Figure #	Figure title One sentence only	Filename This should be the name the file is saved as when it is uploaded to our system. Please include the file extension. i.e.: *Smith_ED_Fig1.jpg*	Figure Legend If you are citing a reference for the first time in these legends, please include all new references in the main text Methods References section, and carry on the numbering from the main References section of the paper. If your paper does not have a Methods section, include all new references at the end of the main Reference list.
Extended Data Fig. 1	Immunophenotyping of CD3+ cells in Pt4 and Pt6	ExtendedData1.tif	Additional FACS plot on CD62L+CD45RA+CD95 expression in CAR^+ T cells, subtype frequencies within total $CD3^+$ cells and relative summary of data on the $CD3^+$ T cell composition overtime for Pt4 (top panels) and Pt6 (bottom panels). Percentages of CD62L/CD45RA compartments are showed inside each gate. Longitudinal contribution of each subpopulations to the $CD3^+$ cell compartment is shown below for each patient according to the color code legend on the right.
Extended Data Fig. 2	Integration sites collected from Pt4 and Pt6	ExtendedData2.tif	Summary of number of IS (grey bars) and relative sequencing reads (white bars) in Pt4 (top panel) and Pt6 (bottom panel) (d = days after treatment)
Extended Data Fig. 3	Distribution of integration sites collected from Pt4 and Pt6	ExtendedData3.tif	Distribution of IS with respect to TSS (left plots) or gene content of the loci (right plots) in the cell product (red) or after infusion at early or late timepoints after treatment (blue) for Pt4 (top panels) and Pt6 (bottom panels).
Extended Data Fig. 4	Gene categories relative to integration sites collected from Pt4 and Pt6	ExtendedData4.tif	Plots on the left show word clouds of hit genes in the cell product (in red) and after infusion at early or late timepoints after treatment (in blue) for Pt4 (top panels) and Pt6 (bottom panels). Relative gene enrichment analysis for top biological processes of hit genes relative to each word cloud in the product (red bars) or after infusion (blue bars) is shown on the right plots (significance reported as -log10 binomial p-value from one-tailed tests decreasing from top to bottom).
Extended Data Fig. 5	Relative abundance and diversity of	ExtendedData5.tif	Scattered dot plots showing relative abundance of IS in the product **(A)** and at early **(B)** or late **(C)** timepoints

	integration sites collected from Pt4 and Pt6		after treatment in Pt4 (plots on the left) and Pt6 (plots on the right) (d = days after treatment, m = months after treatment). Mean percent abundance is shown as a dotted line for each sample. Number of events in each dataset is equal to what reported in Extended Data 2. **D)** Longitudinal plots showing Gini/Simpson Diversity Index (left y-axis) of IS overtime in T_{SCM} (orange lines) and in T_{CM}/T_{EM} (dark blue) and in all T cells (green lines). The grey lines show the percentage of CAR^+ cells overtime (right y-axis).
Extended Data Fig. 6	Correlation of integration sites and number of cells collected from Pt4 and Pt6	ExtendedData6.tif	Longitudinal plots showing number of IS collected **(A)** and number of cells collected **(B)** in T_{SCM} (orange lines) and in T_{CM}/T_{EM} (dark blue) at early timepoints for Pt4 (left panels) and Pt6 (right panels). The plot in **(C)** shows the correlation observed between number of cells collected (y-axis) and clonal diversity for all samples and both patients (x-axis, Shannon Diversity Index) shown as blue dots. Interpolation with best fit curve and R squared value are shown in black.
Extended Data Fig. 7	Integration sites collected from Pt10 and Pt17	ExtendedData7.tif	Summary of number of IS (grey bars) and relative sequencing reads (white bars) in Pt10 (top panel) and Pt17 (bottom panel) (d = days after treatment)
Extended Data Fig. 8	Sharing of integration sites in the product with the CAR T populations in Pt10 and Pt17 after infusion	ExtendedData8.tif	Ring plots showing relative contribution from each subtype (coloured section of the ring) to the pool of IS detected in the product and at early timepoints in Pt10 (left) and Pt6 (right). The total number of IS captured both in the product and after infusion is shown inside each plot. The relative percentage of IS belonging to each T cell subtype of the product that were shared with samples after infusion is shown in white inside each section of each ring plot.

Item	Present?	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. The extension must be .pdf	A brief, numerical description of file contents. i.e.: *Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.*
Supplementary Information	No		
Reporting Summary	Yes	nr-reporting-summary.pdf	

Type	Number If there are multiple files of the same type this should be the numerical indicator. i.e. "1" for Video 1, "2" for Video 2, etc.	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_ Supplementary_Video_1.mov*	Legend or Descriptive Caption Describe the contents of the file
Supplementary Data	Supplementary Data 1	CARPALL_study_protocol.pdf	Clinical study protocol

Parent Figure or Table	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_SourceData_Fig1.xls*, or *Smith_ Unmodified_Gels_Fig1.pdf*	Data description i.e.: Unprocessed Western Blots and/or gels, Statistical Source Data, etc.
Source Data Fig. 1	Biasco_SourceData_Fig.1.xlsx	Raw datasets relative to figure panels
Source Data Fig. 2	Biasco_SourceData_Fig.2.xlsx	Raw datasets relative to figure panels
Source Data Fig. 4	Biasco_SourceData_Fig.4.xlsx	Raw datasets relative to figure panels
Source Data Fig. 5	Biasco_SourceData_Fig.5.xlsx	Raw datasets relative to figure panels
Source Data Fig. 6	Biasco_SourceData_Fig.6.xlsx	Raw datasets relative to figure panels
Source Data Fig. 7	Biasco_SourceData_Fig.7.xlsx	Raw datasets relative to figure panels
Source Data Fig. 8	Biasco_SourceData_Fig.8.xlsx	Raw datasets relative to figure panels
Source Data Extended Data Fig. 1	Biasco_SourceData_ExtendedDataFig.1.xlsx	Raw datasets relative to figure panels
Source Data Extended Data Fig. 2	Biasco_SourceData_ExtendedDataFig.2.xlsx	Raw datasets relative to figure panels
Source Data Extended Data Fig. 5	Biasco_SourceData_ExtendedDataFig.5.xlsx	Raw datasets relative to figure panels
Source Data Extended Data Fig. 6	Biasco_SourceData_ExtendedDataFig.6.xlsx	Raw datasets relative to figure panels
Source Data Extended Data Fig. 7	Biasco_SourceData_ExtendedDataFig.7.xlsx	Raw datasets relative to figure panels

Clonal expansion of T memory stem cells determines early anti-leukaemic responses and long-term CAR T cell persistence in patients

Luca Biasco[1,2], Natalia Izotova[1], Christine Rivat[1], Sara Ghorashian[3], Rachel Richardson[1], Aleks Guvenel[1], Rachael Hough[4], Robert Wynn[5], Bilyana Popova[6], Andre Lopes[6], Martin Pule[7], Adrian J. Thrasher[1] and Persis J. Amrolia[1,8*]

[1]Molecular and Cellular Immunology Section, UCL Great Ormond St Institute of Child Health, London, United Kingdom

