Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Multi-omics DNA and Protein Characterization in CAR T Cell Therapy

Jinxia Ma
National Institutes of Health
Yilong Yang
Mission Bio Inc.
Ruyan Yang
National Institutes of Health
Yihua Cai
National Institutes of Health
Yong Soo Kim
National Institutes of Health
Shu Wang
Mission Bio Inc.
Hua-Jun He
National Institute of Standards and Technology
Zhiyong He
National Institute of Standards and Technology
Nan Zhang
National Institutes of Health
Pamela Garces
National Institutes of Health
Kathryn Martin
National Institutes of Health
Lipei Shao
National Institutes of Health
Anh Dinh
National Institutes of Health
Steven Highfill
National Institutes of Health
Jiaqiang Ren
NIH: National Institutes of Health
Robert Somerville
National Institutes of Health
David Stroncek

National Institutes of Health

Ping Jin

p.jin@mail.cc.nih.gov

National Institutes of Health

Research Article

Keywords: CAR T-cell therapy, Single-cell DNA + Protein sequencing, T-cell subsets, populations within transduced samples

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11529336/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

Background: Transgene copy numbers influence the efficacy and safety of CAR T-cell therapy. Single-cell DNA sequencing, combined with protein analysis, enables high-resolution, simultaneous assessment of genomic and proteomic features at the single-cell level. This study investigates the distribution of vector copy numbers, quantifies transduction efficiency (TE), explores genotype–phenotype relationships, characterizes T-cell subsets and subpopulations in transduced (TR) samples, and evaluates whether distinct T cell subsets differ in their transduction capacity and transgene expression.

Methods: We used the Tapestri multi-omics platform to perform targeted single-cell DNA and protein profiling. A single assay was designed for lentiviral constructs, incorporating60 protein markers, including TE markers. Four non-transduced (UT) and five CD22 CAR transduced (TR) samples were tested. Sequencing data were analyzed using custom workflow based on the Tapestri DNA/DNA + Protein Analysis Pipeline.

Results: Untransduced samples were essentially free of vector signal (≤ 0.1% of cells with VCN >0), while spiked-in Jurkat reference cells carrying four vector copies served as calibration control. Most transduced cells carried one to three copies, with two-copy integrations the largest single fraction, and cells with five or more copies were rare. Surface CAR density increased with vector copy number before plateauing, indicating a saturating rather than linear relationship. Five T-cell subsets were identified in both the UT and the CD22 CAR transduced samples. Transduction efficiency and the ability to express integrated vector copies differed among various T-cell subsets. Transduced samples were defined as double-negative (DN), staining-background (SB), insert-not-expressed (INE), and double-positive (DP). The INE population was the least abundant, the naïve T-cell subset (Tn) showed the highest INE frequency than other T-cell subsets, stem cell memory T-cells (Tscm) had the lowest transduction efficiency; central and effector memory T-cells (Tcm, Tem) exhibited similar levels of transgene-expressing transduced cells.

Comparing DP CAR T-cells with DN, INE, and UT controls shows that reference selection dictates biological interpretation, highlighting the need for multiple complementary controls in phenotypic studies.

Conclusions: Our data offer enhanced resolution and insight into the potential safety of CAR-T cell therapy, and these findings support further investigation of associations between single-cell data and clinical therapeutic response.

Introduction

Chimeric antigen receptor (CAR) T-cell therapy is an advanced form of immunotherapy for the treatment of malignancies that harnesses the power of a patient's own immune system to identify and destroy cancer cells (1). It has shown remarkable success, especially in treating hematologic malignancies such as B-cell acute lymphoblastic leukemia (B-ALL), non-Hodgkin lymphoma, and multiple myeloma, often

when other treatments have failed (2–5). The U.S. Food and Drug Administration (FDA) has approved CAR T-cell therapy for several types of blood cancers. In many cases, CAR T-cell therapy has achieved high remission rates, offering hope to patients with advanced, treatment-refractory cancers (6).

Transgene vector copy number (VCN), which refers to the average number of copies of the viral vector carrying the CAR gene that integrates into the genome of each T cell, is a critical quality control measure that balances the therapy's effectiveness with its potential risks (7). The number of integrated vector copies can influence CAR expression on the T-cell surface, although CAR expression is also affected by vector design and regulatory elements (8). This ensures they can effectively recognize and destroy cancer cells. Conversely, a very high VCN increases the risk of insertional mutagenesis. This occurs when the viral vector integrates into a sensitive area of the host cell's genome, potentially disrupting normal gene function and leading to the development of new cancers (oncogenesis) (9, 10). The FDA's guidelines emphasize the importance of VCN as a critical quality attribute and a key component of the overall safety of the therapy. FDA recommends that the vector copy number (VCN) release criterion be justified based on a product-specific risk assessment, which may include supporting data from insertion-site analysis, clonal dominance assessments, dose, indication, study population, and other relevant factors, as described in FDA's final guidance (11).

A variety of molecular techniques, ranging from classic to cutting-edge, are employed to measure transgene copy numbers of cell therapies. The techniques include quantitative PCR (qPCR), Droplet Digital PCR (ddPCR), and Next-Generation Sequencing (NGS) (12).

qPCR has emerged as a rapid, sensitive, and high-throughput alternative to Southern blotting for transgene copy number analysis (13). This method requires significantly less DNA and can be automated, making it suitable for screening large numbers of samples. It requires a standard curve to calculate the relative quantity of the transgene, which can introduce variability.

ddPCR represents a significant advancement in the precise and absolute quantification of nucleic acids, making it exceptionally well-suited for determining transgene copy number (14). ddPCR provides an absolute quantification of the target DNA concentration without the need for a standard curve (15).

NGS can determine the precise integration site(s) of the transgene within the host genome and the exact number of integrated copies, even if they are arranged in tandem repeats (16). It can also detect structural rearrangements or mutations at the integration site and provide data on the zygosity of the transgene (presence on one or both chromosomes). The field of CAR T-cell therapy is continuously evolving, with ongoing research that is focused on improving the safety and efficacy of these treatments. In addition, new analytical methods have become available. The Tapestri platform enables high-throughput single-cell multi-omics analysis, combining targeted DNA sequencing with protein measurements from the same individual cells (17, 18). This innovative approach provides a powerful tool for dissecting cellular heterogeneity, particularly in complex diseases such as cancer and hematologic disorders. Tapestri's single-cell DNA + protein analysis platform is particularly valuable for characterizing transduced CAR T-cell populations, including cell-to-cell variation in vector copy number, transduction

status, and immunophenotype, offering insights that are not accessible through bulk or single-modality approaches. This approach has recently been qualified for single-cell vector copy number measurement in CAR T products, demonstrating concordance with orthogonal ddPCR across defined reference cell lines and resolving copy number distributions that are indistinguishable by bulk average measurements (19).

We used Tapestri's single-cell protein plus DNA analysis cutting-edge multi-omics approach to conduct single cell DNA and antibody sequencing (DAb-seq) on CD22 CAR T-cell products. Simultaneous DNA and protein profiling enabled us to examine the landscape of T-cell subtypes and transduction efficiency among CAR T-cells and identify four populations among the transduced CAR T-cell samples at the single-cell level. This study established a correlation between VCN and cell surface marker expression, evaluating transduction efficiency using DNA profiles alongside surface markers, and identifying subpopulations within the transduced cell population of CAR-T products at the single cell level.

Materials and Methods

Product sample collection

Autologous peripheral blood mononuclear cells (PBMCs) collected by apheresis were cultured and transduced with a lentiviral vector encoding a CD22 CAR. Samples were collected from the CD22 CAR T-cell final products nine days after initiation of the transduction. Each tube containing 4×10^6 of CAR T-cell product cells was centrifuged at 400 × g for 5 minutes, and the supernatant was removed. The cells were resuspended in 500 µL of 4% HSA Plasma Lyte-A, and then 500 µL of CryoStor® CS10 Freeze Media was added in a drop-wise manner. Following gentle pipette mixing, the cells were transferred to a cryovial, placed in a Corning CoolCell™ cell-freezing container, frozen, and stored at − 80°C overnight. The cells were then transferred to a liquid nitrogen storage tank the next day.

Amplicon design and synthesis

The DNA sequence of the CD19 CAR construct obtained from NIST was used to design a 124-amplicon DNA panel (99 human amplicons, 15 CD19 vector amplicons, 3 CD22 CAR identity amplicons, and 7 vector identity amplicons as negative-controls). Designs were based on hg38 + vector + CAR sequences, with amplicons ranging from 196–267 bp. NIST provided a research Jurkat clonal cell line containing four integrated CD19 CAR copies to use as the VCN control. Due to the difference between the CD19 CAR vector and CD22 CAR vector, 8 out of the 15 CD19 vector amplicons are predicted to be shared by the CD22 CAR vector. The primers were synthesized and validated by Mission Bio.

The samples were analyzed using a copy-number caller to obtain single-cell VCN values, population-average VCN, and percent transduction. Cell identities were determined by combining correlation UMAP with completeness algorithms. Untransduced control cells showed $VCN \leq 0$, transduced cells had *VCN*

$>$ 0, and Jurkat controls had VCN = 4. All runs are summarized in Table S1, with an average of 6,893 cells analyzed per DAb-seq sample.

Antibody panel and antibody-oligonucleotide conjugate (AOC) preparation (titration)

The TotalSeq™-D Heme Oncology Cocktail from BioLegend was used for DNA and protein multi-omics characterization of cancer cells. It contained 45 unique cell surface antigens (including principal lineage antigens) and 3 isotype controls. Ten additional AOCs for T cell subset analysis and five for transduction efficiency analysis were included. A total of 60 antigens were evaluated. Among the AOCs used were CD11a, CD27, CD28, CD57, CD58, CD95, CD122, CD127, CD183, and CD197 (Table S2). Five identity markers (EGFR, Recombinant protein L, Recombinant Human siglec-2, CD19 idiotype, TCR beta) were also included. PBMCs and 4P8P cells were used for titrations. The AOCs were synthesized by BioLegend and titrated in the lab using PE-conjugated antibodies, following BioLegend and Mission Bio recommendations. Serial twofold dilutions (five per antibody) determined final concentrations based on separation of negative and positive populations. Final concentrations were validated on the Tapestri platform before sample testing.

