nature medicine

Article

https://doi.org/10.1038/s41591-023-02415-3

Transcriptional signatures associated with persisting CD19 CAR-T cells in children with leukemia

Received: 22 December 2022

Accepted: 23 May 2023

Published online: 6 July 2023

Check for updates

Nathaniel D. Anderson[1], Jack Birch[2,13], Theo Accogli[2,13], Ignacio Criado[2,13], Eleonora Khabirova[1], Conor Parks[1], Yvette Wood[1], Matthew D. Young[1], Tarryn Porter[1], Rachel Richardson[2,3], Sarah J. Albon[2,3], Bilyana Popova[4], Andre Lopes[4], Robert Wynn[5], Rachael Hough[6], Satyen H. Gohil[7,8], Martin Pule[8], Persis J. Amrolia[3,9], Sam Behjati[1,10,11,14] ✉ & Sara Ghorashian[2,12,14] ✉

In the context of relapsed and refractory childhood pre-B cell acute lymphoblastic leukemia (R/R B-ALL), CD19-targeting chimeric antigen receptor (CAR)-T cells often induce durable remissions, which requires the persistence of CAR-T cells. In this study, we systematically analyzed CD19 CAR-T cells of 10 children with R/R B-ALL enrolled in the CARPALL trial via high-throughput single-cell gene expression and T cell receptor sequencing of infusion products and serial blood and bone marrow samples up to 5 years after infusion. We show that long-lived CAR-T cells developed a CD4/CD8 double-negative phenotype with an exhausted-like memory state and distinct transcriptional signature. This persistence signature was dominant among circulating CAR-T cells in all children with a long-lived treatment response for which sequencing data were sufficient (4/4, 100%). The signature was also present across T cell subsets and clonotypes, indicating that persisting CAR-T cells converge transcriptionally. This persistence signature was also detected in two adult patients with chronic lymphocytic leukemia with decade-long remissions who received a different CD19 CAR-T cell product. Examination of single T cell transcriptomes from a wide range of healthy and diseased tissues across children and adults indicated that the persistence signature may be specific to long-lived CAR-T cells. These findings raise the possibility that a universal transcriptional signature of clinically effective, persistent CD19 CAR-T cells exists.

B-lineage acute lymphoblastic leukemia (B-ALL) is the most common type of childhood cancer and mostly derives from immature B cells that carry the cell surface antigen CD19 (ref. 1). Most children with B-ALL can be cured through first-line treatment comprising combinations of cytotoxic agents. However, relapsed ALL remains a leading cause of childhood death despite intensive cytotoxic chemotherapy often including allogeneic bone marrow transplantation. The advent of CD19 chimeric antigen receptor (CAR)-T cell therapy in recent years has transformed the treatment of intractable ALL[2]. Although a subset of children can be cured, up to 60% of children experience further, typically fatal, disease recurrence due to non-persistence of CAR-T cells or CD19$^-$ leukemic escape[3,4].

A full list of affiliations appears at the end of the paper. ✉e-mail: sb31@sanger.ac.uk; s.ghorashian@ucl.ac.uk

Previously, we generated a novel low-affinity CAR incorporating a CD19-specific single-chain variable fragment (scFv) called CAT, displaying a faster off-rate of interaction than the FMC63 CD19 binder used in prior clinical studies[3]. CAT CAR-T cells showed greater cytotoxicity and proliferative responses in vitro and maintained long-lived molecular remissions in children with relapsed or refractory ALL, as demonstrated in the CARPALL study[3]. The molecular features underpinning CAR-T cell persistence in our study remain unknown. We reasoned that single-cell transcriptomic assays may help elucidate these features. To date, other CAR-T cel products in patients have been studied at the resolution of single cells[5,6]. However, the persistence of CAR T-cellsCAR-T cells in these studies was generally limited to 3 months. An exception was long-lived CAR-T cells in two adult individuals with a different cancer—chronic lymphocytic leukemia (CLL)—in whom anti-CD19 CAR-T cells have persisted for almost a decade thus far[7]. It is unclear whether one can generalize from two adult patients treated for CLL to other hematological malignancies and patient groups, in particular to childhood ALL, or to other CAR-T cell products.

We systematically studied molecular features and clonal dynamics of CAR-T cells in children enrolled in the CARPALL study at serial timepoints, from production to persistence, up to 5 years after infusion.

Results

Overview of study cohort and experiment

We studied 15 consecutive patients with high-risk or relapsed CD19$^+$ B-ALL treated with CD19 CAR-T cell therapy on the CARPALL study (NCT02443831) and in whom adequate CAR-T cells could be isolated for subsequent analyses from cryopreserved samples of blood or bone marrow. Outcomes of the first 14 patients infused were reported[3]; subsequently, a further 18 patients have been treated. Thirteen of 15 (87%) patients studied achieved complete remission; six of these responding patients subsequently relapsed, whereas the other seven achieved long-lived remissions maintained by detectable CAR-T cells and concomitant B cell aplasia (Fig. 1a). We performed detailed phenotyping by flow cytometry in 11 patients, and, in ten children, sufficient CAR-T cells were obtained for further interrogation by single-cell mRNA and T cell receptor (TCR) sequencing (73 patient samples split into 89 gene expression (GEX)/TCR and 64 flow samples; Supplementary Table 1 and Extended Data Fig. 1). Samples were taken from the CAR-T cell product as well as from patients at early (months 1–3), mid (months 4–6) and late (month 7 onwards) timepoints. Early timepoints were defined as 1–3 months, as all patients who achieved molecular complete remission with the absence of measurable residual disease did so within this window. The late timepoints were selected based on the timing of CD19$^-$ relapses, which were generally early events and would have occurred by month 7. The mid timepoints were the interval remaining between early and late. Two patients had samples from the product and at all timepoints (early, mid and late); four patients had all post-infusion timepoints represented; two patients had samples at two of three post-infusion timepoints; and two patients had only the early timepoint interrogated due to early relapse. We isolated CAR-T cells from peripheral blood or bone marrow by flow cytometry using CD3 and CAR expression, before single-cell sequencing (Chromium 10x platform) (Fig. 1b).

Double-negative CAR-T cells delineate late timepoints

In total, we recovered 264,827 single cells that passed quality control, approximately 50,000 of which were CAR-T cells (Supplementary Table 2). We grouped all 264,827 cells using commonly deployed analytical methods and visualized resultant clusters using uniform manifold approximation and projection (UMAP) (Fig. 1c and Extended Data Fig. 2). Clustering segregated CAR-T cells from non-CAR-T cells, with contributions from all patients. Two clusters were completely patient specific; these clusters represented ALL cancer cells from two children, patients P13 and P08, with CD19$^-$ relapses at the time of sampling. Subclustering of CAR-T cells segregated cells transcriptionally into infusion products at month 0 (M0), followed by early (M1–3), mid (M4–M6) and late (M7–M60) timepoints after infusion (Fig. 2a). Cycling cells congregated together from all timepoints, indicating that CAR-T cells remain proliferative several years after infusion. Using a marker-based annotation, CD8$^+$ T cells were the predominant CAR-T cell at all timepoints in most cases, apart from late timepoints where CAR-T cells lacked expression of both CD4 and CD8A transcripts (Fig. 2b, Extended Data Fig. 3 and Supplementary Table 3). CD4 CAR-T cells made minor contributions at this point. Thus, late or persisting CAR-T populations were predominantly double-negative T cells.

Validation of double-negative CAR-T cell state

To confirm the early predominance of a CD8$^+$ subset and the later emergence of a double-negative population, we implemented two orthogonal approaches. First, we analyzed peripheral blood (PB) and bone marrow (BM) samples from seven CARPALL patients collected at late timepoints by flow cytometry (7–72 months after infusion; Supplementary Table 1). We identified CAR-T cells using CD3 expression and use of an anti-idiotype antibody specific for the CAR and assessed expression of CD4 and CD8 on CAR-T cells (Fig. 2c–f and Extended Data Fig. 4a). This analysis confirmed that most cells were double negative at these late timepoints with a smaller contribution from CD8 T cells. This contrasted with the lower proportion of double-negative T cells in the non-CAR-T cell compartment in these patients (Supplementary Table 4). Furthermore, CAR-T cells were also characterised by lack of expression of CD45RA and CCR7, suggesting an effector memory phenotype (Fig. 2g and Extended Data Fig. 4b).

In our second approach, we pursued a cell-marker-independent analysis to assign cell identity to CAR-T cells. We directly compared CAR-T cell transcriptomes to a multi-modal, single-cell atlas of the circulating human immune system[8]. This reference is based on 211,000 human blood mononuclear cells interrogated by single-cell mRNA sequencing and by 228 anti-surface protein antibodies (CITE-seq). Consistent with our initial annotation and protein validation by flow cytometry, we observed that most late-persisting CAR-T cells were classified as double-negative cells, whereas CAR-T cells from earlier timepoints were mainly CD8$^+$ T cells (Fig. 3a,b). The exceptions were patients P09 and P06 in whom an appreciable quantity of early CAR-T cells were double-negative γδ T cells with high expression of *NKG7* and *GNLY* (Extended Data Fig. 5). This is consistent with a previous report that γδ T lymphocytes harbor similarities to CD8 T cells and natural killer (NK) cells[9]. Together, our initial observation with validation by two approaches demonstrates that most persisting CAR-T cells represented double-negative αβ T cells.

Persisting CAR-T cells exhibit a transcriptional signature

Next, we identified differentially expressed genes among CAR-T cells from each timepoint to extract markers of infusion, early, mid and late CAR-T cells. Infusion products were enriched for genes related to cell cycle, nucleosome assembly and glycolysis, plausibly due to in vitro activation during manufacture. Infusion products expressed high levels of genes reflecting naive lymphocyte (that is, *SELL*, *CCR7*, *IL7R* and *LRRN3*) and early memory differentiation status, such as *TCF7* and *LEF1*. The dominant gene expression pattern of post-infusion CARPALL CAR-T cells was defined on a continuum of granzyme gene expression (Fig. 3c). Across post-infusion timepoints, CAR-T cells were skewed toward either higher *GZMH* and *GZMB* expression or higher *GZMK* expression. CAR-T populations that were defined by higher expression of *GZMK* additionally expressed genes related to effector (*LTB*), memory (*CD27* and *IL7R*) and activation (*CD28*) functions, whereas *GZMH*$^+$*GZMB*$^+$ cells expressed *FGFBP2* and *ZEB2*. Unlike the other patients in this study, with one exception (P09), most CAR-T cells at late timepoints expressed *GZMK*. In non-CAR-T cells, the *GZMH/B-GZMK* pattern of expression was also observed; however, CAR-T cells expressed *GZMK* to much higher

https://doi.org/10.1038/s41591-023-02415-3

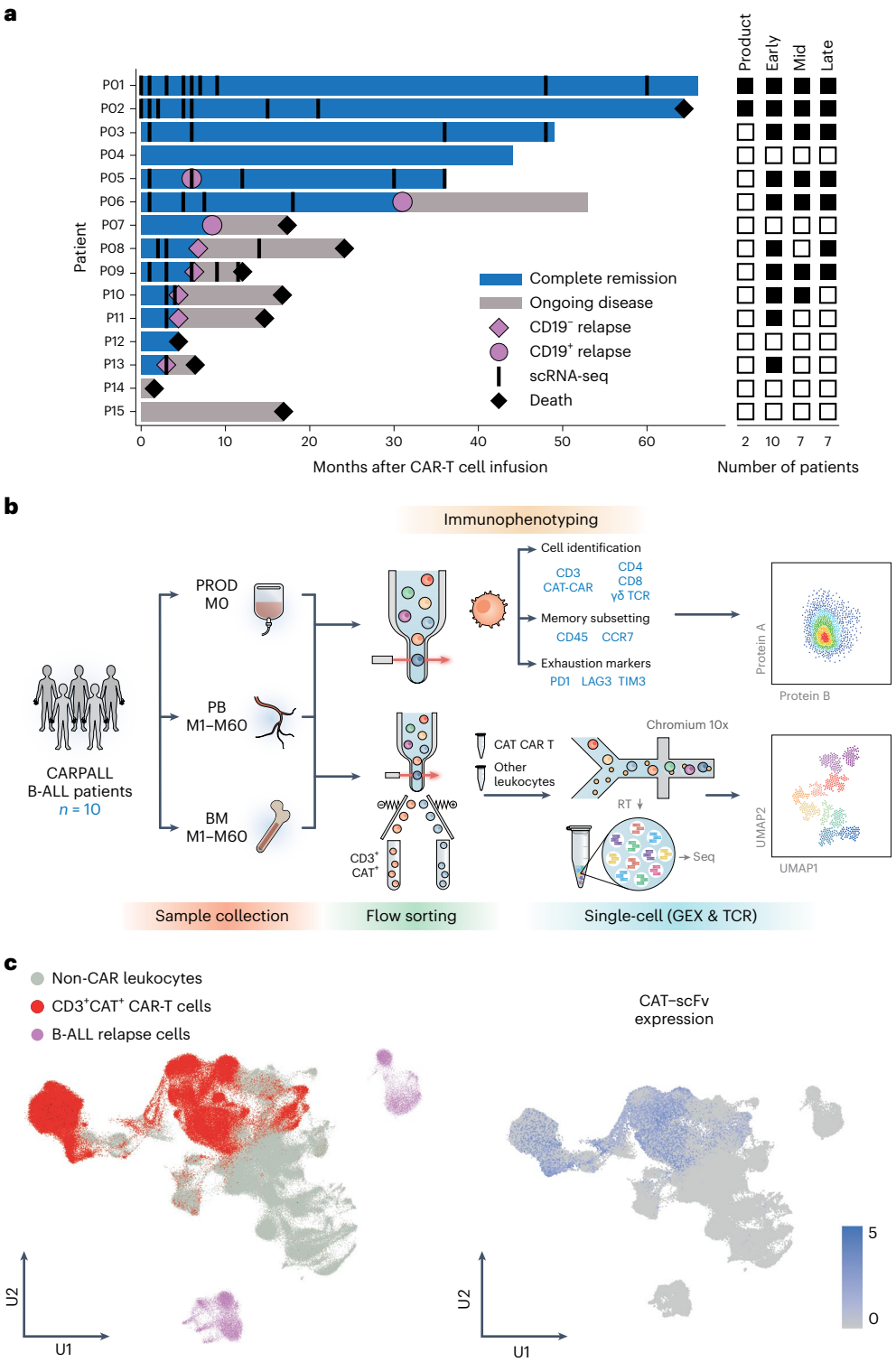


Fig. 1 | Study overview and workflow. a, Swimmer plot illustrating the responses of individual pediatric patients with B-ALL to CAT CAR T-cell therapy and timepoints of sample collection. Attainment of complete remission was associated with attainment of B cell aplasia in all cases. Patient 5 had an isolated unilateral ocular relapse of CD19⁺ leukemia, which was treated with enucleation and remains in ongoing minimal residual disease (MRD) negative remission with no other intervention. Heat map to the right demonstrates timepoint representation per patient. Filled black boxes indicate the presence of the timepoint. Product = infusion products (M0); early = M1–M3; mid = M4–M6; late = M7–M60. **b**, Schematic workflow of study design. Samples were collected from infusion products (PROD), peripheral blood (PB) and bone marrow (BM) between M0 and M60. Samples were used either for flow-based immunophenotyping or for single-cell GEX and TCR sequencing on the Chromium 10x platform. RT, reverse transcription; Seq, sequencing. **c**, UMAP of all cells in the dataset highlighting cell types captured (left) and expression of the CAT-scFv CAR construct (right). CAT-scFv, low-affinity CAR (CAT) incorporating a CD19-specific scFv.