[2]Gene Therapy Program, Dana-Farber/Boston Children's Cancer and Blood Disorders Center, Harvard Medical School, Boston, MA

[3]Molecular Haematology Section, UCL Great Ormond Street Institute of Child Health, London, United Kingdom

[4]Department of Haematology, University College London Hospital, London, United Kingdom

[5]Department of Bone Marrow Transplant, Royal Manchester Children's Hospital, Manchester, United Kingdom

[6]CRUK UCL Cancer Trials Centre, University College London, London, United Kingdom

[7]University College London Cancer Institute, London, United Kingdom

[8]Department of Bone Marrow Transplantation, Great Ormond Street Hospital, London, United Kingdom

*corresponding author: Persis.Amrolia@gosh.nhs.uk

ABSTRACT

Low-affinity CD19 CAR T cells display enhanced expansion and persistence, enabling fate tracking through integration site (IS) analysis. Here we show that IS from early (1 month) and late (>3 years) time-points cluster separately, suggesting different clonal contribution to early responses and prolonged anti-leukemic surveillance. CAR+ T central and effector memory in patients with long-term persistence remained highly polyclonal, whereas diversity dropped rapidly in patients with limited CAR T persistence. Analysis of shared integrants between the CAR T cell product and post-infusion demonstrated that, despite their low frequency, T memory stem cells (TSCM) clones in the product contributed substantially to the circulating CAR T cell pools, both during early expansion and long-term persistence. Our data may help identify patients at risk of early loss of CAR T cells and highlights the critical role of TSCM in both mediating early anti-leukemic responses and long-term surveillance by CAR T cells.

INTRODUCTION

Adoptive T cell transfer shows great promise for the treatment of both viral infections and cancer but much remains to be learnt about the optimal phenotype of T cell products, their clonal dynamics once infused and the origin of both early responding T cells that mediate initial responses as well as long term persisting T cells that provide immunological surveillance. According to the hierarchical model of T cell differentiation, upon antigenic priming naïve T cells (T_N) differentiate in to central memory (T_{CM}) cells which express lymph node homing molecules but have limited effector function which in turn give rise to effector memory cells (T_{EM}) which preferentially traffic to the tissues and differentiate to terminally differentiated effectors (T_{EMRA}) which mediate rapid effector functions[1]. Subsequently, Gattinoni *et al*[2] demonstrated the existence of stem cell memory (T_{SCM}) compartment with a naïve surface phenotype and high proliferative capacity but also attributes of memory T cells which appear endowed with both the potential for multipotent differentiation and long term self-renewal. For durable anti-viral/anti-tumor responses long term immune surveillance is critical but the ontogeny of T cells mediating such long-term responses remains a matter of active debate. Early data from adoptive transfer of gene marked EBV-specific T cells which had a predominantly T_{EM} phenotype at infusion showed these could be detected up to 9 years post-infusion[3]. In a macaque model, CMV-specific CD8 T cell clones derived from T_{CM} (but not T_{EM}) persisted long term after adoptive transfer[4]. Tracking of T cells retrovirally transduced to express ADA using integration site (IS) analysis has demonstrated that genetically engineered T_{SCM} clones persist and preserve their differentiation potential for up to 12 year post-infusion[5]. On the other hand, similar analyses of marked circulating T cells 2-14 years after infusion of T cells retrovirally transduced with the HSV thymidine kinase suicide gene suggested that long term persisting marked cells arose predominantly from both infused T_{SCM} and T_{CM} compartments[6].

CD19 chimeric antigen receptor (CAR) T cells show unprecedented responses in relapsed/refractory acute lymphoblastic leukemia (ALL)[7–10]. Anti-leukemic responses correlate with both the level of early CAR T cell expansion[11] and the duration of CAR T cell engraftment with long-term persistence critical for durable remission if used as a stand-alone therapy. The origin of early expanding and long-term persisting CAR-T cells may differ and has yet to be defined. This will be critical in designing manufacturing protocols to

optimise outcomes. Initial data in a xenogeneic mouse model of mesothelioma where mesothelin CAR transduced human $CD8^+$ $T_{SCM}/T_{CM}/T_{EM}$ were infused with $CD4^+$ CAR T cells demonstrated superior anti-tumor responses with T_{SCM}[2]. In a xenogeneic model of lymphoma, CD19 CAR T cells derived from selected $CD4^+$ T_N and $CD8^+$ T_{CM} populations gave superior anti-tumor activity and a 1:1 combination of bulk $CD4^+$ and $CD8^+$ T_{CM} CD19 CAR T cells showed optimal activity[12], leading the investigators to test this approach clinically in adult ALL[13]. Xu *et al*[14] found that the expansion of CD19 CAR T cells in patients correlated with the frequency of $CD8^+CD45RA^+CCR7^+$ within the infused product whose phenotype approximated T_{SCM} (it should be noted that in this study naïve T cells were not distinguished from T_{SCM} by evaluation of CD95 expression) although CAR T cell expansion/persistence in this study was poor possibly due to the lack of lymphodepletion and the use of a CD28 costimulatory domain respectively. Little is known about the how the phenotype and clonal composition of CAR T cells changes following adoptive transfer. Integration site analysis offers us a unique tool to track the clonal dynamics of genetically modified CAR T cells in patients as each transduced cell is stably marked by a vector integration site acting as a genetic barcode. Using TCRβ and IS analysis, Sheih *et al*[15] have shown that CAR T cells at early time points post-infusion are polyclonal but individual clones display different clonal kinetics after infusion. However, because of limited CAR T cell persistence in this study, the analysis was restricted to 30 days post-infusion. Moreover, it is not clear whether these data are skewed by the use of the infused products derived from selected $CD4^+$ and $CD8^+$ T_{CM}. To date, it has not been possible to isolate long-term (> 1 year) persisting CAR-T cells in patients. In the CARPALL Phase I study, the use of an improved low-affinity CD19 CAR resulted in enhanced expansion and persistence of CAR-T cells post-infusion[16] enabling us for the first time to detect and select long term persisting CAR T cells flow cytometrically. By combining phenotypic analysis after treatment with high-resolution IS analysis of selected CAR T cells, we have analysed the fate of infused CAR-T cells in 2 patients with long-term persistent CAR-T cells in peripheral blood and compared this to 2 patients who had early loss of CAR T cells (Fig. 1A).