Single-cell DNA amplicon and antibody-tag sequencing library preparation using the Mission Bio Tapestri platform

Preparation of cell suspension for staining

Fixed control cells from NIST were spun down, washed twice with DPBS + 1% BSA, then diluted to 2,800–3,200 cells/µL in cell buffer and stored on ice. The cells to be tested were thawed with pre-warmed media, gently mixed, and brought to 10 mL. After centrifugation, cells were washed, counted, resuspended in Cell Stain Buffer (CSB), diluted to 3×10^6 cells/40 µL, and stained with protein L.

Antibody oligonucleotide conjugate (AOC) Staining

For Protein L staining, cells were blocked with Human TruStain FcX and Blocking Buffer and incubated on ice. They were then stained with oligo-conjugated protein L and cell staining buffer (CSB). After incubation, cells were washed twice by centrifugation with pre-chilled CSB.

A custom antibody–oligo conjugate panel was prepared, added to the TotalSeq™-D Heme Oncology Cocktail and used for cell staining. Cells were blocked, stained with the cocktail, washed, filtered, and resuspended. Cell concentration was adjusted, 20% fixed cells were spiked in, and samples were kept on ice until processing.

Single cell DNA and antibody sequencing (DAb-seq)

Single-cell DNA and protein sequencing were performed on the Tapestri platform. Cells were stained, washed, encapsulated, lysed, digested, and barcoded. Targeted DNA was amplified, purified, indexed, and libraries were quantified and normalized. Sequencing was carried out on an Illumina NextSeq 550, with DNA and protein libraries loaded at specific concentrations with PhiX controls.

Single-cell DNA and protein sequence data analysis

Single-cell DNA and protein sequencing data were analyzed using a custom workflow based on Tapestri DNA/DNA + Protein Analysis Pipeline (v3.0) (https://support.missionbio.com/hc/en-us/sections/17820181627415-Tapestri-DNA-DNA-Protein-Analysis-Pipeline), which performs read alignment, cell calling, and generates processed sequencing data in .h5 files. Cell calling was exclusively performed using human amplicons within the panel. For downstream analysis and visualization, processed .h5 files were imported into the Mosaic v3.5 python package (https://missionbio.github.io/mosaic/index.html).

Genotype demultiplexing

To assign cells to their respective cell lines, DNA germline variant genotyping was performed against an in-house curated database. Cells were assigned to a specific genetic background (e.g., Jurkat cell lines) if > 60% of their known germline variants matched the reference, using the Dna.assign from truth function in the Mosaic package. Following initial assignment, the most prominent cell clusters were identified as the populations (designated 'population-1' and 'population-2)' after manual inspection of heterozygosity. Minor clones were subsequently evaluated using a likelihood model to assess their similarity to these identified major populations. Clones that demonstrated the highest likelihood of representing a combinatorial population derived from multiple major clones were classified as 'Mixed' while remaining undefined clusters were labeled as 'Ambiguous.' Both 'Mixed' and 'Ambiguous' populations were excluded from downstream analysis. This demultiplexing strategy effectively removed the majority of doublets or multi-cell clusters with distinct genotypes.

Immunophenotyping using antibody markers

Protein data normalization was performed on a per-sample basis using the Noise-corrected and Scaled Protein counts (NSP) protocol, specifically utilizing the protein.normalize_reads function within the Mosaic package.

Cell type assignment employs a two-round approach:

1. Preliminary Assignment (PACE Module):

The initial assignment employed the PACE module in Mosaic, which applied the default antibody marker expression profile to assign standard PBMC cell types, including T-cell, CD4 + T cell, CD8 + T cell, NK cell, B cell, Monocyte, and Dendritic cell.

Additionally, "sticky cells" are identified by detecting positive signals for the following antibody markers: "IgG1", "IgG2a", "IgG2b", and "CD83".

1. Secondary Assignment (Fixed Threshold):

The second round refined the classification using a fixed threshold (normalized marker read count 0.6) as detailed in the table below. The markers CD45RA, CD45RO, and CD183 initially identify three of the five T-cell subsets. To further differentiate Naïve T-cells and Effector T-cells, the following specific criteria are applied:

Naïve T-cells require both CD62L and CD27 markers to be above the threshold.

Effector T-cells require any one of CD45RA, CD45RO, or CD183 and any one of CD62L, CD27, or CD28 markers to be below the threshold (Table S3).

Vector copy number analysis

Read counts and cell line control information are processed by the in-house developed cnvcaller v1.0.1 pipeline as previously described (Yang, Parikh et al. 2026). Briefly, read counts for each vector amplicon are modeled with a three-parameter negative binomial distribution accounting for amplicon performance, total read depth, and cell-to-cell variability; per-cell copy number is assigned by maximum likelihood summed across amplicons, and the population-level copy number distribution is refined using the weighted mixture method. For Tapestri runs that include VCN control cell lines genetically distinguishable from the test cells, the read counts from the spiked-in control cells are used as control input to model the amplicons as $VCN = 4$.

There are eight vector amplicons which are shared between the CD19 CAR vector and CD22 CAR vector, and were used in this analysis ("CO748_AMP11", "CO748_AMP12", "CO748_AMP13", "CO748_AMP14", "CO748_AMP15", "CO748_AMP16", "CO748_AMP22", and "CO748_AMP23"). The maximum allowed VCN is 9.

Read counts for distinct cell populations, such as those segregated by patient genotypes, T-cell subsets, and/or protein L signal levels, are run through the VCN pipeline individually, using the identical parameters described above.

Assessment of AOC expression using different references

We performed statistical comparisons between the test group and reference groups usingthe two-sided Mann–Whitney U test, with the rank-based effect size Cliff's delta (δ) computed from the same U statistic as $\delta = 2U/(n_1 n_2) - 1$, where n_1 and n_2 denote the number of cells in the test and reference populations, respectively, and U is the rank-sum statistic for the target population; ties were assigned a value of 0.5. This formulation is mathematically equivalent to the proportion of cell pairs in which the test cell exceeds the reference cell minus the proportion in which it is lower, and yields values in [− 1, 1], where positive values indicate higher marker expression in the target population. An AOC marker was

considered significantly altered if it met both statistical and magnitude criteria: the resultingMann–Whitney p-value had to be less than or equal to a predefined significance threshold (0.001), |δ| >= 0.3 and the change between group means had to be substantial. This substantial change was defined as either the means of the two groups completely separating across a predefined boundary value (0.6), or the raw difference between the means exceeding the 0.6 threshold. Markers satisfying these criteria were subsequently classified as "upregulated" or "downregulated" based on whether the mean difference was positive or negative, respectively.

Vector copy number by Droplet Digital PCR (ddPCR)

An Auto DG QX200 ddPCR system (Bio-Rad Laboratories, Hercules, CA) was used for all ddPCR experiments analyzing extracted DNA for vector copy numbers. The final concentrations of forward and reverse primers were 900 nmol/L and 250 nmol/L for probes. Droplets were generated using a QX200 droplet generator followed by PCR using a T100 thermal cycler (Bio-Rad, CA) with PCR parameters of 95°C for 10 min; 40 cycles of 30 s at 94°C and 1 min at 60°C, followed by 98°C for 10 min. The plate was loaded into the QX200 droplet reader (Bio-Rad, CA) after PCR amplification. The data were analyzed using QuantaSoft software (Bio-Rad, CA). rPP30 assay (probe fluorophore: HEX) sequence information was adapted from Bejarano et. al. (20), and the Gag assay (probe fluorophore: FAM) was a courtesy of Lentigen (Miltenyi Biotec, Inc., MD).

Flow cytometry analysis

CAR T-cell product samples were stained with fluorochrome-conjugated anti-human antibodies targeting CD3, CD4, CD8, CCR7, CD45RA, and CD45RO. Antibodies were obtained from Waters™ Biosciences (Franklin Lakes, NJ, USA), except for CD27, CD62L, and CD45RO, which were purchased from BioLegend (San Diego, CA, USA). Fluorochrome-conjugated recombinant human Siglec-2/CD22 protein (R&D Systems, Minneapolis, MN) was used to detect the identity of the CAR T cell product. Cell viability was assessed by staining with 7-aminoactinomycin D (7-AAD; BD Biosciences, NJ). Following a 20-minute incubation at 4°C, samples were washed using the BD FACS™ Lyse Wash Assistant (LWA) and acquired primarily on a BD FACSCanto™ 10-Color flow cytometer, with one sample acquired on a BD FACSLyric™ 12-Color system (Waters™ Biosciences). Data were analyzed using FlowJo software (Waters™ Biosciences, NJ).

For memory and effector differentiation analysis, CD45RA, CD45RO, and CCR7 expression was used to define four phenotypic T-cell subsets: naïve/stem cell memory cells (Naïve/Tscm; $CD45RA^{+}CCR7^{+}CD45RO^{-}$), central memory cells (Tcm; $CD45RA^{-}CCR7^{+}CD45RO^{+}$), effector memory cells (Tem; $CD45RA^{-}CCR7^{-}CD45RO^{+}$), and effector cells (Teff; $CD45RA^{+}CCR7^{-}CD45RO^{-}$).

Results

CD22 CAR T-Cell Samples

Apheresis products from five patients (one female, four males), aged five to 49 years at the time of treatment, were collected between March 2023 and July 2024. Four patients had a diagnosis of Acute Lymphoblastic Leukemia (ALL), and one had Lymphoma. The patients' PBMCs were collected by mononuclear cell apheresis, T cell enriched, cultured, and transduced with a lentiviral vector encoding a CD22 CAR. Final CAR T cell products were collected for the simultaneous measurement of VCN and surface protein markers by Tapestri Single-cell DNA + protein multi-omics platform. The CAR T-cell were also analyzed for VCN by ddPCR and surface protein markers by flow cytometry for comparison with the Single-cell DNA multi-omics data (Table 1).