https://doi.org/10.1038/s41591-023-02415-3

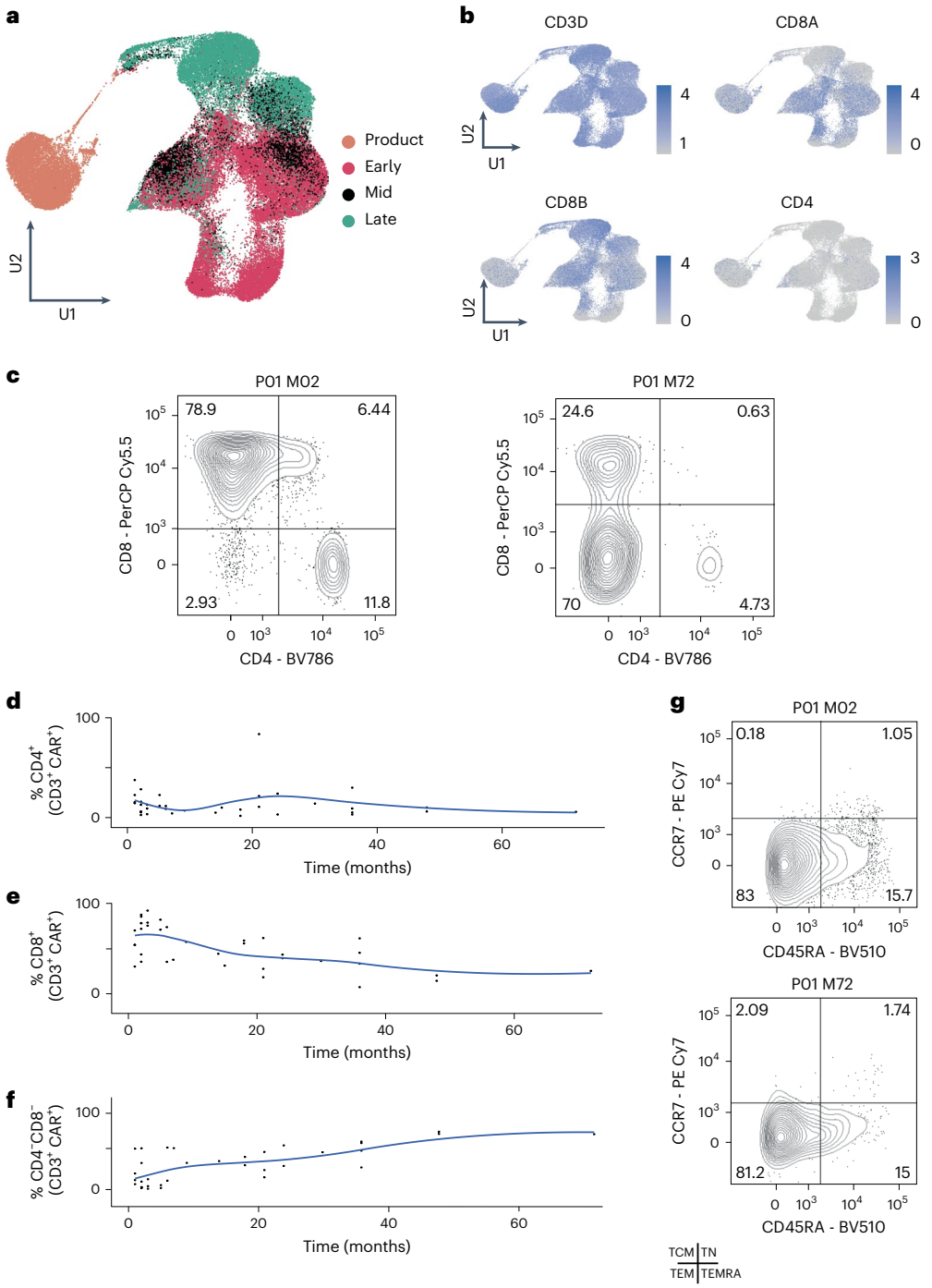


Fig. 2 | Characterization of CARPALL CAR-T cells. a, UMAP of CAT CAR-T cells demonstrates transcriptional clustering based on timepoint. Product = infusion products (M0); early = M1–M3; mid = M4–M6; late = M7–M60. **b**, UMAPs show scaled average expression of lymphocyte markers. **c**, Immunophenotyping CAT CAR-T cells by multi-parameter flow cytometry. Representative example of P01 showing cell identification using CD4 and CD8A at an early (M2) and a late (M72) timepoint. Cells were gated for CD3 and CAT CAR. **d**–**f**, Trajectory scatter plots quantify data from all patient samples for CD4, CD8 and double-negative CAR-T populations. **g**, Representative example of P01 showing immunophenotyping using CD45RA and CCR7 at an early (M2) and a late (M72) timepoint. Cells were gated for CD3 and CAT CAR. TCM, central memory; TEM, effector memory; TEMRA, terminally differentiated effector memory expressing CD45RA; TN, naive; TSCM, stem cell memory. Fluorochromes: BV, brilliant violet; Cy, cyanine dye; PE, R-phycoerythrin; PerCP, peridinin-chlorophyll protein.

levels (Extended Data Fig. 6a). The most recurrent and strongest markers of late CAR-T cells generated a persisting CAR-T signature that was delineated by the expression of bona fide immune-related genes, such as *TIGIT* and *GPR183*, as well as genes with unknown or emerging roles in immune biology (Fig. 3d and Extended Data Fig. 6b,c). The latter genes include *FXYD2*, *HMOX1*, *DENND2D* and *ISG20* (see Supplementary Table 5 for full gene signatures). The top marker of this population of cells was *FXYD2*, which encodes a modulator of the Na^+/K^+ ATPase channel. Of note, *FXYD2* was one of the transcripts expressed in functionally cryptic $CD34^{low}CD3^-CD4^+CD8^-$ intrathymic T progenitors that have been described in the human thymus[10]. In aggregate, our data reveal that, within and across patients, thousands of CAR-T cells converge on a double-negative cellular phenotype that displays a common and distinct gene signature.

https://doi.org/10.1038/s41591-023-02415-3

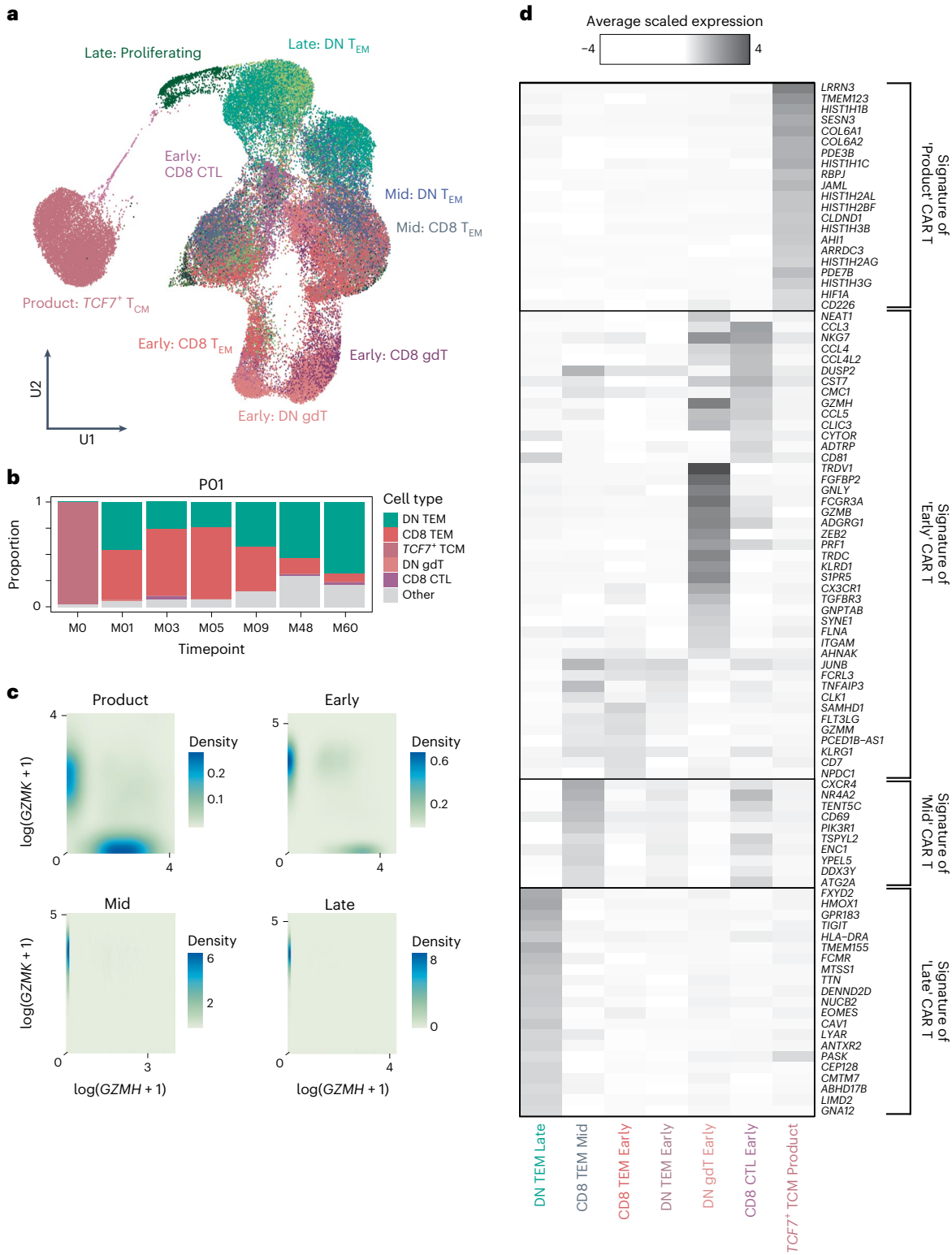


Fig. 3 | Cell typing and defining a persisting CAR-T cell transcriptional signature. a, UMAP of CAT CAR-T cells shows closest cell type matching using a PBMC reference and cell-marker-based annotation. Late (persisting, M7–M60) CAR-T cells are in green hues; mid (M4–6) CAR-T cells are in black/blue hues; and early (M1–3) CAR-T cells are in pink/red hues. **b**, Stacked bar plots show dynamic cell type proportions over time for a representative patient (P01). The top five abundant cell types in the dataset are shown. **c**, Contour plots quantify expression of *GZMK* against *GZMH* across patients per timepoint. **d**, Heat map shows differential gene expression results creating gene signatures of each timepoint. DN, double-negative; TCM, central memory; TEM, effector memory.

In recent years, the classical dogma of a dichotomy between memory and exhausted T cells has been challenged with the description of functionally active memory cells that bear an imprint of prior exhaustion[11]. One of the most highly expressed genes in the persistence signature was the exhaustion marker *TIGIT*. We, therefore, assessed the co-expression of exhaustion markers in our CAR-T cells. We found that

https://doi.org/10.1038/s41591-023-02415-3

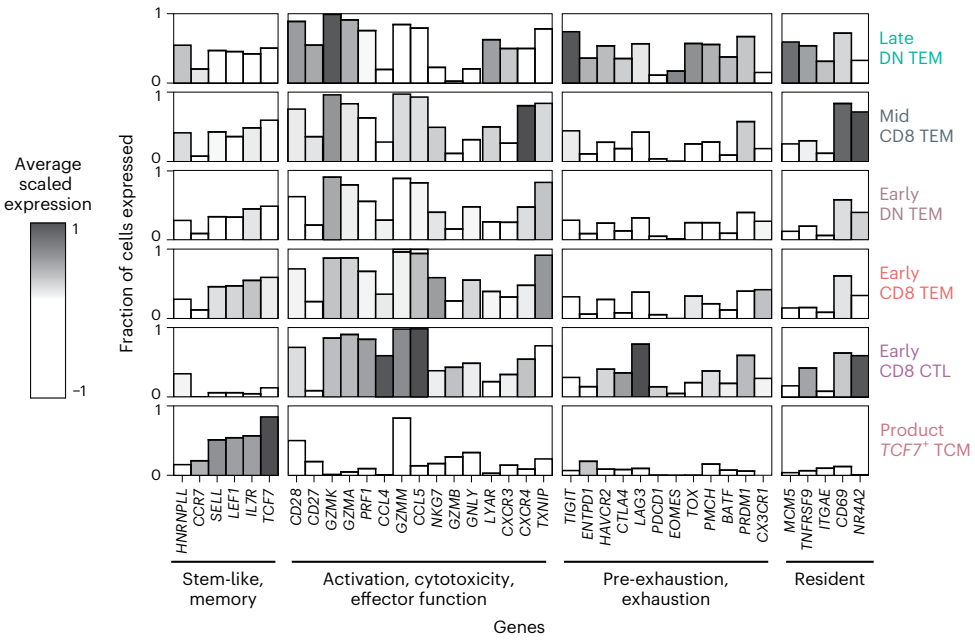


Fig. 4 | CAR-T cell immunophenotyping. Bar plots show custom gene modules that functionally characterize CAR-T cells. The height of the bar refers to the fraction of cells per cell type that express the gene. Higher bars indicate that more cells of that cell type are expressing that gene. The shading refers to the average scaled expression of those genes for that cell type. Darker shading indicates that the expression of that gene is above the average expression of that gene across all cell types. All CAR-T cells (product, early, mid and late) from all patients ($n = 10$) are analyzed. DN, double-negative; TCM, central memory; TEM, effector memory.

late CAR-T cells expressed canonical co-inhibitory receptors, such as *HAVCR2* and *LAG3*, but to a lesser extent *PDCD1* (Fig. 4). We, therefore, interrogated the gene and matched protein expression (flow cytometry) of these mediators related to exhausted and precursor exhausted T cells (Extended Data Fig. 7a). Precursor exhausted T cells have been isolated in human cancer, where their presence has been associated with response to immunotherapy. Like precursor exhausted T cells, late CAR-T cells bore expression of relevant effector genes, such as *GZMK* and *PRF1*, as well as transcription factors associated with T cell exhaustion, including *TOX*[12–14], *NFATC1* (ref. 15), *BATF*[16] and *PRDM1* (ref. 17). Although exhibiting features of exhaustion, the late CAR-T cells did not appear terminally differentiated, as supported by low expression of *B3GAT1* (CD57) and by expressing no more than two exhaustion markers by flow cytometry (Extended Data Fig. 7b). Unlike CAR-T cells within the products that we evaluated, late CAR-T cells did not express high levels of *TCF7*, which orchestrates a state of memory stemness in precursor exhausted T cells in other contexts[18]. However, instead, there was robust expression of *JUN*, an AP-1-associated transcription factor that can mediate the reversal of T cell exhaustion and maintenance of cells with stem cell memory properties[19]. Thus, although long-persisting CAR-T cells did not exactly phenocopy precursor exhausted T cell populations described previously, this cell type would best describe their effector memory, exhaustion-imprinted status determined both transcriptionally and by flow cytometry.

Polyclonal population structures of persisting CAR-T cells

Within each patient, we had observed that, irrespective of T cell subset, thousands of cells converged on the same transcriptional state at late timepoints, raising the question of whether expansions of specific clones underpinned this functional convergence. We, therefore, interrogated TCR sequences of CAR-T cells and obtained readouts from 88 of 89 samples with concomitant gene expression data (Extended Data Fig. 1 and Supplementary Table 1). We found that the vast majority of cells across timepoints harbored unique clonotypes not observed at other timepoints. This indicates that the underlying gene pool remained sufficiently diverse to preclude the capture and tracking of individual clones (Fig. 5a and Extended Data Fig. 8a,b). An important consideration of this analysis is that the frequency of CAR-T cells diminishes over time, such that, by late timepoints, the frequency is as little as 0.05% of total $CD3^+$ cells in circulating blood (Extended Data Fig. 8c). Nevertheless, if the population structure were monoclonal, we would capture the same clone on each blood draw. Of the few trackable clonotypes, the top 10 clonotypes at early timepoints remained among the relative majority at later timepoints but decreased in frequency over time. We observed an extensive variability in cell type composition among clonotypes, irrespective of whether they were unique or observed across timepoints. Clonotypes that were observed across timepoints were predominantly $CD8^+$ T cells, whereas unique clonotypes tended to be double-negative T cells (Fig. 5b and Extended Data Fig. 8a,b). For two patients (P02 and P01), for whom we have infusion product TCR data, we were able to track 1.7% and 0.5% of clones across from infusion products to 2 years and 5 years, respectively. In aggregate, these clonal structures indicated that, at all timepoints, CAR-T cell populations were genetically diverse, consistent with insertion site analyses previously performed on CARPALL CAR-T cells[20]. In particular, there was no evidence of the dominance of one or more clones at late timepoints. Overall, these findings indicate that functional convergence of the persistence signature was not driven by clonal expansion.