RESULTS

CAR T cells kinetics and phenotype post infusion

In order to study the composition of CAR T cells in our patients overtime we performed a comprehensive immunophenotypic analysis for the detection of CAR^+ cells and for the characterization of different T cell subtypes (Fig. 1B,C). All 4 patients analyzed showed similar kinetics of CAR^+ T cell engraftment showing relative contribution in the $CD3^+$ compartment as high as 14-80% at the time of peak expansion (day 14) followed by contraction to 6-30% by day 30 (reflecting elimination of $CD19^+$ targets) and then low level long term persistence in Pt4 and Pt6 (0.1% at 36 and 28 months) (Fig. 1D,E). We then assessed the naïve/memory phenotype of CAR T cells in both the infused product and in the peripheral blood post infusion flow cytometrically using CD45RA, CD62L and CD95 as markers (Extended Data 1). In the patients with long term CAR T cell persistence, the infused CAR T cell product was primarily composed of T_{CM} (78-81% of CAR^+ T cells) with a lower contribution of T_{EM} (17-21%), but also a small but detectable pool (1-2%) of T_{SCM} (Fig. 1F). Following infusion, in both these patients the composition of CAR^+ T cells switched substantially towards the T_{EM} phenotype, which became the predominant fraction at both early and late timepoints (Fig. 2E). There was a small increase in the percentage of terminally differentiated effector CAR T cells (T_{EMRA}) in both patients up to the 6 months timepoint, but this declined thereafter. Interestingly, the small fraction of T_{SCM} present upon infusion became increasingly represented over time, implying that the CAR^+T_{SCM} detected in the cell product are indeed endowed with high self-renewing potential after infusion.

For comparison, we performed similar phenotyping analysis on CAR T cells from 2 patients (Pt10 and Pt17) who lost circulating CAR T cells at 2 and 4 months post-infusion to investigate whether differences in product composition or relative engraftment of T cell subsets might correlate with lack of persistence in patients. As shown in Figure 2, the CAR T cell products (particularly Pt17) had a more differentiated phenotype with a lower proportion of T_{SCM} and higher T_{EM} than the patients with long term persistence (Fig. 2A,B,D). Similarly, by day 30 post-infusion, CAR T cells from these patients showed a predominantly T_{EM} phenotype whereas in Pt4 and 6 we observed a substantially higher preservation of the T_{SCM} and T_{CM} compartments (Fig. 2A,C).

Integration site analysis of CAR T cells

Next, we characterized CAR T cells of these patients at single clone level to address a) what are the CAR T cell clonal dynamics occurring during acute anti-tumor response phase *vs* steady state, b) what is the clonal composition of the long-surviving CAR^+ T cells and c) which fraction of the cell product contributes the most to the long-term maintenance of CAR^+ cells in these patients. To this aim we collected, through our well-established protocol that combines FACS sorting of blood cell subpopulations, whole genome amplification, Linear Amplification Mediated PCR (LAM-PCR) combined with high throughput sequencing (Illumina)[5], a total of 9,881 and 3,667 integration sites (IS) from T_{SCM}, T_{CM} and T_{EM} isolated from the cell products of Pt4 and Pt6 respectively. Similarly, we were also able to collect 4,417 and 6,646 IS in Pt4 and Pt6 respectively, from FACS-sorted CAR^+ T_{SCM} and T_{CM}/T_{EM} at 14 and 30 days after infusion as well as 1,676 and 1,930 unique IS from the total circulating CAR^+ cells of both individuals respectively at 180 and 720/840 days after treatment (Extended Data 2). It should be noted it was not possible to select subpopulations at these later time points due to the low fraction of CAR^+ cells in circulation. The number of IS collected at long-term follow ups is well above that previously reported at similar time points after CAR T cell treatment[17] and suggest that while clonal diversity decreases somewhat post-infusion, CAR T cells remain highly polyclonal even at late time points post-infusion. We then analyzed the distribution of our IS in the genome of CAR^+ T cells from the infused products and isolated post-infusion at both early and late timepoints. As shown in Fig. 3 (top panels) and Extended Data 3, in the cell products the lentiviral vector integrated into (on average 69% of IS in both patients) or in proximity of a variety of different genes most of which are classical targets of lentiviral integrations in gene therapy[18] with a prevalence of genes related to biological processes such as DNA metabolism and repair (Extended Data 4). Notably, there was no evidence of clonal selection after infusion with respect with the general insertional profile (Fig. 3, bottom panels and Extended Data 3) and of clones with lentiviral genomic insertions in proximity of genes exerting a specific cellular function (Extended Data 4).

Clonal diversity is preserved in long-lasting CAR T cells

We next assessed the diversity of CAR^+ T cells in the CAR T products by means of number and relative contribution of IS collected starting from FACS-sorted T_{SCM}, T_{CM} or T_{EM}. Because the frequency of these populations differed in the cell product, to normalize for this we used equal amounts of genomic DNA from each subpopulation for performing IS analysis. All subpopulations showed high diversity in both patients with the biggest clones ranging from 2.1% to 14.7% of each population (with the exception of one clone with IS contributing to 30.4% of the reads collected from the T_{EM} of Pt6) (Fig. 4 top panels, Fig.5 top panels and Extended Data 5A). Of note, the T_{CM} fraction of both individuals showed a higher number of integrants and higher diversity as compared to the other 2 populations. This suggests that T_{CM} may be more permissive to transduction as compared to T_{SCM} and T_{EM}, possibly as a result of a higher activation/ expansion during *in vitro* transduction. Based on our initial observation on the increasing relative contribution of T_{SCM} in circulation as compared to that infused in the cell product, we investigated the clonal dynamics of this population after infusion in more detail. To do this, we isolated the T_{SCM} from the rest of the memory/effector cell fractions in the early response phase (14 and 30 days post-infusion) when the number of circulating CAR^+ cells in circulation permitted FACS sorting of CAR^+ T cell subfractions. In Fig.4 and Fig.5 (bottom panels), we compared the IS profile of sorted CAR^+ T_{SCM} with the rest of the CAR^+ population at the same timepoints. We observed that in both patients the clonal diversity of T_{SCM} (measured either as Shannon or Gini/Simpson diversity and shown in Fig. 4,5 bottom panels and Extended Data 5D respectively) dropped substantially during the early response phase compared with the infused product. Most of the T_{SCM} isolated from patients early after treatment were in fact composed of individual clones bearing IS whose reads ranged from 89.0% to 96.9% of total read count (Extended Data 5B). It is important to note that neither the number of sorted cells nor the amount of IS collected per samples would alone explain this finding (Extended Data 6), and that the same experimental conditions have applied with different results to the study of small subfractions of progenitor cells in gene therapy patients with different results[19]. This observation is instead likely a reflection of *bona fide* clonal bursts and oligoclonal expansion occurring in T_{SCM} during the early response phase (Fig. 4 and Fig. 5, bottom panels). At the same timepoints, the clonal composition of the remaining CAR^+ cells (T_{CM}/T_{EM}) remained highly polyclonal despite the major shift from T_{CM} to T_{EM} phenotype occurring upon infusion described

above. This would suggest that the CAR^+ T_{CM} and T_{EM} cells infused in these patients have expanded in a more "homeostatic" fashion as compared to their T_{SCM} counterparts.

We then analyzed in both patients the diversity of the total CAR^+ cell composition at later timepoints (180 and 720/840 days after infusion). While the frequency of CAR^+ cells dropped substantially from the initial phase, we observed that the clonal diversity of long-term persisting CAR T cells remained high (Fig. 4 and Fig. 5, bottom panels and Extended Data 5C). This suggests that CAR T cells detected at late time points have persisted through "homeostatic" survival and that long-term maintenance of CAR^+ cells in our patients was not dependent on the expansion of a few large clones carrying non-physiological/aberrant insertional profiles.