Amplicon panel design and VCN evaluation

Targeted amplicon sequencing was used to evaluate total variants and the specificity of vector-derived reads in the CD22 CAR transduced (TR) and untransduced (UT) T-cell samples. A total of ten samples, including both UT samples and TR samples, were analyzed. For the CD22 CAR T-cell samples, a vector transduced Jurkat cell line was spiked in. All heatmaps had the same set of sorted variants, which meant that UT and TR samples shared the same subclones in patient CAR T-cells and viral transduction had no effect (Figure S1A & S1B). Samples TR3 and TR5 had three different genotype clones, including one from the spiked-in Jurkat cell line and another that may have resulted from prior allogeneic CAR T-cell therapy. The other 4 TR samples had a single genotype clone identified. Across the UT samples (UT3–UT6) (Figure S1A), read counts mapping to vector regions were absent in nearly all cells with 97.9–100% of the cells having zero vector reads, while endogenous genomic regions showed consistent baseline coverage (Mann–Whitney U test, Benjamini–Hochberg FDR > 0.05 for all 99 human amplicons). For the Jurkat cell line all the expected vector amplicons were detected. This pattern confirmed that the assay selectively amplified vector-associated sequences, while minimizing nonspecific amplification. In the CD22 CAR T-cell product samples (TR3 to TR7) (Fig. 1A and Figure S1B), distinct clusters of enriched amplicons corresponding to vector components were observed. The distribution of signals across vector elements was consistent between replicates (Spearman rho: minimum = 0.97, median = 0.99) in CD22 CAR T cells, demonstrating reproducibility of the assay. Vector composition analysis further confirmed that vector reads mapped predominantly to expected vector regions (3 CAR1 amplicons and 8 vector amplicons, Figure S1B). More specifically, across the five transduced samples, an average of99.89 % (rnge: 99.83% – 99.90%) of all vector-aligned reads mapped to these 11 target regions, supporting the specificity of amplification and accurate detection of vector integration components.

VCN analysis revealed negligible copy numbers in UT samples (< = 0.1% of cell with VCN > 0 in each UT sample), consistent with the absence of vector integration. In contrast, the CD22 CAR T-cells product samples showed measurable vector copy numbers with variability between samples, reflecting differences in transduction efficiency (Fig. 1B and 1C). Analysis of vector copy number distribution showed that most transduced cells carried between one to three vector copies per cell, with two-copy integrations representing the largest fraction across samples. Average VCN values ranged approximately around one copy per cell (average VCN 0.69–1.24 in TR samples), with some variation across individual samples. Using the UT cells as a negative control and transduced Jurkat cells as a positive control, these

results demonstrate that the amplicon-based sequencing approach effectively detects vector-derived sequences, distinguishes transduced from untransduced samples and enable accurate estimation of vector copy number.

Protein panel and antibody oligonucleotide conjugate (AOC) expression

To validate antibody-derived oligonucleotide detection and confirm expected T-cell marker expression, antibody marker read distributions and marker expression profiles were analyzed across the CD22 CAR T-cell and UT samples.

Antibody marker read distribution analysis demonstrated high concordance between the two groups (Pearson r = 0.864, p = 9.93e-18; Spearman rho = 0.937) (Fig. 2A). The results indicated a similar marker distribution pattern and suggested that lentiviral transduction did not substantially alter overall antibody capture efficiency or marker detection. Several canonical T-cell markers, including CD45RA, CD45RO, CD62L, CD27, CD183, and CD28, showed substantial read representation (average 0.3–11.5% protein reads per sample), indicating successful antibody capture and sequencing.

To further examine marker expression patterns, density distributions of selected markers were compared across individual samples (Fig. 2B). Canonical T-cell markers (CD3, CD4 and CD8) exhibited consistent expression patterns across both the CD22 CAR T-cell and UT samples (e.g. CD3: average 2.46% in CD22 CAR T-cell vs 3.67% in UT samples; Mann–Whitney FDR > 0.05 for all three markers). Markers associated with naïve and memory T-cell phenotypes, including CD45RA, CD45RO, CD62L, and CD27, displayed the expected wide range of expression values (from low to high) that was required to accurately resolve and assign different T-cell subsets in the subsequent analysis based on their specific expression patterns (Table S3). Similarly, chemokine receptor CD183 and the costimulatory molecule CD28 showed comparable expression across groups (Mann–Whitney, FDR > 0.05) and low replicate variability (CV < 41%) in both groups.

Spiked-in control antibodies, including Recombinant siglec-2-CD22 and recombinant protein L, were detected with distinct distributions across samples, confirming successful incorporation and detection of custom AOC reagents. The Average Recombinant Protein L percentage was 0.48% in UT group samples and 10.99% in transduced group samples. One TR sample stained with Recombinant siglec-2-CD22 marker had 1.94% protein reads assigned. Protein L was served as a transduction-efficiency marker, and siglec-2 was included for comparison.

Together, these results demonstrate that the custom AOC panel performs reliably across samples and that key T-cell markers are expressed as expected in both transduced and UT samples.

Characterization of VCN and its correlation with Protein L

VCN and transduction efficiency (TE) were evaluated across samples TR3–TR7 using ddPCR and DAb-seq with three distinct TE metrics: TE^1 (flow cytometry), TE^2 (CAR sequencing reads), and TE^3 (protein L

or siglec-2 sequencing reads) (Table 2). ddPCR showed vector copies per cell ranged from 0.495 (TR7) to 1.042 (TR3). TE^1 values followed a similar trend, with the highest efficiency observed in the TR3 sample (54.58%) and the lowest in the TR7-P sample (25.79%), which was comparable to that of the TR7-S sample (28.10%). Vector copies per transduced cell were relatively consistent across samples, ranging from 1.59 (TR6) to 1.91 (TR3), indicating stable integration levels among successfully transduced cells.

Dab-seq CAR read analysis (TE^2) showed slightly higher copies per cell compared to ddPCR, ranging from 0.64 (TR7-P) to 1.23 (TR3). TE^2 values were also modestly elevated relative to TE^1, with TR3 reaching 56.8% and TR7-P reaching 33.6%. Copies per transduced cell estimated by TE^2 were consistently higher than ddPCR, ranging from 1.88 to 2.15.

DAb-seq-based protein-associated read analysis (TE^3) further supported these trends. TE^3 values were generally higher than TE^2 for most samples, with TR3 showing the highest efficiency (60.30%) and TR7 variants remaining the lowest (31.7%). Copies per transduced cell derived from TE^3 were comparable to TE^2, ranging from 1.70 to 2.13.

Across all methods, sample TR3 consistently demonstrated the highest VCN and transduction efficiency, while sample TR7 variants showed the lowest TE despite maintaining comparable copies per transduced cell. Intermediate samples (TR4–TR6) displayed moderate and consistent values across platforms.

Correlation between VCN and protein L and surface maker expression

To evaluate whether Recombinant Protein L binding reflects transgene copy number or surface marker expression, we analyzed its relationship with VCN and other AOCs at single-cell resolution across TR and UT samples (Fig. 3A). In CD22 CAR transduced samples, per-cell Protein L levels increased progressively with VCN, with sample-level median expression rising from low levels (normalized read count = 0.28) at VCNs of 0–1 to substantially higher levels of 1.51 at intermediate VCNs of 2–4 (Mann-Whitney p = 0.005). The trend was positive in every replicate (per-sample Spearman rho > 0 in 5/5 TR samples with protein L, median rho = 0.78; sign test p = 0.219), with an overall rank correlation of rho = 0.55 across the 6 samples. At higher VCN values (4–6), median sample-level Protein L signal did not differ further (Kruskal-Wallis p = 0.372), consistent with a saturation plateau. In contrast, untransduced samples showed low and relatively uniform Protein L signal across all VCN categories (only 50 of 68,752 UT cells (0.07%) carried VCN > = 1), so the apparent high-VCN UT categories reflect a handful of rare events rather than a biological trend.

Per-sample, median Protein-L was markedly higher in CD22 CAR transduced cells than UT cells (0.761 vs 0.03; Mann-Whitney p = 0.016). Together, these results demonstrate that Protein L signal is specific to successful transduction and scales with vector copy number up to a saturation threshold, whereas the low residual signal in untransduced cells represents non-specific background.

Assessment of surface markers showed stable expression across all VCN categories in transduced samples, with no evidence of VCN-dependent changes. Here we presented the representative surface marker, CD11a and CD45RA as examples. In transduced cells, CD11a levels remained consistent across all VCN categories (Spearman rho = -0.09, 95% CI [-0.40, 0.39], spanning zero), with median expression values showing minimal variation (category medians varied by 0.32 units across VCN bins with CV = 6.3%) and no discernible VCN-dependent trend (Fig. 3B). Similarly, CD11a expression profiles were observed in untransduced controls, indicating that baseline expression of this marker is independent of vector copy number. Notably, CD11a levels in transduced samples were comparable to those in untransduced cells, even at higher VCN categories (difference = -0.04; Mann-Whitney $p > 0.05$). These findings demonstrate that transduction and increasing vector integration do not alter CD11a expression, supporting preservation of cellular phenotype following gene transfer for this marker. CD45RA expression remained low across all vector copy number (VCN) categories in transduced samples, with no evidence of VCN-dependent changes (Spearman rho = -0.21, 95% CI [-0.55, 0.07], spanning zero). Comparable CD45RA levels were observed in untransduced controls (difference = -0.11; Mann-Whitney $p > 0.05$), indicating that CD45RA expression is independent of transduction and vector integration. Together, these results demonstrate that increasing VCN does not alter CD45RA expression, supporting preservation of cellular phenotype following gene transfer (Fig. 3C).

Correlation between protein L and AOC binding

To assess the relationship between Recombinant Protein L and AOC expression, normalized single-cell read counts were compared across CD22 CAR transduced and UT samples. The results of analysis of CD11a expression are shown as an example in Fig. 3D. Across all conditions, cells displayed predominantly vertical distributions, indicating stable CD11a expression over a wide range of protein L signals. Consistent with this observation, linear regression analysis revealed negligible correlation between protein L and CD11a in every sample, with R^2 values ranging from 0.00 to 0.02. While CD22 CAR-transduced samples exhibited broader and, in some cases, bimodal distributions of protein L binding, UT samples showed more compact signal distributions; however, these differences did not translate into coordinated changes in CD11a expression. Collectively, these data demonstrate that Recombinant protein L binding occurs independently of CD11a expression across all tested conditions. Similar to the results with CD11a, no other custom AOCs associated with protein L with $R^2 > 0.2$ were found. Similar results were obtained with CD28, as shown in Fig. 3E.

Collectively, these results indicate that Recombinant Protein L binding reflects transduction status and, to a limited extent, vector copy number, but occurs independently of surface expression.