Evaluation of the persistence signature across T cells

As we had observed a transcriptional convergence of CAR-T cells across thousands of cells within and across patients, we speculated that the persistence signature may be pervasive across different CAR-T cell products. To date, one further single-cell transcriptomic study of persistent CAR-T cells has been reported—of two adult patients with CLL treated with anti-CD19 CAR-T cells (CTL019 cells) that have persisted for one decade thus far[7]. We interrogated CAR-T cell data from these two patients by assigning a persistence signature score to each cell (the AddModuleScore function in Seurat[21]). Remarkably, the module was expressed in CTL019-persisting CD4 CAR-T cells in almost its entirety

https://doi.org/10.1038/s41591-023-02415-3

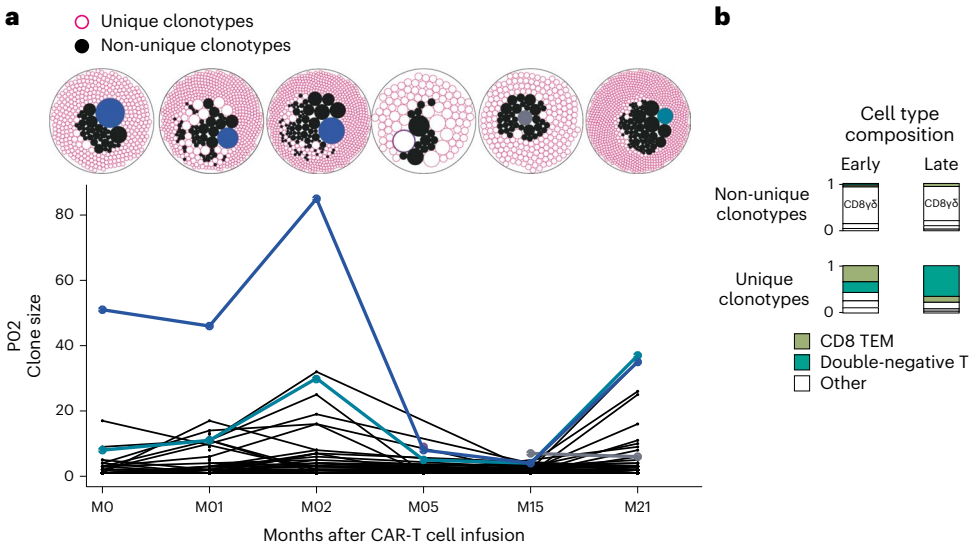


Fig. 5 | Population structures of CAR-T cells. a, Representative sample (P02) illustrating the changes in clonal architecture of CAR-T cells over time. Packed circle plots show the size of each clonotype. Filled-in black circles represent clonotypes that are not unique, as they are observed across timepoints. Conversely, pink donut circles represent clonotypes that are unique to that timepoint (and not observed across time). Blue/purple/gray colored circles represent the dominant clonotype at that timepoint that corresponds with the clonal trajectories below. **b**, Cell type composition stacked bar plots demonstrate the shift in cell type abundances between early and late timepoints and between unique and non-unique clonotypes. TEM, effector memory.

(17/22 genes) (Fig. 6a). To compare our CARPALL CAR-T cell signal with CTL019 cells in an unbiased, quantitative manner, we used a method of cell-to-cell matching based on logistic regression[22]. We found that the strongest match of persisting CTL019 CD4$^+$ CAR-T cells was to persisting double-negative T cells in the CARPALL data (Fig. 6b). It should be noted that persisting CTL019 cells were primarily derived from patient 1 (541/959, 56%), although, reassuringly, the persistence signature was also evident in a small number of cells from patient 2 (40/959, 4%). Overall, the similarity of persisting CARPALL and CTL019 CAR-T cells was not confined to gene sets but extended to the entire transcriptome.

We then questioned whether the persistence signature of CAR-T cells may have a physiological correlate. To this end, we scored T cells from a variety of healthy tissues, other CAR-T studies to date[5,6,23] and cancer cell single-cell atlases, including normal peripheral blood[8], human fetal bone marrow[24], human fetal thymus[25] and as many as 16 types of human cancers[26–34], including tumors that are considered to be immunogenic and had long-term response to immune checkpoint inhibitors (for example, lung cancer and melanoma) (see Supplementary Table 6 for datasets analyzed). We were unable to detect T cells harboring the persistence signature at an appreciable frequency in any one tissue, barring occasional cells (Fig. 6c and Extended Data Fig. 9). The median frequency across tissues was −0.04 (−0.6 to 1.4). These observations indicate that the CAR-T cell persistence signature is rarely found in other biological contexts.

Discussion

A lack of CAR-T cell persistence leading to CD19$^+$ relapse is the main cause of therapy failure after licensed CAR-T cell therapy for ALL[35,36] and contributes to relapse in other B cell malignancies, such as myeloma[37]. Therefore, a key question of CAR-T cell biology is why some cells persist whereas others perish. With this knowledge, we might better understand how to select patients, modify treatment phasing and optimize manufacturing protocols to support greater persistence and improve outcomes. To date, robust biomarkers of persistence have not been identified and can be validated only after directly demonstrating successful long-term persistence in patients. A key requirement of this is to systematically examine the biological status of long-lived CAR-T cells. Currently, there is a paucity of these datasets, as persisting CAR-T cells have been probed in only a very limited number of patients to date[7]. As such, we generated a single-cell RNA sequencing (scRNA-seq) dataset of cells from 10 patients with B-ALL treated with a CD19-targeting CAR-T cell product. Although our work represents, to our knowledge, the largest study of persisting single CAR-T cell transcriptomes, it still represents a modest cohort. Our key finding of a recurrent transcriptional state representing persistence is unlikely to be affected by the size of the cohort. Within each individual, every long-lived CAR-T cell represents a biological replicate of the signal. Accordingly, our finding has been reproduced multiple times within patients and has then been validated across individuals, including in the different clinical context of CLL. However, the size of our study precludes our ability to capture nuances of CAR-T cell transcription as well as any associations between CAR-T cell states and clinical subsets of patients, for which larger studies are required.

We found that late-persisting CAR-T cells mainly comprised a population that did not express CD8-α or CD4 co-receptors transcriptionally or via surface expression. In healthy individuals, double-negative cells typically comprise a minor population of all T cells, and we observed similar proportions in non-CAR T-cells from the same patient. In general, there was a steady reduction in CD8$^+$ CAR-T cells over time, which matched a progressive increase in double-negative populations. This contrasts the long-lived CAR-T cells from adult CAR-treated patients with CLL, where double-negative CAR-T cell populations were noted at earlier timepoints and, on further investigation, were determined to be γδ T cells. Although γδ T cells were also observed in four patients in our cohort, they did not contribute to early CAR-T cell populations in the other patients. In our cohort, the predominance of double-negative CAR-T cells was particularly noted at later timepoints in all patients. We verified that, at the later timepoints, double-negative T cells were not contributed to by γδ CAR T cells or CAR NK cells. Late CAR-T cells in both cohorts showed evidence of an activated, proliferative and effector status with strong expression of *GZMA* and *GZMK*. Because the double-negative phenotype observed in late-persisting CAR-T cells is reminiscent of early thymocyte differentiation and the fact that we noted high expression levels of *GPR183*, an oxysterol receptor that provides survival and migratory signals to thymocytes and CD4$^+$ T follicular helper cells[38], we hypothesized a link between persisting

https://doi.org/10.1038/s41591-023-02415-3

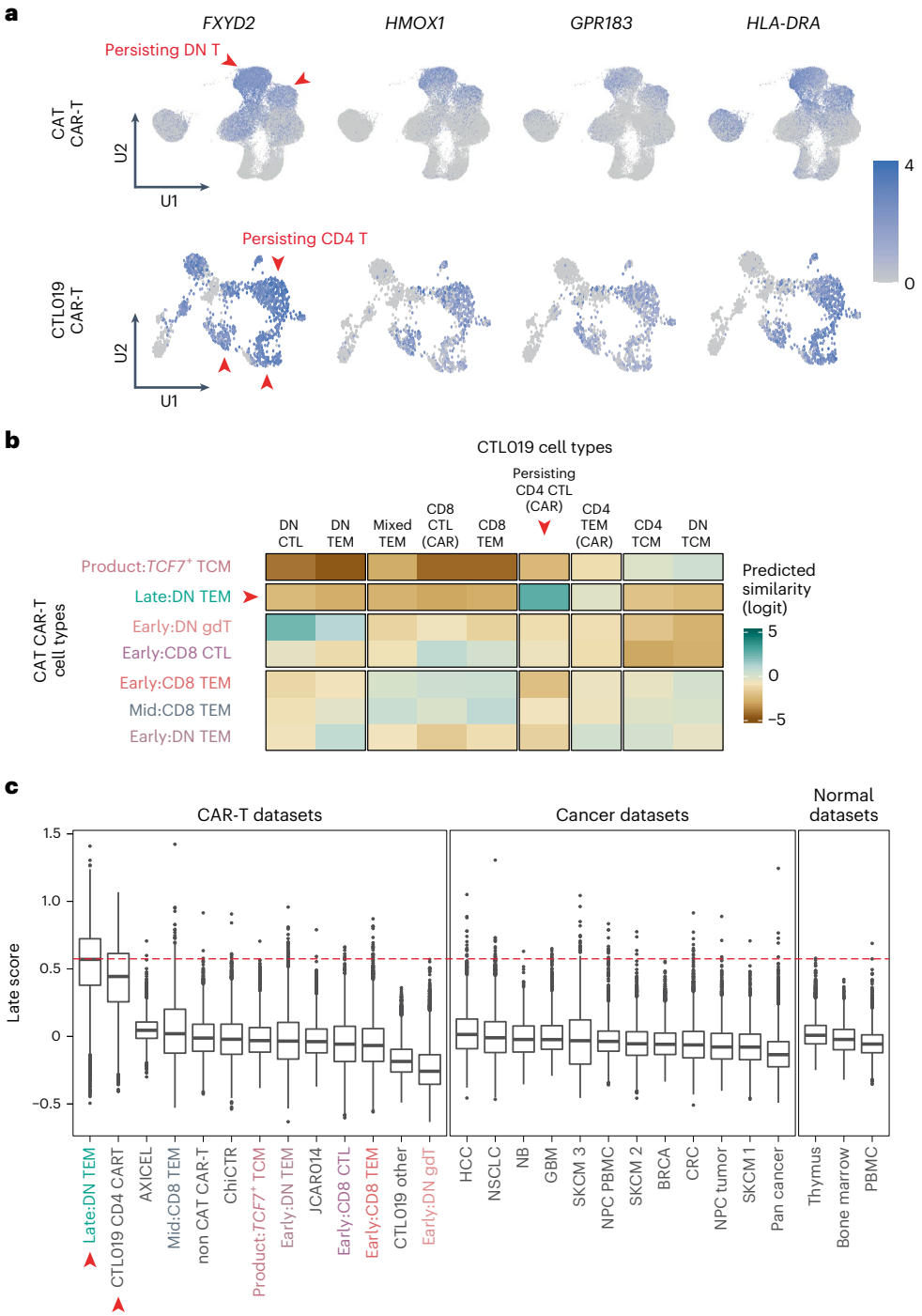


Fig. 6 | Evaluating the persisting transcriptional signature across T cells, including adult long-lived CAR-T cells. a, UMAP plots show expression of the strongest markers of the late-persisting CAR-T signature. Red arrows highlight persisting CAR-T cells between the CARPALL (CAT CAR-T) and CTL019 studies. **b**, Heat map demonstrates logistic regression cell-to-cell matching. CTL019-persisting CD4 CAR-T cells match strongly to late double-negative CAT CAR-T cells from the CARPALL study, as indicated by the red arrows. **c**, Box plots show the per-cell distribution of the late score as defined by the late-persisting CAR-T signature. Scores are shown for CAR-T, cancer and normal datasets. Publicly available datasets analyzed in this study are described in Supplementary Table 6. The red line represents the median of the late double-negative (DN TEM) cells from the CARPALL study. Colored cell types refer to CARPALL CAR-T cells. Red arrows indicate persisting CAT CAR-T cells from CARPALL and persisting CTL019 cells. Box plots show the first quartile (the lower end of the box) and the third quartile (the upper other end of the box) as well as the median values (center line) per dataset. The 'whiskers' extend from the ends of the box to a maximum and minimum of 1.5 times the interquartile range beyond the box. Outliers are shown as dots. AXICEL, axicabtagene ciloleucel CAR-T cells (infusion products); BRCA, breast cancer; ChiCTR, BCMA CAR-T (ChiCTR, 1800017404); CRC, colorectal cancer; DN, double-negative; GBM, glioblastoma; HCC, hepatocellular carcinoma; JCAR014, Fred Hutchinson Cancer Center CAR-T cells; NB, neuroblastoma; NPC, nasopharyngeal carcinoma; NSCLC, non-small lung cell carcinoma; SKCM, skin cutaneous melanoma (1 = Li; 2 = Yost; 3 = Jerby-Anon); TCM, central memory; TEM, effector memory. The numbers of cells and samples used in this figure are described in Supplementary Table 6.

CAR-T cells and thymic cell development. Ultimately, we found that the transcriptional status of the late, double-negative CAR-T cells did not map to any thymocyte subset in T cell development but, rather, to mature T cells.

Late-persisting CAR-T cells did not conform to quiescent early memory T cell populations but expressed genes associated with effector function and an activated state. These cells also maintained their proliferative capacity. In keeping with recent reports of precursor exhausted T cells bearing the hallmark of activation, late CAR-T cell populations expressed markers and transcription factors associated with exhaustion, including *TOX*[12–14] and *BATF*[39], among others. With reference to this highly activated status, one might speculate that these are circulating effector cells differentiated from rarer memory precursor populations after exposure to antigen. As these patients had no evidence of circulating B cells or existing CD19-expressing hematogones in the bone marrow, it is plausible that these cells were activated as the result of an emerging CD19-expressing hematogone population. However, the existence of minute central or stem cell memory CAR-T cells at this time-point may count against this hypothesis.

Late-persisting CAR-T cells, although activated and imprinted with markers of exhaustion, did not express *FOXO3* or *B3GAT1*, which are associated with terminal differentiation in the context of exhaustion. Instead, they expressed markers associated with memory-like characteristics, including *Jun*, *BCL2* and *IL7R*. Thus, they more closely matched precursor exhausted T cell populations as have been described in chronic viral infections[11,39,40], cancer-infiltrating T cells[41,42] and early post-infusion of CD19 CAR-T cells. Unlike previous reports of precursor exhausted T cell populations, however, they did not express high levels of *TCF7* (refs. 39,40) or *FOXO1* (refs. 43,44), confounding the suggested centrality of such transcription factors in driving long-lived CAR-T cell persistence and overcoming terminal exhaustion. In a previous report, c-Jun overexpression was sufficient to restore antigen responsiveness, memory function and long-term proliferative capacity in CAR-T cells exhausted due to tonic CAR signaling. The high expression of *JUN* in most late-persisting CAR-T cells in this study points to a plausible mechanism for long-lived persistence in these cells. Overall, our data instead support that, although previously exhausted CAR-T cells may indeed give rise to long-persisting populations, the predominance of transcription factors driving memory status is likely specific to the characteristics of the CAR, disease and model. Despite some similarities to precursor exhausted T cells, the persistence module was rarely expressed by T cells in a range of pathophysiological contexts. Within the signature, there were a number of genes with little-known roles in T cells, including *FXYD2*, *DENND2D* and *HMOX1*. Overall, further work is needed to elucidate their function in T cells and how they may contribute to persistence.

A key finding of this study was a transcriptional signature of persistence that was reproducible across thousands of cells in every patient with long-lived CAR-T cells and durable anti-B-ALL responses. This signature was not identified when interrogating non-CAR-T cells from the same patients, T cell populations in normal development, T cells from a range of cancer datasets or other CAR-T cell studies. The persistence signature and underlying cell state were detected in an independent dataset of long-lived CAR-T cells from adults with CLL who had received a different CD19 CAR-T cell product. Of note, persisting CAR-T cells from the independent dataset were cytotoxic CD4 T cells with oligoclonal population structures, in contrast to double-negative CAR-T cells that were polyclonal, as reported in this study. These differences could arise from the different techniques used for TCR analyses (integration site analysis versus single-cell TCR analysis), the number of patients evaluated or the fundamental differences in the CAR product. Although we noted these differences among CAR-T cells at different timepoints, the late CAR-T cell signature that we defined here was reproducible across both studies, indicating that it may represent a surrogate marker of longevity. Although this falls short of an easily measured biomarker of persistence with which to test CAR-T cell products, this understanding brings us a step closer to identifying such assays. That the transcriptional status noted was so pervasive in long-persisting CAR-T cells lends weight to the possibility that the signature may not only be a surrogate marker of longevity but, notably, may also provide a basis to investigate underlying cell-intrinsic or cell-extrinsic factors that drive CAR-T cell persistence. Given the data presented here, the longevity of CAR-T cells is likely not based on clonal selection and expansion. Rather, it is possible that the ongoing interplay with the environment shapes the resultant phenotype of long-lived CAR-T cells and supports functional diversity. With this knowledge, we will be primed in our ability to engineer this key characteristic into CAR-T cell therapies for hematological malignancies of the future.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-023-02415-3.