To study the clonal dynamics of CAR^+ T cells over time we first compared all datasets from early or late timepoints and performed unsupervised clustering according to IS similarities (Fig. 6A). We found that IS isolated at 14 and 30 days after infusion showed the highest similarities and clustered separately from the clonal datasets of 180 and 720/840 days, suggesting that clones active in the early response phases were less likely to be recaptured at later times after treatment. Because we have multiple capture timepoints we could estimate the lower bound number of clones composing the CAR^+ population by means of mark-recapture statistics as previously described [20]. The results suggest that around 34,000 and 55,000 CAR^+ clones might still be in circulation in Pt4 and Pt6 respectively up to 24/28 months after infusion (Fig. 6B). We then assessed the recapture probability of the expanded T_{SCM} clones observed at early timepoints (Fig. 6C). We found that these clones were different at each follow up and some of them were also detected in the early CAR^+ T_{CM}/T_{EM} population, suggesting that this expansion was transient and accompanied by differentiation of the T_{SCM} to memory and effector cells. In contrast, these large clones do not seem to have contributed substantially to the pool of long-term surviving CAR T cells (Fig. 6C).

Importantly with regard to oncogenic potential of integrated viral vector sequences, we did not observe any major skewing in the distribution of IS towards gene categories related to T cell growth, activation or survival. In this regard, we identified and tracked 12 clones in our patients bearing an integration in the TET2 gene (Fig. 6D), a locus previously associated with clonal dominance in a single patient with CLL treated with CD19 CAR T cells[17]. Of note, none of these clones showed signs of expansion either in the product or

after infusion, contributing to a maximum of 0.044% of sequencing reads out of the total collected in each sample/timepoint (Fig. 6E). Collectively, these data show no evidence of insertional mutagenesis resulting in clonal selection of CAR T cells in these patients either during the early response phase or in long-term persisting CAR T cells.

Lastly, we compared these data with results from Pts10 and 17. The number of unique integration sites retrieved in these individuals were broadly similar to those seen in Pt4 and Pt6. In Pts 10 and 17 we collected 4,633 and 2,823 IS respectively in T_{SCM}, T_{CM} and T_{EM} belonging to the cell products as well as 2,295 and 1,020 IS respectively in T_{SCM} and T_{CM}/T_{EM} during the first month of follow up (Extended Data 7). We first measured the relative abundance of each integration site (Fig. 7A) and analyzed the clonal diversity of the sorted CAR^+ populations from the CAR T cell product and after treatment. As shown in Fig. 7B (top panels), these individuals displayed an overall lower level of clonal diversity in all populations composing the drug product as compared the patients with long surviving CAR^+ T cells and this was particularly evident when analysing the T_{SCM} compartment. After infusion, we could again observe a clonal burst of CAR^+ T_{SCM} during the first month. However, in contrast to the patients with long-term persistence of CAR T cells, the diversity of the T_{CM}/T_{EM} compartments rapidly dropped over the first month in these patients before CAR T cell disappearance (Fig. 7B, bottom panels).

Origin and maintenance of long-term persisting CAR T cells

Having observed that the long-term surviving CAR^+ population was composed of a large and stable number of clones of equivalent sizes, we investigated whether we could establish a link between the IS of the cell product and the ones detected after infusion. Our goal was to understand which of the infused CAR^+ T cell subtypes gave rise to the early expanding and long-term persisting CAR T cells post-infusion. The depth of our analysis and the number of IS collected in the product and from the patients allowed us to detect a substantial number of identical IS in samples collected before and after infusion for the first time. The sharing of IS between the CAR T cell product subsets and samples obtained at different timepoints post-infusion is represented in the correlation maps of Fig. 8A. These plots quantify the IS sharing between all samples analyzed showing in blue dots the paired comparisons where IS sharing was more

significant (positive correlation). In both patients with long term persistence, we observed a relatively high frequency of IS shared among the T subpopulations composing the cell product suggesting that differentiation has occurred during *in vitro* expansion at the time of manufacture. Interestingly, we also found a relatively high proportion of IS shared between the T_{SCM} of the product and the T_{SCM} isolated at 14 and 30 days after infusion in both patients. This would imply that CAR^+ T_{SCM} active during the early CAR T cell response phase are predominantly derived from T_{SCM} in the product that have expanded during *in vitro* manufacture. Notably, T_{SCM} sharing with T_{CM}/T_{EM} up to Day 30 after treatment ranged between 0.3% to 23% percent averaging 8.3% suggesting active differentiation in circulation of T_{SCM} originally contained in the drug product. In contrast, T_{CM} and T_{EM} in the product shared on average a maximum of 2.4% of IS with the same subtypes up to early timepoints after therapy (range 0.09- 7.1), a result in line with the more limited differentiation potential of these memory cell types as compared to T_{SCM}.

Despite the detection of identical IS between the cell product and late timepoints, the IS numbers were not sufficient to reach significance, possibly due to inherent sampling limitations combined with the paucity of CAR^+ T cells still left in circulation in these patients at 6 to 24/28 months after treatment. Nonetheless, the numbers of shared IS allowed us to investigate the relative contribution of the product subsets to the early and late phases after infusion. As shown in Fig. 8B and Extended Figure 8, in all 4 patients, regardless of long-term persistence at early time points the majority of the shared integrants were derived from the T_{SCM} CAR^+ T cells in the product suggesting this compartment has a key role in early expansion. Similarly, in the patients with long-term persistence, the product subpopulation contributing the most to the clonal pool at late time points was the T_{SCM}, from which we collected 46.6% to 60.5% of the clones detected both *in vitro* and after infusion. Such a contribution is particularly relevant when one considers the extremely low frequency of T_{SCM} observed at the time of infusion (1-2% of total CAR^+ T cells) as compared to T_{CM} or T_{EM}.

DISCUSSION

In summary, by combining immunophenotyping and molecular tracking we have gained important novel insights into the nature and dynamics of both early expanding CAR T cells that mediate initial responses as well as long-term surviving CAR T cells providing

immune surveillance against leukaemic relapse in 2 patients up to 3 years after infusion. The IS methodology has been widely used by our group and others as a reliable method for tracking clones showing added value compared to high throughput TCR sequencing. This is because a) it allows a specific tracking of engineered T-cell clones even when they are highly diluted in a population of vector negative cells and therefore in conditions that prevent FACS sorting of CAR+ cells (such as at late timepoints) and b) differently from high throughput TCR sequencing, it provides a set of stable genomic markers whose relative contribution is independent from the effect of thymic involution, immune tolerance or extra thymic rearrangements. Nonetheless, it should be noted that it is currently impossible to precisely discriminate individual clones based on the assumption 1 IS = 1 clone, since each genetically engineered cell may carry multiple copies of integrated vector so that interpretation must be based on data trends rather than on the tracking of specific individual IS or clones.

We show that CAR T cells in the infused product display a polyclonal integration site profile and that CAR T cells remain clonally diverse at early and late time points post-infusion with no evidence of clonal dominance. Individual clones showed different patterns of expansion and contraction with time, with different clones responsible for early expansion and long-term persistence. Interestingly, despite having observed several IS in the TET2 locus, we could not find any sign of clonal expansion of the relevant CAR T cell clones. It should be noted that the TET2 insertional mutagenesis event described by Fraietta *et al*[17] was accompanied by a hypomorphic mutation in the second TET2 allele, an extremely rare occurrence which is unlikely to have taken place in the patients analysed in the present study. It would seem therefore that bi-allelic disruption of the TET2 locus is needed for clonal expansion.

