM} cells originating from self-renewing T_{SCM} proliferation computed as $(p_s T_{SCM})/(2^k \Delta T_N + p_s T_{SCM})$; **C**) and **G**) show the variation in T_{SCM} half-lives for $CD4^+$ and $CD8^+$ T cells respectively; and **D**) and **H**) the variation in T_{SCM} antigen-specific precursor lifespans (time until the last cell specific for a given antigen dies or differentiates). In all cases k is plotted on the x axis. All results depicted are provided in S1 Data.

In summary, we were not able to estimate the size of the clonal expansion *k* upon differentiation of naïve cells to T_{SCM} cells and instead utilised a strategy to investigate T_{SCM} dynamics despite uncertainty in this parameter.

T_{SCM} clonal longevity.

Next we quantified T_{SCM} clonal longevity. We fitted the mathematical model (implicit heterogeneous version) simultaneously to the telomere length and isotope labelling data with *k* fixed sequentially at different values in the range 0 to 20. We found that, with the exception of one dataset ($CD8^+$ T cells in DW01), the contribution of naïve cells to T_{SCM} replacement was never less than 20% and could be as much as 90% (Fig 4B and 4F). Correspondingly, the average half-life of a T_{SCM} clone was short: the maximum ever observed (across all values of *k* and across all individuals) was 4 years, but typically it was much shorter and in the range 0-500 days (Fig 4C and 4G). These conclusions about short clonal longevity were robust to assumptions regarding the activity of telomerase. Specifically, for all values of telomerase compensation considered in the range 0 to *k,* the estimated average clonal half-lives were never higher than those estimates reported above (in which compensation was a free parameter). Importantly, this half-life represents the duration of memory to an antigen (Methods), not the conventional population half-life.

It is possible that a very small number of surviving T_{SCM} cells is sufficient to generate a substantial recall response and so a short clonal half-life is not necessarily incompatible with long-lived recall responses. Moreover, extinction of some clones specific for a given antigen is not necessarily problematic if other clones (bearing different TCR) specific for the same antigen survive. To assess this possibility we used the exact Gillespie algorithm (Methods), to quantify the time for the last cell of an antigen-specific precursor population to disappear. This is reported as the precursor lifespan in Fig 4D and 4H. Whilst this did lead to a considerable increase in longevity (stochastic estimates of total antigen-specific precursor lifespan were typically three times longer than the deterministic clonal half-life) maximum estimates were still only of the order of 2,000 days (about 5yrs) for most individuals.

In summary, although individual parameters were poorly identifiable, all parameter combinations able to describe the experimental data were associated with average clonal half-life estimates which were much lower than the 8-15 years half-life of the recall response [3, 4]. Even total precursor lifespans (times until the last cell of an antigen-specific precursor population disappears) were lower than those values in most cases. We conclude that the average T_{SCM} population is replaced too rapidly for it to be the stem cell population responsible for maintaining memory.

Subpopulation kinetics: yellow fever virus-specific responses

The model used above allows for heterogeneity but nevertheless reports population averages (i.e. the proliferation rate and clonal half-life averaged across the whole T_{SCM} population). This averaging could be hiding a small, long-lived population within the bulk short-lived population. Expanding our model which deals with heterogeneity implicitly (and thus averages across the population) to one which deals with heterogeneity explicitly (and thus provides estimates for the half-lives of all subpopulations) is problematic, as even the simplest version of the explicit heterogeneity model suffers from severe identifiability issues [30] and fails to deliver the parameters of interest when fitted to labelling data. We confirmed that, for our more complex system with both naïve and T_{SCM} cells, an explicit description of heterogeneity provided no information. To address this problem we therefore sought an alternative class of data.

We analysed published data of the vaccine-induced (YFV)-specific T_{SCM} response in humans from Fuertes Marraco *et al* [21]. In brief, the magnitude of the $CD8^+$ T_{SCM} cell response to the HLA-A*02-restricted YFV $NS4b^{214-222}$ epitope was measured by HLA class I tetramer at different time points (range 0.27 years-35.02 years) post-vaccination in a cross-sectional study of 37 recipients of the YF-17D YFV vaccine.

We fitted the explicit heterogeneity version of the naïve (T_N) and T_{SCM} model to all three types of $CD8^+$ T cell data (isotope labelling, telomere length, and YFV) simultaneously (SI Methods). The fits are shown in Fig 5. As for the implicit heterogeneity model, the fraction of new T_{SCM} cells originating from naïve cells was high (min 10%, median 44%). We found evidence for at

least two subpopulations of $CD8^+$ T_{SCM} cells (designated T_{SCM1} and T_{SCM2} for the purposes of this discussion). The majority of the T_{SCM} cells generated upon clonal expansion of naïve cells differentiated into the T_{SCM1} subpopulation, characterized by a short half-life (≤ 1 year) and a high replacement rate (median 0.02 per day, inter-quantile range, 0.024-0.045 per day), slightly higher than the average rates estimated by the previous model. The remaining fraction of the generated clone was observed to enter a long lived subpopulation with a median half-life of 9 years (Table 1, S2 Table). Surprisingly, although the fraction of naïve cells entering the T_{SCM2} pool was low, because of its low death/differentiation rate, the number of long-lived T_{SCM2} cells in the circulation at any given time could be as high, or even higher, than the number of rapidly proliferating T_{SCM1} cells. Results are summarised schematically in Fig 5C. Four other weighting strategies (of the different types of data) yielded the same conclusions in all cases (S1 Fig, S3 Table).

Fig 5. Label incorporation, telomere length, and YFV predictions for $CD8^+$ naïve and T_{SCM} cells.

A) Fit of the explicit heterogeneity model to isotope labelling data from $CD8^+$ naïve (T_N) and T_{SCM} cells, and to the YFV-specific T_{SCM} data. Experimental data is represented by solid black symbols. **B)** fits to the average telomere length differences (Θ) between the $CD8^+$ T_N and T_{SCM} pools, experimental data shown in red. The number of base pairs (bps) lost per division was taken to be δ =50bp/division. Isotope labelling, telomere length, and the YFV data were fitted simultaneously. **C)** Schematic summary of estimated T_N and T_{SCM} dynamics. Estimates are the medians across subjects obtained from the fits shown in panels A and B. Parameter estimates for each subject are provided in Table 1 and S2 Table. The size of the squares is proportional to the population size. Labelling data and telomere data are provided in S1 Data, the YFV tetramer data was previously published [21].

Table 1. Parameter estimates for $CD8^+$ T_{SCM} cells from the explicit heterogeneity model

Parameter estimates with 95% confidence intervals (in parentheses) obtained by fitting the explicit heterogeneity model to isotope labelling, telomere length and YFV datasets simultaneously. Table shows estimated half-life of the two subpopulations, the relative size of the long-lived T_{SCM2} subpopulation (T_{SCM2}/T_{SCM}) and the fraction of cells from each clonal burst that enter the T_{SCM2} subpopulation (f). Additional parameters are given in S2 Table.

	half-life T_{SCM1} [years]	half-life T_{SCM2} [years]	$T_{SCM2}/T_{SCM\ [TOTAL]}$	f
DW01	0.02 (0.02-6.74)	13.92 (2.26-20.68)	0.25 (0.14-0.89)	$5.6x10^{-4}$ ($7.6x10^{-4}$-$7.2x10^{-1}$)
DW04	0.14 (0.01-2.79)	4.59 (2.13-20.41)	0.58 (0.20-0.98)	$4.1x10^{-2}$ ($2.0x10^{-2}$-$7.6x10^{-1}$)
DW10	0.69 (0.05-0.77)	9.09 (2.33-16.50)	0.51 (0.37-0.74)	$7.4x10^{-2}$ ($1.9x10^{-2}$-$1.5x10^{-1}$)
DW11	0.90 (0.03-3.98)	8.39 (3.75-17.01)	0.82 (0.46-0.95)	$3.3\ x10^{-1}$ ($2.1x10^{-2}$-$7.3x10^{-1}$)
MEDIAN	0.41 (0.02-3.39)	8.74 (2.30-18.71)	0.55 (0.28-0.92)	$5.8x10^{-2}$ ($1.9x10^{-2}$-$7.3x10^{-1}$)

Long-lived $CD8^+$ T_{SCM} cells: degree of self-renewal & clonal stability.

The long-lived T_{SCM} subpopulation identified above is a potential candidate for the stem cell population responsible for the maintenance of immune memory. We therefore investigated the degree of self-renewal and clonal stability within this long-lived T_{SCM} compartment. The degree of self-renewal of a population at steady state, 1/(death rate + differentiation rate – proliferation rate), quantifies the upstream input necessary to maintain a population. If there is a large upstream contribution then the degree of self-renewal will be low [34]. A perfectly self-renewing population (e.g. hematopoietic stem cells) will have an infinite degree of self-renewal. We quantified the degree of self-renewal for the long-lived population and found a median of 4600 days, range 2400-7300d (S2 Table). This implies that, on average, a T_{SCM} cell (or its progeny) from the long-lived subpopulation resides in the T_{SCM} compartment without dying or differentiating for 4600 days.

The total $CD8^+$ T_{SCM} population is small (2-3% of circulating $CD8^+$ lymphocytes [19], 1-5% of lymph node-resident $CD8^+$ T cells [35]). If only a proportion of this already small population is responsible for maintaining memory then this raises the issue that, although the precursor population specific for a given antigen may have a long-half life, the small size of that population could mean that its dynamics are highly stochastic. That is, there may be wide ranges in the length of memory and some antigen-specific precursor populations would be predicted to be lost by stochastic extinction soon after generation i.e. memory would be erratic and fallible. To investigate the stochasticity of the length of memory within the long lived T_{SCM} pool we performed Gillespie simulations of the size of the antigen-specific precursor population based on the parameter estimates derived from model fitting for each of the 4 individuals with $CD8^+$ T cell data. Although there was stochasticity in the half-lives of antigen-specific precursor across different runs the variation was not large (Fig 6A), and the different trajectories were tightly clustered (Fig 6B).

Fig 6. Stochastic dynamics of the long-lived antigen-specific T_{SCM} populations

Gillespie simulations of the change in size of antigen-specific precursor populations within the long-lived T_{SCM} pool were performed for each individual. **A)** Distribution of antigen-specific precursor population half-lives **B)** Five randomly chosen Gillespie simulations for the two subjects exhibiting the most inter-realisation variation (DW01 and DW10). Simulations were performed using the T_{SCM} parameters estimated by fitting the explicit heterogeneity model (Table 1 and S2 Table). All results depicted are provided in S1 Data.

Subpopulation kinetics: compatibility with long half-lives between 5 and 15 years.

Concerned that the YFV vaccine, which is known to generate an exceptional $CD8^+$ T cell response, may not be representative of a typical antigen we also sought to study T_{SCM} dynamics independent of the YFV dataset. Guided by the concept that there may be a long-lived T_{SCM} subpopulation we fitted the explicit kinetic heterogeneity model to the isotope labelling and telomere length data, ignoring the YFV data but imposing a half-life greater than 5 years on the long-lived T_{SCM} subpopulation. As expected, parameters could no longer be reliably identified but we were able to conclude that the dynamics of $CD8^+$ T_{SCM} cells were compatible with long subpopulation half-lives between 5 and 15 years (S2 Fig). This approach also allowed us to study $CD4^+$ T_{SCM} cells (which were not measured in the YFV study). Again, we found that the dynamics of $CD4^+$ T_{SCM} cells were compatible with the presence of a similarly long-lived subpopulation with a half-life of between 5 and 15 years (S3 Fig).

Discussion

One leading explanation for the maintenance of long-term immunological memory is the existence of a stem cell-like population of memory T cells, able to both self-renew and to differentiate into all other subsets of the T cell memory pool [12, 17, 21]. It has been suggested that the recently discovered T_{SCM} population is the main stem cell population responsible for maintaining T cell memory [18, 36].

In a previous study we used stable isotope labelling to investigate T_{SCM} dynamics at equilibrium in healthy subjects [26]. This revealed unexpectedly high rates of turnover in the $CD4^+$ and $CD8^+$ T_{SCM} compartments. Whilst supporting the concept that the T_{SCM} population as a whole is stable, it did not establish whether new T_{SCM} cells were generated from naïve cells or by T_{SCM} proliferation. Moreover, the key parameters of clonal longevity and self-renewal, which are prerequisites for stemness, could not be deconvoluted. Given that T_{SCM} cells were found to die and be replaced rapidly, the source of the replacing cell becomes critical. Even a small contribution from naïve cells can result in a weakly self-renewing

population and loss of memory as existing clones are replaced by cells specific for a different antigen.

In the present study, we overcame these limitations by simultaneously fitting telomere length and tetramer data to constrain the space of possible models and improve parameter identifiability. We report that firstly, the T_{SCM} population is kinetically heterogeneous, with at least two kinetically distinct subpopulations turning over at different rates; and secondly, that the dynamics of a subpopulation of T_{SCM} cells are compatible with their hypothesized role as the main stem cell-like T cell precursor responsible for the maintenance of T cell memory. The best description of the data is one in which the kinetically heterogeneous T_{SCM} population, despite its high average replenishment rate of approximately 2% per day, contained a fraction of long-lived T_{SCM} cells. The half-life of this slower subpopulation was approximately 9 years, consistent with the 8-15 years half-life estimated for the recall response to a given antigen [3, 4]. Furthermore, although this subpopulation was small, the dynamic behaviour of individual clones was not excessively stochastic and the half-life of a given antigen-specific precursor population was tightly distributed. Finally, we estimated that the degree of self-renewal of the long-lived T_{SCM} subpopulation was approximately 4600 days.