References

1. Maloney, K. W. et al. Outcome in children with standard-risk B-cell acute lymphoblastic leukemia: results of Children's Oncology Group trial AALL0331. *J. Clin. Oncol.* **38**, 602–612 (2020).
2. Maude, S. L. et al. Chimeric antigen receptor T cells for sustained remissions in leukemia. *N. Engl. J. Med.* **371**, 1507–1517 (2014).
3. 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.* **25**, 1408–1414 (2019).
4. Xu, X. et al. Mechanisms of relapse after CD19 CAR T-cell therapy for acute lymphoblastic leukemia and its prevention and treatment strategies. *Front. Immunol.* **10**, 2664 (2019).
5. 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).
6. Deng, Q. et al. Characteristics of anti-CD19 CAR T cell infusion products associated with efficacy and toxicity in patients with large B cell lymphomas. *Nat. Med.* **26**, 1878–1887 (2020).
7. Melenhorst, J. J. et al. Decade-long leukaemia remissions with persistence of CD4$^+$ CAR T cells. *Nature* **602**, 503–509 (2022).
8. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
9. Pizzolato, G. et al. Single-cell RNA sequencing unveils the shared and the distinct cytotoxic hallmarks of human TCRVδ1 and TCRVδ2 γδ T lymphocytes. *Proc. Natl Acad. Sci. USA* **116**, 11906–11915 (2019).
10. Lee, M. S., Hanspers, K., Barker, C. S., Korn, A. P. & McCune, J. M. Gene expression profiles during human CD4$^+$ T cell differentiation. *Int. Immunol.* **16**, 1109–1124 (2004).
11. Galletti, G. et al. Two subsets of stem-like CD8$^+$ memory T cell progenitors with distinct fate commitments in humans. *Nat. Immunol.* **21**, 1552–1562 (2020).
12. Khan, O. et al. TOX transcriptionally and epigenetically programs CD8$^+$ T cell exhaustion. *Nature* **571**, 211–218 (2019).
13. Scott, A. C. et al. TOX is a critical regulator of tumour-specific T cell differentiation. *Nature* **571**, 270–274 (2019).
14. Alfei, F. et al. TOX reinforces the phenotype and longevity of exhausted T cells in chronic viral infection. *Nature* **571**, 565–569 (2019).
15. Martinez, G. J. et al. The transcription factor NFAT promotes exhaustion of activated CD8$^+$ T cells. *Immunity* **42**, 265–278 (2015).
16. Quigley, M. et al. Transcriptional analysis of HIV-specific CD8$^+$ T cells shows that PD-1 inhibits T cell function by upregulating BATF. *Nat. Med.* **16**, 1147–1151 (2010).

17. Shin, H. et al. A role for the transcriptional repressor Blimp-1 in CD8$^+$ T cell exhaustion during chronic viral infection. *Immunity* **31**, 309–320 (2009).
18. Chen, Z. et al. TCF-1-centered transcriptional network drives an effector versus exhausted CD8 T cell-fate decision. *Immunity* **51**, 840–855 (2019).
19. Lynn, R. C. et al. c-Jun overexpression in CAR T cells induces exhaustion resistance. *Nature* **576**, 293–300 (2019).
20. Biasco, L. et al. Clonal expansion of T memory stem cells determines early anti-leukemic responses and long-term CAR T cell persistence in patients. *Nat. Cancer* **2**, 629–642 (2021).
21. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
22. Young, M. D. et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594–599 (2018).
23. Li, X. et al. Single-cell transcriptomic analysis reveals BCMA CAR-T cell dynamics in a patient with refractory primary plasma cell leukemia. *Mol. Ther.* **29**, 645–657 (2021).
24. Jardine, L. et al. Blood and immune development in human fetal bone marrow and Down syndrome. *Nature* **598**, 327–331 (2021).
25. Park, J. E. et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, eaay3224 (2020).
26. Kildisiute, G. et al. Tumor to normal single-cell mRNA comparisons reveal a pan-neuroblastoma cancer cell. *Sci. Adv.* **7**, eabd3311 (2021).
27. Li, H. et al. Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma. *Cell* **176**, 775–789 (2019).
28. Yost, K. E. et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat. Med.* **25**, 1251–1259 (2019).
29. Jerby-Arnon, L. et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell* **175**, 984–997 (2018).
30. Azizi, E. et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308 (2018).
31. Guo, X. et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24**, 978–985 (2018).
32. Lee, H. O. et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat. Genet.* **52**, 594–603 (2020).
33. Liu, Y. et al. Tumour heterogeneity and intercellular networks of nasopharyngeal carcinoma at single cell resolution. *Nat. Commun.* **12**, 741 (2021).
34. Zheng, L. et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* **374**, abe6474 (2021).
35. Dourthe, M. E. et al. Determinants of CD19-positive vs CD19-negative relapse after tisagenlecleucel for B-cell acute lymphoblastic leukemia. *Leukemia* **35**, 3383–3393 (2021).
36. van Waart, H. et al. Effect of low-intensity physical activity and moderate- to high-intensity physical exercise during adjuvant chemotherapy on physical fitness, fatigue, and chemotherapy completion rates: results of the PACES randomized clinical trial. *J. Clin. Oncol.* **33**, 1918–1927 (2015).
37. Munshi, N. C. et al. Idecabtagene vicleucel in relapsed and refractory multiple myeloma. *N. Engl. J. Med.* **384**, 705–716 (2021).
38. Li, J., Lu, E., Yi, T. & Cyster, J. G. EBI2 augments Tfh cell fate by promoting interaction with IL-2-quenching dendritic cells. *Nature* **533**, 110–114 (2016).
39. Utzschneider, D. T. et al. Early precursor T cells establish and propagate T cell exhaustion in chronic infection. *Nat. Immunol.* **21**, 1256–1266 (2020).
40. Utzschneider, D. T. et al. T cell factor 1-expressing memory-like CD8$^+$ T cells sustain the immune response to chronic viral infections. *Immunity* **45**, 415–427 (2016).
41. Siddiqui, I. et al. Intratumoral Tcf1$^+$PD-1$^+$CD8$^+$ T cells with stem-like properties promote tumor control in response to vaccination and checkpoint blockade immunotherapy. *Immunity* **50**, 195–211 (2019).
42. Miller, B. C. et al. Subsets of exhausted CD8$^+$ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019).
43. Delpoux, A., Lai, C. Y., Hedrick, S. M. & Doedens, A. L. FOXO1 opposition of CD8$^+$ T cell effector programming confers early memory properties and phenotypic diversity. *Proc. Natl Acad. Sci. USA* **114**, E8865–E8874 (2017).
44. Michelini, R. H., Doedens, A. L., Goldrath, A. W. & Hedrick, S. M. Differentiation of CD8 memory T cells depends on Foxo1. *J. Exp. Med.* **210**, 1189–1200 (2013).

[1]Wellcome Sanger Institute, Hinxton, UK. [2]Developmental Biology and Cancer, UCL Great Ormond Street Institute of Child Health, London, UK. [3]Molecular and Cellular Immunology, UCL Great Ormond Street Institute of Child Health, London, UK. [4]Cancer Research UK & UCL Cancer Trials Centre, London, UK. [5]Department of Bone Marrow Transplantation, Royal Manchester Children's Hospital, Manchester, UK. [6]Children and Young People's Cancer Service, University College London Hospitals NHS Foundation Trust, London, UK. [7]Department of Haematology, University College London Hospitals NHS Foundation Trust, London, UK. [8]Department of Haematology, UCL Cancer Institute, London, UK. [9]Department of Bone Marrow Transplantation, Great Ormond Street Hospital for Children, London, UK. [10]Department of Paediatrics, University of Cambridge, Cambridge, UK. [11]Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [12]Department of Haematology, Great Ormond Street Hospital for Children, London, UK. [13]These authors contributed equally: Jack Birch, Theo Accogli, Ignacio Criado. [14]These authors jointly supervised this work: Sam Behjati, Sara Ghorashian. ✉e-mail: sb31@sanger.ac.uk; s.ghorashian@ucl.ac.uk

Methods

Sample acquisition, ethics and patient consent

Data from this study were generated from patients enrolled in the CARPALL study (NCT02443831). CARPALL was a multi-center, non-randomized, open-label, phase 1, single-stage clinical study designed to evaluate the safety, efficacy and response of CD19 CAR-T cells in children and young adults (≤24 years of age) with high-risk relapsed $CD19^+$ malignancies. Patient data were collected at Great Ormond Street Hospital (GOSH) and the University College London (UCL)-GOSH Institute of Child Health (ICH), and laboratory data were generated in the study central laboratories at GOSH, UCL-GOSH ICH as well as the Sanger Institute. Patient recruitment occurred from 2016 to 2019. Data collection, sequencing and analysis were from 2016 to 2023. The study protocol and outcomes are available here: http://www.ctc.ucl.ac.uk/TrialDetails.aspx?Trial=116&term=carpall. Key clinical factors for this cohort are described in Supplementary Table 7. All patients who took part in this study were diagnosed with B-ALL. Written informed consent was obtained from all patients or their parents/guardians before study entry. Patients did not receive compensation for participation in the study. Patient sex was reported by patients or parents and confirmed upon (external) examination. Study results do not apply to any one sex or gender. Sex or gender were not considered in the study design, as all children and young aduts with high-risk B-ALL, independent of sex/gender, were considered. The sex of patients was noted, and this is described in the table of patient characteristics (Supplementary Table 7). This trial was approved by the UK Medicines and Healthcare Products Regulatory Agency (clinical trial authorization no. 20363/0361/001). Ethical approval was obtained from the London–West London Gene Therapy Advisory Committee (GTAC) Research Ethics Committee (REC ref. no. 16/LO/0283). Note that the CARPALL study initially used monospecific low-affinity CD19 CAR-T cells for therapy of B-ALL; however, a study amendment allowing investigation of dual CD19 and CD22 CAR targeting is currently displayed on the ClinicalTrials.gov website. Historic versions of this trial before November 2020 can be viewed using the following link: https://clinicaltrials.gov/ct2/history/NCT02443831 (compare any version before November 2020). The analyses included here were not pre-specified in the clinical trial protocol.

Flow cytometry

CAR-T cells were isolated from either fresh peripheral blood or cryopreserved aliquots of the infusion product (IP), peripheral blood mononuclear cells (PBMCs) or bone marrow mononuclear cells (BMMCs). For fresh peripheral blood, PBMCs were isolated via density gradient centrifugation with Lymphopure (BioLegend). For cryopreserved samples, aliquots were rapidly thawed and washed in complete RPMI (10% FCS and 1% L-glutamine, Gibco). Flow cytometry was performed with a BD LSR II and cell sorting with a FACSAria III (BD Biosciences). Data analysis was performed using FlowJo version 10 (Tree Star) or FACS DIVA 8.0.1. Expression of CAR was detected by a CAR anti-idiotype antibody (bespoke product, Evitria, 1/200) and goat anti-rat IgG PE antibody (Poly4054, BioLegend, 1/400). The following reagents were used for phenotypic analysis of CAR-T cells: PD-1 BV421 (EH12.2H7, BioLegend, 1/20), CD45RA BV510 (HI100, BD Biosciences, 1/100), Lag3 BV605 (11C3C65, BioLegend, 1/20), TCRgd BV650 (B1, BD Biosciences, 1/20), CD127 BV711 (HIL-7R-M21, BD Biosciences, 1/20), CD4 BV784 (SK3, BioLegend, 1/100), CD25 VioBright FITC (4E3, Miltenyi Biotec, 1/100), Tim3 PECF594 (7D3, BD Biosciences, 1/20), CD8 PerCP-Cy5.5 (SK1, BioLegend, 1/40), CCR7 PE/Cy7 (G043H7, BioLegend, 1/40), CD95 APC (581, BioLegend, 1/10), CD3 AF700 (SK7, BioLegend, 1/40), CD27 APC/Cy7 (M-T271, BioLegend, 1/20), TIGIT BV605 (741182, BD Biosciences, 1/40), GPR183 PE/Dazzle594 (SA313E4, BioLegend, 1/40) and GZMK APC (GM26E7, BioLegend, 1/40). DAPI and Fixable Viability Dye eFluor 455UV (eBioscience) were used to discriminate viable cells. For intracellular markers, cells were fixed (Fixation Buffer, BioLegend) and permeabilized (Intracellular Staining Permeabilization Wash Buffer 10×, BioLegend) before staining. Human BD Fc Block (BD Biosciences) was used as a blocking reagent. Fluorescence minus one (FMO) controls were used to determine expression thresholds where required. The full list of antibodies can be found in Supplementary Table 8. The flow cytometry gating strategy for immunophenotyping can be found in Extended Data Fig. 10.

CAR-T cell isolation and scRNA-seq using the 10x Chromium platform

Patient cells were harvested as described above for flow cytometry. Cryopreserved samples for 10x were rapidly thawed and washed with complete RPMI containing 50 U ml^{-1} of benzonase (Merck Life Science Limited). Cells were then stained with CAR anti-idiotype, followed by goat anti-rat IgG PE antibody and antibodies to CD3 APC (UCHT1, BioLegend, 1/20) and CD45 FITC (2D1, BioLegend, 1/20). DAPI was used to distinguish viable cells. CAR-T cells were isolated as $CD45^+CD3^+CAR^+$ events in a live singlet leukocyte forward-scatter (FSC)/side-scatter (SSC) gate using a BD FACSAria III flow sorter. The flow cytometry gating strategy for CAR sorting can be found in Extended Data Fig. 10. CAR and non-CAR populations were sorted simultaneously and then immediately used downstream for the 10x workflow. Flow-sorted cells (CAR and non-CAR) were loaded according to the standard protocol of the Chromium Single Cell 5′ Kit (v2 chemistry). A TCR single-cell library was subsequently prepared from the same cells with the Chromium Single Cell V(D)J Enrichment Kit. The 5′ gene expression library and the TCR single-cell library were pooled with a molar ratio 10:1 for sequencing on Illumina NovaSeq S4 with 28 × 90 bp, aiming for an average of 300,000 reads per cell for the 5′ gene expression library and 30,000 reads per cell for the TCR single-cell library.

Raw sequencing data processing, data filtering and normalization

The raw scRNA-seq data were demultiplexed and mapped to reference genome GRCh38, with the CAT-scFv sequence inserted, using Cell Ranger (10x Genomics, version 5.0.0). To filter lower-quality cells, we removed any cell with fewer than 300 genes, fewer than 1,000 unique molecular identifiers (UMIs) or where more than 10% of the read counts were derived from the mitochondrial genome. We excluded nuclear mitochondrial genes, heat shock proteins and ribosomal genes from our analysis.

Feature counts for each cell were divided by the total counts for that cell and multiplied by 10,000, followed by natural-log transformation using log1p. Counts data were then scaled such that each feature will be centered to have a mean of 0 and an s.d. of 1 for each gene. Principal component analysis was performed using the top 2,000 highly variable genes, and data were grouped into clusters using a community detection finding algorithm taking the first 75 principal components as inputs. Using these principal components, we calculated a UMAP for data visualization and calculated clusters using the *k*-nearest neighbors approach with resolution parameter set to 1. This was performed using the Seurat package in R (R version 4.0.3 and Seurat version 4.0.6).

Cluster annotation and multi-modal reference mapping

CAR-T cells were defined as cells sorted for CD3 and the CAR by flow cytometry and belonging to clusters expressing the 'CAT-scFv' gene. CAR-T cells were clustered separately and labeled with their timepoint bins: product (M0), early (M1–M3), mid (M4–M6) and late (M7–M60). Clusters were subsequently annotated using lymphoid markers (that is, *CD8A*, *CD8B* and *CD4*) and established markers of T cell states curated from literature (Extended Data Fig. 3 and Supplementary Table 2). To supplement cell type annotation, the PBMC multi-modal reference was downloaded and processed using the instructions from the vignette. CAR-T cells were projected into the multi-modal reference using the FindTransferAnchors() and MapQuery() functions available in Seurat.