Our results, summarized in the scheme of Fig. 8C, suggest that following infusion, CAR^+ T cells undergo a rapid switch from a predominantly T_{CM} to T_{EM} phenotype during the early response phases without major clonal imbalance in the memory compartment but with a self-limiting clonal burst phase occurring in the T_{SCM} population. T_{SCM} from the infused product can potentially differentiate into other subsets post-infusion. Analysis of shared integrants between the CAR T cell product and after treatment at early (< 1 month) time points demonstrated that despite the low frequency of T_{SCM} in the products (0.5-2%), a substantial fraction of these clones (10-73%) were responsible for the circulating CAR T cells detected during this early expansion phase in all patients regardless of long-term

persistence, implying that differentiation of the oligoclonally expanded T_{SCM} in to T_{EM} may be responsible for much of the early anti-leukemic response. The patients with early loss of CAR T cells appeared to have a more differentiated phenotype with a lower proportion of T_{SCM} and higher T_{EM} as well as reduced clonal diversity in their products. Patients with poor CAR T cell persistence showed markedly reduced clonal diversity in CAR^+ T_{CM} and T_{EM} compared to those with long term persistence. We speculate that when the drug product is more differentiated and less polyclonal, CAR T cells may not be able to preserve a clonal repertoire with enough diversity to avoid exhaustion during the clonal burst associated with the early anti-leukemic response. Further studies in larger patient cohorts are needed but if these findings are confirmed these data have important clinical implications, suggesting that assessment of the phenotype and clonal diversity of the product and at early time points may identify patients at high risk of early CAR T cell loss who may benefit from consolidative transplant.

In the patients with long term CAR T cell persistence, CAR T cells did not terminally differentiate into T_{EMRA} cells upon exposure to CD19 antigen but instead the T_{SCM} phenotype became more prevalent over time concomitantly with long-term maintenance of clonal diversity. We can therefore speculate that long-term maintenance of CAR^+ T cells is supported by self-renewing cells endowed with homeostatic proliferative capacity like T_{SCM}. Critically, we could estimate that clones belonging to the small fraction of T_{SCM} present in the cell product at the time of infusion had the highest recapture probability within the pool of long-lived CAR^+ T cells in circulation, proving for the first time the crucial role of T_{SCM} in long-term anti-leukemic surveillance by CAR T cells.

Our working hypothesis is that the remarkable early expansion and persistence of our CAR T cell product achieved in Pts 4 and 6 reflects a combination of 2 key factors. Firstly, our previous preclinical data[16] both *in vitro* and in a xenogeneic mouse model suggested that signaling through our low affinity CAT CAR results in enhanced CAR T cell expansion which may in part be mediated through higher expression of IL-7R and Bcl-2, promoting homeostatic proliferation and preventing apoptosis. Moreover, the faster off rate with this CAR could potentially have reduced differentiation of CAR T cells during the early anti-leukemic response, preventing terminal differentiation and allowing self-renewal of CAR^+ T_{SCM} capable of long-term survival. Our CAR construct incorporated a 4-1BB costimulatory domain, which may confer prolonged persistence compared with CD28 by preventing exhaustion through tonic

signaling[21,22]. Secondly, our manufacturing protocol was relatively short (7-8 days) and did not utilise exogenous cytokines in the medium which may favor less *in vitro* differentiation and preservation of high levels of T_{CM} cells whose expansion potential could have contributed to an enhanced early response. These culture conditions allowed a small fraction of T_{SCM} to be infused in those patients in large enough doses to survive the initial clonal burst phases and to maintain a pool of CAR^+ cells long term.

There are clearly some unavoidable technical constraints to the data reported in this study, including the limited number of patients in whom material was available, the stability of the CD62L marker upon cell culture that could introduce a bias in the estimation of T_{SCM}/T_{CM} in the drug product, the low frequency of circulating CAR^+ T cells at late time points, the potential sampling biases and the surrogate nature of clonal size quantification based on sequencing abundance of each IS, as a consequence of uneven amplifications of vector-genome junctions[23].

Despite these limitations, our data provide evidence for the first time of the critical role of T_{SCM} in both mediating early anti-leukemic responses and long-term persistence of CAR T cells after infusion. Future efforts will focus on validating these findings in larger patient cohorts and improvement of CAR design and manufacturing methodology to preserve T_{SCM} and T_{CM} in the infused cell product.

REFERENCES

1. Sallusto, F., Lenig, D., Förster, R., Lipp, M. & Lanzavecchia, A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* (1999) doi:10.1038/35005534.
2. Gattinoni, L. *et al.* A human memory T cell subset with stem cell-like properties. *Nat. Med.* **17**, 1290–1297 (2011).
3. Heslop, H. E. *et al.* Long-term outcome of EBV-specific T-cell infusions to prevent or treat EBV-related lymphoproliferative disease in transplant recipients. *Blood* (2010) doi:10.1182/blood-2009-08-239186.
4. Berger, C. *et al.* Adoptive transfer of effector CD8+ T cells derived from central memory cells establishes persistent T cell memory in primates. *J. Clin. Invest.* (2008) doi:10.1172/JCI32103.
5. Biasco, L. *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.* **7**, 273ra13 (2015).
6. Oliveira, G. *et al.* Tracking genetically engineered lymphocytes long-term reveals the dynamics of t cell immunological memory. *Sci. Transl. Med.* **7**, 1–14 (2015).
7. Maude, S. L. *et al.* Chimeric antigen receptor T cells for sustained remissions in leukemia. *N. Engl. J. Med.* (2014) doi:10.1056/NEJMoa1407222.
8. Lee, D. W. *et al.* T cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukaemia in children and young adults: A phase 1 dose-escalation trial. *Lancet* (2015) doi:10.1016/S0140-6736(14)61403-3.
9. Gardner, R. A. *et al.* Intent-to-treat leukemia remission by CD19 CAR T cells of defined formulation and dose in children and young adults. *Blood* (2017) doi:10.1182/blood-2017-02-769208.
10. Maude, S. L. *et al.* Tisagenlecleucel in children and young adults with B-cell lymphoblastic leukemia. *N. Engl. J. Med.* (2018) doi:10.1056/NEJMoa1709866.
11. Mueller, K. T. *et al.* Cellular kinetics of CTL019 in relapsed/refractory B-cell acute lymphoblastic leukemia and chronic lymphocytic leukemia. *Blood* (2017) doi:10.1182/blood-2017-06-786129.