The quantification of the dynamics of T_{SCM} cells is not easily addressed in humans and different studies invariably involve different compromises. An advantage of our study is that it examines the natural dynamics of T_{SCM} cells in healthy, lymphocyte-replete individuals. A disadvantage is that many of the model parameters were poorly identifiable; nevertheless firm conclusions about clonal longevity and self-renewal can be drawn. A second disadvantage is that it utilises vaccination data generated using a vaccine (the YFV-17D vaccine) known for its ability to generate an exceptional $CD8^+$ T cell response and which may not be representative of average immunity. To address this potential caveat we repeated our analysis without reference to the YFV-vaccination data. Whilst this reduced our ability to estimate individual parameters it did confirm the conclusion that the T_{SCM} population consists of subpopulations with different kinetics and that high turnover and short half-lives of the bulk T_{SCM} population does not rule out the existence of a slowly turning over subpopulation with the dynamic properties required for T_{SCM} cells to maintain both $CD4^+$ and $CD8^+$ T cell memory. Finally, it should be noted that our study was confined to circulating T_{SCM} cells. In mice, it has been shown that memory T cell dynamics can vary across different anatomical compartments (spleen versus bone marrow) [37]. For human studies, ethical and technical considerations mean that repeated sampling of cells must be limited to the peripheral blood compartment. However, this does enable direct comparison with previously reported data, including the seminal description of human T_{SCM} cells [16] which was based on circulating cells.

Our findings of kinetic heterogeneity in the human T_{SCM} population, are reminiscent of the proliferative heterogeneity described for transplanted hematopoietic stem cells (HSCs) in lethally irradiated mice [38] in which levels of Kit receptor distinguished cell subpopulations

with different expansion capacities. Similarly, studies of human $CD4^+$ and $CD8^+$ T_{SCM} cells cultured in the presence of cytokines (either IL-15 or IL-7+IL-15) *in vitro* have also reported that a fraction of cells proliferated rapidly while the majority remained quiescent [39] [19]. Finally, in rhesus macaques, circulating T_{SCM} cells had a level of Ki-67 expression that the author remarked was "unexpectedly large" (mean>10%) [36]; this is consistent with our finding that, on average, the T_{SCM} population in peripheral blood turns over rapidly. Our proliferation rate estimates for naïve and T_{SCM} cells can be compared with proliferation rate estimates of other T cell subsets obtained using stable isotope labelling. We found that the proliferation of naïve T cells is slowest, $p=0.0005d^{-1}$, and comparable with previous estimates [40], though in this latter study the gating strategy would have inadvertently included T_{SCM} cells in the naive cell gate. Next is the slow T_{SCM} subpopulation with a proliferation rate an order of magnitude faster than naïve cells ($p=0.002d^{-1}$). Then the fast T_{SCM} subpopulation with a median proliferation rate of $p=0.015d^{-1}$ comparable with that of memory cells ($0.006d^{-1}$ – $0.02d^{-1}$ [29, 41]). Giving the following rank order of proliferation rates: naïve < slow T_{SCM} < fast $T_{SCM} \lesssim$ memory. Our degree of self-renewal estimates are more difficult to place in context. To the best of our knowledge, this parameter has only been quantified previously for murine HSCs. Our estimate of 4,600 days for the degree of self-renewal of the long-lived T_{SCM} subpopulation is naturally less than the corresponding estimate for murine HSC which is, by definition, infinite but is greater than the degree of self-renewal of compartments immediately adjacent to HSC in the differentiation pathway, namely short-term-HSC (degree of self-renewal of 90-150 days in mice) and multipotent progenitors (MPP) (degree of self-renewal of 7-28 days in mice) [23, 34]. If we convert to "animal-lifespans" (80yr for human, 2yr for mice; a scaling which appears valid for T cell kinetics [41]) then we see that the degree of self-renewal of T_{SCM} cells is 0.15 lives, i.e. a long lived T_{SCM} cell resides without dying or differentiating for approximately 15% of our lifespan. This is similar to the degree of self-renewal of short term HSCs (0.12-0.2 lives) and greater than the degree of self-renewal of MPPs (0.01-0.04 lives). It is remarkable that a peripheral cell population that is towards the end of the haematopoietic differentiation pathway should have a degree of self-renewal that is comparable with short-term HSCs.

Interestingly, we found strong evidence for continual differentiation of naïve T cells into the T_{SCM} cell pool despite the study volunteers being healthy with no symptomatic infection. For both the implicit and the explicit heterogeneity models, the contribution of naïve cells to T_{SCM} replacement was typically about 50% and never less than 10% (Fig 4B and 4F). This may represent differentiation of naïve cells in response to continual low level exposure to novel environmental antigen and/or to persistent antigen. Considerable recruitment of naïve cells to the memory pool in the apparent absence of novel antigen has been previously described for mice persistently infected with Polyoma virus [42] or lymphocytic choriomeningitis virus [42] and for healthy mice [43].

The role of the short-lived T_{SCM} subpopulation which we identify is unclear. Potentially it is activated naïve cells rapidly transitioning to effectors whilst others are retained to form the

long-lived T_{SCM} pool that is the basis of memory. This is consistent with recent evidence that T_{SCM} cells may pass through a phase in which they express effector molecules [44].

This work suggests a number of future directions. One important direction is to establish phenotypic markers to distinguish the "true" T_{SCM} subpopulation. A second is to develop a model to predict the T cell receptor repertoire of the true T_{SCM} subpopulation and whether this differs from bulk T_{SCM} cells and the functional consequences of any such difference. Finally, it is important to know whether or not murine T_{SCM} populations are similarly heterogeneous since this would facilitate a whole range of experiments not possible in humans.

Our results show that substantial kinetic heterogeneity exists within the T_{SCM} pool, encompassing a long-lived subpopulation with the dynamic properties required to maintain both $CD4^+$ and $CD8^+$ T cell memory. Further characterization of these *bona fide* T_{SCM} cells may illuminate the mechanistic basis of durable immune protection and facilitate translational efforts to develop more effective vaccines and immunotherapies.

Methods

EXPERIMENTAL DATA

Ethics statement.

Approval was granted by the Cardiff University School of Medicine and London-Chelsea Research Ethics Committees (REC: 13/LO/0022. IRAS: 109455). All studies were conducted according to the Principles of the Declaration of Helsinki and all subjects gave written consent.

Study participants.

Five healthy adults were studied (DW01, age 32; DW02, age 64; DW04, age 83; DW10, age 34; DW11, age 29). All subjects were CMV-seropositive and HIV-1-seronegative.

Stable isotope labelling *in vivo*.

We have previously described the labelling protocol in detail [26]. Briefly, participants were given oral doses of 70% deuterated water (^2H_2O) over a 7-week period (50 ml three times daily for one week, then twice daily thereafter). Saliva samples were collected for evaluation of body water labelling. Venous blood was drawn at successive time-points during and after labelling. Peripheral blood mononuclear cells were sorted at high purity using a custom-modified BD FACSAria II flow cytometer into $CD4^+$ and $CD8^+$ naïve and T_{SCM} cells on the basis of cell surface expression (naïve: $CD45RO^-CD27^{bright}CCR7^+CD95^-$ T_{SCM}: $CD45RO^-CD27^{bright}CCR7^+CD95^+$). Both subsets were further assessed for expression of other cell surface markers; T_{SCM} cells were found to be $CD45RA^+$, $CD28^+$, $CD127^+$ and $CD57^-$ (Fig 1). Deuterium enrichment in the DNA of the sorted T cell subsets was measured by gas chromatography/mass spectrometry [45].

Single chromosome telomere length analysis.

DNA from $CD4^+$ and $CD8^+$ naïve and T_{SCM} cells (sorted as above) was extracted and single telomere length analysis was carried out at the XpYp telomere as described previously [46].

Yellow Fever Vaccine data.

Published data were acquired from a cross-sectional study of 37 healthy adults who received a single dose of the yellow fever vaccine YF-17D [21]. Time since vaccination ranged from 3 months to 35 years. Four subjects who received multiple YFV vaccinations were excluded from the analysis. $CD8^+CD45RA^+CCR7^{INT}$ T_{SCM} cells specific for the HLA-A*02-restricted YFV $NS4b^{214-222}$ epitope were quantified by tetramer staining and flow cytometry.

MATHEMATICAL MODELLING

Homogeneous and implicit heterogeneous models.

To study the dynamics of T_{SCM} clones, we developed an ordinary differential equation (ODE) model of the linear differentiation pathway between naïve and T_{SCM} cells (Fig 2). Non-linear models of differentiation were not considered [47]. In the classical model of T_{SCM} formation, de-differentiation of T_{CM}, T_{EFF} and T_{EM} cells to T_{SCM} is infrequent [48-50] so we assumed a one-way differentiation pathway. Immunological memory has been shown to be generated by clonal expansion, in which naïve cells encountering antigen in lymphoid tissue, undergo several rounds of division [1, 51-53]. We assume that newly generated T_{SCM} cells can arise in two ways: either they are the product of self-renewal (proliferation of T_{SCM} cells, p_s), or they result from the differentiation of naïve cells following antigen exposure at a rate Δ. We assume that naïve and T_{SCM} cells are in constant recirculation between lymph and blood [15, 19] giving the following equations:

$$\dot{T}_N = (p_n - d_n - \Delta)T_N \quad (1)$$

$$\dot{T}_{SCM} = \Delta 2^k T_N + (p_s - d_s)T_{SCM} \quad (2)$$

where T_N and T_{SCM} are the total number of cells in the naïve and T_{SCM} populations respectively; p_n, d_n, p_s and d_s the proliferation and disappearance rates of naïve and T_{SCM} cells respectively; Δ is the fraction of naïve cells activated by antigen exposure per day and k is the number of divisions occurring during clonal expansion in the differentiation from naïve to T_{SCM} cells (resulting in 2^k T_{SCM} cells generated from each naïve cell). A value of k equal to one indicates that naïve cells divide only once after priming, and a value of k equal to zero indicates that no divisions occur after antigen exposure. Deuterium labelling experiments measure the fraction of deoxyadenosine nucleosides (dAs) with incorporated deuterium, so we construct the model in terms of dAs. The absolute numbers of labelled dAs derived from equations (1) and (2) are:

$$\dot{T}_N^* = p_n c U(t) T_N - (d_n^* + \Delta)T_N^* \quad (3)$$

$$\dot{T}_{SCM}^* = \Delta(2^k - 1)cU(t)T_N + \Delta T_N^* + p_s c U(t) T_{SCM} - d_s^* T_{SCM}^* \quad (4)$$

where T_N^* and T_{SCM}^* represent the absolute numbers of labelled dAs from naïve and T_{SCM} cells, respectively; d_n^* and d_s^* are the disappearance rates of labelled naïve and T_{SCM} cells respectively, c is the amplification factor for enrichment and $U(t)$ is an empirical function describing label availability as measured in saliva [40]:

$$U(t) = f_r\left(1 - e^{-\delta t}\right) + \beta e^{-\delta t} \quad (5)$$

$$U(t) = [f_r\left(1 - e^{-\delta \tau}\right) + \beta e^{-\delta \tau}]e^{-\delta(\tau - t)} \quad (6)$$

where τ represents the time at which the administration of label is stopped, f_r represents the fraction of deuterium in water, δ the turnover rate per day of body water, and β the baseline saliva enrichment. Parameters f_r, β and δ are known to vary between individuals, and their values were obtained by fitting equations 5 and 6 to successive measurements of label enrichment in saliva from each subject. The measured body water enrichments with best fits are shown in S4 Fig; estimates of f_r, β and δ are provided in S4 Table. The equations for the fraction of labelled dAs are then:

$$\dot{F}_{TN} = p_n cU(t) - (d_n^* + \Delta)F_{TN} \quad (7)$$

$$\dot{F}_{TSCM} = (2^k - 1)cU(t)\frac{\Delta T_N}{T_{SCM}} + \frac{\Delta T_N}{T_{SCM}}F_{TN} + p_s cU(t) - d_s^* F_{TSCM} \quad (8)$$

While the naïve pool has been observed to be kinetically homogeneous [40], the heterogeneity of the T_{SCM} population has not yet been explored. We have previously argued [28] that if a cell population is kinetically heterogeneous, the rate of label uptake during a labelling experiment will not be equal to the rate at which the label is lost. This scenario arises because the labelled population (and therefore the disappearance rate estimated from it) will not be representative of the whole population, as subpopulations with faster kinetics divide faster and will be overrepresented in the labelled cells. We can therefore impose kinetic homogeneity on the T_{SCM} pool ("**homogenous model**") by constraining $d_s^* = d_s$ where $2^k\Delta T_N + p_s T_{SCM} = d_s T_{SCM}$. Removal of this constraint is equivalent to allowing the T_{SCM} pool to be heterogeneous ("**implicit heterogeneity model**"). Assuming kinetic homogeneity in the naïve pool, the equations for the fractions of labelled DNA at steady state are therefore:

$$\dot{F}_{TN} = p_n(cU(t) - F_{TN}) \quad (9)$$

$$\dot{F}_{TSCM} = \frac{\Delta T_N}{T_{SCM}}((2^k - 1)cU(t) + F_{TN}) + p_s cU(t) - d_s^* F_{TSCM} \quad (10)$$

where $d_s^* = d_s$ yields the homogeneous version of the model and, if d_s^* is free, then this yields the implicit heterogeneity version of the model. We also tested a version of the model in which the assumption of homogeneity for naïve cells ($d_n^* = d_n$) was relaxed.

The explicit heterogeneous model.