Assessment of CAR-T transduction efficiency using protein L and siglec-2

To evaluate the efficiency of protein L and siglec-2 as transduction markers, the TR7 sample was divided into two portions, which were stained separately with protein L (TR7-P) and siglec-2 (TR7-S). Protein L

and siglec-2 read distributions were compared between the TR7-P and TR7-S samples to assess overall marker representation and relative abundance (Fig. 4A). Overall, both samples showed similar antibody read profiles with several markers demonstrating relatively high read percentages. Among the catalog markers, CD4, CD71, and CD11a showed the highest read proportions in both samples. CD4 displayed strong signals, accounting for approximately 9.9–10.3% of reads in both TR7-P and TR7-S. CD71 was another prominent marker, with read percentages reaching ~ 13.7% in TR7-P and ~ 12.2% in TR7-S. CD11a also exhibited relatively high abundance (~ 7–8%).

Several other markers showed moderate representation, including CD5, CD7, CD8, CD25, CD44, CD49d, CD62L and HLA-DR. The reads for these markers generally ranged between ~ 2–7%. Some markers demonstrated modest differences between the two samples; for example, CD49d and HLA-DR appeared slightly higher in TR7-S compared with TR7-P.

Custom AOC markers also contributed to the read distribution. Notably, protein L showed a substantial signal in TR7-P (~ 12.6%), and siglec-2 had unique signal (1.94%) in TR7-S samples. Other custom markers, such as CD183 and CD95, exhibited moderate read levels in both samples. Overall, the antibody read distributions suggest largely comparable immune marker representation between TR7-P and TR7-S, with a few sample-specific differences between Protein L and Siglec-2 signals.

VCN distributions were analyzed for TR7-P and TR7-S to evaluate transduction efficiency and genomic integration profiles (Fig. 4B). In both samples, the majority of cells exhibited VCN = 0, indicating a substantial non-transduced population. TR7-P showed an average copy number of 0.74 across all cells and 2.06 among transduced cells, with an overall transduction rate of 35.7%. Similarly, TR7-S displayed an average VCN of 0.69 across all cells and 1.99 in transduced cells, with a transduction rate of 34.7%. Low to moderate copy numbers (VCN = 1–3) accounted for most (97–99%) transduced cells in both conditions while higher copy numbers were rare (0.9–3.1% of transduced cells). These results indicate comparable transduction efficiency and copy number distributions between TR7-P and TR7-S, with no evidence of excessive high-copy integrations.

We further analyzed the correlation between VCN and siglec-2 expression. In sample TR7-S, siglec-2 expression increased from low levels (0.10) at VCN 0 to substantially higher levels of 1.12 at VCN > 0 (Fig. 4C). Median siglec-2 rose monotonically across VCN bins within this sample (Spearman $\rho = 0.93$ across 8 VCN categories). As these data were derived from a single transduced sample without biological replicates, this VCN-dependent increase is reported as a descriptive observation consistent with the trend seen for Protein-L, rather than as a statistically powered result.

Consistency of CD4 + and CD8 + T-cell percentages measured with DAb-seq and flow cytometry

The distribution of $CD4^+$ and $CD8^+$ T cells was assessed across UT and CD22 CAR T transduced samples using DAb-seq analysis (Fig. 5A). UT3 and TR3 samples were predominantly $CD4^+$, with $CD4^+$ cells

comprising 83.12% and 86.83% of the population, respectively, and relatively low proportions of $CD8^+$ cells (16.88% and 13.17% for UT3 and TR3 samples, respectively). In contrast, UT4, TR4, UT5, and TR5 samples displayed a more balanced $CD4^+/CD8^+$ composition, with $CD8^+$ cells representing approximately 51–58% of the population. UT6 and TR6 samples had a higher $CD4^+$ fraction, with $CD4^+$ cells accounting for 56.17% and 65.72%, respectively. Notably, TR7-P and TR7-S samples were strongly enriched for $CD4^+$ cells (76.84% and 75.03%, respectively), with corresponding reductions in $CD8^+$ cell percentages. Overall, these data demonstrate substantial variability in $CD4^+$ and $CD8^+$ T-cell composition across UT and TR conditions, with TR7 samples showing a pronounced $CD4^+$ bias.

The proportions of $CD4^+$ and $CD8^+$ T cells across UT and CD22 CAR-transduced samples were also determined by flow cytometry (Fig. 5B). UT3 and TR3 samples were predominantly $CD4^+$, with $CD4^+$ cells comprising 83.29% and 85.78% of the populations, respectively. UT4 and TR4 samples showed a more balanced distribution, with $CD8^+$ cells representing 55.59% and 52.16%, respectively. In UT5 and TR5 samples, $CD4^+$ cells were modestly enriched (56.50% and 53.88%). UT6 and TR6 samples demonstrated an increase in $CD4^+$ cells, reaching 58.57% and 64.47%, respectively. The TR7 sample exhibited a strong $CD4^+$ bias, with $CD4^+$ cells accounting for 72.37% of the population. Overall, flow cytometry revealed variable $CD4^+/CD8^+$ composition across samples, with some TR samples showing a consistent enrichment of $CD4^+$ T cells.

Quantitatively, DAb-seq and flow cytometry measurements differ only modestly, with $CD4^+$ and $CD8^+$ frequencies differ by a median of 0.17% and 1.26% in corresponding samples examined. Minor deviations are observed in some samples (< 12% difference for CD4 + and < 13.5% difference for CD8 + for UT5, UT6, TR5 and TR6 samples), but these do not alter the overall classification of samples as $CD4^+$ or $CD8^+$ enriched. Importantly, both platforms capture similar T-cell compositions across the UT and CD22 CAR transduced samples, indicating strong agreement between single-cell DNA/protein profiling by DAb-seq and conventional flow cytometry for immunophenotyping.

Overall, these results support the reliability of the DAb-seq assay for quantifying $CD4^+$ and $CD8^+$ T-cell populations in comparison to the established flow cytometry assay.

DAb-seq reveals T-cell subset differentiation and variable transduction efficiency

After antigen priming, naive T (Tn) cells progressively differentiate into diverse memory T-cell subpopulations, and ultimately, into terminally differentiated effector T (Teff) cells. Based on the description by Gattinoni and colleagues (Gattinoni, Speiser et al. 2017), T-cells can be separated into five subsets: Naïve T cells (Tn), T memory stem cells (Tscm), Central memory Tcells (Tcm), Effector memory T cells (Tem), and T Effector cells (Teff). We included all the protein markers mentioned in that paper in our panel. These markers were: CD3, CD45RA, CD45RO, and CD62L, which were from the 45-plex panel from Mission Bio, and CD11a, CD27, CD28, CD57, CD58, and CD95 which were from our AOCs.

DAb-seq transduction efficiency varied across T cell subsets and samples (TR3–TR7). In general, memory T cell populations (TR3, TR4, TR5, and TR6), exhibited higher transduction efficiencies (40.8–67.0%) than Tn cells (21.7% – 48.3%). Tcm and Tem subsets consistently showed the highest levels of transduction, reaching peaks of approximately 64–67% in TR3 and TR6. Tn cells demonstrated moderate transduction efficiencies across all samples, while Tscm showed the lowest transduction levels overall (12.5–41.2%). Teff cells displayed intermediate transduction efficiencies (22.2–49.7%) that were generally lower than Tcm (35.8–61.5%) and Tem (27.9–67.0%) but higher than Tscm (12.5–41.2%). Notably, samples TR7-P and TR7-S exhibited overall reduced transduction efficiencies across all subsets (median 31.85%) compared with other samples (median 46.25%), indicating sample-dependent variability (Fig. 6A and Figure S2).

The average VCN varied across T-cell subsets and among individual samples. Across all samples, effector memory T-cells consistently exhibited the highest VCN (VCN = 0.57–1.52), with peak values observed in samples TR3 (VCN = 1.52), TR5 (VCN = 1.23), and TR6 (VCN = 1.34). Central memory T-cells also showed relatively high VCN levels (VCN = 0.79–1.34), whereas effector T-cells displayed intermediate values (VCN = 0.47–1.03). In contrast, naïve and stem cell memory T-cells generally demonstrated lower VCNs (VCN = 0.28–0.96), with stem cell memory T-cells showing the lowest values across most samples (VCN = 0.28–0.8). Notably, inter-sample variability was evident, particularly within the effector memory and central memory compartments. These data suggest preferential enrichment or maintenance of vector integration within memory T-cell subsets, especially effector memory cells (Fig. 6B).

VCN-to-CAR expression discordance is associated with T-cell subset

To assess whether the efficiency of converting integrated vector copies into surface CAR protein depends on T-cell differentiation state, we modeled the per-cell relationship between copy number and CAR expression abundance across the five transduced samples (TR3, TR4, TR5, TR6 and TR7-P) with protein-L. For each cell we related $\log_2$-transformed surface CAR signal (background noise from 1_DN population-subtracted protein-L) to $\log_2$-transformed vector copy number in 4_DP and 3_INE populations. The slope of this relationship is the copy-to-CAR conversion rate, expressed as the fractional increase in surface CAR per doubling of integrated copies. Conversion rates were estimated in a linear mixed model with a per-sample random intercept, and the analysis was restricted to the three subsets with sufficient transduced cells (Central memory, effector, and effector memory T-cells); but naive and stem cell memory subsets were excluded, with 239 and 412 transduced cells across transduced samples, respectively, which was insufficient for reliable analysis.

The conversion rate differed significantly across subsets (VCN × subset interaction, joint Wald $F_{(2)}$ = 11.58, $p = 9.43 \times 10^{-6}$). Effector T-cells converted copies to surface CAR at a substantially higher rate (slope 0.82) than either central memory (0.45) or effector memory T-cells (0.51), whereas the two memory subsets were indistinguishable. Pairwise contrasts confirmed that effector T-cells differed from

both central memory (Δslope 0.35 ± 0.08, Holm-adjusted $p = 1.94 \times 10^{-5}$) and effector memory cells (Δslope 0.30 ± 0.07, adjusted $p = 7.32 \times 10^{-5}$), while central and effector memory cells did not differ (Δslope 0.05 ± 0.07, adjusted p = 0.42). This ranking was reproducible across biological replicates: Effector T cells showed the steepest conversion slope in all five samples, indicating a consistent property of the Effector state rather than an effect driven by individual outlying samples.

T cell subset analysis in untransduced vs. transduced samples

We were curious if the T cell subset percentage would have changed during transduction and the following cell culture. We compared T-cell subset composition between UT and CD22 CAR transduced samples using two analytical approaches: DAb-seq and flow cytometry.