12. Sommermeyer, D. *et al.* Chimeric antigen receptor-modified T cells derived from defined CD8+ and CD4+ subsets confer superior antitumor reactivity in vivo. *Leukemia* (2016) doi:10.1038/leu.2015.247.
13. Turtle, C. J. *et al.* CD19 CAR-T cells of defined CD4+:CD8+ composition in adult B cell ALL patients. *J. Clin. Invest.* (2016) doi:10.1172/JCI85309.
14. Xu, Y. *et al.* Closely related T-memory stem cells correlate with in vivo expansion of CAR.CD19-T cells and are preserved by IL-7 and IL-15. *Blood* (2014) doi:10.1182/blood-2014-01-552174.
15. Sheih, A. *et al.* Clonal kinetics and single-cell transcriptional profiling of CAR-T cells in patients undergoing CD19 CAR-T immunotherapy. *Nat. Commun.* **11**, 219 (2020).
16. Ghorashian, S. *et al.* Enhanced CAR T cell expansion and prolonged persistence in pediatric patients with ALL treated with a low-affinity CD19 CAR. *Nat. Med.* (2019) doi:10.1038/s41591-019-0549-5.
17. Fraietta, J. A. *et al.* Disruption of TET2 promotes the therapeutic efficacy of CD19-targeted T cells. *Nature* (2018) doi:10.1038/s41586-018-0178-z.
18. Aiuti, A. *et al.* Lentiviral hematopoietic stem cell gene therapy in patients with wiskott-aldrich syndrome. *Science (80-.).* **341**, (2013).
19. Scala, S. *et al.* Dynamics of genetically engineered hematopoietic stem and progenitor cells after autologous transplantation in humans. *Nat. Med.* **24**, 1683–1690 (2018).
20. Biasco, L. *et al.* In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases. *Cell Stem Cell* **19**, 107–119 (2016).
21. Milone, M. C. *et al.* Chimeric receptors containing CD137 signal transduction domains mediate enhanced survival of T cells and increased antileukemic efficacy in vivo. *Mol. Ther.* (2009) doi:10.1038/mt.2009.83.
22. Long, A. H. *et al.* 4-1BB costimulation ameliorates T cell exhaustion induced by tonic signaling of chimeric antigen receptors. *Nat. Med.* (2015) doi:10.1038/nm.3838.

23. Biasco, L. Integration Site Analysis in Gene Therapy Patients: Expectations and Reality. *Hum. Gene Ther.* **28**, 1122–1129 (2017).

ACKNOWLEDGEMENTS

This work was funded the JP Moulton Foundation and supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London. P.J.A. is a recipient of an NIHR Research Professorship which also supported SG. MP is supported by the National Institute of Health Research University College London Hospital Biomedical Research Centre. A.T. is a recipient of a Wellcome Trust Senior Fellowship. The work of L.B. was supported by The Wellcome Trust (104807/Z/14/Z, Principal Research Fellowship awarded to A.J Thrasher), and the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK. L.B is also currently Director of R&D for the gene therapy company AVROBIO located in Cambridge, MA, US (none of what presented in this work is supported by or relates to AVROBIO). The work of L.B. was also in part performed utilizing the resources of the Gene Therapy Program of the Dana Farber/Boston Children's Cancer and Blood Disorders Center.

AUTHORS CONTRIBUTIONS STATEMENT

L.B. designed, co-supervised the study, performed computational analyses and wrote the manuscript. N.I. and C.R. performed cells isolation and molecular insertion sites retrieval. S.G. and R.R. collected and provided clinical sample material for analysis. A.G. cryopreserved study samples and performed flow cytometric staining. R.H. and R.W. were Principal Investigators for the clinical study. B.P. wrote study documentation and provided trial management. A.L. provided statistical analysis for the study. M.P. generated the CAR construct and participated in its preclinical characterisation. A.J.T. conceived the idea and participated in the experimental design and data analysis. P.J.A. supervised the study as PI and wrote the manuscript.

COMPETING INTERESTS STATEMENT

No competing interests to declare.

FIGURE LEGENDS

Figure 1 Immunophenotypic characterization of patients with long-lasting CAR$^+$ T cells in the product and after infusion overtime. **A)** Schematic summarizing the manufacture of CAR T cells and key aspects of CARPALL clinical study design. The days reported in this schematic are reflective of the days of *in vitro* expansion since peripheral blood mononuclear cells (PBMC) isolation. **B)** Schematic representation of survival and expansion potential of the main T cell subpopulations (TN = naïve T cells, TSCM = T memory stem cells, TCM = T central memory cells, TEM = T effector memory cells, TEMRA = T effector memory CD45RA$^+$ cells); **C)** Gating scheme used for the identification of T cell subtypes used for following up CAR$^+$ T cells and total CD3$^+$ cells. **D)** FACS plot showing CAR$^+$ cells within CD3$^+$ cells as well as T cell composition inside the CAR$^+$CD3$^+$ CD45RA$^+$CD62L$^+$ population for Pt4 (top panels) and Pt6 (bottom panels). Percentages of CAR$^+$ cells and CD62L/CD45RA compartments are showed inside each gate. Timepoints of analysis are shown on top of each set of panels (d = days after CAR$^+$ cells infusion). **E)** Percentage of CAR$^+$ cells in the CD3$^+$ cell compartment overtime in Pt4 (top panel) and Pt6 (bottom panel). **F)** Summary of data shown in panels C for CAR$^+$ T cell composition overtime for Pt4 (left panel) and Pt6 (right panel). Pie charts show composition of CAR$^+$ T cells at each timepoint. Longitudinal contribution of each subpopulations to the CAR$^+$ cell compartment is shown in the relative plot below.

Figure 2 Immunophenotypic characterization of patients with short-living CAR$^+$ T cells in the product and after infusion overtime. **A)** FACS plot showing CAR$^+$ cells within CD3$^+$ cells as well as T cell composition inside the CAR$^+$CD3$^+$ CD45RA$^+$CD62L$^+$ population for Pt10 (top panels) and Pt17 (bottom panels). Percentages of CAR$^+$ cells and CD62L/CD45RA compartments are showed inside each gate. Timepoints of analysis are shown on top of each set of panels (d = days after CAR$^+$ cells infusion). **B)** Pie charts showing the composition of the drug product in Pt10 and Pt17. **C)** Longitudinal contribution of each subpopulations to the CAR$^+$ cell compartment in Pt10 and Pt17 vs Pt4 and Pt6. **D)** Bar plot showing the relative fraction of T_{SCM} detected in the drug product in Pt4, Pt6 vs Pt10 and Pt17.

Figure 3 Distribution and abundance of IS collected in the product and after infusion overtime in Pt4 and Pt6. Circos plot showing the genomic distribution of IS in the product (top panels) and after infusion (bottom panels) in Pt4 (left panels) and Pt6 (right panels). Description of each concentric data track and timepoint (d = days after CAR^+ cells infusion) is shown in the legends on the left of the circus plots. Description of the content of these plots is reported in the grey legend on the top right. Worclouds showing top hit genes in the product and after infusion are shown at the center of each plot (the bigger the gene name the higher the number of IS detected in its locus).