The implicit heterogeneous model (above) describes the average population dynamics of a heterogeneous T_{SCM} pool. However, to estimate the sizes, the proliferation and the disappearance rates of each of the T_{SCM} subpopulations, these subpopulations need to be modelled explicitly. The explicit heterogeneous model describes two kinetically distinct subpopulations of T_{SCM} cells. Additional subpopulations could not be resolved with the available data. If there are more than 2 subpopulations with distinct kinetics then subpopulation 1 and subpopulation 2 which we measure will have kinetics that represent of the average of their smaller constituent subpopulations. The equations for the total numbers of cells in each of the modelled pools are:

$$\dot{T}_N = (p_n - d_n - \Delta)T_N \quad (11)$$

$$\dot{T}_{SCM1} = \Delta(1 - f)2^k T_N + (p_{s1} - d_{s1})T_{SCM1} \quad (12)$$

$$\dot{T}_{SCM2} = \Delta f 2^k T_N + (p_{s2} - d_{s2})T_{SCM2} \quad (13)$$

where f is the proportion of cells from the clonal burst (of size 2^k) that differentiate into the second T_{SCM} subpopulation; p_{s1}, d_{s1}, p_{s2}, and d_{s2} are the proliferation and disappearance rates of T_{SCM1} and T_{SCM2} cells respectively; and the remaining parameters are as described above. The absolute number of labelled dAs derived from equations 11, 12 and 13 are:

$$\dot{T}_N^* = p_n cU(t)T_N - (d_n + \Delta)T_N^* \quad (14)$$

$$\dot{T}_{SCM1}^* = \Delta(1-f)(2^k - 1)cU(t)T_N + \Delta(1-f)T_N^* + p_{s1}cU(t)T_{SCM1} - d_{s1}T_{SCM1}^* \quad (15)$$

$$\dot{T}_{SCM2}^* = \Delta f(2^k - 1)cU(t)T_N + \Delta f T_N^* + p_{s2}cU(t)T_{SCM2} - d_{s2}T_{SCM2}^* \quad (16)$$

and the fractions of labelled DNA assuming steady state for cell population sizes are:

$$\dot{F}_{TN} = p_n(cU(t) - F_{TN}) \quad (17)$$

$$\dot{F}_{TSCM1} = \Delta(1-f)\frac{T_N}{T_{SCM1}}((2^k - 1)cU(t) + F_{TN} - 2^k F_{TSCM1}) + p_{s1}(cU(t) - F_{TSCM1}) \quad (18)$$

$$\dot{F}_{TSCM2} = \Delta f \frac{T_N}{T_{SCM2}}((2^k - 1)cU(t) + F_{TN} - 2^k F_{TSCM2}) + p_{s2}(cU(t) - F_{TSCM2}) \quad (19)$$

Telomere length model (for homogeneous & implicit heterogeneity models).

Following de Boer et al. [54], we derived an ODE model to describe the progressive shortening with division of the average telomere lengths in the naïve and T_{SCM} cell populations (Fig 2B) for the homogenous and implicit heterogeneity models. The equivalent derivation for the explicit heterogeneity model follows the same pattern and is presented in the SI.

T_N and T_{SCM} cell populations were modelled as a series of *n* compartments, with T_{Ni} (or T_{SCMi}) representing the number of T_N (or T_{SCM}) cells that have divided *i* times. If δ is the number of base pairs lost per cell division, then the cells in T_{Ni} (or T_{SCMi}) have decreased their telomere lengths by δi base pairs. From equations (1) and (2) above, the number of cells in the T_{Ni} and T_{SCMi} compartments are:

$$\dot{T}_{N_i} = 2p_n T_{N_{i-1}} - (p_n + \Delta + d_n)T_{N_i} \quad (20)$$

$$\dot{T}_{SCM_i} = 2p_s T_{SCM_{i-1}} - (p_s + d_s)T_{SCM_i} + \Delta 2^k T_{N_{i-C}} \quad (21)$$

where p_n, Δ, d_n, p_s, d_s, and k are as in equations (1) and (2), and *C* is a parameter to allow for the impact of telomerase. If *C* = 0 then no shortening of telomeres occurs during clonal expansion (total compensation by telomerase); if *C*=*k* then there is no compensation by telomerase. Experimental evidences suggests that, both for HSCs and peripheral T cells, telomerase attenuates but does not prevent telomere loss [55] i.e. 0<C<k. The equations for the average number of divisions undergone by the cells in the naïve and the T_{SCM} pools in our model (equations 20 & 21), follows trivially from the derivation in [54] and are given by:

$$\dot{\mu}_{TN} = 2p_n \quad (22)$$

$$\dot{\mu}_{TSCM} = 2p_s - \Delta 2^k \frac{T_N}{T_{SCM}}(\mu_{T_{SCM}} - \mu_{T_N} - C) \quad (23)$$

Average telomere lengths are obtained by multiplying μ_{TN} and μ_{TSCM} by δ. We define $\Theta := \delta(\mu_{T_{SCM}} - \mu_{T_N})$, the difference between the average telomere lengths in the T_N and the T_{SCM} pool (in units of base pairs). The dynamics of Θ are then given by:

$$\dot{\Theta} = 2(p_s - p_n)\delta - \Delta 2^k \frac{T_N}{T_{SCM}}(\Theta - C\delta). \quad (24)$$

Finally, when equation 24 reaches steady state (shown numerically to have occurred for all subjects) then:

$$\Theta = C\delta + \frac{(p_s - p_n)\delta T_{SCM}}{2^{(k-1)}\Delta T_N}. \quad (25)$$

We use an estimate of δ=50 bp/division, within the reported range of 35-70 bp/division [54]. Including the telomere data helps to constrain the model fit by placing a bound on the maximum number of divisions occurring between an average naïve T cell and an average T_{SCM} cell.

Computing clonal lifespans and half-lives.

We computed the half-life of a clone deterministically from the T_{SCM} parameters estimated by fitting the ODE models. We are interested in the half-life of the memory (rather than the classically reported population half-life) and so this is defined as

$$\tau_{1/2} = \frac{\ln(2)}{d_s - p_s}. \quad (26)$$

Antigen-specific precursor lifespans (time until the last cell of an antigen-specific memory precursor population disappears) were computed stochastically using the exact Gillespie algorithm [56, 57]. At each step of the algorithm, either a division or a disappearance event is chosen, with respective probabilities of $x(t)p/S(t)$ and $x(t)d/S(t)$, where *p* and *d* are the proliferation and disappearance rates of the T_{SCM} population; *x(t)* is population size at time *t*; and *S(t)* is the sum of *x(t)p* and *x(t)d*. At the end of each step, time *t* is incremented by a number of days sampled from an exponential distribution with rate *S(t)*. As the naïve T cell population in an adult human has an approximate size of 10^{11} cells [58], we estimate the initial size of a T_{SCM} clone in the long-lived subpopulation as:

$$size_0 = 10^{11} \times 2^k \times \Delta \times f \quad (27)$$

where *k* is the number of divisions that occur during clonal expansion in the differentiation from T_N to T_{SCM}, Δ is the fraction of naïve cells activated by the same antigen and *f* is the fraction of a newly generated clone that enters the long-lived pool. Antigen-specific T cell precursor frequency has been estimated in the naïve cell pool at < 1 to 352 per 10^5 naïve $CD8^+$ cells [59]. For estimates of precursor lifespan in the implicit heterogeneity model we fixed Δ to a representative value of $1x10^{-5}$ (and as we are considering average lifespan, *f*=1). For estimates of precursor lifespan of the long-lived T_{SCM} subpopulation obtained using the explicit heterogeneity model the values of Δ and *f* estimated by model fitting were used. Calculations of the initial size of a long-lived T_{SCM} clone are provided in S5 Table.

Model of the frequency of YFV-specific $CD8^+$ T_{SCM} cells.

The proportion of tetramer$^+$ T_{SCM} cells expressed as a fraction of $CD8^+$ $CD16^-$ lymphocytes as a function of time since vaccination was modelled allowing for two kinetically distinct subpopulations with exponential decay kinetics:

$$F = re^{-\alpha t} + (1 - r)e^{-\beta t} \quad (28)$$

Degree of self-renewal.

The degree of self-renewal of the long-lived TSCM subpopulation (T_{SCM2}) is defined as

$$\begin{aligned}\text{degree of self-renewal} &= \frac{1}{d_{s2} - p_{s2}} \\ &= \frac{T_{SCM2}}{\Delta 2^k \, f T_N}\end{aligned} \tag{29}$$

Fitting procedures

Models were fitted to the experimental data by minimizing the sum of squared residuals using the pseudoOptim algorithm from the FME package in R [60, 61]. Details of the fitting procedure, the different models, data fitted and rationale are summarised in the SI Methods.

Script availability

The scripts used for fitting the homogenous model, the implicit heterogeneous model and the explicit heterogeneous model as well as the script for running the Gillespie simulation are all freely available at Zenodo DOI 10.5281/zenodo.1253178 [27].

STATISTICS

The fit of the homogenous and heterogeneous model was compared using Fisher's F-test for nested models. This test compares the goodness of fit of nested models to data, taking into account the different number of parameters in the models [62].

Acknowledgements

This work was facilitated by the Imperial College High Performance Computing Service.

References

1. Farber DL, Yudanin NA, Restifo NP. Human memory T cells: generation, compartmentalization and homeostasis. Nat Rev Immunol. 2014;14(1):24-35.

2. Ahmed R, Gray D. Immunological memory and protective immunity: understanding their relation. Science. 1996;272(5258):54-60.

3. Crotty S, Felgner P, Davies H, Glidewell J, Villarreal L, Ahmed R. Cutting edge: long-term B cell memory in humans after smallpox vaccination. J Immunol. 2003;171(10):4969-73.

4. Hammarlund E, Lewis MW, Hansen SG, Strelow LI, Nelson JA, Sexton GJ, et al. Duration of antiviral immunity after smallpox vaccination. Nat Med. 2003;9(9):1131-7.

5. Fearon DT, Manders P, Wagner SD. Arrested differentiation, the self-renewing memory lymphocyte, and vaccination. Science. 2001;293(5528):248-50.

6. Macallan DC, Borghans JA, Asquith B. Human T Cell Memory: A Dynamic View. Vaccines. 2017;5(1):E5-57.

7. Mackay LK, Stock AT, Ma JZ, Jones CM, Kent SJ, Mueller SN, et al. Long-lived epithelial immunity by tissue-resident memory T (TRM) cells in the absence of persisting local antigen presentation. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(18):7037-42.

8. Jiang X, Clark RA, Liu L, Wagers AJ, Fuhlbrigge RC, Kupper TS. Skin infection generates non-migratory memory CD8+ T(RM) cells providing global skin immunity. Nature. 2012;483(7388):227-31.

9. Di Rosa F. Maintenance of memory T cells in the bone marrow: survival or homeostatic proliferation? Nat Rev Immunol. 2016;16(4):271.

10. Sallusto F, Geginat J, Lanzavecchia A. Central memory and effector memory T cell subsets: function, generation, and maintenance. Annu Rev Immunol. 2004;22:745-63.

11. Stemberger C, Neuenhahn M, Gebhardt FE, Schiemann M, Buchholz VR, Busch DH. Stem cell-like plasticity of naive and distinct memory CD8+ T cell subsets. Semin Immunol. 2009;21(2):62-8.

12. Graef P, Buchholz VR, Stemberger C, Flossdorf M, Henkel L, Schiemann M, et al. Serial transfer of single-cell-derived immunocompetence reveals stemness of CD8(+) central memory T cells. Immunity. 2014;41(1):116-26.

13. Gerlach C, Rohr JC, Perie L, van Rooij N, van Heijst JW, Velds A, et al. Heterogeneous differentiation patterns of individual CD8+ T cells. Science (New York, NY. 2013;340(6132):635-9.

14. Zhang Y, Joe G, Hexner E, Zhu J, Emerson SG. Host-reactive CD8+ memory stem cells in graft-versus-host disease. Nat Med. 2005;11(12):1299-305.

15. Lugli E, Gattinoni L, Roberto A, Mavilio D, Price DA, Restifo NP, et al. Identification, isolation and in vitro expansion of human and nonhuman primate T stem cell memory cells. Nature protocols. 2013;8(1):33-42.

16. Gattinoni L, Lugli E, Ji Y, Pos Z, Paulos CM, Quigley MF, et al. A human memory T cell subset with stem cell-like properties. Nature medicine. 2011;17(10):1290-7.

17. Biasco L, Scala S, Basso Ricci L, Dionisio F, Baricordi C, Calabria A, et al. In vivo tracking of T cells in humans unveils decade-long survival and activity of genetically modified T memory stem cells. Sci Transl Med. 2015;7(273):273ra13.

18. Gattinoni L, Speiser DE, Lichterfeld M, Bonini C. T memory stem cells in health and disease. Nat Med. 2017;23(1):18-27.

19. Gattinoni L, Lugli E, Ji Y, Pos Z, Paulos CM, Quigley MF, et al. A human memory T cell subset with stem cell-like properties. Nat Med. 2011;17(10):1290-7.

20. Fisman DN, Savage R, Gubbay J, Achonu C, Akwar H, Farrell DJ, et al. Older age and a reduced likelihood of 2009 H1N1 virus infection. N Engl J Med. 2009;361(20):2000-1.

21. Fuertes Marraco SA, Soneson C, Cagnon L, Gannon PO, Allard M, Abed Maillard S, et al. Long-lasting stem cell-like memory CD8+ T cells with a naive-like profile upon yellow fever vaccination. Sci Transl Med. 2015;7(282):282ra48.

22. Oliveira G, Ruggiero E, Stanghellini MT, Cieri N, D'Agostino M, Fronza R, et al. Tracking genetically engineered lymphocytes long-term reveals the dynamics of T cell immunological memory. Sci Transl Med. 2015;7(317):317ra198.

23. Busch K, Klapproth K, Barile M, Flossdorf M, Holland-Letz T, Schlenner SM, et al. Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature. 2015;518(7540):542-6.

24. Sun J, Ramos A, Chapman B, Johnnidis JB, Le L, Ho Y-J, et al. Clonal dynamics of native haematopoiesis. Nature. 2014;514(7522):322-7.

25. Sawen P, Lang S, Mandal P, Rossi DJ, Soneji S, Bryder D. Mitotic History Reveals Distinct Stem Cell Populations and Their Contributions to Hematopoiesis. Cell reports. 2016;14(12):2809-18.

26. Ahmed R, Roger L, Costa Del Amo P, Miners KL, Jones RE, Boelen L, et al. Human Stem Cell-like Memory T Cells Are Maintained in a State of Dynamic Flux. Cell reports. 2016;17(11):2811-8.

27. Costa Del Amo P. Scripts from "Human TSCM cell dynamics in vivo are compatible with long-lived immunological memory and stemness". Openly available from Zenodo DOI 105281/zenodo1253178. 2018.

28. Asquith B, Debacq C, Macallan DC, Willems L, Bangham CR. Lymphocyte kinetics: the interpretation of labelling data. Trends in immunology. 2002;23(12):596-601.

29. Macallan DC, Asquith B, Irvine AJ, Wallace DL, Worth A, Ghattas H, et al. Measurement and modeling of human T cell kinetics. European journal of immunology. 2003;33(8):2316-26.

30. Ganusov VV, Borghans JA, De Boer RJ. Explicit kinetic heterogeneity: mathematical models for interpretation of deuterium labeling of heterogeneous cell populations. PLoS Comput Biol. 2010;6(2):e1000666.