Differential gene expression and immunophenotyping of CAR-T cells

CAR-T cells were clustered separately at a global (across patients) level and per patient. Seurat's FindAllMarkers() function was used to identify differentially expressed genes from cells across patients (global clustering) and within a patient using previously annotated cell types and timepoint bins (product, early, mid or late) as the label (that is, late: CD8 TEM). These analyses were performed using the two-sided Wilcoxon rank-sum test with Bonferroni multiple testing correction. Only genes with an average $\log_2$ fold change above 0.5 were considered. For the per-patient analysis, markers were tallied and ordered from most to least recurrent across labels (timepoint bin: cell type). With the exception of the product, where only two samples were available, markers were considered recurrent if present in more than two patients. Gene signatures were derived from the intersection of the top 20 recurrent (across patients) marker genes and the global markers. For immunophenotyping analysis presented in Fig. 4, gene modules were curated from literature. The average scaled expression and percentage of cells expressing the gene were determined using the input derived from the data slot of the DotPlot() function in Seurat and replotted as shaded bar plots.

TCR analysis

Chromium 10x V(D)J single-cell sequencing data were mapped and quantified using the software package cellranger vdj (version 5.0.0) using the GRCh38 reference (vdj_GRCh38_alts_ensembl-5.0.0). The consensus annotation files were generated per sample and used for downstream analyses. Clonotypes were defined per experimental sample based on unique TCR VJ sequences and complementarity-determining region (CDR3) motifs. Basic TCR statistics, such as the number of clones and the distribution of lengths and counts, were computed using Immunarch (version 0.7.0). For clonal tracking analyses, entries with a single or more than two alpha or beta chain(s) were considered one clone. Clonal population circles were created using the ggraph and igraph packages in R (version 2.0.5 and version 1.2.6, respectively). Unique clonotypes were defined as cells with shared TCR alpha and beta sequences that were not observed across timepoints but were uniquely observed at only one timepoint within the patient. Conversely, non-unique clonotypes are cells with shared TCR alpha and beta sequences that are present across at least two timepoints within a patient. The population circle plots were created by defining a 'root' and specifying the clonotype names and sizes as 'branches' on the same level of the tree.

Cell-to-cell matching: logistic regression

To determine the probability that the transcriptome of each CARPALL CAT CAR-T cell was similar to CTL019 (tisagenlecleucel) CAR-T cells from two adult patients with CLL[7], logistic regression was used in R, as previously described[22,26,45,46]. CTL019 raw counts data were processed as described above, using the same parameters as the CARPALL dataset. CTL019 cells were re-annotated using marker-based approaches, as described above. We trained logistic regression models with CTL019 cells using our cell type annotation.

Gene module scoring

Published datasets from CAR-T cells, cancer and normal development were downloaded, and T cells were identified using *CD3D* and *CD3E* expression. T cell partitioned datasets were randomly downsampled to 10,000 cells, if exceeding this threshold. T cell clusters were processed and re-clustered, as described above. Module scores were calculated using the AddModuleScore() function available in Seurat using Seurat clusters as labels (Louvain algorithm). The average expression level of each cell type (or cluster) was calculated on a single-cell level and then subtracted by the aggregated expression of control feature sets. Gene modules were defined based on differential gene expression of CAT CAR-T cells.

Reporting summary

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

Data availability

Raw sequencing data produced in this study have been deposited at the European Genome-phenome Archive (accession number EGAD00001010018). These data are available under restricted access. Sequencing data requests will be reviewed by the Independent Data Monitoring Committee and the Trial Management Group of the CARPALL study and will be subject to patient confidentiality. After approval, a data access agreement with University College London (UCL) will be required. All requests for raw materials will be reviewed by UCL Business (UCLB) to verify whether the request is subject to any intellectual property or confidentiality obligations. All requests will be processed within 8 weeks. Processed data have been uploaded to Zenodo[47]. Publicly available datasets analyzed in this study are described in Supplementary Table 6. The GRCh38 reference genome was downloaded from the 10x Genomics website: https://support.10xgenomics.com/single-cell-gene-expression/software/release-notes/build. Source data are provided with this paper.

Code availability

We have included the source code used to generate the figures and tables presented in this analysis. Code used in this study can be found here: https://github.com/natedandy/CAT_CART_paper_2023. Logistic regression was performed using code found here: https://github.com/constantAmateur/scKidneyTumors.

References

45. Khabirova, E. et al. Single-cell transcriptomics reveals a distinct developmental state of *KMT2A*-rearranged infant B-cell acute lymphoblastic leukemia. *Nat. Med.* **28**, 743–751 (2022).
46. Custers, L. et al. Somatic mutations and single-cell transcriptomes reveal the root of malignant rhabdoid tumours. *Nat. Commun.* **12**, 1407 (2021).
47. Anderson, N. D. Transcriptional signatures associated with persisting CD19 CAR T-cells in children with leukaemia. https://doi.org/10.5281/ZENODO.7937878 (2023).

Acknowledgements

We thank the clinical staff who contributed to the care of the children enrolled in the CARPALL study. Figure 1b was produced by A. Garcia, scientific illustrator from Bio-Graphics. We are indebted to the children and their families who participated in our research. This work was funded by Cancer Research UK (CRUK). The CARPALL study was supported by Children with Cancer, Great Ormond Street Hospital Children's Charity and the JP Moulton Trust. Later cohorts were supported by Autolus LTD. The study was also supported by the National Institute for Health Research Biomedical Research Centres at Great Ormond Street Hospital for Children NHS Foundation Trust, University College London Hospital, King's Health Partners, Great Ormond Street Hospital, University College London Hospital, the Wellcome Trust (institutional grant 108413/A/15/D) and a personal fellowship to S.B. (223135/Z/21/Z). S.G. is the recipient of a CRUK-funded Cancer Immunotherapy Accelerator Award, which supported J.B., T.A. and I.C. N.D.A. is supported by a Marie Sklodowska-Curie Individual Fellowship.

Author contributions

S.B. and S.G. designed the study. N.D.A., E.K., I.C., J.B., T.A., T.P. and Y.W. performed experiments. N.D.A., E.K., I.C., T.A., J.B., R.R., S.A., R.W., B.P., R.H. and A.L. collected and analyzed data. S.G. contributed reagents, tissue and clinical data. N.D.A., S.B. and S.G. wrote the manuscript. S.H.G., M.D.Y., C.P., M.A.P., P.J.A., S.G. and S.B. provided technical support and conceptual advice. S.B. and S.G. oversaw the study. All authors have approved the manuscript.

Competing interests

M.A.P. is employed by and owns stock in Autolus, Ltd., which has licensed the CD19 CAR. S.G., M.A.P. and P.J.A. have patent rights for CAT CAR in targeting CD19 and may receive royalties from Autolus Therapeutics PLC (patent application: World Intellectual Property Organization, WO 2016/139487 Al). N.D.A., S.B. and S.G. have filed a patent application related to the use of a transcriptional signature to determine a persistence phenotype in CAR-T cells. The remaining authors declare no competing financial interests.

Additional information

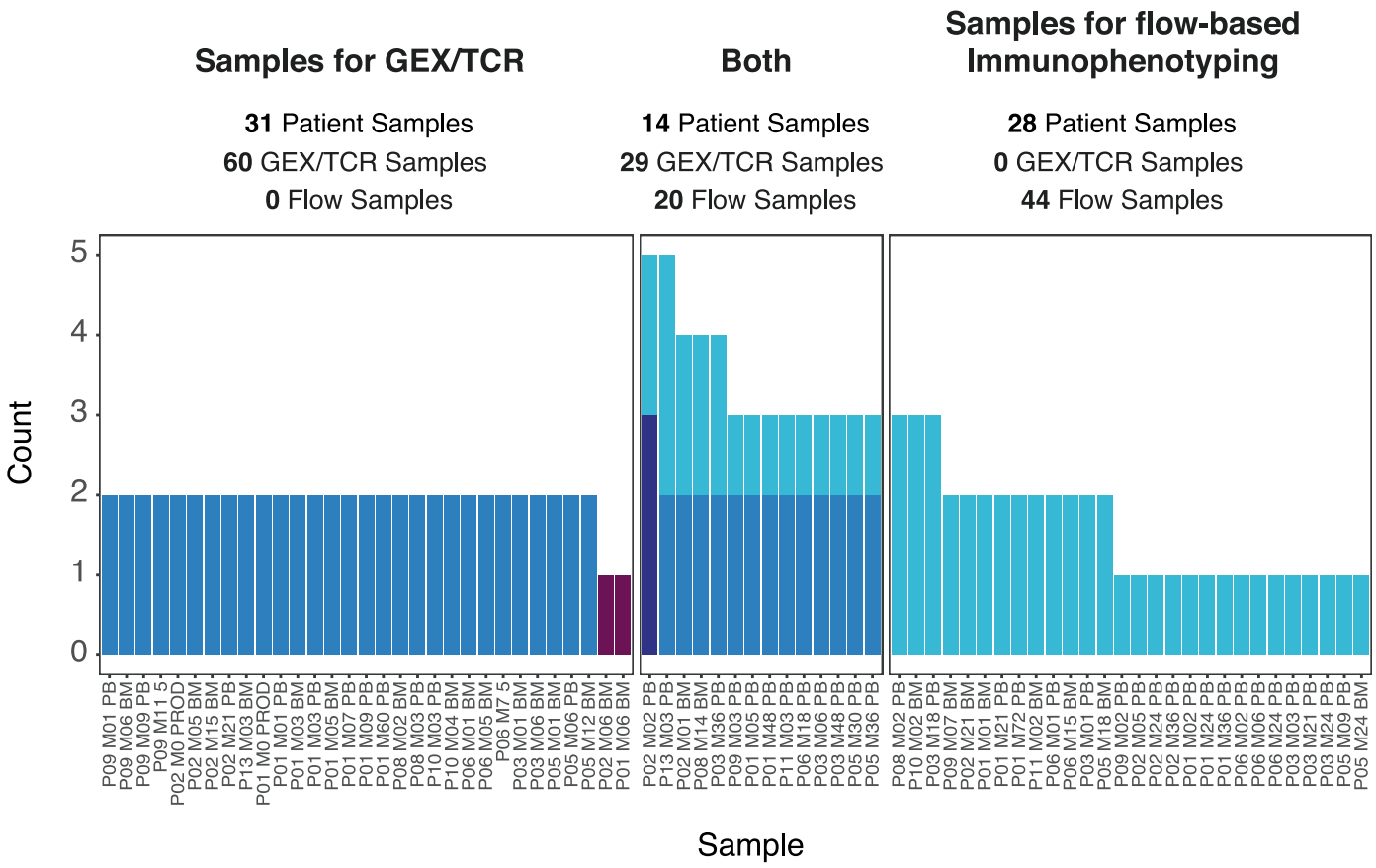
Extended data is available for this paper at https://doi.org/10.1038/s41591-023-02415-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-023-02415-3.

Correspondence and requests for materials should be addressed to Sam Behjati or Sara Ghorashian.

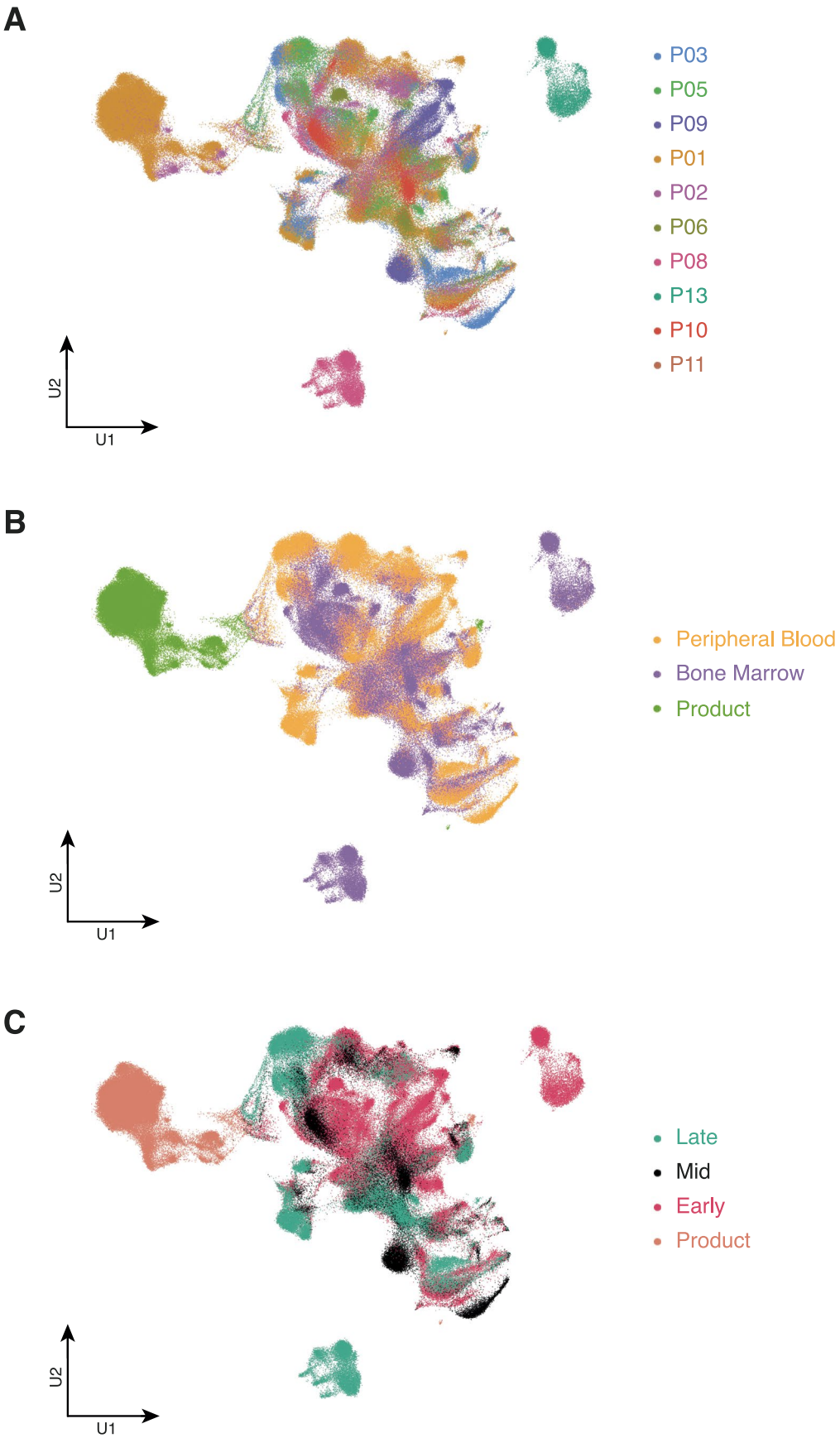
Peer review information *Nature Medicine* thanks Andrea Schmidts and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary handling editor: Ulrike Harjes, in collaboration with the *Nature Medicine* team.