Figure 4 Diversity of IS in different T cell subtypes overtime from Pt4. Top panels: for each product subpopulation analysed coloured circles (T_{SCM} = orange, T_{CM} = light blue, T_{EM} = blue) contain bubble plots of clones contributing >0.01% to the total population. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside the bubble. The plots on the right show Shannon Diversity Index for each subpopulation. Bottom panels: plots showing Shannon Diversity Index (left y-axis) of IS overtime in T_{SCM} (orange lines) and in T_{CM}/T_{EM} (dark blue) and in all T cells (green lines). The grey lines show the percentage of CAR^+ cells overtime (right y-axis) For each timepoint analyzed colored circles contain bubble plots of clones contributing >0.01% to the total population at each timepoint. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside each bubble.

Figure 5 Diversity of IS in different T cell subtypes overtime from Pt6. Top panels: for each product subpopulation analysed coloured circles (T_{SCM} = orange, T_{CM} = light blue, T_{EM} = blue) contain bubble plots of clones contributing >0.01% to the total population. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside the bubble. The plots on the right show Shannon Diversity Index for each subpopulation. Bottom panels: plots showing Shannon Diversity Index (left y-axis) of IS overtime in T_{SCM} (orange lines) and in T_{CM}/T_{EM} (dark blue) and in all T cells (green lines). The grey lines show

the percentage of CAR^+ cells overtime (right y-axis) For each timepoint analyzed colored circles contain bubble plots of clones contributing >0.01% to the total population at each timepoint. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside each bubble. Analysis performed as in Figure 4.

Figure 6 Tracking of IS in different T cell subtypes overtime in Pt4 and Pt6. A) Unsupervised clustering based on IS sharing among subpopulations overtime in Pt4 (left panel) and Pt6 (right panel). **B)** Estimation of the number of clones composing the CAR^+ cell population in each patient based on the Mth Chao (LB) model applied to IS recapture probability. Shown are individual values and relative standard errors from the abundance model fit estimation (for Pt4 abundance=34,083 and standard error = 1,699, for Pt6 abundance = 55,888 and standard error = 2,795). **C)** Heatmaps showing tracking and relative contribution of T_{SCM} clones (IS, rows) at d14 and d30 recaptured in other T subtypes and timepoints (columns) in Pt4 (top panel) and Pt6 (bottom panel). Intensity of blue is proportional to relative clone size (IS abundance). **D)** Distribution of IS in proximity of TET2 in both patients combined. The x-axis displays chromosomal coordinates (bp) while the y-axis the IS abundance (log10 of percentage) calculated within the population and timepoint where it has been observed. The dark blue bars show IS detected in the patients while red bars IS detected in the cell product. The transcription start site (TSS), exons (grey boxes) and introns (grey lines) of the TET2 gene are shown on top of the plot in their respective chromosomal localization. **E)** Heatmap showing tracking and relative contribution of the IS shown in panel D for each patient in the product (in red) or after infusion (in blue). Relative IS abundance (percentage) calculated within the population and timepoint where it has been observed is reported inside each colored field (white field = not detected).

Figure 7 Diversity of IS in different T cell subtypes overtime from Pt10 and Pt17. A) Scattered dot plots showing relative abundance of IS in the product (top section) and at early timepoints (bottom section) after treatment in Pt10 (plots on the top) and Pt17 (plots on the bottom) (d = days after treatment, m = months after treatment). Mean percent abundance is shown as a dotted line for each sample. Number of events in each dataset is equal to what reported in Extended Data 7. **B)** Diversity of drug product and CAR+ cells after

infusion in Pt10 (top panels) and Pt17 (bottom panels). Top section: for each product subpopulation analysed coloured circles (T_{SCM} = orange, T_{CM} = light blue, T_{EM} = blue) contain bubble plots of clones contributing >0.01% to the total population. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside the bubble. The plots on the right show Shannon Diversity Index for each subpopulation. Bottom section: plot showing Shannon Diversity Index (left y-axis) of IS overtime in T_{SCM} (orange lines) and in T_{CM}/T_{EM} (dark blue). The grey lines show the percentage of CAR^+ cells overtime (right y-axis) For each timepoint analyzed colored circles contain bubble plots of clones contributing >0.01% to the total population at each timepoint. Dimension of the bubble is proportional to the size of the clone. The name of the gene closest to the relative IS is reported inside each bubble.

Figure 8 Comparison of IS distribution pre and post infusion. A) Correlograms showing upper triangle of correlation matrix among IS datasets from Pt4 (left panel) and Pt6 (right panel). Positive correlations are shown in blue while negative in red. Size of circles is proportional to the correlation value. **B)** Ring plots showing relative contribution from each subtype (colored section of the ring) to the pool of IS detected in the product and at early (left plots) or late (right plots) timepoints in Pt4 (top panels) and Pt6 (bottom panel). The total number of IS captured both in the product and after infusion is shown inside each plot. The relative percentage of IS belonging to each T cell subtype of the product that were shared with samples after infusion is shown in white inside each section of each ring plot. **C)** Model summarizing our working hypothesis for CAR^+ T cell dynamics occurring in the patients of this study.

METHODS

Patients

The CARPALL study (NCT02443831) was approved by the UK Medicines and Healthcare Products Regulatory Agency (clinical trial authorization no. 20363/0361/001) and ethical approval (including the research in this manuscript) was obtained from the London West London & GTAC Research Ethics Committee (REC ref no. 16/LO/0283). Written informed consent was obtained from patients or their parents/guardians prior to study entry. All patients had relapsed CD19+ ALL and 3 had relapsed post-BMT. Patients 4 and 6 had long term persistence of CAR T cells associated with ongoing durable complete remissions whereas patients 10 and 17 had early loss of CAR T cells. Patient 4 is a 21-year-old man who had a combined bone marrow and CNS relapse 1 year post matched unrelated donor peripheral blood stem cell transplant and was treated with 10^6 CAR^+ T cells/kg in 3rd CR (molecular MRD 2.6 x 10^{-5}) on 23/11/16. He is currently well in molecular CR with detectable circulating CAR T cells 3.5 years post infusion. Patient 6 is an 11-year-old boy who had a combined bone marrow and CNS relapse 14 months after a matched unrelated donor BMT for ALL and relapsed again in the CSF with molecular disease in the bone marrow 7 months after 2nd BMT. He was treated with 10^6 CAR^+ T cells/kg (molecular MRD 5 x 10^{-4}) on 5/10/16 and remains well in molecular CR with detectable circulating CAR T cells 3.5 years post infusion. Patient 10 was a 10-year-old boy with an isolated bone marrow relapse 17 months after matched family donor BMT who was treated with 10^6 CAR^+ T cells/kg in 3rd CR (molecular MRD 3 x 10^{-5}) on 30/8/17 with molecular CR and excellent initial CAR T cell expansion at 1 month but CAR T cells were undetectable by flow and qPCR from 2 months post-infusion. He relapsed with CD19 disease 7 months post CAR T cell therapy and subsequently died of a fungal infection. Patient 17 is a 15-year-old boy with late isolated CNS relapse who relapsed again in the CNS despite cranial irradiation and was treated with 10^6 CAR^+ T cells/kg in 3rd CR (molecular MRD 5 x 10^{-5} negative) on 25/7/18 with continuing molecular CR and excellent initial CAR T cell expansion at 1 month but CAR T cells were undetectable by flow and qPCR from 4 months. He was consolidated with a mismatched unrelated donor BMT and remains in CR 2 years post-transplant. The average Viral Copy Number/cell for the infused CAR T cell product for patients 4, 6, 10 and 17 was 8, 3.8, 4.9 and 5.6 respectively.