31. De Boer RJ, Perelson AS, Ribeiro RM. Modelling deuterium labelling of lymphocytes with temporal and/or kinetic heterogeneity. Journal of the Royal Society, Interface / the Royal Society. 2012;9(74):2191-200.

32. Buchholz VR, Flossdorf M, Hensel I, Kretschmer L, Weissbrich B, Graf P, et al. Disparate individual fates compose robust CD8+ T cell immunity. Science (New York, NY. 2013;340(6132):630-5.

33. Kaech SM, Ahmed R. Memory CD8+ T cell differentiation: initial antigen encounter triggers a developmental program in naive cells. Nature immunology. 2001;2(5):415-22.

34. Hofer T, Busch K, Klapproth K, Rodewald HR. Fate Mapping and Quantitation of Hematopoiesis In Vivo. Annual review of immunology. 2016;34:449-78.

35. Hong H, Gu Y, Sheng SY, Lu CG, Zou JY, Wu CY. The Distribution of Human Stem Cell-like Memory T Cell in Lung Cancer. Journal of immunotherapy (Hagerstown, Md : 1997). 2016;39(6):233-40.

36. Lugli E, Dominguez MH, Gattinoni L, Chattopadhyay PK, Bolton DL, Song K, et al. Superior T memory stem cell persistence supports long-lived T cell memory. The Journal of clinical investigation. 2013;123(2):594-9.

37. Siracusa F, Alp OS, Maschmeyer P, McGrath M, Mashreghi MF, Hojyo S, et al. Maintenance of CD8(+) memory T lymphocytes in the spleen but not in the bone marrow is dependent on proliferation. European journal of immunology. 2017;47(11):1900-5.

38. Grinenko T, Arndt K, Portz M, Mende N, Gunther M, Cosgun KN, et al. Clonal expansion capacity defines two consecutive developmental stages of long-term hematopoietic stem cells. The Journal of experimental medicine. 2014;211(2):209-15.

39. Abdelsamed HA, Moustaki A, Fan Y, Dogra P, Ghoneim HE, Zebley CC, et al. Human memory CD8 T cell effector potential is epigenetically preserved during in vivo homeostasis. The Journal of experimental medicine. 2017;214(6):1593-606.

40. Vrisekoop N, den Braber I, de Boer AB, Ruiter AF, Ackermans MT, van der Crabben SN, et al. Sparse production but preferential incorporation of recently produced naive T cells in the human peripheral pool. Proc Natl Acad Sci U S A. 2008;105(16):6115-20.

41. Westera L, Drylewicz J, den Braber I, Mugwagwa T, van der Maas I, Kwast L, et al. Closing the gap between T-cell life span estimates from stable isotope-labeling studies in mice and humans. Blood. 2013;122(13):2205-12.

42. Vezys V, Masopust D, Kemball CC, Barber DL, O'Mara LA, Larsen CP, et al. Continuous recruitment of naive T cells contributes to heterogeneity of antiviral CD8 T cells during persistent infection. The Journal of experimental medicine. 2006;203(10):2263-9.

43. Gossel G, Hogan T, Cownden D, Seddon B, Yates AJ. Memory CD4 T cell subsets are kinetically heterogeneous and replenished from naive T cells at high levels. eLife. 2017;6.

44. Akondy RS, Fitch M, Edupuganti S, Yang S, Kissick HT, Li KW, et al. Origin and differentiation of human memory CD8 T cells after vaccination. Nature. 2017.

45. Busch R, Neese RA, Awada M, Hayes GM, Hellerstein MK. Measurement of cell proliferation by heavy water labeling. Nature protocols. 2007;2(12):3045-57.

46. Capper R, Britt-Compton B, Tankimanova M, Rowson J, Letsolo B, Man S, et al. The nature of telomere fusion and a definition of the critical telomere length in human cells. Genes Dev. 2007;21(19):2495-508.

47. Wherry EJ, Teichgraber V, Becker TC, Masopust D, Kaech SM, Antia R, et al. Lineage relationship and protective immunity of memory CD8 T cell subsets. Nature immunology. 2003;4(3):225-34.

48. Appay V, van Lier RA, Sallusto F, Roederer M. Phenotype and function of human T lymphocyte subsets: consensus and issues. Cytometry Part A : the journal of the International Society for Analytical Cytology. 2008;73(11):975-83.

49. Mahnke YD, Brodie TM, Sallusto F, Roederer M, Lugli E. The who's who of T-cell differentiation: human memory T-cell subsets. Eur J Immunol. 2013;43(11):2797-809.

50. Cieri N, Oliveira G, Greco R, Forcato M, Taccioli C, Cianciotti B, et al. Generation of human memory stem T cells after haploidentical T-replete hematopoietic stem cell transplantation. Blood. 2015;125(18):2865-74.

51. B. Alberts AJ, J. Lewis, M. Raff, K. Roberts, P. Walter. Molecular Biology of the Cell. 4 ed: Garland Science; 2002.

52. Murphy K. Janeway's Immunobiology. 8 ed: Garland Science; 2012.

53. von Andrian UH, Mempel TR. Homing and cellular traffic in lymph nodes. Nat Rev Immunol. 2003;3(11):867-78.

54. De Boer RJ, Noest AJ. T cell renewal rates, telomerase, and telomere length shortening. J Immunol. 1998;160(12):5832-7.

55. Rufer N, Brummendorf TH, Kolvraa S, Bischoff C, Christensen K, Wadsworth L, et al. Telomere fluorescence measurements in granulocytes and T lymphocyte subsets point to a high turnover of hematopoietic stem cells and memory T cells in early childhood. The Journal of experimental medicine. 1999;190(2):157-67.

56. Renshaw E. Stochastic population processes : analysis, approximations, simulations. Oxford: Oxford University Press; 2011. xii, 652 p. p.

57. Wilkinson DJ. Stochastic modelling for systems biology. Boca Raton, FL ; London: Chapman & Hall/CRC; 2006. 254 p p.

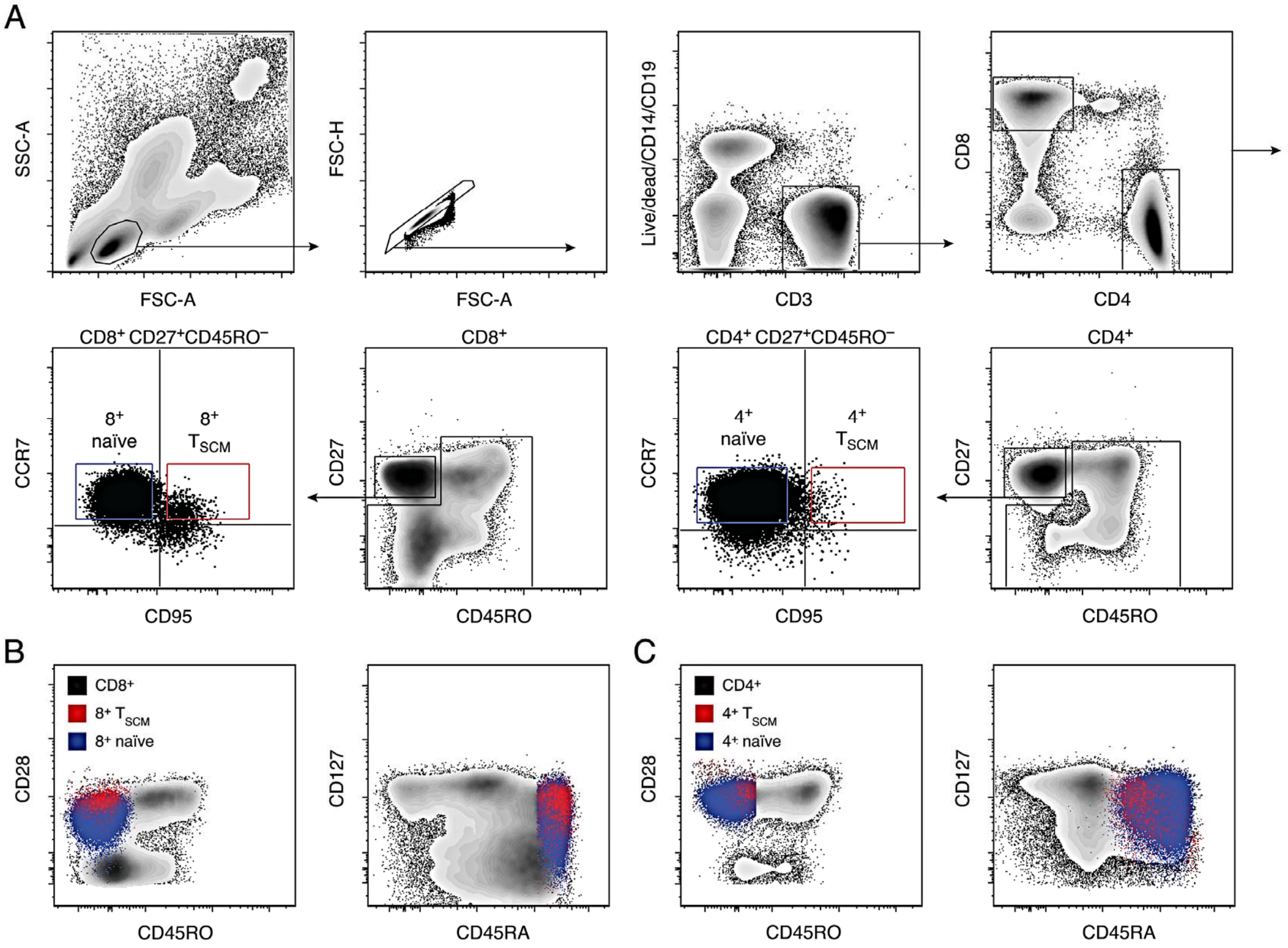
58. Bains I, Antia R, Callard R, Yates AJ. Quantifying the development of the peripheral naive CD4(+) T-cell pool in humans. Blood. 2009;113(22):5480-7.

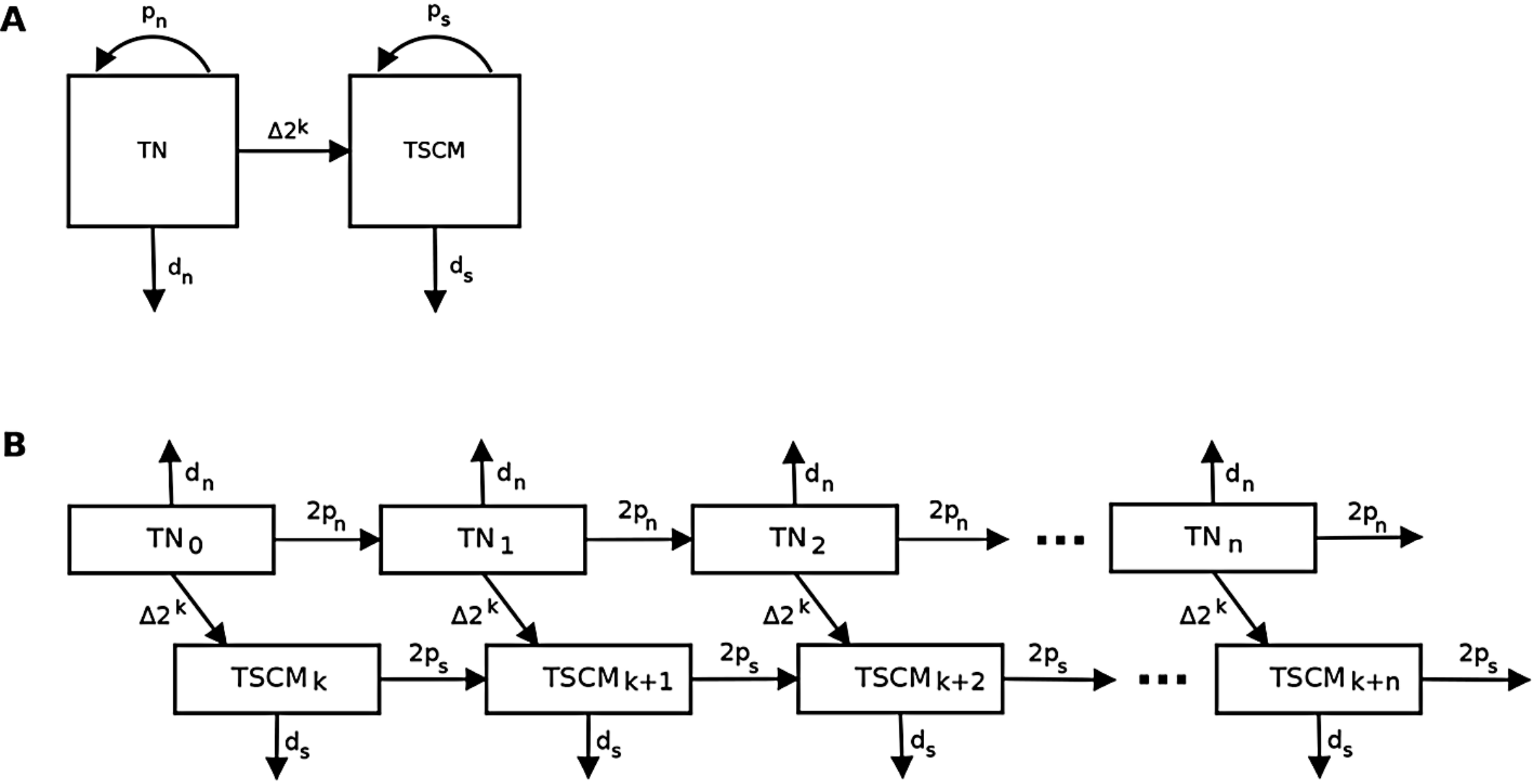
59. Neller MA, Ladell K, McLaren JE, Matthews KK, Gostick E, Pentier JM, et al. Naive CD8(+) T-cell precursors display structured TCR repertoires and composite antigen-driven selection dynamics. Immunol Cell Biol. 2015;93(7):625-33.