Across both platforms, CD22 CAR transduced samples showed a general shift toward more differentiated T-cell subsets compared with UT samples. In the DAb-seq analysis (Fig. 7A), UT samples displayed relatively higher proportions of Tn (average percentage of 5.8% in UT samples vs. 2.5% in transduced samples) and Tscm (average percentage of 17.6% in UT samples vs. 6.9% in transduced samples), whereas CD22 CAR transduced samples showed enrichment in Tcm (average percentage of 24.3% in TR samples vs. 17.2% in UT samples) and Tem (average percentage of 39.2% in transduced samples vs. 25.9% in UT samples) subsets. Teff cells remained present in both groups but showed only modest differences between UT and the CD22 CAR transduced samples (average percentage of 33.5% in UT samples vs. 27.1% in TR samples) (Fig. 7B). Averaged data indicated that the transduced samples had higher percentages of Tcm cells (24.27% in transduced vs. 17.18% in UT samples) and Tem cells (39.17% in transduced vs. 25.89% in UT samples) compared with UT samples, while UT samples retained slightly higher Tn cells (2.53% in CD22 CAR transduced samples vs. 5.82% in UT samples) and Tscm cells (6.90% in transduced vs. 17.63% in UT samples) proportions.

Flow cytometry analysis (Fig. 7C, D) revealed a similar overall trend but with some quantitative differences. UT samples showed higher levels of Naïve/Tscm (average 14.98% in UT vs. 9.45% in TR samples) and Teff cells (average 31.77% in UT vs. 22.88% in TR samples) relative to CD22 CAR transduced samples, whereas transduced samples displayed increased Tem cells (average 17.19% in UT vs. 31.69% in TR samples) and Tcm cell (average 3.42% in UT vs. 5.75% in transduced samples) populations. The increase in Tem was particularly notable in transduced samples compared with UT samples. Overall, the flow cytometry data demonstrated a stronger shift toward effector memory phenotypes following transduction.

When comparing the two analytical platforms, both DAb-seq and flow cytometry analysis showed consistent trends in the relative distribution of T-cell subsets, with transduction associated with increased differentiation toward central memory and effector memory phenotypes. However, DAb-seq detected a clearer representation of naïve T-cells (average 3.84% in all samples) and provided higher estimates of central memory cells (average 21.44% in all samples by DAb-seq vs average 4.71% by flow cytometry) and effector memory cells (average 33.86% in all samples by DAb-seq vs average 25.24% by

flow cytometry), while flow cytometry showed relatively higher effector and effector memory percentages. Despite these differences in magnitude, the overall patterns of T-cell subset distribution between untransduced and transduced samples were broadly consistent across the two methods.

Characterization of four distinct cell populations within transduced samples based on TE marker expression

Four populations from the transduced samples could be separated based on the VCN and transduction efficiency marker expression: 1-double -negative (1_DN, VCN negative & Protein L / siglec-2 negative); 2-staining background (2_SB, VCN negative & protein L / siglec-2 positive); 3-insert-not-expressed (3_INE, VCN positive & protein L / siglec-2 negative); 4-double positive (4_DP, VCN positive & protein L / siglec-2 positive). For this analysis, VCN positive was defined as VCN > 0, while protein L / siglec-2 positive was defined as normalized reads ≥ 0.6. The four populations (1_DN, 2_SB, 3_INE, and 4_DP) were analyzed across individual CD22 CAR transduced samples (TR3, TR4, TR5, TR6, TR7-P, and TR7-S) (Fig. 8 and figure S3). Across all transduced samples, the DN population represented the predominant fraction (ranging from 37.29–64.93%) (Fig. 8A). The DN frequencies were highest in TR7-P (64.93%) and TR7-S (62.24%) samples, followed by TR4 (56.47%), whereas the TR3 sample showed the lowest DN proportion (37.29%) that means it had the highest transduction efficiency. These results revealed that different patient samples have different capacities to acquire and express vector. The 2_SB population varied among samples, with relatively higher representation in the TR4 (9.86%) and TR5 (7.41%) samples compared with the other samples (3.58–5.87%). The 3_INE population was present at low frequencies across all samples (1.95–5.39%), showing only minor variability. The same trends can be seen from the average cell percentages (Fig. 8B). Variation among the cell populations was moderate for 1_DN (Std = 9.98%) and 4_DP (Std = 10.81%), as indicated by the spread of individual data points and error bars, whereas 2_SB (Std = 2.07%) and 3_INE (Std = 1.13%) displayed consistently low values with relatively small standard deviations. Overall, populations 1_DN (average 51.96%) and 4_DP (average 38.95%) demonstrated markedly higher cell percentage values compared to 2_SB (average 6.04%) and 3_INE (average 3.05%), and inter-sample variability was observed in the relative contributions of 1_DN and 4_DP populations, particularly in TR4 and the paired TR7 samples (percentage difference between 1_DN and 4_DP > 24.74% in TR4/TR7 samples vs < 0.81% in TR3/TR5/TR6 samples).

Among the four VCN and transduction efficiency populations, the distribution of T cell subsets varied (Fig. 8C). Tn cells were present at low frequencies (median frequency of 3.01% of cells within each population per sample) across all VCN and transduction efficiency populations, with only minor differences observed (frequency range 0.26–8.46%). Tscm cells showed modest variation, with 1_DN (median frequency of 13.14%), 2_SB (median frequency of 10.74%) and 3_INE (median frequency of 9.23%) displaying slightly higher percentages compared to 4_DP (median frequency of 2.96%). For Tcm cells, all VCN and transduction efficiency populations demonstrated moderate frequencies. The population 2_SB exhibited relatively higher Tcm levels (median frequency 35.23%) compared to the others, while 1_DN (median frequency 22.60%) and 4_DP populations (median frequency 28.26%) showed intermediate values and 3_INE (median frequency 20.29%) displayed slightly lower levels. Teff

cells were similar to Tcm cells, with higher abundance than Tn and Tscm cells in all groups. Population 3_INE (median frequency 31.68%) showed comparatively higher Teff percentages, followed by 1_DN (median frequency 29.02%) and 2_SB (median frequency 22.42%), whereas 4_DP (median frequency 20.37%) demonstrated relatively lower values. Tem cells represented the highest frequency subset overall. Population 4_DP (median frequency 44.68%) showed the greatest Tem cell percentages, followed by 3_INE (median frequency 35.72%) and 2_SB (median frequency 32.06%), while 1_DN (median frequency 28.78%) exhibited comparatively lower values. Overall, memory and effector cell subsets (Tcm, Tem, and Teff) predominated over naïve subsets across all populations, with notable inter-population differences particularly in Tem and Teff cell compartments.

The distribution of the four VCN and transduction efficiency populations varied across the T-cell subsets (Tn, Tscm, Tcm, Tem, and Teff) (Fig. 8D). Most of the Tn cell population likely could not be expressed as 1_DN (58.41%) was the most frequent population. Tscm cell showed the lowest proportion of 4_DP (19.70%) cells, while Tcm cells contained a relatively lower percentage of 3_INE (2.23%) cells and a higher proportion of 2_SB (8.55%) cells. Tem cells showed relative high proportion of 4_DP (40.49%) and moderate proportion of 3_INE (2.97%) cells. Teff cells exhibited a moderate enrichment of all four VCN and transduction efficiency populations compared with the other subsets. Overall, the relative abundance of the four populations differed among the T cell subsets, indicating distinct population compositions across differentiation states.

We further analyzed the distribution of T-cell subsets across the four VCN and transduction efficiency populations (1_DN, 2_SB, 3_INE, and 4_DP) (Fig. 8E). The results showed broadly similar patterns with notable shifts in specific compartments. The Tem cell population constituted the largest proportion in most groups, remaining relatively stable across conditions (32.39–44.28% of all transduced samples), although a slight decrease was observed in the 2_SB cell subset. The percentage of Teff and Tcm cells varied more substantially between groups, showing the proportion of Tcm cells from 18.81–35.64% and the proportion of Teff cells ranging from 20.69–30.34%. Tscm and Tn cells consistently comprise smaller fractions of the total population. Tscm cells appear slightly enriched in 1_DN (9.19%) and 2_SB (7.45%) populations compared to 4_DP (3.52%), while Tn cells remain minimal across all groups with only minor fluctuations (1.80–4.00%).

Overall, while Tem cells dominance was maintained, the relative balance between Teff and Tcm cells differs among populations, suggesting shifts in memory T-cell composition depending on the group. While the relative proportions of memory T-cell subsets varied, the dominance of differentiated effector cell phenotypes (Teff/Tem) suggests a high level of antigen experience or activation across all four experimental conditions.

AOC expression and control comparison

The four cell populations distinguished within the transduced samples provide two additional useful controls 1_DN and 3_INE, beyond the untransduced sample, for AOC expression analysis (table 3).

Comparison of 4-DP vs 1_DN cell populations as control

Comparison of the 4_DP with the 1_DN cell populations control revealed consistent upregulation of CD141 and HLA-DR markers in several of the samples (Figure S4). Sample TR3 demonstrated increased expression of CD141 (1.58-fold change, Cliff's delta = 0.46; Mann-Whitney $p < 0.001$), and HLA-DR (2.13-fold change, Cliff's delta = 0.67; Mann-Whitney $p < 0.001$) without detectable downregulation of any of the markers evaluated. A similar pattern was observed in sample TR5, which in addition, had increased expression of the markers CD25 (1.73-fold change, Cliff's delta = 0.34; Mann-Whitney $p < 0.001$) along with CD141 (1.59-fold change, Cliff's delta = 0.40; Mann-Whitney $p < 0.001$), and HLA-DR (1.86-fold change, Cliff's delta = 0.62; Mann-Whitney $p < 0.001$). Sample TR6 exhibited upregulation of CD25 (1.47-fold change, Cliff's delta = 0.35; Mann-Whitney $p < 0.001$) and CD141 (1.70-fold change, Cliff's delta = 0.49; Mann-Whitney $p < 0.001$). In contrast, sample TR4 showed a more restricted response characterized without any statistically detectable changes in marker expression. Overall, CD141 was the marker consistently increased across most of TR samples in this comparison.

Comparison of 4_DP vs 3_INE as control

Relative to the 3_INE control samples, TR3, TR5, and TR6 demonstrated highly concordant expression profiles, with coordinated upregulation of CD141 (fold change 1.48–1.91), and HLA-DR (fold change 1.85–2.58). In addition to these two markers, TR5 had increased expression of the markers CD25 (1.64-fold change, Cliff's delta = 0.32; Mann-Whitney $p < 0.001$) along with CD45RO (1.50-fold change, Cliff's delta = 0.37; Mann-Whitney $p < 0.001$). No downregulated markers were identified in these groups. Sample TR4 again displayed a distinct pattern, and no expression change was observed. Among all three control comparisons, this comparison demonstrated the greatest overall consistency between TR3, TR5, and TR6 samples, suggesting a shared activation-associated phenotype in the 4-double-positive condition (Figure S5).