Characterisation and isolation of lymphoid and myeloid cells

We performed immunophenotyping on whole blood EDTA samples while cell sorting was performed on PBMCs isolated from the whole blood by density gradient centrifugation using LymphoPrep (Sigma) after dextran sedimentation. For sorting PBMCs were thawed in RPMI (10% FCS) pre-warmed with DNase 200 u/ml (also used TexMACS 5% human AB serum). Cells were then washed, counted anempty trash frod stained on ice in 100 ul PBS for every 5x10^6 cells. Then Fc block (BD 564220, 5ul) was applied for 10 mins and without washing, anti-idiotype (2.5ul or 3.5ug) was added for 20 mins. Cells (5x10^6 cells in 100ul) were then washed and stained for 30 mins with secondary antibody mix containing: CD19 CAT19 B12RIgG2a Evitria followed by secondary staining with Anti rat IgG PE for CAR detection (Biolegend 405406, 0.5ul), CD3 APC-Cy7 (Biolegend 300426, 5ul), 7-AAD (BD 555816, 5ul), CD95 BV711 (BD 563132, 5ul), CD45RA v450 (BD 8053598, 2ul) and CD62L APC (Biolegend 304810, 3ul). Cells were washed, resuspended in sort buffer (PBS (1%FBS) + 1mM EDTA), passed through a cell strainer (30-40uM) and then sorted using FASCAriaIII cells sorter. Raw FACS data was collected using DIVA software (BD Biosciences) and analysed with FlowJo (TreeStar). When feasible, an aliquot of the sorted cells was re-run through the cell sorter to check fraction purity.

DNA extraction and genome amplification

We performed DNA extraction using QIAamp DNA Blood Mini Kit (Qiagen) for samples of 10,000 to 5 x 10^6 cells and MN NucleoSpin® Tissue XS kit (manufacturer MN) for up to 10,000 cells, according to manufacturers' instructions. Genomic DNA was quantified using Nanodrop One Spectrophotometer (Thermofisher). For integration site analysis, a whole-genome amplification (WGA) was performed if the DNA yield after extraction was lower than 300 ng using the Qiagen Repli-G Mini Kit (Manufacturer), according to manufacturer's instruction.

IS collection and library preparation

For vector integration site analysis, we used LAM-PCR combined with high-throughput sequencing as previously described [5,20]. Briefly, two rounds of linear PCR were performed (50 cycles each) to enrich for vector genome junctions using biotinylated primers specific for

vector LTRs. Streptavidin-coupled magnetic beads (Invitrogen Dynabeads Kilobase Binder Kit) were added to each sample, to capture linearly amplified fragments, followed by complementary strand synthesis. Samples then underwent restriction enzyme digest using 3 different enzymes - MluCI, AciI or HpyCH4IV, to minimize bias and improve genome coverage. Linker cassettes were ligated to these fragments, then samples underwent 2 rounds of exponential amplification PCR to amplify fragments containing vector LTR and linker cassette sequences. PCR fragment size depends on the distance between the known vector sequence and the closest enzyme recognition site. To visualize amplified fragments, we used high-resolution gel electrophoresis (ElchromScientific). Final PCR products were purified with QIAQUICK PCR purification kit (QIAGEN). Fusion PCR was used to add customized sequence-specific Illumina adaptors to the final PCR products. Different combinations of sample barcodes included in both LTR adaptors and LC adaptors were used to differentiate between different samples.

Computational analyses of IS data

IS identification and analysis were performed through a custom analytical pipeline extensively described in previous publications[5,18–20]. Briefly, raw integration sites data sets underwent series of different bioinformatics filtering procedures according to the type of analysis to be performed. All data sets were processed with a "collision detection filter" to univocally assign each IS to a patient and to one or more T cell subpopulations by applying a 10-fold rule for contamination identification as previously reported. A final matrix M was generated where each row r represented an individual integration site while each column c an individual cell type/sample and timepoint. Each entry of M contained the abundance of each r for each c in terms of sequencing reads. The data shown in Fig. 3-7 and Extended Data 2-8 were generated on the basis of the IS databases attached as Source data. Graphical representations of IS analyses were generated with Prism8 (GraphPad Software) unless otherwise specified below. Panels of Fig. 3 were generated using the Circos software (http://circos.ca). Word clouds of Fig. 3 and Extended Data 4 were generated on the basis the relative incidence of the single closest gene to each IS and plotted using the online suite *WordClouds* (*https://www.wordclouds.com*). Panels with lines of Fig. 4,5,7B and Extended Data 1 were created plotting IS diversity overtime calculated as Shannon Diversity or Gini-Simpson Index through the R

package *BiodiversityR (https://cran.r-project.org/web/packages/BiodiversityR/index.html)*. The "bubble" plots of Fig. 4,5,7B were created on the basis of the IS with abundance >0.01% relative to each subpopulation and timepoint using the R package *packcircles* (*https://cran.r-project.org/web/packages/packcircles/index.html*). The dendrograms of Fig. 6A were generated based on the similarity among IS datasets through the *hclust* function (*https://www.rdocumentation.org/packages/fastcluster/versions/1.1.25/topics/hclust*). Estimation of clonal abundance and standard errors shown in Fig. 6B was calculated by the conversion of *M* to a *M(0,1)* matrix of incidence and by the application to *M(0,1)* of log-linear models for closed populations through the R package *Rcapture (https://cran.r-project.org/web/packages/Rcapture/index.html, function = closedp.t)*. The Mth Chao (LB) method was selected for visualization in Fig. 6B being the most conservative estimation among the ones with the lowest BIC (Bayesian Information content). The heatmaps of Fig. 6C,E were generated through the use of the R package *gplots* (*https://cran.r-project.org/web/packages/gplots/index.html*). The correlograms of Fig. 8A were generated using the R package *corrplot (https://cran.r-project.org/web/packages/corrplot/corrplot.pdf)*. Correlation values for both plots were generated through the function *Mcorr*. The IS distribution around TSS, gene content of IS loci shown in Extended Data 3 and the Gene Ontology analysis of hit genes shown in Extended Data 3 (right panels) were generated through the online suite *GREAT* (*Genomic Regions Enrichment of Annotations Tool*) (*http://great.stanford.edu/public/html/)* using BED data from IS coordinates, associating genomic regions with the rule of single nearest gene within 1000Kb to each IS and applying the gene annotations to these regions.

Code availability

No new code was generated or used in this manuscript.

Data availability

The fastq files relative to sequencing of IS amplicons generated for this study are available through the NCBI repository (https://submit.ncbi.nlm.nih.gov/) BioProject ID PRJNA718947 and accession number generation is underway. Source data for

Fig.1,2,4-8 and Extended Data Fig. 1,2,5-7 have been provided as Source Data files. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Statistics & Reproducibility

No statistical method was used to predetermine sample size. The experiments were not randomized, and no data were excluded from the analyses. Further information on research design is available in the Nature Research Reporting Summary linked to this article.