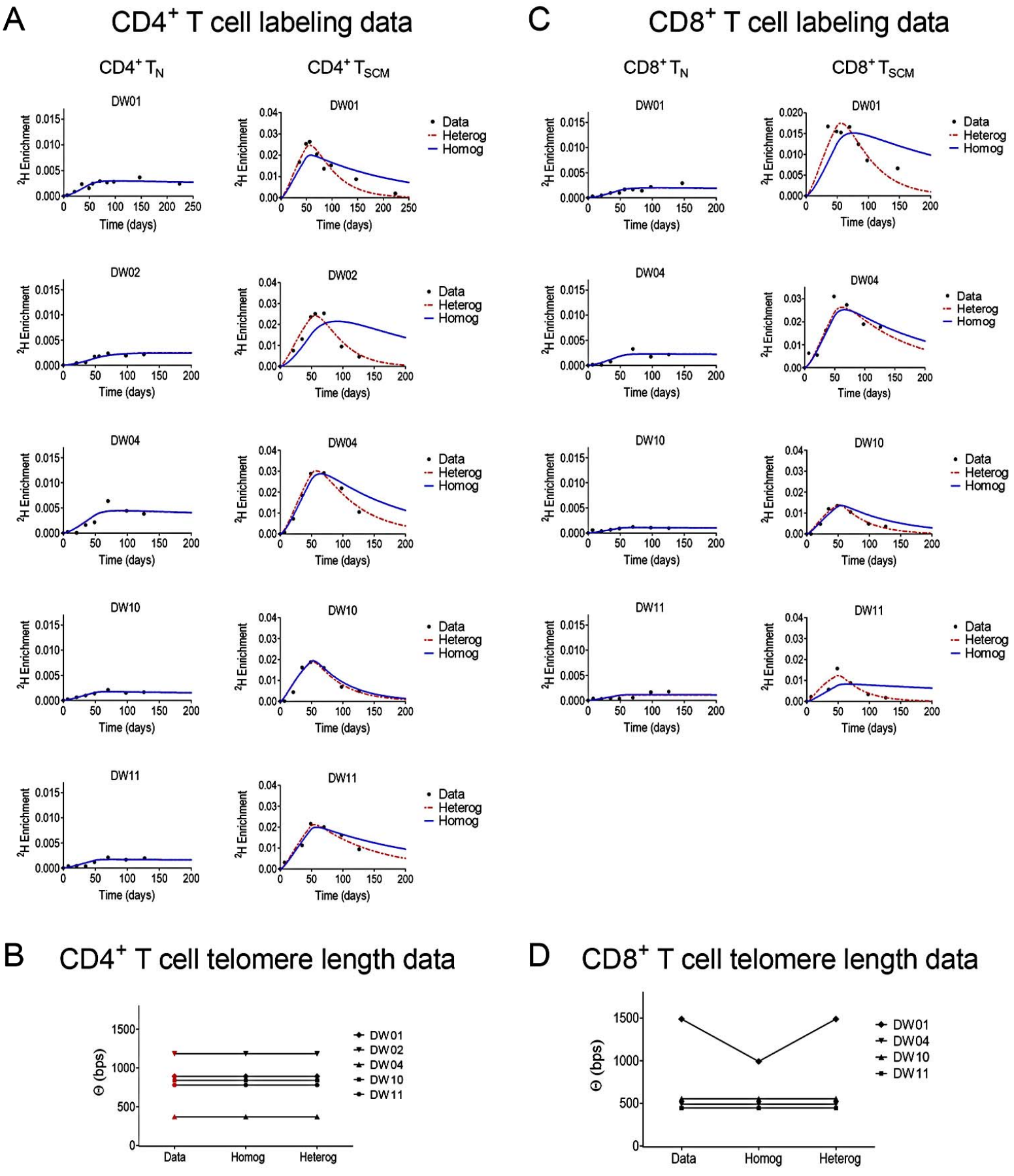
60. R: A language and environment for statistical computing. URL http://www.R-project.org/. 2014.

61. Soetaert K. R Package FME : Inverse Modelling, Sensitivity,Monte Carlo – Applied to a Dynamic Simulation Model. URL https://cran.r-project.org/web/packages/FME/.

62. Stauffer HB. Contemporary Bayesian and frequentist statistical research methods for natural resource scientists. Hoboken, N.J.: Wiley-Interscience; 2008. xv, 400 p. p.

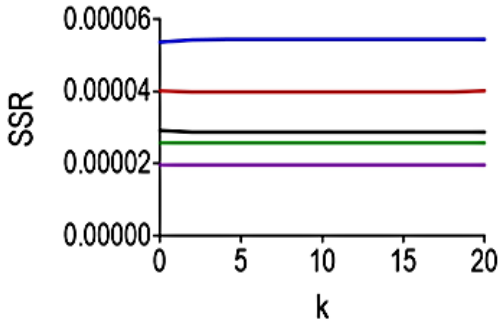

A
SSC-A
FSC-A
FSC-H
FSC-A
Live/dead/CD14/CD19
CD3
CD8
CD4
CD8+ CD27+CD45RO−
8+ naïve
8+ TSCM
CCR7
CD95
CD8+
CD27
CD45RO
CD4+ CD27+CD45RO−
4+ naïve
4+ TSCM
CCR7
CD95
CD4+
CD27
CD45RO
B
CD8+
8+ TSCM
8+ naïve
CD28
CD45RO
CD127
CD45RA
C
CD4+
4+ TSCM
4+ naïve
CD28
CD45RO
CD127
CD45RA


A
p_n
p_s
TN
$\Delta 2^k$
TSCM
d_n
d_s
B
d_n
$2p_n$
TN_0
TN_1
TN_2
TN_n
$\Delta 2^k$
$2p_s$
$TSCM_k$
$TSCM_{k+1}$
$TSCM_{k+2}$
$TSCM_{k+n}$
d_s

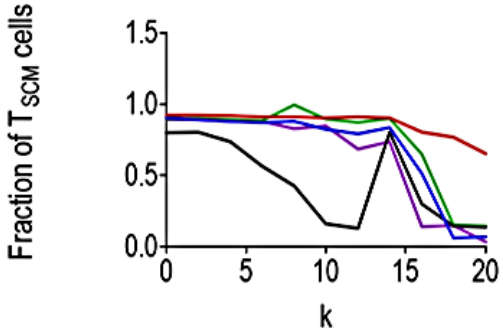

A
CD4+ T cell labeling data
CD4+ TN
CD4+ TSCM
DW01
DW02
DW04
DW10
DW11
2H Enrichment
Time (days)
Data
Heterog
Homog
C
CD8+ T cell labeling data
CD8+ TN
CD8+ TSCM
B
CD4+ T cell telomere length data
Θ (bps)
Data
Homog
Heterog
D
CD8+ T cell telomere length data