Comparison of 4_DP vs UT samples

Comparison of DP populations with UT samples revealed broader but more variable marker-expression changes (Figure S6). Sample TR3 showed increased expression of CD28 (1.45 fold change, Cliff's delta = 0.36; Mann-Whitney $p < 0.001$), CD45RO (1.64 fold change, Cliff's delta = 0.44; Mann-Whitney $p < 0.001$), CD95 (1.35 fold change, Cliff's delta = 0.40; Mann-Whitney $p < 0.001$), CD141 (2.34 fold change, Cliff's delta = 0.61; Mann-Whitney $p < 0.001$), and HLA-DR (6.84 fold change, Cliff's delta = 0.87; Mann-Whitney $p < 0.001$) without detectable downregulated markers, representing the broadest activation-associated profile among all groups. The transduced sample TR5 demonstrated upregulation of CD141 (2.10-fold change, Cliff's delta = 0.55; Mann-Whitney $p < 0.001$), and HLA-DR (3.23-fold change, Cliff's delta = 0.76; Mann-Whitney $p < 0.001$) together with downregulation of CD28 (0.74-fold change, Cliff's delta =- 0.31; Mann-Whitney $p < 0.001$). Sample TR6 similarly exhibited increased CD141 (1.57-fold change, Cliff's delta = 0.43; Mann-Whitney $p < 0.001$), and HLA-DR (2.24-fold change, Cliff's delta = 0.65; Mann-Whitney $p < 0.001$) expression and additional up-regulation of CD25 (1.82-fold change, Cliff's delta = 0.46; Mann-Whitney $p < 0.001$). In contrast, sample TR4 displayed reduced CD38 expression (0.61-fold change, Cliff's

delta = -0.40; Mann-Whitney p < 0.001). Compared with the other control analyses, comparison to the UT samples revealed greater heterogeneity in both upregulated and downregulated markers, indicating more complex phenotypic differences between untreated and 4-double-positive conditions.

Overall, the double-negative population serves as an internal reference for cells exposed to the transduction process but lacking vector integration, controlling for process-associated (non-integration) effects, whereas the INE population is a stronger control for determining whether protein expression drives the observed phenotypic shifts. The UT sample reflects the overall impact of transduction and expansion on T-cell quality and activation state relative to the starting material.

Discussion

CAR T-cell therapy involves genetically engineering a patient's T-cells to express a Chimeric CAR, which enables them to recognize and kill cancer cells (21). The transgene copy number and its expression level are critical factors in CAR T cell therapy because higher transgene copy numbers generally correlate with higher CAR expression levels on the surface of the T-cells, which can lead to more potent anti-tumor activity (22). In addition, higher copy numbers could also potentially lead to overexpression, which may contribute to off-target toxicities or T-cell exhaustion (23). On the other hand, too few copies might result in insufficient expression and anti-tumor activity. As such, rather than simply serving as a quality control measure, transgene copy number results may have implications for the safety and efficacy of CAR T-cell products, which may ultimately lead to improved patient outcomes (24). CAR-T cell vector copy numbers are usually measured by qPCR or ddPCR, and the transduction efficiency is usually measured by cell surface markers using flow cytometry. These assays are performed on bulk samples and cannot link the genotype and phenotype. Also, DNA copy count may not always map precisely to the number of functional cells because of multiple vector copies per cell, silenced transgenes, variability in expression(16). Flow cytometry is effective for assessing surface-marker expression in bulk populations of transduced cells; however, unlike ddPCR and qPCR, it does not directly measure transgene copy number or quantify transgene expression levels.

We used a DNA and protein sequencing system to simultaneously detect both vector copy number and protein expression on the same CAR T-cells at single cell resolution. We also separated five different T-cell subsets using protein markers. Furthermore, by using vector amplicon and surface markers, we distinguished four different populations in transduced cells and gained better insight into the composition of the transduced samples.

CAR and vector sequences detected in transduced cells and T cell subsets were identified using protein markers at the single cell level

The Tapestri DNA + Protein system utilizes droplet microfluidics and DNA-barcoded antibodies to provide a high-resolution, multi-omics profile of individual cells. By encapsulating single cells to amplify and barcode their genomes, the platform moves beyond the averaged data of bulk sequencing to definitively detect transgenes within specific cells. This approach enables simultaneous phenotypic

validation; because the system measures protein expression alongside DNA sequencing, researchers can confirm not only the presence of a transgene but also whether it is actively producing a protein product. Ultimately, this integrated workflow bridges the gap between genotype and phenotype to offer a comprehensive picture of cellular activity at the single-cell level. Furthermore, T-cell subset characterization is critical for insights into composition, function, and activation state. By using the Tapestri system, we identified all five T cell subsets described in a paper published in Nature Medicine (25) using all mentioned T-cell markers.

We used the CD19 gene backbone from NIST to design one assay for multiple lentiviral constructs. This assay includes ninety-nine human amplicons, fifteen of the vector amplicons based on the NIST CD19 gene sequence, three identity amplicons based on the CAR sequence, and seven identity amplicons based on the vector sequence for another protocol as negative controls. The specific CAR sequence could be detected in the transduced cells and all the negative controls, and untransduced samples show no signals. The TotalSeq™-D Heme Oncology Cocktail from BioLegend has forty-five unique cell surface antigens which includes principal lineage antigens and three isotype control antibodies to aid in the DNA and protein multi-omics characterization of cancer cells. In addition to this panel, we included ten anti-human AOCs to characterize T cell subsets and five AOCs for transduction-efficiency markers, for a total of sixty antigens in the panel. We used Protein L as the transduction efficiency marker, which distinguished transduced cells from untransduced cells. By using a CAR-specific amplicon and Protein L cell surface marker, we successfully identified transduced cells in the transduced samples.

Correlation between Protein L expression and vector copy number

Transgenes can insert into multiple loci in the genome; therefore, it is common to have more than one transgene copy per genome of transduced cells (16). If a cell has more copies of a specific gene, there are more potential "factories" producing RNA from that gene. Since RNA-seq reads are a proxy for RNA abundance, an increased amount of RNA will naturally lead to more reads mapping to that gene, although sequencing depth, transcript length, library preparation, and other technical factors can influence this relationship. This relationship forms the basis of RNA-seq–based gene expression quantification and has been extensively characterized and validated in numerous studies. Many studies on CAR T cell manufacturing have shown data where CAR-T cells are sorted based on surface CAR expression and analyzed for genomic integration sites or quantification of transgene DNA (26, 27). Concurrently, they might use RNA-seq to confirm that cells with higher CAR protein (and implicitly, higher vector copy number) also have higher CAR RNA reads. A study comparing CAR-T cells generated via PiggyBac transposon (PB-CAR-T) versus lentiviral transfection found that PB-CAR-T cells exhibited much higher CAR expression and stronger upregulation of cytokines, chemokines, and receptors in RNA-seq data (28).

In our study, the observation that protein L expression levels increased with VCN over the low-to-intermediate range and then plateaued, rather than scaling proportionally with transgene dosage across

the full range. While VCN provides a measure of the average number of integrated lentiviral or retroviral cassettes per cell, several biological and technical factors may decouple this genomic metric from surface protein density, and these plausibly account for the saturation we observed at higher copy numbers.

Transcriptional Silencing and Integration Site Effects. The genomic locus of integration significantly influences transgene potency (29). While VCN measures the quantity of integrated vectors, it does not account for the "quality" of the integration site. Vectors integrated into heterochromatin or regions prone to epigenetic silencing may remain transcriptionally inactive, whereas single integrations in euchromatic "hotspots" can yield disproportionately high expression.

Translational Saturation and Metabolic Burden. There may be a biological "ceiling" on CAR protein production. As VCN increases, cellular machinery (ribosomes, chaperones, and the Golgi apparatus) can become saturated. High-copy-number cells may face significant metabolic stress, leading to a plateau in protein folding and trafficking efficiency despite abundant mRNA transcripts.

Post-Translational Regulation and Turnover. CAR expression is also governed by its stability on the cell membrane (30). Factors such as protein internalization, degradation rates, and shedding can normalize surface density across a population. If the rate of protein turnover increases in response to high intracellular concentrations, steady-state surface expression will appear independent of the underlying VCN.

In the context of cellular therapies, these findings suggest that VCN alone is an insufficient predictor of biological function. Beyond a threshold of roughly two to four copies, a higher VCN does not guarantee a more "potent" cell if protein L remains constant. On the other hand, this lack of correlation may actually be favorable. It suggests a level of biological buffering that prevents extreme overexpression, which could otherwise lead to increased exhaustion in engineered T-cells. Future efforts should focus on optimizing transduction efficiency rather than simply increasing the average copy number per cell. Additional strategies may include engineering vectors to achieve more consistent expression or using targeted knock-in approaches to integrate vectors at defined genomic sites that support stable and predictable transgene expression.

Different transduction efficiency across T cell subsets

T-cells are not a monolithic group; they exist in various states (Naive, Stem Cell Memory, Central Memory, Effector). Certain T-cell subsets are naturally more resistant to viral entry or transgene expression (31). More differentiated T-cells are more metabolically active, dividing and permissive to viral entry, so they tend to be easier to transduce. Less differentiated T-cells tend to be harder to transduce, but they yield products with better persistence and long-term antitumor function. Our data showed that most transgenes integrated into the genome and could be expressed, and different T-cell subtypes had different transduction efficiency. As expected, Tscm cells exhibited the lowest transduction efficiency. In

addition to their lower transduction efficiency, transduced Tn cells also expressed the CAR construct less efficiently than other T-cell subsets. The observation that VCN varies significantly across distinct T-cell subtypes—specifically between Tn, Tscm, central memory Tcm, and Teff cell populations, highlights the biological heterogeneity that governs lentiviral or retroviral transduction. This variance suggests that the efficiency of gene transfer is not a uniform process among all cell types but is instead dictated by the metabolic and physiological state of the target cell at the time of transduction. The differential susceptibility of T-cell subsets to viral entry and integration can be attributed to several intrinsic factors:

Cell Cycle Status and Proliferation. Lentiviral vectors can transduce non-dividing cells, but efficiency is markedly higher in actively cycling cells. Memory cell subsets, Tcm and Tscm, often exhibit different kinetics of activation and cell cycle entry compared to terminally differentiated effectors. If certain subsets enter the S-phase more rapidly following stimulation (e.g., via CD3/CD28 beads), they present a more receptive window for stable viral integration (32).