https://doi.org/10.1038/s41591-023-02415-3



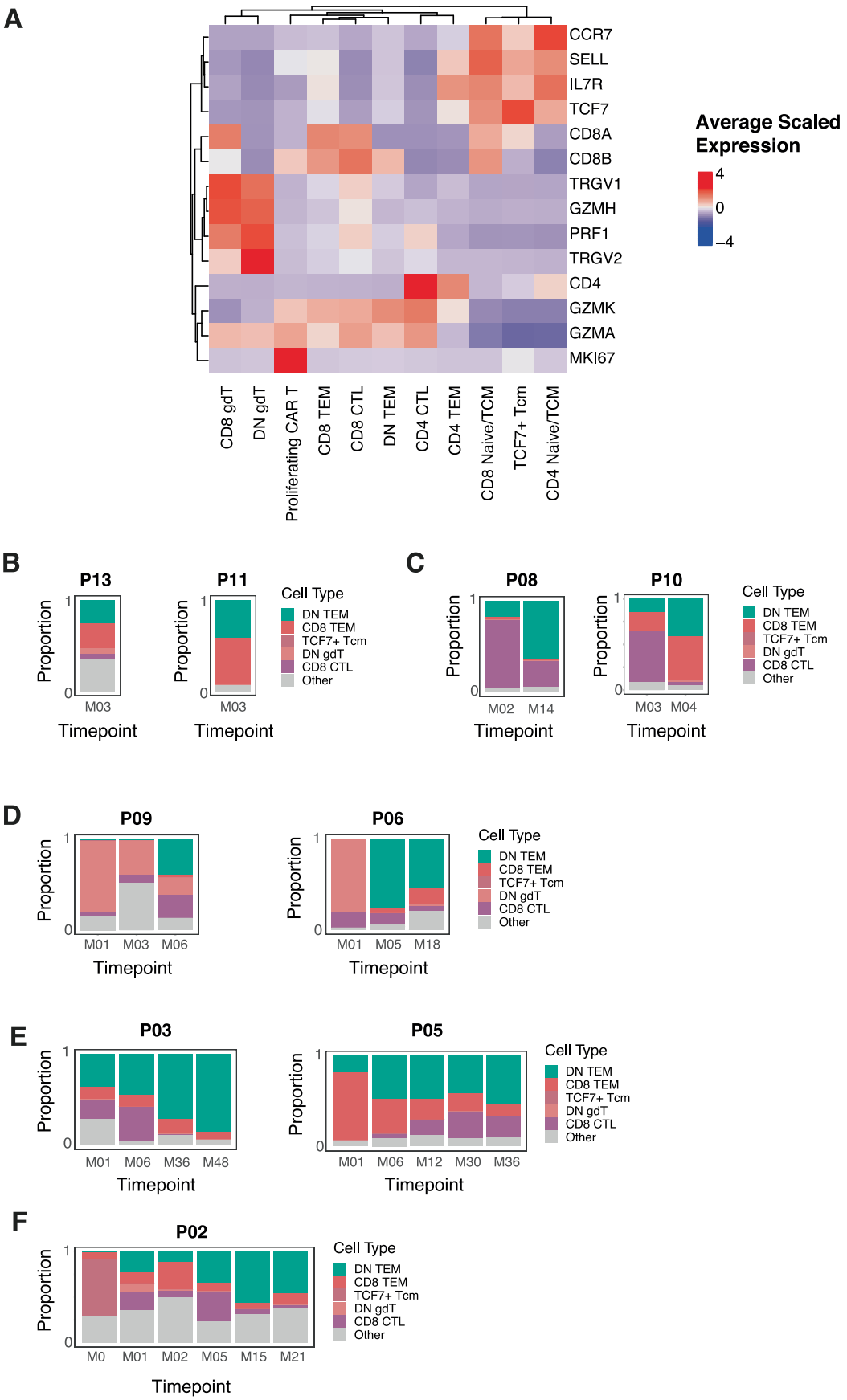
Extended Data Fig. 1 | Sample breakdown. Bar charts show the number of experimental samples derived from each patient sample. Each patient sample, defined as an individual, timepoint and sample source (that is P01 M01 PB) can be used for gene expression (GEX) and T-cell receptor (TCR) sequencing and/or flow-based immunophenotyping. For those samples with GEX/TCR sequencing, they are flow-sorted by CD3 and CAR prior to sequencing, with the exception of P02 M02 PB which was sorted for CAR-T cells, T cells and other leukocytes, thus they only contribute 1 experimental sample each. For two samples (P02 M06 BM and P01 M06 BM), no cells were recovered from the CAR T channel. M=month, PB = peripheral blood, BM = bone marrow.

https://doi.org/10.1038/s41591-023-02415-3



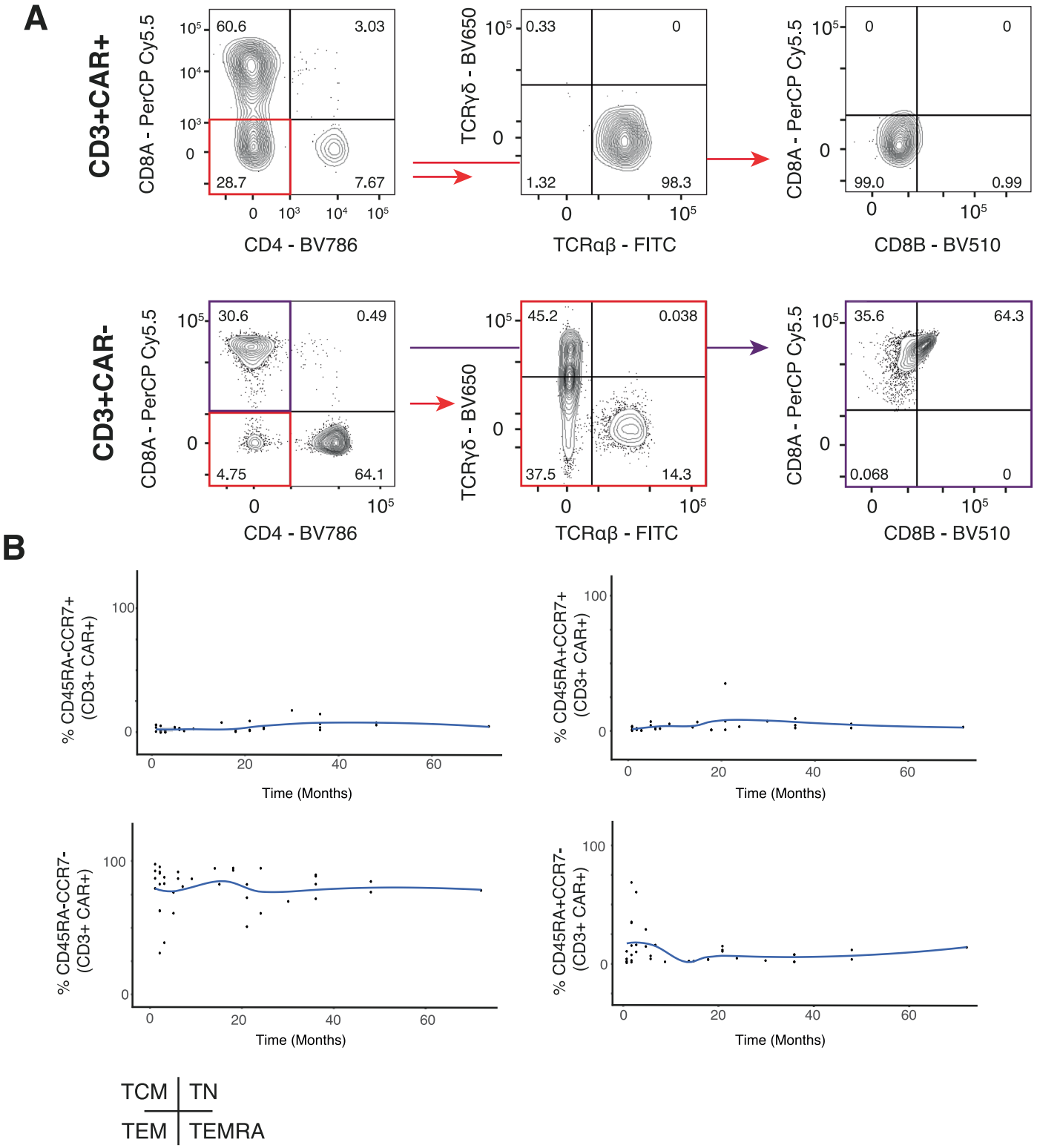
Extended Data Fig. 2 | Global cluster annotation. Uniform Manifold Approximation Projections (UMAPs) show all cells in the dataset coloured by **(a)** patient of origin, (**b**) sample source, and **(c)** timepoint.

https://doi.org/10.1038/s41591-023-02415-3



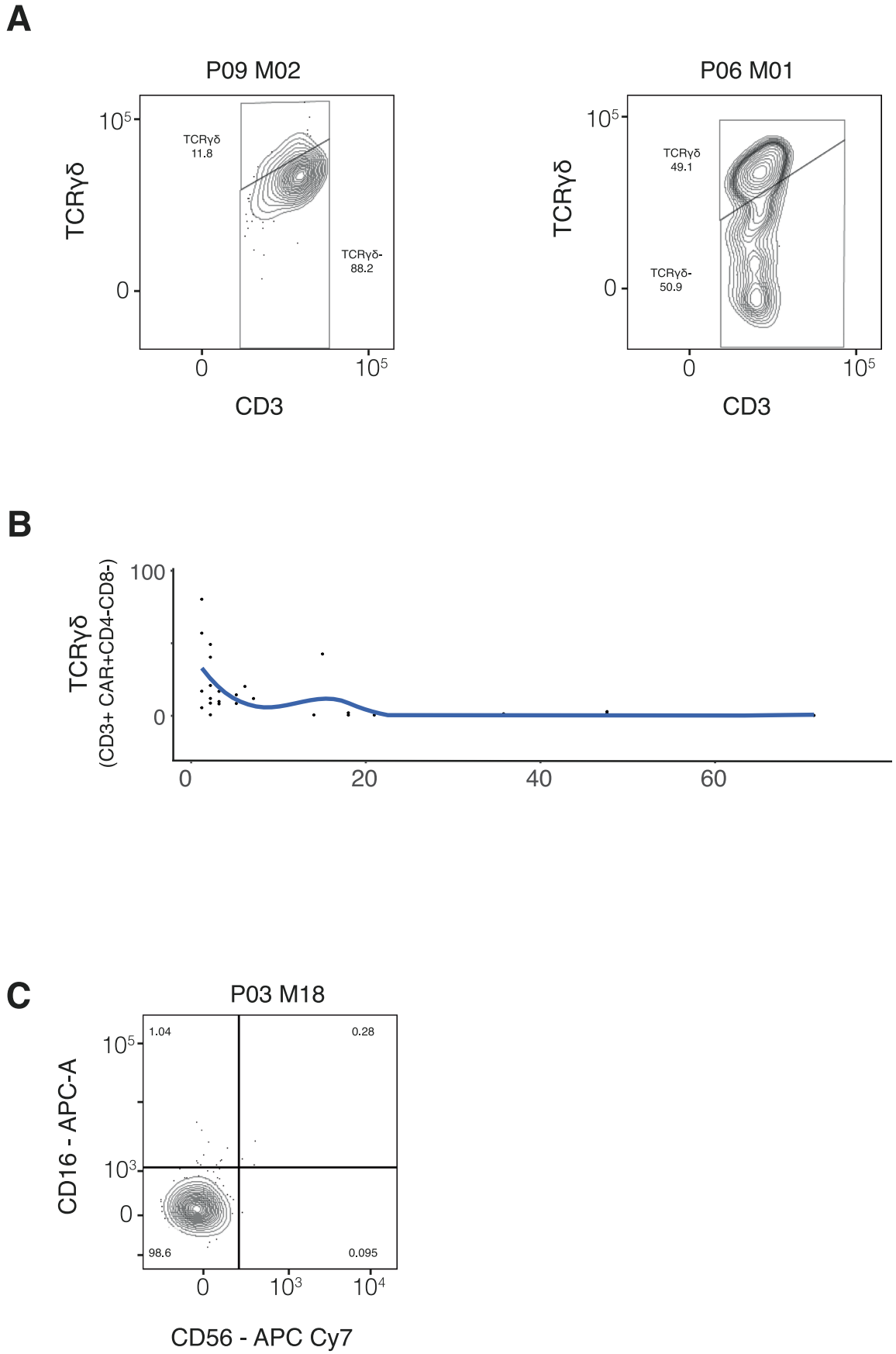
Extended Data Fig. 3 | Cluster markers and cell type trajectories. (a) Heatmap shows the average scaled expression of marker genes used for cell type annotation. **(b–f)** Stacked bar plots show dynamic cell type proportions over time for all patients. Top 5 abundant cell types in the dataset are shown.

https://doi.org/10.1038/s41591-023-02415-3



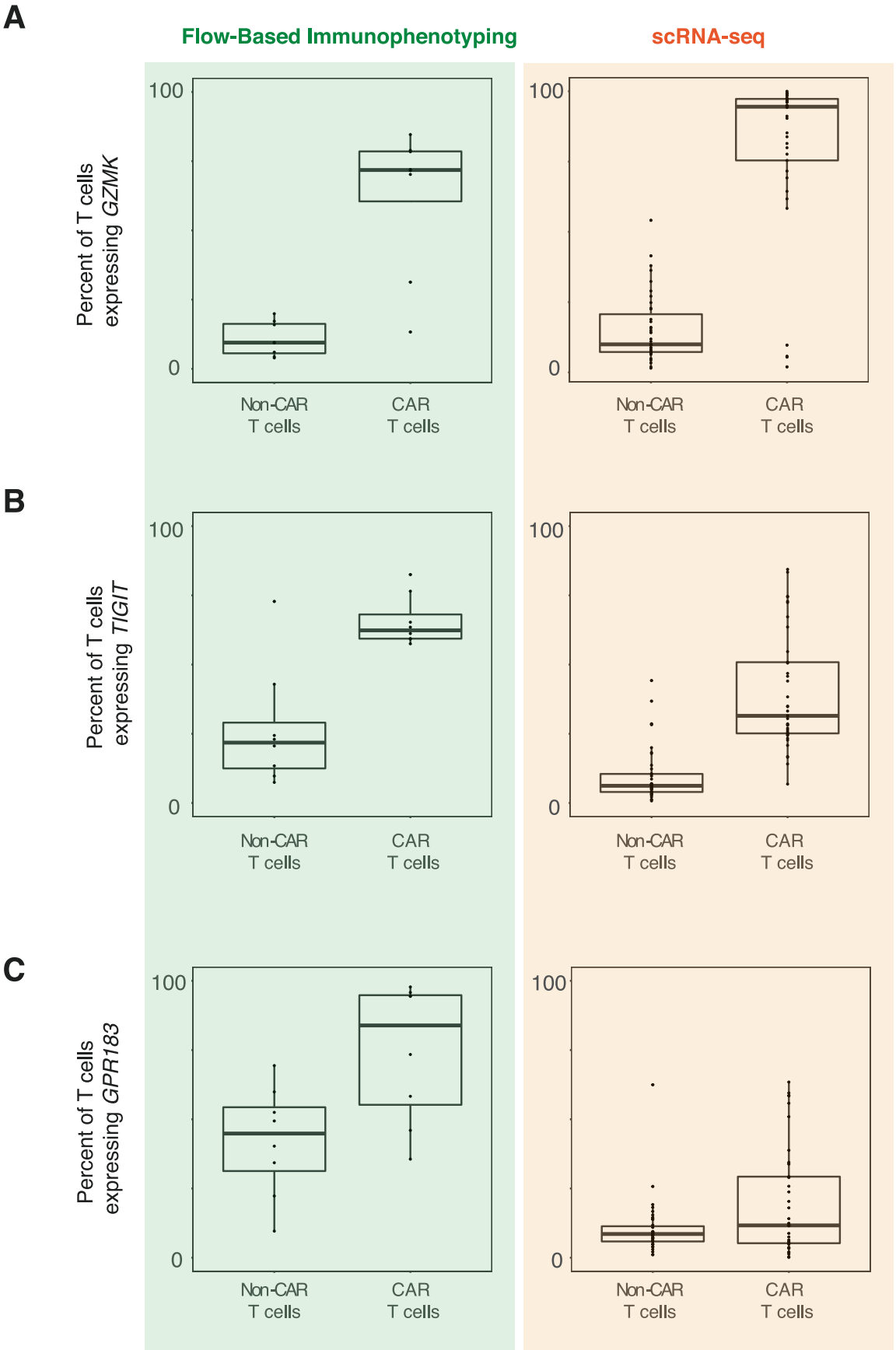
Extended Data Fig. 4 | General flow-based immunophenotyping. (a) Immunophenotyping non-CAR and CAR-T cells by multi-parameter flow cytometry. All cells were initially gated for CD3 and the CAR. Red boxes and arrows highlight subsequent gating strategies for representative examples illustrating double-negative CAR-T cells are αβ T cells and CD8A/CD8B negative, in contrast to non-CAR-T cells. Fluorochromes: PerCP-Cy5.5 = Peridinin chlorophyll protein-Cyanine5.5, FITC = Fluorescein isothiocyanate, Brilliant Violet (BV) 650, BV510, BV786. **(b)** Trajectory scatter plots quantify flow plots signals across patient samples for CD45RA and CCR7. Cells were gated for CD3 and the CAR. TN = Naïve, TSCM = Stem Cell Memory, TCM = Central Memory, TEM = Effector Memory, TEMRA = Terminally Differentiated Effector Memory expressing CD45RA.

https://doi.org/10.1038/s41591-023-02415-3



Extended Data Fig. 5 | Gamma-Delta and NK Immunophenotyping CAR-T cells by multi-parameter flow cytometry. All cells were initially gated for CD3, CAR, CD4 and CD8A. **(a)** Flow plots for representative samples (P06 and P09) with high populations of double-negative TCRγδ cells. **(b)** Trajectory scatter plots quantify flow plots signals across patient samples for TCRγδ. Cells were previously gated for CD3, CAR, CD4 and CD8. **(c)** Flow plots show gating for CD16 and CD56. Prior gating includes CD3, CAR, CD4, CD8. Fluorochromes: APC = Allophycocyanin, Cy = Cyanin.