$CD4^+$ T cells

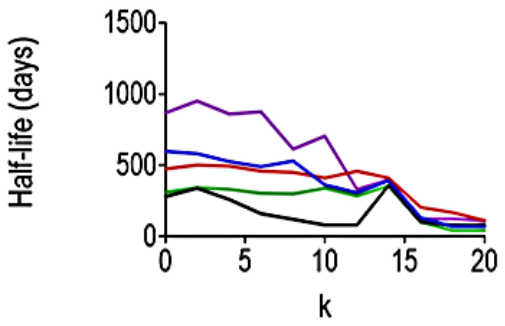
A Sum of squared residuals



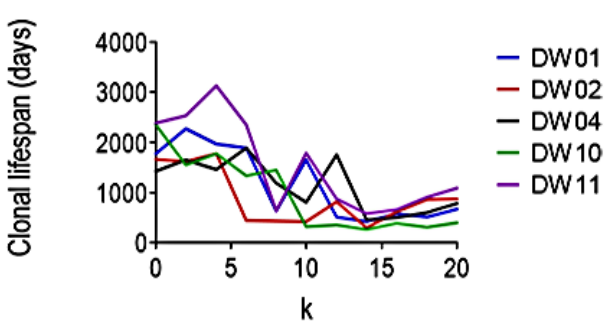
B Fraction of T_{SCM} from self-renewal



C Average half-life of T_{SCM}

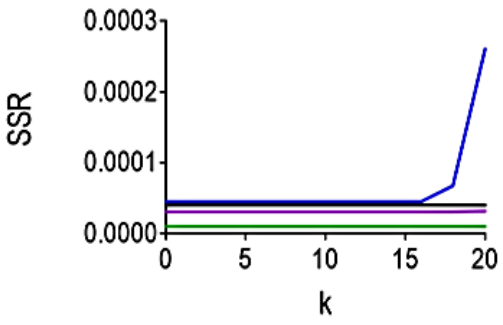


D Average total lifespan

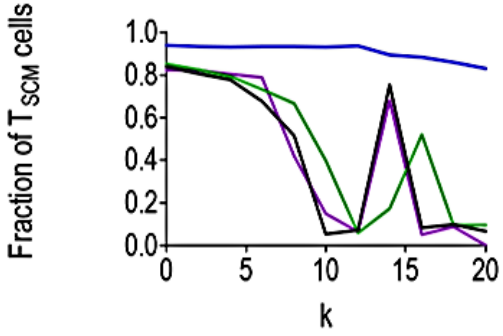


$CD8^+$ T cells

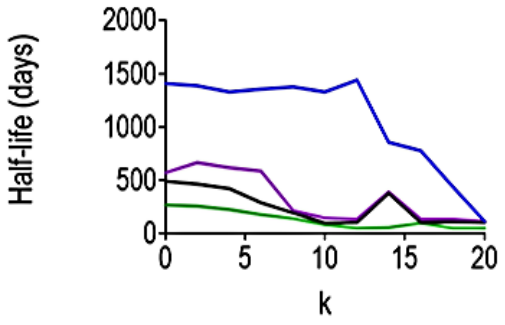
E Sum of squared residuals



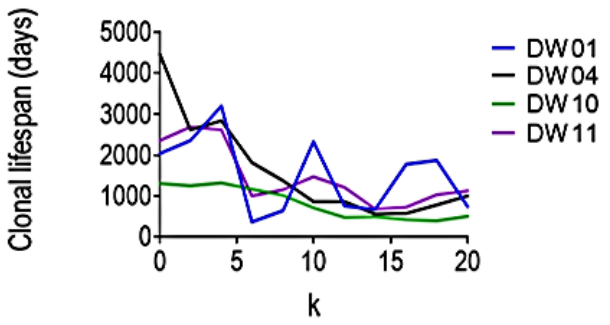
F Fraction of T_{SCM} from self-renewal

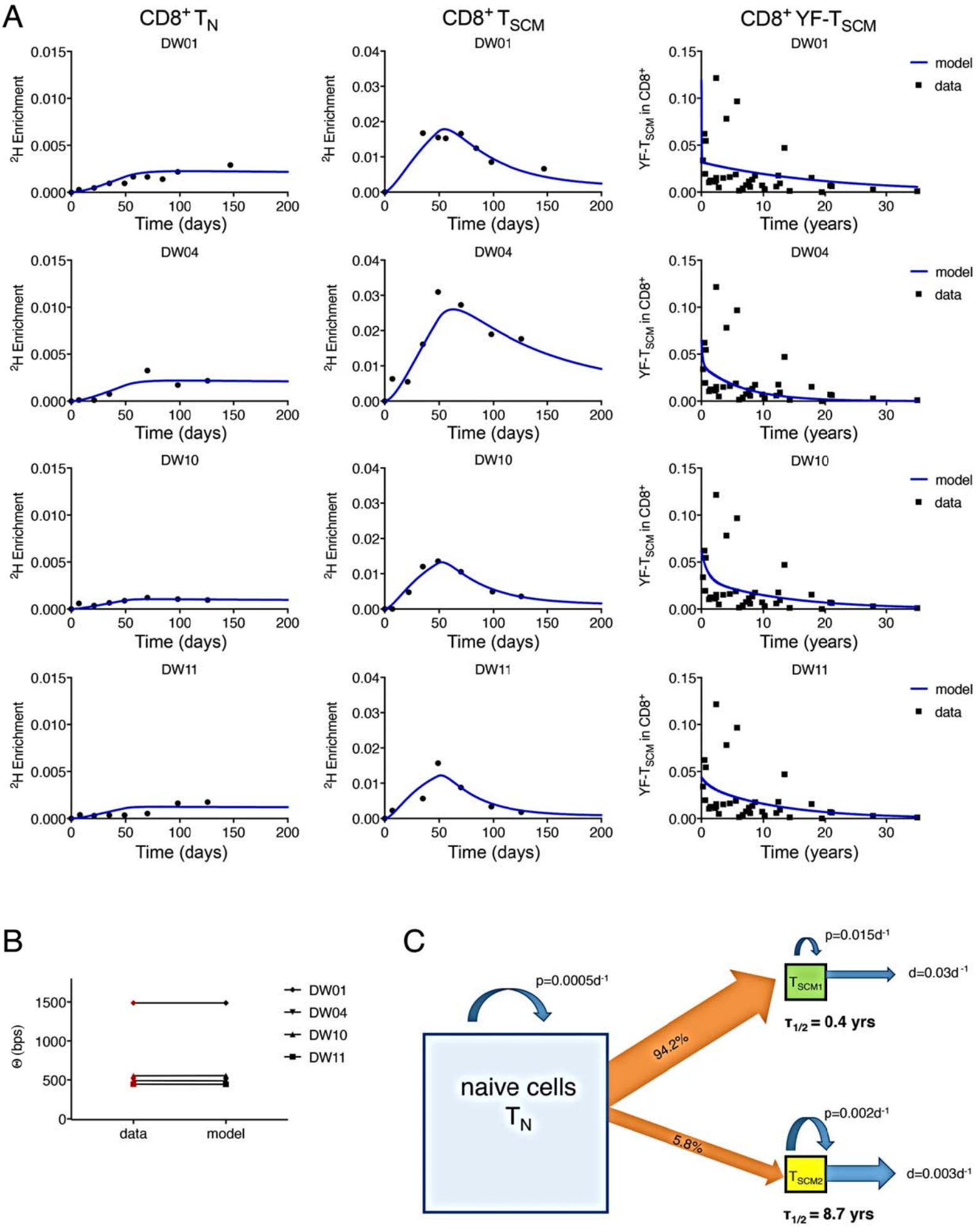


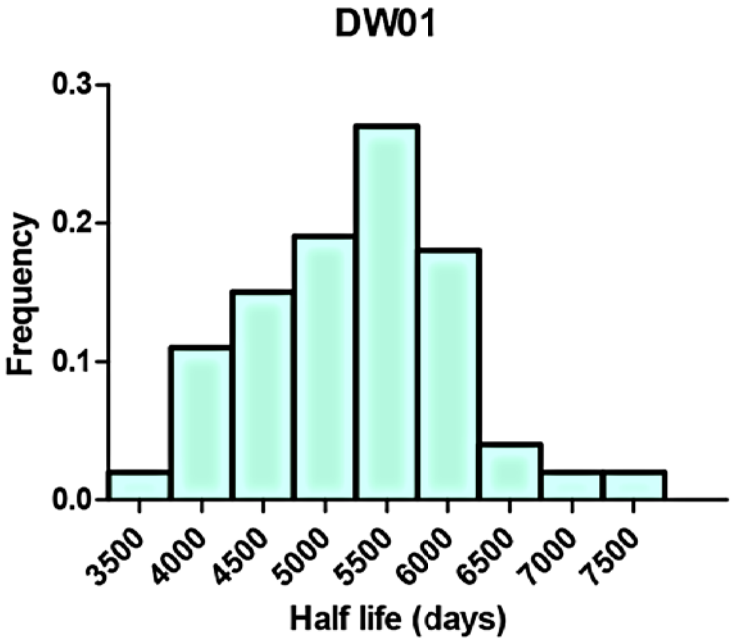
G Average half-life of T_{SCM}

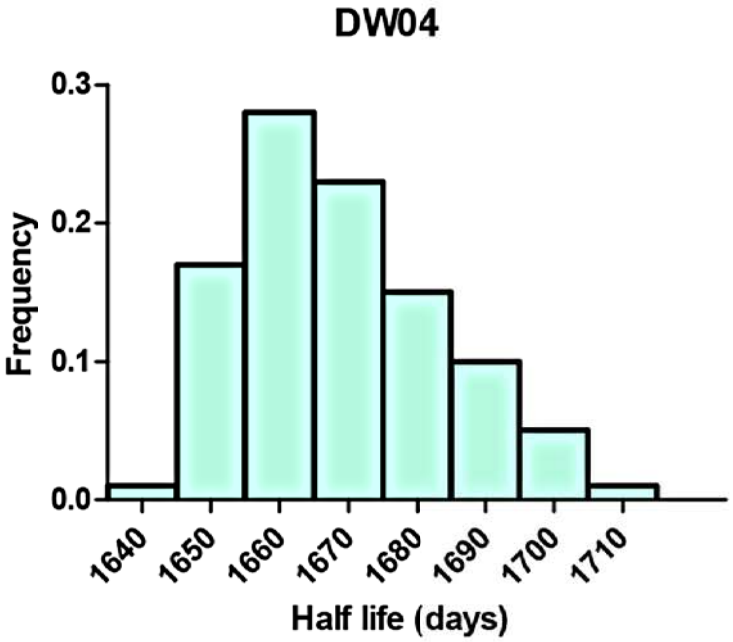


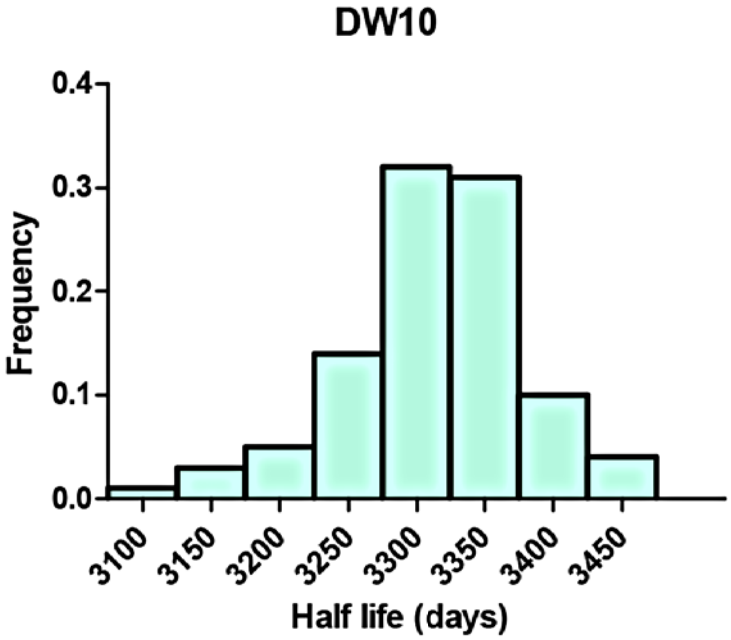
H Average total lifespan

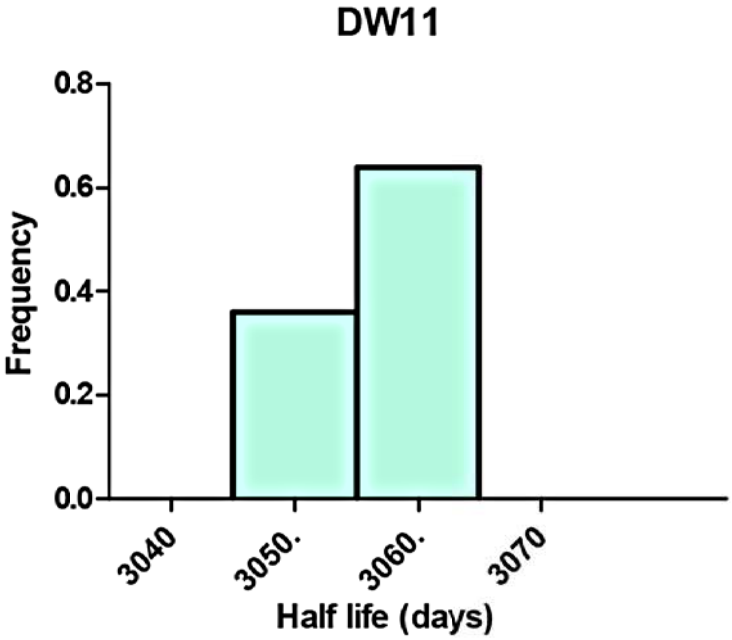


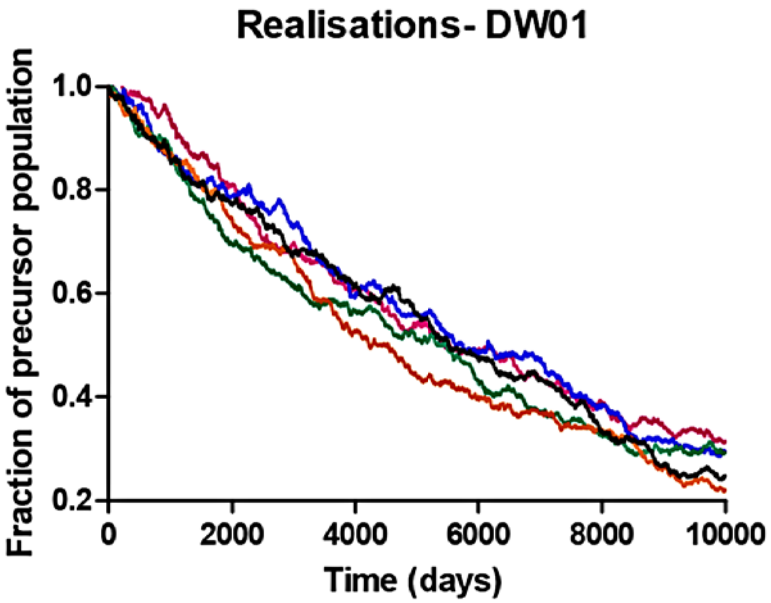

A
CD8+ TN
CD8+ TSCM
CD8+ YF-TSCM
DW01
DW04
DW10
DW11
2H Enrichment
YF-TSCM in CD8+
Time (days)
Time (years)
model
data
B
Θ (bps)
DW01
DW04
DW10
DW11
data
model
C
p=0.0005d-1
naive cells
TN
94.2%
5.8%
p=0.015d-1
TSCM1
d=0.03d-1
τ1/2 = 0.4 yrs
p=0.002d-1
TSCM2
d=0.003d-1
τ1/2 = 8.7 yrs

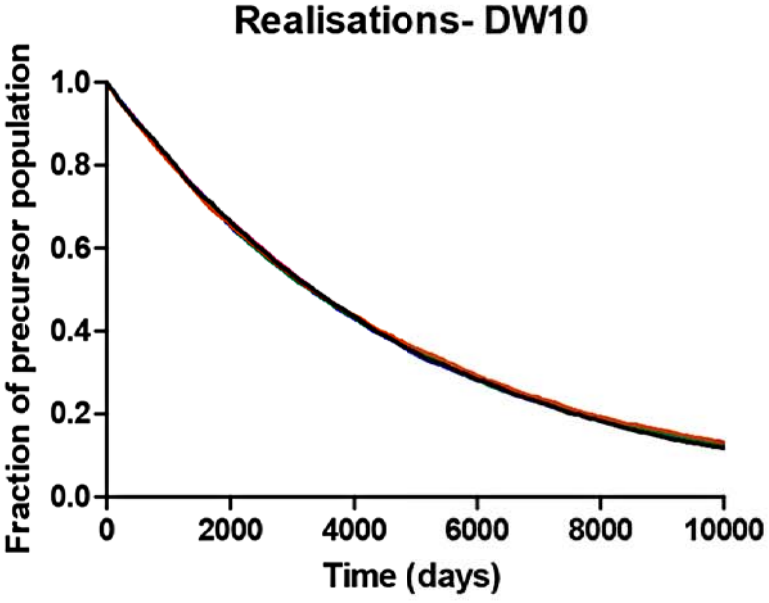

DW01
Frequency
0.3
0.2
0.1
0.0
3500
4000
4500
5000
5500
6000
6500
7000
7500
Half life (days)


DW04
Frequency
0.3
0.2
0.1
0.0
1640
1650
1660
1670
1680
1690
1700
1710
Half life (days)


DW10
Frequency
0.4
0.3
0.2
0.1
0.0
3100
3150
3200
3250
3300
3350
3400
3450
Half life (days)


DW11
Frequency
0.8
0.6
0.4
0.2
0.0
3040
3050.
3060.
3070
Half life (days)


Realisations- DW01
Fraction of precursor population
1.0
0.8
0.6
0.4
0.2
0
2000
4000
6000
8000
10000
Time (days)


Realisations- DW10
Fraction of precursor population
1.0
0.8
0.6
0.4
0.2
0.0
0
2000
4000
6000
8000
10000
Time (days)

SUPPLEMENTARY INFORMATION

Contents

Supplementary Methods

The explicit kinetic heterogeneity model for telomere length data

Following de Boer et al. [1], we derived an ODE model to describe the change in the average telomere lengths in the T_N population and in each of the T_{SCM} subpopulations (T_{SCM1} and T_{SCM2}) under the explicit kinetic heterogeneity model. Each of the three populations considered (T_N, T_{SCM1} and T_{SCM2}) was modelled as a series of n compartments. The equations for the number of cells in the T_{Ni}, T_{SCM1i} and T_{SCM2i} compartments are:

$$\dot{T}_{N_i} = 2p_n T_{N_{i-1}} - (p_n + \Delta + d_n)T_{N_i} \quad \text{(S1)}$$

$$\dot{T}_{SCM1_i} = 2p_{s1} T_{SCM1_{i-1}} - (p_{s1} + d_{s1})T_{SCM1_i} + \Delta f 2^k T_{N_{i-C}} \quad \text{(S2)}$$

$$\dot{T}_{SCM2_i} = 2p_{s2} T_{SCM2_{i-1}} - (p_{s2} + d_{s2})T_{SCM2_i} + \Delta(1-f) 2^k T_{N_{i-C}} \quad \text{(S3)}$$

where p_n, f, Δ, d_n, p_s, d_s, k and C are as in Methods. The equations for the average number of divisions undergone by cells in the T_N, T_{SCM1} and the T_{SCM2} pools (following the derivation in [1]) are:

$$\dot{\mu}_{TN} = 2p_n \quad \text{(S4)}$$

$$\dot{\mu}_{TSCM1} = 2p_{s1} - \Delta f 2^k \frac{T_N}{T_{SCM1}}(\mu_{T_{SCM1}} - \mu_{T_N} - C) \quad \text{(S5)}$$

$$\dot{\mu}_{TSCM2} = 2p_{s2} - \Delta(1-f) 2^k \frac{T_N}{T_{SCM2}}(\mu_{T_{SCM2}} - \mu_{T_N} - C) \quad \text{(S6)}$$

If δ is the number of base pairs (bp) lost in each division then, by multiplying μ_{TN}, μ_{TSCM1} and μ_{TSCM2} by δ, the average telomere lengths (in units of bp) are obtained. So the difference between the average telomere lengths of the T_N and the T_{SCM} pool is described by:

$$\dot{\Theta} = \delta\left(\frac{T_{SCM1}}{T_{SCM}}\dot{\mu}_{T_{SCM1}} + \frac{T_{SCM2}}{T_{SCM}}\dot{\mu}_{T_{SCM2}} - \dot{\mu}_{T_N}\right) \quad \text{(S7)}$$

Fitting procedure

In the first part of the study, stable isotope labelling and telomere length data were fitted simultaneously. Isotope labelling residuals were normalized by the sum of the observations to guarantee equal contribution of all data points, and the combined sum of squares score was computed giving the same weight to both isotope labelling and telomere length data. The free parameters in the implicit heterogeneous model (equations 9, 10 and 25) were p_n, Δ, k, p_s, d_s^*, the ratio of T_N to T_{SCM} cells (T_N/ T_{SCM}) and C. During the fitting procedure, the ratio T_N/ T_{SCM} was assumed to lie within the range 6 to 120. These bounds were estimated given that the fractions of peripheral blood $CD4^+$ and $CD8^+$ naïve T cells are typically 35% and 40% respectively [2], the fraction of peripheral blood T_{SCM} cells is typically 2-3% [2, 3], and the ratio

of the total number of T_N to T_{SCM} cells in the lymph nodes can be double the ratio observed in peripheral blood (at least in macaques [2]).