Expression of Viral Entry Receptors. The density of receptors required for viral attachment (such as LDLR for VSV-G pseudotyped vectors) may fluctuate across T-cell differentiation states. A higher surface density of these receptors on less-differentiated cells (Tn/Tscm) could facilitate increased viral entry, resulting in a higher VCN compared to more mature T cell subsets.

Intracellular Restriction Factors. T-cells possess innate antiviral mechanisms, such as SAMHD1 or TRIM5 alpha, which can restrict viral reverse transcription or integration (33) The expression and activity of these restriction factors are known to vary based on the cell's differentiation stage and metabolic activation, potentially creating "bottlenecks" that lower VCN in specific subtypes.

The uneven distribution of VCN across these subsets has profound implications for the final cellular product. If the highly potent, long-lived memory T-cell populations (Tscm) have a higher VCN, the overall therapeutic "fitness" of the product may improve. Conversely, if transduction skews toward short-lived effector cells, treatment durability may be compromised.

Methodological Implications for VCN Analysis. A population-level VCN measured by ddPCR or qPCR represents a weighted average across the entire product and may therefore conceal important subset-specific imbalances. For example, a product may have an apparently acceptable bulk VCN of 2.0 while the memory compartment is under-transduced, with an average VCN of 0.5, and the effector compartment is over-transduced, with an average VCN of 5.0. Individual cells may have a VCN greater than 1, particularly when they are efficiently transduced and retain the capacity to proliferate and differentiate. The primary concern is the potential for over-transduction and excessive transgene expression. However, clinical-trial data have not demonstrated an increased incidence of secondary malignancies, despite the black box warning.

To achieve a more granular understanding, manufacturing processes increasingly use multi-parametric flow cytometry combined with subset-specific sorting prior to VCN analysis, and previous studies have investigated strategies to enhance cellular transduction permissiveness while minimizing differentiation

(31, 34). This ensures the transgene is distributed effectively across the cells best suited for long-term antitumor surveillance.

Beyond Vector Copy Number: Cellular Determinants of Surface Protein Expression

The effectiveness of different T cell subsets in expressing integrated vector copies is a critical determinant of the functional output of engineered T-cell products. Our findings show that conversion efficiency from VCN to surface CAR protein is not uniform across subsets but is shaped by cellular state. Specifically, effector T-cells exhibit a markedly higher capacity to translate integrated copies into surface CAR expression, with a conversion slope of 0.82, compared to Central memory (0.45) and Effector memory T cells (0.51). This suggests that effector T-cells are more proficient at utilizing each integrated transgene for protein production, likely reflecting their heightened metabolic activity and transcriptional permissiveness. In contrast, memory subsets, despite their recognized persistence and anti-tumor potential, display a lower conversion rate, indicating that more integrated copies are required to achieve equivalent surface CAR levels. Importantly, these differences are consistent across biological replicates, emphasizing that conversion capacity is an intrinsic property of the T cell differentiation state, not an artifact of sample variability. This subset-dependent conversion efficiency has substantial implications for CAR T-cell manufacturing: optimizing both transduction and expression in memory populations may be necessary to maximize the longevity and potency of therapeutic cells, since over-transduction of effector cells could risk excessive expression and associated toxicity. Therefore, understanding and manipulating the factors that govern copy-to-CAR conversion in each subset is pivotal for designing next-generation CAR T-cell therapies that balance efficacy, durability, and safety. Indeed, the development of rheostat-like CAR systems, in which CAR surface abundance can be reversibly tuned to modulate downstream signaling and cytotoxic function, highlights the therapeutic importance of controlling receptor density independently of vector copy number (35).

Simultaneous evaluation of DNA and protein expression can separate transduced samples into four populations.

DN population in transduced samples.

The ability of a viral vector (typically lentivirus or retrovirus in CAR T-cell production) to successfully transduce a cell and integrate its genetic payload is a multi-step process. If any one of these "gates" is closed, the cell remains in the 1-DN (Double-Negative) population.

Cells that are rapidly dividing are generally easier to transduce than quiescent (resting) cells (36). The DN group may represent a subpopulation of T cells that were less metabolically active during the transduction window. Lentiviral vectors are often pseudotyped with VSV-G (Vesicular Stomatitis Virus G protein), which binds to the LDLR (Low-Density Lipoprotein Receptor) on T cells (37). If a T-cell has low expression of LDLR (common in resting or quiescent cells), the virus cannot "dock." Even if it binds, the

viral envelope must fuse with the cell membrane or endosomal membrane. Factors like cholesterol content or membrane tension can inhibit this fusion.

The lack of transduction in the 1_DN population may be further exacerbated by the presence of intrinsic restriction factors such as SAMHD1, which restricts reverse transcription by depleting intracellular dNTP pools (38), and APOBEC3G, which mediates viral genome degradation. These act as molecular "bouncers" that degrade viral RNA or block reverse transcription before DNA can be made.

Even if the vector enters the cell, it must enter the nucleus and stitch itself into the host's chromosomes. For lentiviruses, the "Pre-integration Complex" (PIC) must be actively transported through the nuclear pore (39). If T-cells aren't actively cycling, integration won't occur. The enzyme Integrase prefers "open" chromatin (euchromatin) that is being actively transcribed. If the DNA lands in "closed" or "silent" regions (heterochromatin), it may fail to integrate or be silenced.

In the transduced samples, the presence of a healthy DP population suggests the virus was functional, but the DN group likely represents cells that were either quiescent (not dividing) at the time of infection or were "missed" due to the stochastic nature of the MOI.

SB - Different T cell subsets acquired different staining backgrounds

Since protein L binds to the kappa light chain of immunoglobulin (often used to detect the scFv of a CAR), it is highly sensitive to any other structure on the cell surface that resembles an antibody or contains a light chain (40).

Surface Area and Receptor Density. As T cells differentiate from Tn to Teff cells, they generally increase in size and surface area during the activation/blast phase. Teff and Tem cells are larger and have a higher density of surface receptors and adhesion molecules. More surface area means more non-specific binding of the TE marker reagent. Tn and Tscm cells are typically smaller and more quiescent, leading to lower overall background signals (41, 42).

Endogenous BCR/Ig Interaction. While T cells don't produce antibodies, they can "pick them up" from the culture media (e.g., if using human serum or certain FBS batches) or via Fc receptors. More differentiated cells (Tem/Teff) often have higher expression of various Fc receptors or "sticky" glycans that can trap proteins from the media, which Protein L then detects as background.

Non-Specific Binding to Activation Markers. The process of transduction drives cells toward the Teff phenotype. These cells express high levels of activation markers (CD25, CD69, HLA-DR). Protein L is notoriously "sticky." The highly active, highly protein-dense membrane of a Teff cell is more likely to engage in non-specific protein-protein interactions than the "smoother" membrane of a Tn or Tscm cell.

INE in the transduced samples

INE (Insert-Not-Expressed) cells were DNA-positive but Protein-negative. These cells are a critical "silent" group. In some cells, the CAR gene might integrate into a region of heterochromatin (tightly packed, "silenced" DNA), but the cell's machinery cannot access it to make mRNA and protein. Also, T-cells have internal defense mechanisms to shut down foreign DNA (43). Over the 9 days of expansion, some cells that initially expressed the CAR may have methylated the promoter region of the viral vector. These cells were once positive but became "INE" by the time the final product stage is reached. By Day 9, the cells have undergone massive expansion, with some reaching a state of metabolic insufficiency. Severely exhausted or over-differentiated cells may prioritize basic survival over transgene expression, leading to a loss of surface marker signals.

Double positive (DP) in transduced samples

DP has the terminally differentiated Teff phenotype, which accounts for about 29% of the total population. Tscm cells are critical for long-term persistence and the ability to regenerate other T-cell subsets, and this Tscm compartment has a noticeably contracted DP population. Interestingly, the DP population shows a slightly higher proportion of Tcm cells than the DN cell population. This suggests that the DP cell population is highly effector-oriented, it also retains a distinct central memory pool, potentially for lymph node homing. The DP cell population is highly specialized for immediate cytotoxicity. The high Teff cell counts among DP cells indicate a state of high activation or perhaps chronic antigen exposure. While DP cells are likely more potent in the short term, the loss of the Tscm "stemness" may limit their long-term persistence compared to the 1_DN group. These differences may result from transduction-induced metabolic reprogramming and concomitant bystander signaling. The possible reasons for DP phenotype are:

Activation-Induced differentiation. The process of transduction (typically via viral vectors) requires the T-cells to be in a highly proliferative state. To achieve this, cells are activated using anti-CD3/CD28 beads or antibodies. The DP cells that successfully integrate the transgene and express the CAR are often those that responded most robustly to the initial activation. This intense proliferation drives them further down the differentiation pathway toward a Teff phenotype.

Metabolic stress and resource competition. Transduced cells are under higher metabolic demand because they are actively synthesizing the new protein (the transgene) in addition to their endogenous proteins. This metabolic "heaviness" can favor glycolytic pathways over oxidative phosphorylation (44). Glycolysis is strongly associated with the Teff and Tem phenotypes Helps to generate the cytotoxic function of the cells.

Differential proliferation rates. It is common for transduced cells to have a slight proliferative advantage or disadvantage (45). If the DP cells divided more times than the other cells during the expansion phase, they would naturally appear more differentiated. Each cell division "clocks" the T cell closer to the effector stage and further from the naive/stem stage.

Impact on quality control (QC). In a manufacturing setting, the ratio of 1_DN to 4_DP phenotypes can provide insight concerning the health of the culture: high 4_DP Teff / Low 1_DN Teff indicates the transduction and subsequent signaling are specifically driving the differentiation state of the cells; high Teff in both populations suggests the general culture conditions (e.g., too much IL-2, high seeding density, or over-activation) are driving differentiation, rather than the transgene itself.

Based on our results, the DP population is noticeably more differentiated than the other populations, suggesting the transgene expression or the transduction process is a primary driver of this shift.