https://doi.org/10.1038/s41591-023-02415-3

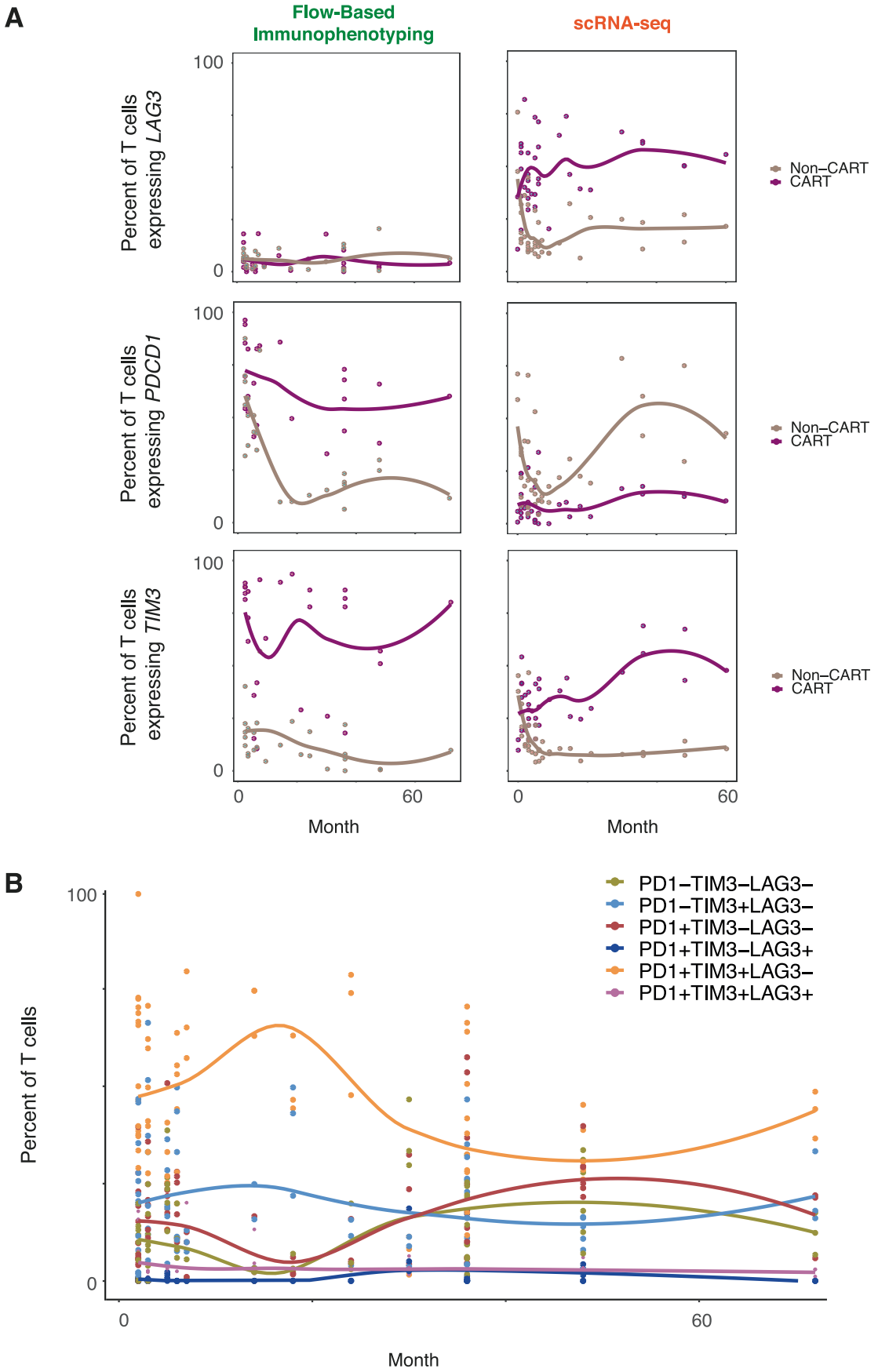


Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s41591-023-02415-3

Extended Data Fig. 6 | Expression of key genes in the persisting CAR T-cell signature. Boxplots show the percent of T cell expression for **(a)** *GZMK*, **(b)** *TIGIT*, and **(c)** *GPR183* for flow-based immunophenotyping (left) and single-cell RNA sequencing (scRNA-seq) data. CAR-T cell scRNA-seq, data is derived from 60188 cells from 43 independent samples. For non-CAR T-cell scRNAseq, data is derived from 98003 cells from 43 independent samples. Flow data is derived from 15 independent samples. Box plots show the first quartile (the lower end of the box) and the third quartile (the upper other end of the box), as well as the median values (centre line) per dataset. The 'whiskers' extend from the ends of the box to a maximum and minimum of 1.5 times the interquartile range beyond the box. Outliers are shown as dots.

https://doi.org/10.1038/s41591-023-02415-3

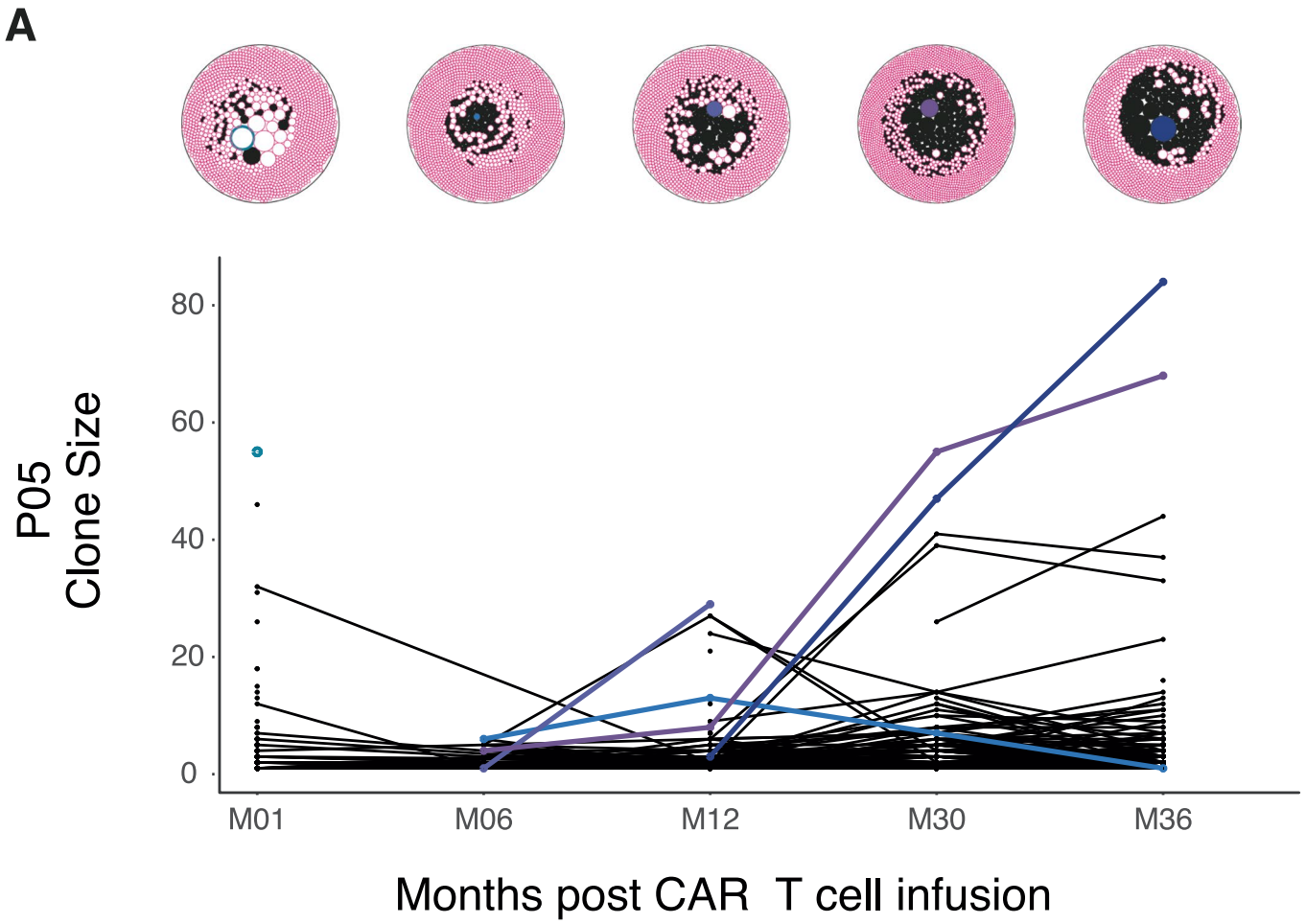


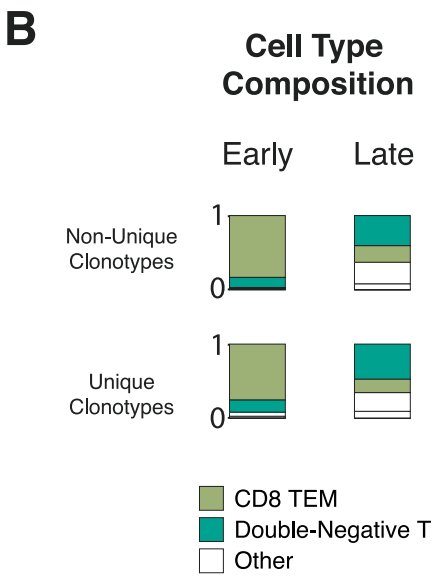
Extended Data Fig. 7 | See next page for caption.

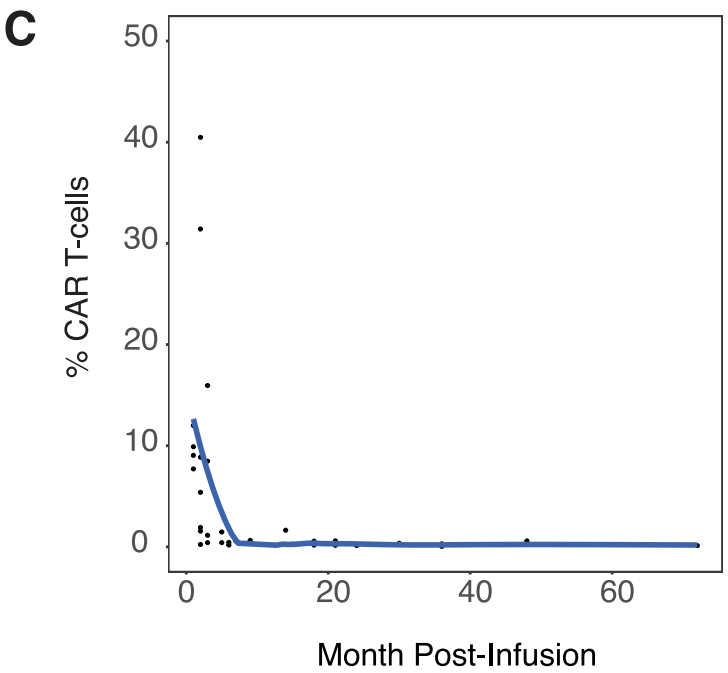
https://doi.org/10.1038/s41591-023-02415-3

Extended Data Fig. 7 | Exhaustion panel expression and flow-based immunophenotyping. (a) Trajectory scatter plots show the percent of T cells expression for *LAG3*, *PDCD1*, and *TIM3* across time for flow-based immunophenotyping (left) and single-cell RNA sequencing (scRNA-seq) data. Ribbons refer to the 95% confidence level interval for predictions using a 'loess' model to fit the line. **(b)** Trajectory scatter plots show the percent of T cells expression for the combination of *LAG3*, *PDCD1*, and *TIM3* across time for flow-based immunophenotyping. Ribbons refer to the 95% confidence level interval for predictions using a 'loess' model to fit the line.

https://doi.org/10.1038/s41591-023-02415-3





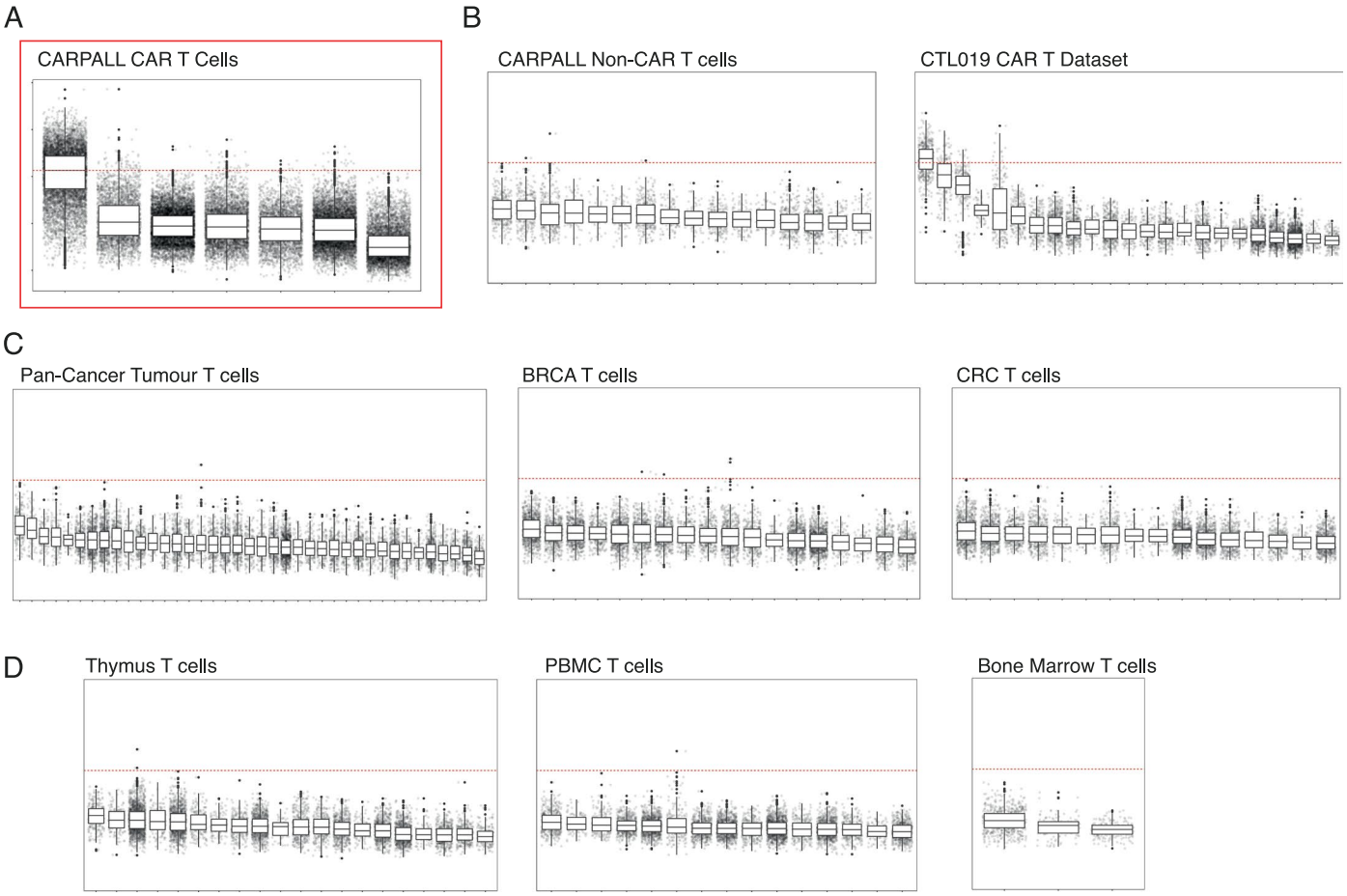


Extended Data Fig. 8 | See next page for caption.