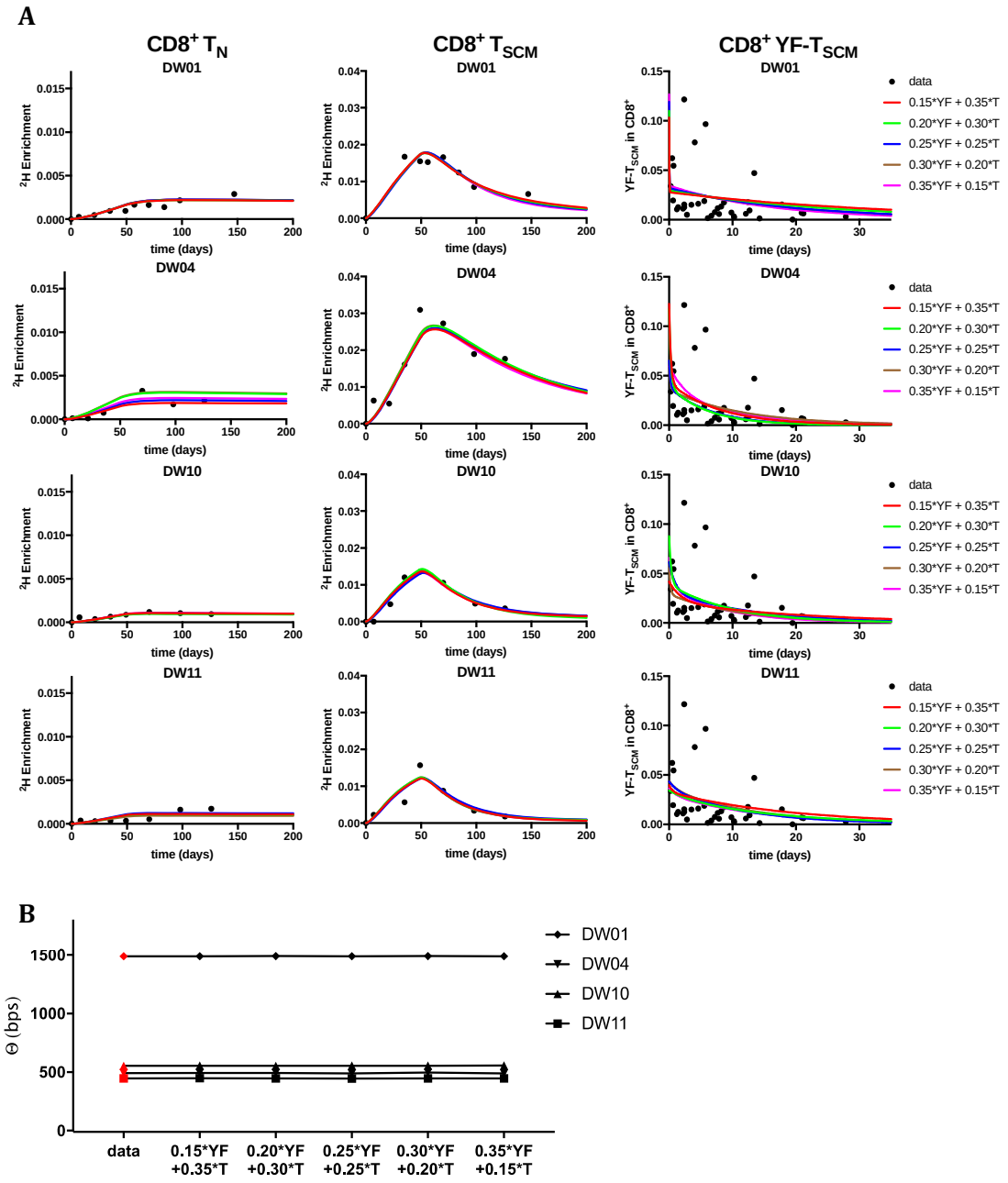
In the second part of the study, isotope labelling, telomere length data and YFV tetramer data were fitted simultaneously. The parameters α, β, and *r* of the two exponentials YFV model in equation 28 were equated to the net disappearance rates d_{s1}-p_{s1}, d_{s2}-p_{s2}, and to the ratio T_{SCM2}/ T_{SCM} in equations 18, 19 and S7. In the first instance the sum of squares for labelling, telomere length and YFV were weighted in the ratio 0.5 : 0.25 : 0.25. This is essentially a subjective choice that was made because we wanted the labelling data to be the most informative with 50% of the weight (the telomere length data represented only a single timepoint, and the YFV data were cross-sectional and obtained from unrelated individuals). Other weighting schemes were also considered and discussed in the text. The free parameters were p_n, Δ, f, k, p_{s1}, p_{s2}, the ratio of T_N to T_{SCM} cells (T_N/ T_{SCM}), the relative size of the two T_{SCM} subpopulations (T_{SCM2}/T_{SCM}) and C. d_{s1} and d_{s2} were eliminated using the steady state constraints $(1-f)2^k\Delta T_N+p_{s1}T_{SCM1}=d_{s1}T_{SCM1}$ and $f2^k\Delta T_N+p_{s2}T_{SCM2}=d_{s2}T_{SCM2}$.

The 95% confidence intervals of parameter estimates were calculated by creating 1000 bootstrap datasets (i.e. sampling from the data with replacement), fitting the model of interest to each dataset, estimating the parameters and then taking the upper and lower 2.5% percentiles of the resulting 1000 parameter estimates.

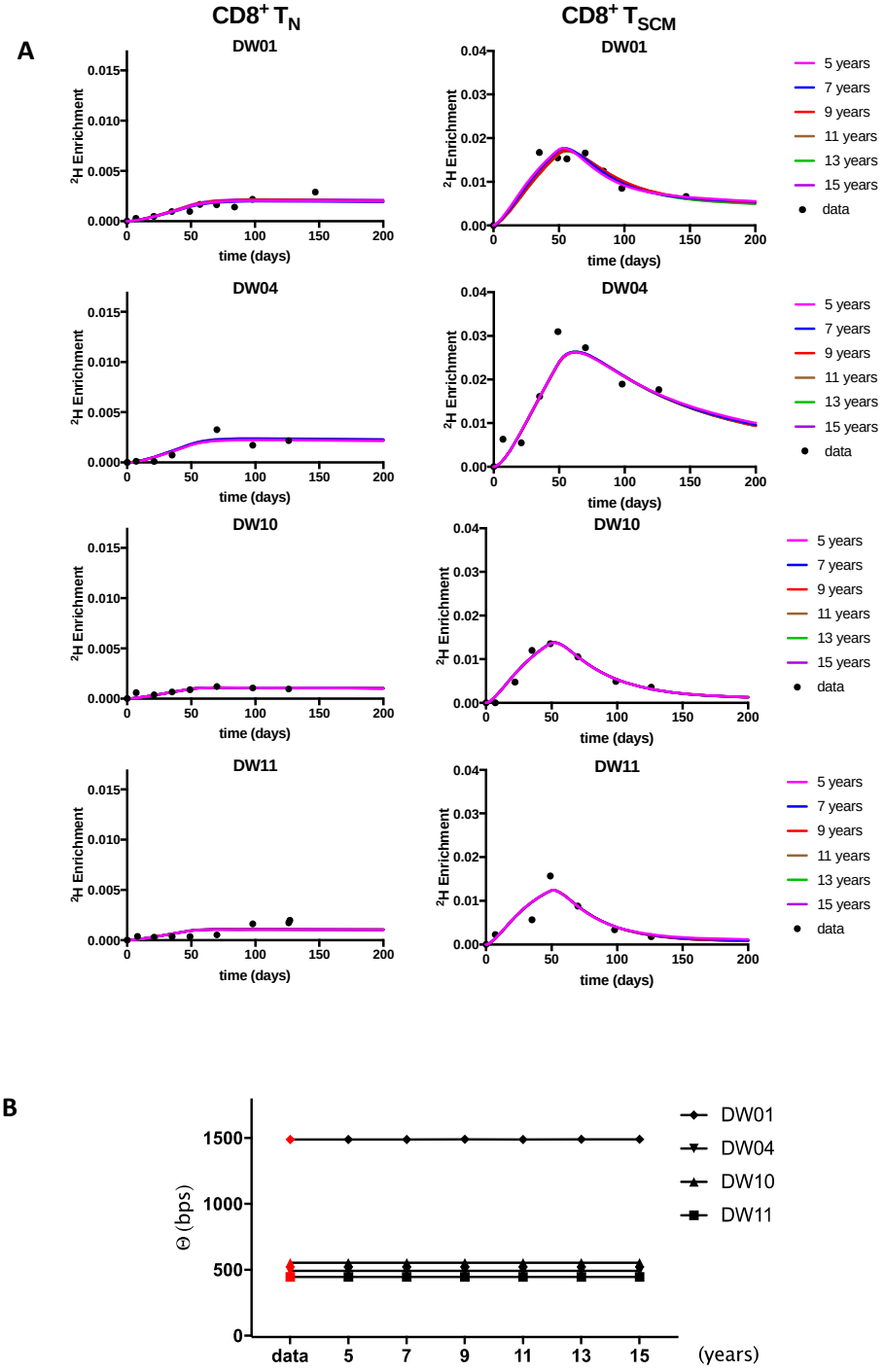
Summary of different modelling approaches

MODEL	DATA FITTED	RATIONALE	RESULTS REPORTED
Implicit heterogeneous	CD4 & CD8 labelling & telomere length	Test whether the T_{SCM} population is homogeneous by comparing fits of homogeneous and heterogeneous versions of the basic model (chosen because models are nested and have similar numbers of free parameters) & estimate average population parameters	Figures 2, 3 Supp Table S1
Explicit heterogeneous	CD8 labelling, telomere length, YFV	Estimate subpopulation parameters	Figures 4, 5 Table 1 Supp Fig S1 Supp Table S2, S3
	CD4 & CD8 labelling, telomere length	Test whether a slow and a rapid T_{SCM} subpopulation are compatible with the data independent of YFV data.	Supp Fig S2, S3

Supplementary Figures

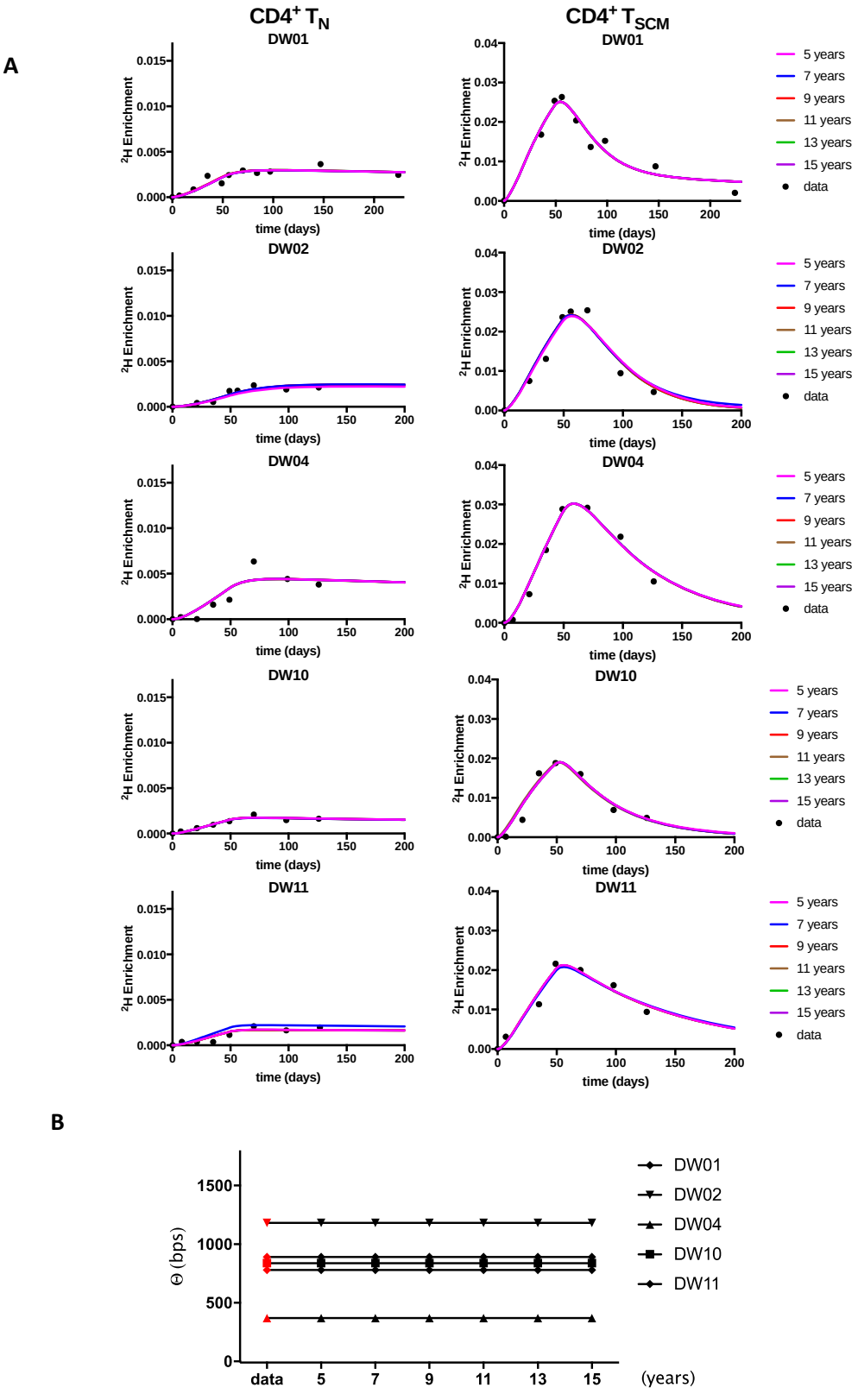


Supplementary Figure S1. Impact of different weightings when fitting CD8⁺ T cell data.
A) Fits to the labelling data for CD8⁺ naïve and T_{SCM} cells and to the YFV T_{SCM} data, when the labelling, telomere length, and YFV datasets were fitted simultaneously with the explicit heterogenity model. Each fit was generated with a different combination of weights to assess the importance of each dataset in the computation of the sum of squares during the fitting process. The weight corresponding to the labelling dataset was always fixed to 0.5. The experimental data is represented by solid black symbols. **B)** The corresponding fits to the average telomere length differences (Θ) between the T_N and T_{SCM} pools, with experimental data points in red. The number of base pairs lost in each division was taken to be δ =50bp/division.