https://doi.org/10.1038/s41591-023-02415-3

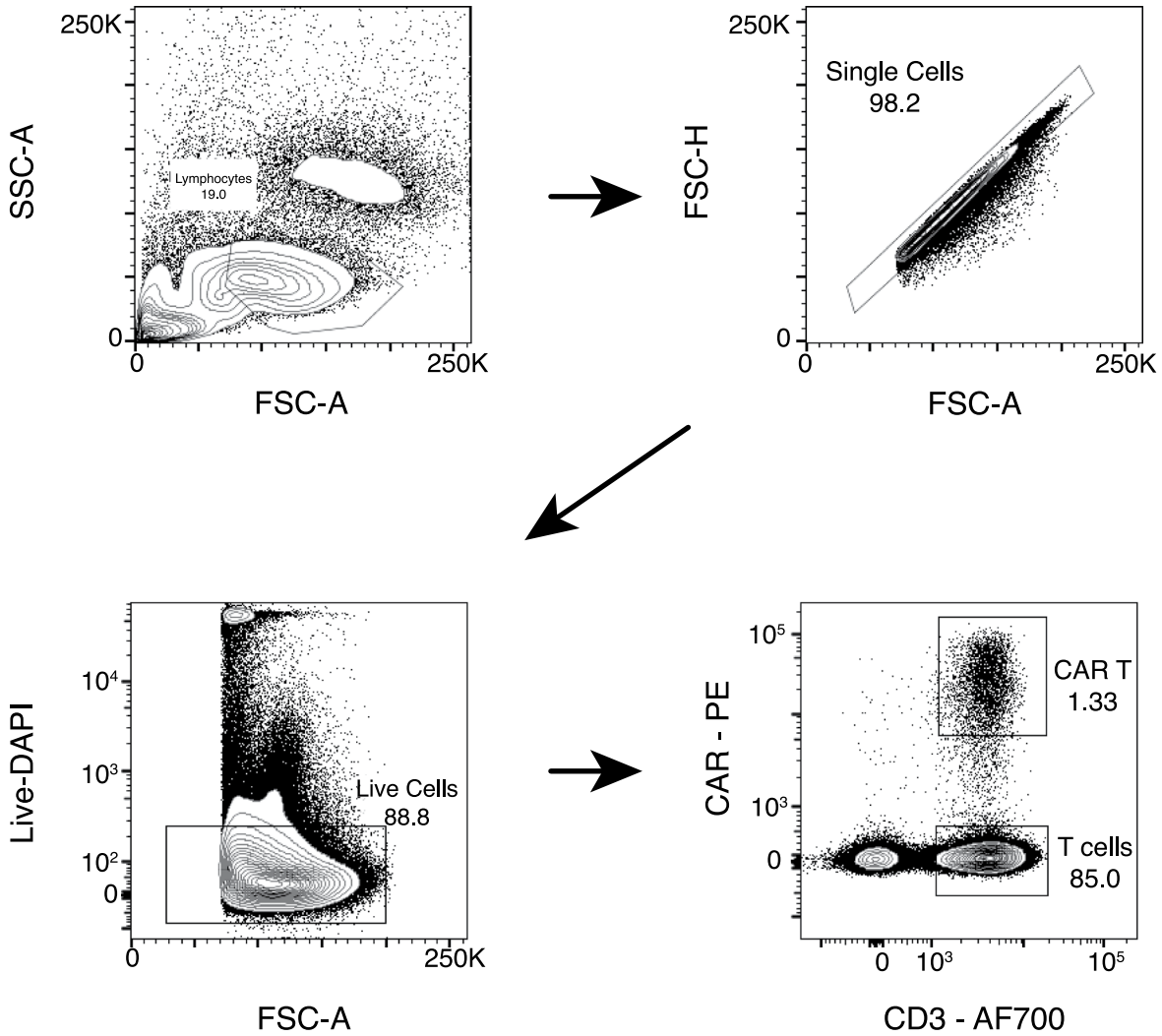
Extended Data Fig. 8 | Clonal structure of CAR-T cells. (a) Clonal trajectory illustrates the changes in clonal architecture of CAR-T cells over time across a second representative patients with more than 3 timepoints sequenced. Packed circle plots show the size of each clonotype. Filled-in black circles represent clonotypes that are not unique, as they are observed across timepoints. Conversely, pink donut circles represent clonotypes that are unique to that timepoint (and not observed across time). Blue/purple/grey coloured circles represent the dominant clonotype at that timepoint that correspond with the clonal trajectories below. **(b)** Cell type composition stacked bar plots demonstrate the shift in cell type abundances between early timepoints and late timepoints between unique and non-unique clonotypes (right). **(c)** Trajectory scatter plots quantify data from all patient samples for CAR expression across time.

https://doi.org/10.1038/s41591-023-02415-3



Extended Data Fig. 9 | Evaluation of the persisting CAR-T cell signature across T-cells: expanded view. Boxplots show the per cell distribution of the late score as defined by the late, persisting CAR-T signature. Scores are shown for representative datasets encompassing (**a**) the CARPALL CAR-T cells, (**b**) CARPALL Non-CAR-T cells, CTL019 CAR-T cells, (**c**) Tumour-infiltrating T cells, and (**d**) T cells from normal datasets. The red dashed line represents the median of the late double-negative (DN TEM) cells from the CARPALL study. Across datasets, the x-axis ticks refer to Seurat-based clusters per dataset, with the exception of the CARPALL CAR-T cell dataset where the x-axis refers to annotated cell types. Box plots show the first quartile (the lower end of the box) and the third quartile (the upper other end of the box), as well as the median values (centre line) per dataset. The 'whiskers' extend from the ends of the box to a maximum and minimum of 1.5 times the interquartile range beyond the box. Outliers are shown as dots. BRCA = Breast Cancer, CRC=Colorectal Cancer, PBMC = Peripheral Blood Mononuclear Cells. Publicly available datasets analysed in this study and the number of cells and samples used in this figure are described in Supplementary Table 6.

https://doi.org/10.1038/s41591-023-02415-3



Extended Data Fig. 10 | Flow cytometry gating strategy for CD3$^+$ CAR$^+$ T cells sorting. SSC = side scatter, FSC = forward scatter. Fluorochromes: FITC = Fluorescein isothiocyanate, PE = R-phycoerythrin, APC = Allophycocyanin.

nature portfolio

Corresponding author(s): Sam Behjati and Sara Ghorashian

Last updated by author(s): May 16, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Flow data analysis was performed using FlowJo v10 (Tree Star, Inc., Ashland OR), or FACs DIVA 8.0.1. Single-cell analyses was performed using the Seurat package in R (R version 4.0.3, Seurat version 4.0.6). The raw scRNA-seq data were demultiplexed and mapped to reference genome GRCh38, with the CAT-scFv sequence inserted, using Cell Ranger (10x Genomics, version 5.0.0). Chromium 10× V(D)J single-cell sequencing data were mapped and quantified using the software package cellranger vdj (v5.0.0) using the GRCh38 reference (vdj_GRCh38_alts_ensembl-5.0.0). Basic TCR statistics, such as the number of clones and the distribution of lengths and counts, were computed using Immunarch (version 0.7.0). Clonal population circles were created using the ggraph and igraph packages in R (version 2.0.5 and version 1.2.6, respectively). Logistic regression was performed using custom code found here: similarity.R in https://github.com/constantAmateur/scKidneyTumors.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw sequencing data produced in this study has been deposited at the European Genome-phenome Archive (accession number EGAD00001010018). This data is available under restricted access. Sequencing data requests will be reviewed by the Independent Data Monitoring Committee and Trial Management Group of the CARPALL study and will be subject to patient confidentiality. After approval, a data-access agreement with UCL will be required. All requests for raw materials will be reviewed by UCL Business (UCLB) to verify whether the request is subject to any intellectual property or confidentiality obligations. All requests will be processed within 8 weeks. Processed data has been uploaded to Zenodo. Publicly available datasets analysed in this study are described in Supplementary Table 6. The GRCh38 reference genome was downloaded from the 10X Genomics website: https://support.10xgenomics.com/single-cell-gene-expression/software/release-notes/build.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Study results do not apply to any one sex or gender. Sex or gender were not considered in the study design, as all high risk B-ALL children and young adults, independent of sex/gender, were considered. The sex of patients was noted and these are described in the table of patient characteristics (Supplementary Table 7).
Population characteristics	Children and young adults (age 24 years or younger) with high risk/relapsed CD19+ haematological malignancy.
Recruitment	Children meeting the inclusion criteria were recruited at Great Ormond Street Hospital, University College Hospital and Manchester Royal Children's Hospital as part of a previously reported clinical trial. Written, informed consent was obtained from all patients or their parents/guardians prior to study entry. The present secondary investigation did not actively recruit any new patients.
Ethics oversight	The present study is an investigation using samples collected from an existing clinical trial. Ethical approval was obtained from the London West London & Gene Therapy Advisory Committee (GTAC) Research Ethics Committee (REC ref. no. 16/LO/0283).

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

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Data from this study were generated from patients enrolled in the CARPALL study (NCT02443831). We performed detailed phenotyping by flow cytometry in 11 of 15 patients, and for ten children, sufficient CAR T-cells were obtained from serial time points for further interrogation by single-cell mRNA and TCR sequencing. Sample size calculation was not performed as samples were derived from whichever patients were enrolled in the study.
Data exclusions	No data were excluded
Replication	Single-cell RNA sequencing findings in this study were orthogonally validated using flow-based immunophenotyping and by comparing results to similar publicly available datasets. Experiments were not replicated and performed independently (each sample was sequenced once).
Randomization	Randomization is not relevant in this study, as we characterized long-lived CAR T-cells from children with leukemia. The study design does not include allocation of samples into experimental groups. All samples with sufficient material were sequenced and analyzed.
Blinding	Blinding was not relevant in this study as patients were not allocated to groups during data collection or analysis. The purpose of this study was to characterize long-lived CAR T-cells from children with leukemia.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

The following antibodies were used for identification and phenotypic analysis of CAR T cells in accordance with manufacturer's instructions or as indicated: CAR anti-idiotype antibody (bespoke product, Evitria, 1/200), PD-1 BV421 (EH12.2H7, Biolegend #329920, 1/20), CD45RA BV510 (HI100, BD #563031, 1/100), Lag3 BV605 (11C3C65, Biolegend #369324, 1/20), TCRgd BV650 (B1, BD #564156, 1/20), CD127 BV711 (HIL-7R-M21, BD #563165, 1/20), CD4 BV784 (SK3, Biolegend #344642, 1/100), CD25 VioBright FITC (4E3, Miltenyi #130-113-282, 1/100), Goat anti-Rat IgG PE (Poly4054, Biolegend #405406, 1/400), Tim3 PECF594 (7D3, BD #565560, 1/20), CD8 PerCP-Cy5.5 (SK1, Biolegend #344710, 1/40), CCR7 PE/Cy7 (G043H7, Biolegend #353226, 1/40), CD95 APC (581, Biolegend #305612, 1/10), CD3 AF700 (SK7, Biolegend #344822, 1/40), and CD27 APC/Cy7 (M-T271, Biolegend #356424, 1/20); TIGIT BV605 (741182, BD #747841, 1/40), GPR183 PE/Dazzle594 (SA313E4, Biolegend #368917, 1/40) and GZMK APC (GM26E7, Biolegend #370510, 1/40); CD45 FITC (2D1, Biolegend #368508, 1/20) and CD3 APC (UCHT1, Biolegend #300439, 1/20).

Validation

Primary antibodies were validated and titrated with appropriate antigen positive and negative controls in order to determine the optimal stain concentration for each test. CAR T cell detection with the CAR anti-idiotype antibody were validated against healthy donor controls and pure populations of CAR T cells in accordance with standard operating protocols in UK NEQAS and ISO accredited laboratories. The following list is an overview of the validations performed for each antibody:

Target ; Fluorochrome ; Clone ; Validation
PD-1 ; BV421 ; EH12.2H7 ; Resting and anti-CD3/CD28 activated PBMC measured at 24-, 48- and 72-hours post-activation
CD45RA ; BV510 ; HI100 ; Healthy donor PBMC T-cell population phenotyping
Lag3 ; BV605 ; 11C3C65 ; Resting and anti-CD3/CD28 activated PBMC measured at 24-, 48- and 72-hours post-activation
TCRgd ; BV650 ; B1 ; Healthy donor PBMC T-cell population phenotyping
CD127 ; BV711 ; HIL-7R-M21 ; Healthy donor PBMC T-cell population phenotyping
CD4 ; BV784 ; SK3 ; Healthy donor PBMC T-cell population phenotyping
CD25 ; VioBright FITC ; 4E3 ; Healthy donor PBMC T-cell population phenotyping
CAR anti-idiotype ; N/A ; Bespoke ; In vitro non-transduced and transduced CAR T cells with transgene distal 2A mCherry
Goat anti-rat IgG ; PE ; Poly4054 ; In vitro non-transduced and transduced CAR T cells with transgene distal 2A mCherry
Tim3 ; PECF594 ; 7D3 ; Resting and anti-CD3/CD28 activated PBMC measured at 24-, 48- and 72-hours post-activation
CD8 ; PerCP-Cy5.5 ; SK1 ; Healthy donor PBMC T-cell population phenotyping
CCR7 ; PE/Cy7 ; G043H7 ; Healthy donor PBMC T-cell population phenotyping
CD95 ; APC ; 581 ; Healthy donor PBMC T-cell population phenotyping
CD3 ; AF700 ; SK7 ; Healthy donor PBMC T-cell population phenotyping
CD27 ; APC/Cy7 ; M-T271 ; Healthy donor PBMC T-cell population phenotyping
TIGIT ; BV605 ; 741182 ; in vitro non-transduced and CAR T cells repetitively stimulated with CD3/CD28 dynabead and Nalm6 target cells
GPR183 ; PE/Dazzle594 ; SA313E4 ; Healthy donor PBMC T-cell population phenotyping and in vitro CAR T cells
GZMK ; APC ; GM26E7 ; Healthy donor PBMC T-cell population phenotyping and in vitro CAR T cells
CD45 ; FITC ; 2D1 ; Healthy donor PBMC T-cell population phenotyping
CD3 ; APC ; UCHT1 ; Healthy donor PBMC T-cell population phenotyping

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration

NCT02443831

Study protocol

http://www.ctc.ucl.ac.uk/TrialDetails.aspx?Trial=116&term=carpall

Data collection

The present study is not a clinical study but uses samples from a previous clinical trial. Patient data were collected at Great Ormond Street Hospital (GOSH) and UCL-GOSH Institute of Child Health (ICH), whereas laboratory data were generated in the study central

	laboratories at GOSH, UCL-GOSH ICH, as well as the Sanger Institute. Patient recruitment occurred from 2016-2019. Data collection, sequencing, and analysis was from 2016-2023.
Outcomes	The present study is not a clinical trial, but uses samples from a previous clinical trial. Primary and secondary outcomes from the initial trial can be found in the above protocol. There were no primary or secondary outcomes reported in this manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	For flow cytometry-based immunophenotyping, patient cells were acquired from either fresh peripheral blood or cryopreserved aliquots of the infusion product (IP), peripheral blood mononuclear cells (PBMC) or bone-marrow mononuclear cells (BMMC). For fresh peripheral blood, PBMC were isolated via density gradient centrifugation with Lymphopure (Biolegend). For cryopreserved samples, aliquots were rapidly thawed and washed in complete RPMI (10% FCS and 1% L-glutamine, Gibco). Up to 10x10^6 cells were stained in 100ul of Cell Staining Buffer (Biolegend). For intracellular markers, cells were fixed (Fixation Buffer, Biolegend) and permeabilised (Intracellular Staining Permeabilization Wash Buffer 10X, Biolegend) prior to staining. Human BD Fc Block (BD) was used as a blocking reagent to reduce non-specific interactions. For fluorescence-activated cell sorting, patient cells were acquired from either fresh peripheral blood or cryopreserved aliquots of the IP, PBMC or BMMC. Fresh peripheral blood was processed as described and cryopreserved samples were rapidly thawed and washed with complete RPMI containing 50U/mL of Benzonase Nuclease (Merck Life Science Limited). Up to 10x10^6 cells were stained in 100uL of Cell Staining Buffer (Biolegend). Human BD Fc Block (BD) was used as a blocking reagent to reduce non-specific interactions.
Instrument	Flow cytometry was performed with a BD LSR II (4-laser configuration; 355nm, 405nm, 488nm and 633nm) and cell sorting with a FACS Aria III (4-laser configuration; 405nm, 488nm, 561nm and 647nm) (BD Biosciences)
Software	Data analysis was performed using FlowJo v10 (Tree Star, Inc., Ashland OR), or FACs DIVA 8.0.1.
Cell population abundance	Cell population abundances are shown in the flow cytometry plots and tables of raw data.
Gating strategy	For immunophenotyping, CAR-T cells were isolated as viable CD3+/CAR+ events in a singlet leukocyte forward-scatter (FSC)/ side-scatter (SSC) gate For cell sorting, CAR-T cells were isolated as viable CD45+/CD3+/CAR+ events in a singlet leukocyte forward-scatter (FSC)/ side-scatter (SSC) gate.

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