Supplementary Figure S2. CD8⁺ T cell data are compatible with a long subpopulation half-life independent of yellow fever virus vaccine data.

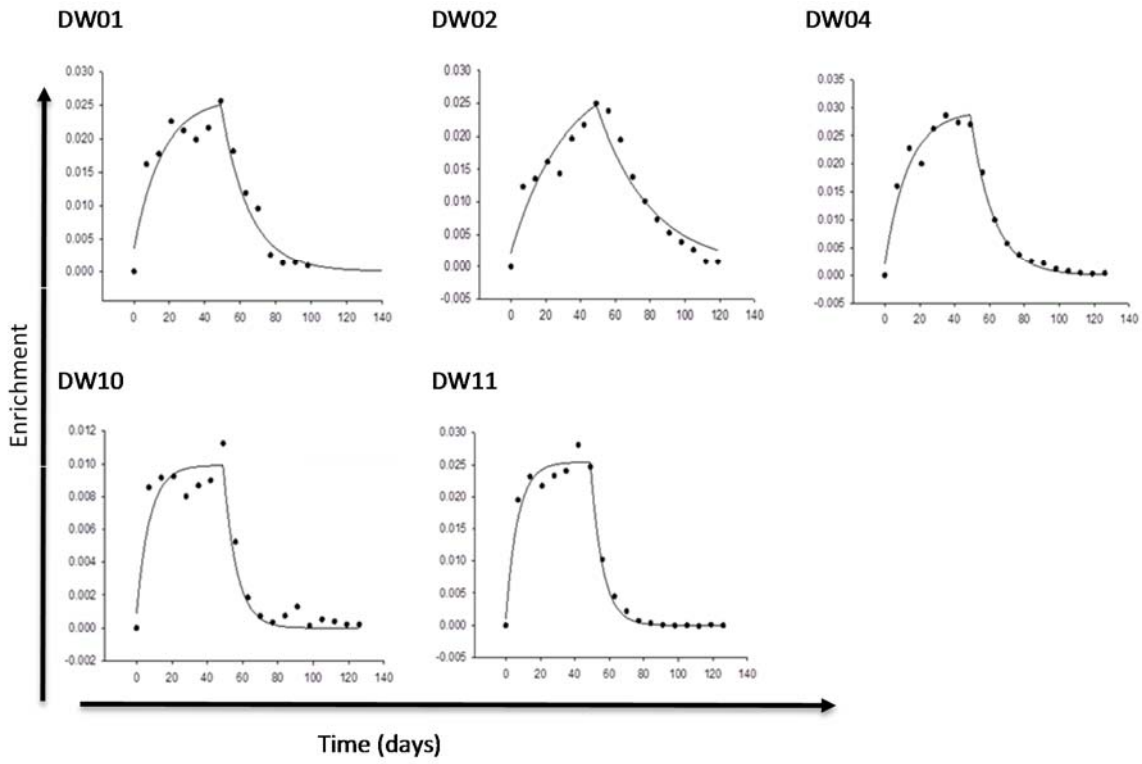
A) Observed and fitted label incorporation in CD8⁺ naïve and T_{SCM} cells when the explicit heterogeneity model was fitted to isotope labelling and telomere length datasets simultaneously, constraining the half-life of the slower subpopulation to lie between 5-6, 7-8, 9-10, 11-12, 13-14 or 14-15 years. Note that the predictions for many of the different half-lives overlie each other and so cannot be distinguished. **B)** Corresponding fits to the average telomere length differences (Θ) between the T_N and T_{SCM} pools, with experimental data points shown in red. The number of base pairs (bp) lost in each division was taken to be δ =50bp/division.



Supplementary Figure S3. CD4⁺ T cell data are compatible with a long subpopulation half life.

A) Observed and fitted label incorporation in CD4⁺ naïve and T_{SCM} cells when the explicit heterogeneity model was fitted to the labelling and telomere length datasets simultaneously, constraining the half-life of the slower subpopulation to lie between 5-6, 7-8, 9-10, 11-12, 13-14 or 14-15 years. Note, the predictions for many of the different half-lives overlie each other

and so cannot be distinguished. **B)** Corresponding fits to the average telomere length differences (Θ) between the T_N and T_{SCM} pools, with experimental data points shown in red. The number of base pairs (bp) lost in each division was taken to be δ =50bp/division.



Supplementary Figure S4. Deuterium enrichment in body water.

Body water enrichment as measured in saliva (dots) with best fit curves (solid line) in all volunteers. Parameter estimates from the model fits are given in SI Table S4.

Supplementary Tables

	T_{SCM} population (one-tailed p-values)		T_N population (one-tailed p-values)	
id	**$CD4^+$**	**$CD8^+$**	**$CD4^+$**	**$CD8^+$**
DW01	2.75e-08	2.53e-06	0.93	0.36
DW02	2.61e-11		0.21	
DW04	5.80e-07	6.50e-02	1.00	0.48
DW10	1.09e-01	4.86e-07	0.63	0.52
DW11	9.55e-04	5.65e-06	0.53	0.66
median	**5.80e-07**	**4.09e-06**	**0.63**	**0.50**
pooled	**3.50e-23**	**6.10e-15**	**0.86**	**0.68**

Supplementary Table S1. Statistical comparison of the quality of fit of the homogeneous and heterogeneous models to labelling and telomere length data for T_N and T_{SCM} cells.
The table shows one-tailed p-values obtained using Fisher's F-test comparison of the homogenous model with the (implicit) heterogeneous model. Fisher's F-test for nested models takes into account the different number of parameters in the models. The null hypothesis, that the population is kinetically homogeneous, was confidently rejected in 7/9 cases for the T_{SCM} population but in 0/9 cases for the T_N population. P values were pooled using Fisher's combined p (bottom row).

id	P_n [day^{-1}]	p_{s1} [day^{-1}]	p_{s2} [day^{-1}]	d_{s1}[day^{-1}]	d_{s2} [day^{-1}]	**Self-renewal** [d]
	(95% CI)	**95% CI**	**(95% CI)**	**95% CI**	**95% CI**	**95% CI**
DW01	0.0005	0.005	0.021	0.087	0.0214	7300
	(0.0003,0.0020)	(0,0.2)	(0.0015,0.0661)	(0.003,0.20)	(0.0016,0.0666)	(1800,12500)
DW04	0.0004	0	0.003	0.013	0.0033	2400
	(0.0003,0.0008)	(0,0.2)	(0.0001,0.0078)	(0.008,0.35)	(0.0005,0.0085)	(1200,5300)
DW10	0.0006	0.025	0.002	0.028	0.0022	4800
	(0.0005,0.0011)	(0,0.1055)	(0,0.0083)	(0.02,0.15)	(0.0003,0.0094)	(1400,8400)
DW11	0.0003	0.029	0	0.031	0.0002	4400
	(0.0001,0.0006)	(0.0032,0.183)	(0,0.0081)	(0.02,0.31)	(0.0001,0.0086)	(1700,9500)
MEDIAN	**0.0005**	**0.015**	**0.002**	**0.029**	**0.0027**	**4600**
	(0.0003,0.001)	**(0,0.19)**	**(0.0001,0.008)**	**(0.01,0.26)**	**(0.0004,0.009)**	**(1500,8900)**

Supplementary Table S2. Parameter estimates for CD8$^+$ T_{SCM} cells from the explicit heterogeneity model.

Parameter estimates with 95% CI (in parentheses) obtained by fitting the explicit heterogeneity model to the labelling, telomere length and YFV datasets for CD8$^+$ T cells simultaneously. The table shows the fitted parameters p_n (the proliferation rate of the naïve population), p_{s1} and p_{s2} (the proliferation rates of the two T_{SCM} subpopulations) as well as the derived parameters d_{s1} and d_{s2} (the disappearance rates of the two T_{SCM} subpopulations) and the degree of self-renewal of the long-lived subpopulation T_{SCM2}. d_{s1} and d_{s2} were calculated using the steady state constraints $(1-f)2^k\Delta T_N+p_{s1}T_{SCM1}=d_{s1}T_{SCM1}$ and $f2^k\Delta T_N+p_{s2}T_{SCM2}=d_{s2}T_{SCM2}$ and the degree of self-renewal of the T_{SCM2} subpopulation was calculated as $1/(d_{s2}-p_{s2})$. This table is a continuation of the table of parameters in the main text (Table 1).

id	weight combination	half-live T_{SCM1} [years]	half-live T_{SCM2} [years]	T_{SCM2}/T_{SCM}	f
DW01	0.15*YF+0.35*T	0.02	23.87	0.25	0.000
	0.20*YF+0.30*T	0.02	18.67	0.25	0.000
	0.25*YF+0.25*T	0.02	13.92	0.25	0.001
	0.30*YF+0.20*T	0.02	14.15	0.26	0.001
	0.35*YF+0.15*T	0.02	11.20	0.25	0.001
	median	**0.02**	**14.15**	**0.25**	**0.001**
DW04	0.15*YF+0.35*T	0.18	5.49	0.35	0.017
	0.20*YF+0.30*T	0.18	4.26	0.33	0.021
	0.25*YF+0.25*T	0.14	4.59	0.58	0.041
	0.30*YF+0.20*T	0.14	7.93	0.58	0.023
	0.35*YF+0.15*T	1.84	7.88	0.51	0.196
	median	**0.18**	**5.49**	**0.51**	**0.023**
DW10	0.15*YF+0.35*T	1.29	14.45	0.52	0.087
	0.20*YF+0.30*T	0.23	6.92	0.44	0.026
	0.25*YF+0.25*T	0.69	9.09	0.51	0.075
	0.30*YF+0.20*T	0.15	7.84	0.52	0.020
	0.35*YF+0.15*T	0.50	6.27	0.53	0.083
	median	**0.50**	**7.84**	**0.52**	**0.075**
DW11	0.15*YF+0.35*T	0.38	13.43	0.82	0.116
	0.20*YF+0.30*T	2.31	11.20	0.83	0.500
	0.25*YF+0.25*T	0.90	8.39	0.25	0.001
	0.30*YF+0.20*T	0.91	9.36	0.26	0.001
	0.35*YF+0.15*T	0.78	9.87	0.25	0.001
	median	**0.90**	**9.87**	**0.82**	**0.307**

Supplementary Table S3. Impact of relative weighting of telomere length and YFV data.
Parameter estimates (and medians) estimated using the explicit heterogeneity model for the five weighting strategies in Supplementary Figure S1. The table shows half-life estimates for the T_{SCM1} and T_{SCM2} populations respectively, the relative size of the T_{SCM2} subpopulation (T_{SCM2}/T_{SCM}), and the fraction of cells generated from each clonal burst that enter the long-lived T_{SCM2} subpopulation (*f*). Different rows (for a given individual) correspond to different weighting strategies, where the weights applied to the yellow fever virus (YF) and telomere length (T) datasets are indicated in the second column "weight combination", and the weight applied to the labelling dataset was in all cases 0.5.

ID	*fr*	*δ*	β
DW01	0.026	0.063	0.0035
DW02	0.031	0.032	0.0022
DW04	0.029	0.075	0.0021
DW10	0.010	0.135	0.0009
DW11	0.026	0.145	0.0011

Supplementary Table S4. Estimates of body water parameters.

Parameter estimates obtained by fitting an empirical model (equations 5 & 6 main text) to the saliva enrichment in each individual. *fr* represents the fraction of deuterium in water, *δ* the turnover rate per day of body water, and β the baseline saliva enrichment. The corresponding fits are provided in SI Figure S4.

Id	Δ	k	f	Initial size
DW01	5.05E-04	0.025	5.6E-04	2.9E+04
DW04	1.05E-07	9.753	4.1E-02	3.7E+05
DW10	1.65E-07	6.536	7.4E-02	1.1E+05
DW11	9.70E-08	9.906	3.3E-01	3.1E+06

Supplementary Table S5. Estimates of the initial size of a long lived T_{SCM} clone.

The initial size (*size*$_0$) of a long-lived T_{SCM} clone is

$$size_0 = T_N \times 2^k \times \Delta \times f$$

where T_N is the number of naïve $CD8^+$ T cells (taken to be approximately 10^{11} in an adult human [4]), *k* is the number of divisions occurring during clonal expansion in the differentiation from T_N to T_{SCM}, *Δ* is the fraction of naïve cells activated by the same antigen and *f* is the fraction of a newly generated clone that enters the long-lived T_{SCM} pool. The parameters *k*, *Δ* and *f* were obtained by fitting the explicit heterogeneity model simultaneously to the labelling, telomere length and YFV tetramer data; the initial clone size (*size*$_0$) was then calculated from these estimated parameters.

Antigen-specific naïve T cell precursor frequency (*Δ*) has been estimated at < 1 to 352 per 10^5 naïve $CD8^+$ T cells [5], consistent with our estimates (0.01–50 per 10^5 naïve $CD8^+$ T cells).

Supplementary References

1. De Boer RJ, Noest AJ. T cell renewal rates, telomerase, and telomere length shortening. J Immunol. 1998;160(12):5832-7.

2. Lugli E, Dominguez MH, Gattinoni L, Chattopadhyay PK, Bolton DL, Song K, et al. Superior T memory stem cell persistence supports long-lived T cell memory. The Journal of clinical investigation. 2013;123(2):594-9.

3. Gattinoni L, Lugli E, Ji Y, Pos Z, Paulos CM, Quigley MF, et al. A human memory T cell subset with stem cell-like properties. Nat Med. 2011;17(10):1290-7.

4. Bains I, Antia R, Callard R, Yates AJ. Quantifying the development of the peripheral naive CD4(+) T-cell pool in humans. Blood. 2009;113(22):5480-7.

5. Neller MA, Ladell K, McLaren JE, Matthews KK, Gostick E, Pentier JM, et al. Naive CD8(+) T-cell precursors display structured TCR repertoires and composite antigen-driven selection dynamics. Immunol Cell Biol. 2015;93(7):625-33.