Cytokine environment adjustments to preserve the stemness in CAR T cell culture

To preserve the "stemness" of the DP population and prevent the shift toward the Teff phenotype, adjusting the cytokine environment during expansion is one of the most effective strategies.

IL-2 strongly activates the mTOR pathway and increases the expression of Blimp-1, a transcription factor that promotes the Teff and Tem cell phenotypes (46). IL-2 drives heavy reliance on glycolysis, and it results in rapid expansion, which likely contributes to the high Teff cell percentage seen in the 4_DP group. IL-7 promotes the expression of anti-apoptotic proteins such as Bcl-2 which is essential for the homeostatic maintenance of naive and memory T-cells (47). It also allows for cell survival and low-level proliferation without forcing the cell into an activated effector state. IL-15 promotes the expression of Tcf7 (TCF-1), a master regulator of T cell "stemness" and shifts the cell toward Fatty Acid Oxidation (FAO) and mitochondrial spare respiratory capacity. Cells expanded in IL-15 retain a much higher percentage of Tscm and Tcm markers, even after transduction (31). IL-21 has earned its reputation as the "anti-exhaustion" cytokine primarily due to its unique ability to sustain the effector function and proliferative capacity of CD8 + T-cells during chronic antigen exposure (48). The interplay between these four cytokines was analyzed to identify a combinatorial approach that favors the Tscm phenotype over terminal differentiation (49, 50).

For manufacturing CAR T-cells, promoting a Tcm/Tscm profile for the DP can be achieved by replacing IL-2 with an IL-7/IL-15 cocktail which often results in a product with superior persistence (51). The combination of IL-7 + IL-15 + IL-21 can result in the generation of CAR T-cells with > 50% T ~ SCM~ enrichment producing less exhausted, longer-lived CAR T cells with potent anti-tumor activity without requiring conventional αCD3/CD28 activation (50). If the gap between the proportion DN and DP cells in the CAR T cell product is too wide, it may indicate the initial CD3/CD28 stimulus is too "hot." Reducing the bead-to-cell ratio or the duration of the stimulus can help preserve the early memory phenotype. Some CAR T cell manufacturing protocols now incorporate small molecule inhibitors (like Akt inhibitors) during the expansion phase to "lock" the cells in a Tscm state, preventing the DP cell population from maturing into Teff cells (52).

Mechanistic and translational implications of control selection in CAR T-cell expression

DN population as the control

If using the DN as the control, CD141 and HLA-DR were consistently upregulated in the DP populations, suggesting a distinct activation or differentiation state associated with CAR expression and/or manufacturing. Increased HLA-DR is consistent with T-cell activation and has been reported in CAR T cells, particularly in association with 4-1BB signaling and tonic CAR signaling (53, 54). In contrast, the significance of CD141 upregulation in CAR T cells is less well established. Although CD141 is commonly associated with cDC1, it is not lineage-specific (54). Thus, increased CD141 may reflect a distinct T-cell state generated or enriched during CAR T-cell manufacturing; further studies are needed to determine its functional significance. The limited response observed in TR4 indicates that this subgroup may be less responsive or functionally distinct compared with the other TR populations.

INE population as the control

The comparison with the insert-not-expressed control revealed the most uniform response pattern across samples TR3, TR5, and TR6. The repeated upregulation of CD141, and HLA-DR across these groups suggests that the 4-double-positive condition consistently promotes immune activation and differentiation-related signaling relative to the insert-not-expressed state. The absence of detectable downregulated markers further supports the interpretation that the dominant biological effect in this comparison is activation-associated enhancement rather than suppression. Sample TR4 again showed a distinct phenotype, without any regulation change detected, which may reflect alternative signaling pathways unique to this subgroup.

UT sample as the control

When using UT samples as the control, the DP cell population showed broader and more heterogeneous immune-marker modulation than in the other control analyses. Sample TR3 exhibited the strongest activation-associated phenotype, including increased CD28 and CD95 expression in addition to CD45RO, CD141, and HLA-DR. This pattern suggests enhanced co-stimulatory signaling and activation-state regulation in the 4-double-positive cell population. In contrast, samples TR4 and TR5 demonstrated downregulation of genes, indicating altered co-stimulatory capacity relative to untreated samples. Reduced CD38 expression in TR4 and decreased CD28 expression in TR5 further suggest subgroup-specific modulation of activation and trafficking pathways. Sample TR6 maintained a predominantly associated signature, with no detectable suppression of additional markers. Collectively, when comparing the transduced cells to the UT cells it was found that the 4-double-positive condition induces substantial phenotypic remodeling relative to untreated samples, while also highlighting distinct regulatory behaviors among the individual transduced cell populations.

Overall interpretation across the three control comparisons

Across all control comparisons, CD141 and HLA-DR emerged as two of the most consistently upregulated markers, indicating a distinct CAR-associated cellular phenotype. Increased HLA-DR is

consistent with enhanced T-cell activation, whereas the biological significance of CD141 enrichment in CAR-T cells remains less well established. Together, these findings suggest an association between CAR expression and phenotypic changes in the transduced cells. CD25 and CD45RO were also recurrently elevated, particularly in the TR3, TR5, and TR6 samples, further supporting a coordinated activation-associated phenotype. Among the transduced samples, TR3, TR5, and TR6 exhibited broadly similar response profiles, whereas TR4 consistently displayed a more restricted and distinct phenotype. Collectively, these results suggest that the double-positive population is associated with selective immune activation and phenotypic remodeling, while also revealing sample-specific differences in the magnitude and composition of these responses.

Considerations for control selection in CAR T-Cell studies: DN vs. INE vs. UT

An important consideration in the interpretation of CAR T-cell phenotypic data is the selection of an appropriate population for comparison. In this study, three distinct comparison groups were used, including 1-double-negative (DN), 3-insert-not-expressed (INE), and untransduced (UT) samples, each providing different biological and translational insights into the effects of CAR engineering and expression.

Comparison against the DN cell population is particularly useful for identifying phenotypic features associated with successful acquisition of the CAR-positive state. Because DN cells generally undergo similar activation, transduction, and culture procedures as CAR-positive cells but fail to acquire the final double-positive phenotype, this control serves as an internal manufacturing-matched control. Therefore, marker differences observed between the DP and DN cell populations are more likely to reflect biological changes associated with productive CAR generation rather than nonspecific effects of ex vivo manipulation.

Using the INE cell population for comparisons provides an additional mechanistic advantage because these cells contain genetic inserts but do not express the CAR-associated phenotype. As a result, comparison with INE samples allows discrimination between effects caused by vector insertion itself and those specifically driven by functional CAR expression and signaling. The highly consistent upregulation of CD141, and HLA-DR in the TR3, TR5, and TR6 samples relative to INE indicates that these changes are likely associated with active CAR expression rather than transduction alone. Consequently, the INE comparison may represent the most informative control for studying CAR-specific signaling and expression-dependent immune remodeling.

In contrast, UT samples represent non-engineered cells that have not undergone transduction. Comparison with UT cells therefore captures the cumulative effects of the CAR T-cell transduction, including vector exposure and CAR-associated activation. The broader phenotypic changes observed relative to UT samples, including increased expression of activation and co-stimulatory markers such as CD25, CD28, CD45RO, CD95, CD141, and HLA-DR, suggest that CAR engineering induces substantial immune-state remodeling beyond control cellular phenotypes.

Together, these various comparisons provide complementary biological perspectives. Comparison with the DN cell population is most appropriate for evaluating productive CAR acquisition within engineered populations, comparison with INE cells is most informative for isolating CAR-expression-dependent biological effects, and comparison with UT cells best reflects the transduction impact of the CAR T-cell manufacturing process. Consideration of these distinct comparisons is therefore essential for accurate interpretation of CAR T-cell phenotypic and functional changes.

Conclusions

This study demonstrates that DAb-seq-based single-cell multi-omics enables characterization of CAR T-cell products by integrating VCN, transgene expression, and surface phenotype at single-cell resolution. DAb-seq VCN estimates were more closely aligned with ddPCR results. Distinct T-cell subsets and transduced populations exhibited unique transduction and phenotypic profiles. CD141 and HLA-DR were consistently enriched in CAR+ cells, indicating a CAR-associated phenotype, with HLA-DR supporting an activated T-cell state while the significance of CD141 remains to be determined. Control selection substantially influenced phenotypic interpretation, highlighting the importance of complementary controls. These findings support further investigation of single-cell multi-omics for CAR T-cell product characterization and its relationship to therapeutic response.

Abbreviations

AOC
antibody-oligonucleotide conjugate
B
ALL-B-cell acute lymphoblastic leukemia
CAR
T-chimeric antigen receptor T-cell
cDNA
complementary DNA
CGT
cell and gene therapy
CSB
cell staining buffer
DAb
seq-DNA and antibody sequencing
ddPCR
Digital Droplet PCR
DN
double negative
DNA
deoxyribonucleic acid

dNTPs
deoxynucleotide triphosphates
DP
double positive
ELISA
Enzyme-Linked Immunosorbent Assay
INE
insert-not-expressed
NGS
Next-Generation Sequencing
PBMC
peripheral blood mononuclear cell
PCR
Polymerase Chain Reaction
qPCR
Quantitative PCR
RNA
Ribonucleic acid
RNA
seq-RNA Sequencing
RT
PCR-Reverse Transcription PCR
SB
staining-background
SDS
sodium dodecyl sulfate
SDS
PAGE-Sodium dodecylsulfate polyacrylamide gel electrophoresis
SITE
seq-Selective enrichment and Identification of Tagged genomic DNA Ends by sequencing
TE
transduction efficiency
TR
transduced
Tcm
central memory T-cell
Tem
effector memory T-cell
Teff
effector T-cell

Tn
naïve T-cell
Tscm
stem cell memory T-cell
UT
untransduced
VCN
vector copy number
VSV
G-Vesicular Stomatitis Virus G protein
WGS
Whole Genome Sequencing

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki and complied with the U.S. Federal Policy for the Protection of Human Subjects (the 'Common Rule' under 45 CFR part 46). The study was reviewed and approved by the Institutional Review Board (IRB) of the National Institutes of Health (NIH) Intramural Research Program, and written informed consent was obtained from all participants prior to enrollment.

Consent for publication

All co-authors have read, approved, and consented to the submission and publication of this manuscript.

Availability of data and materials

Patient-related data not included in the paper were generated as part of completed or ongoing clinical trials and may be subject to patient confidentiality; therefore, access to these data may be restricted.

Single-cell DNA and protein sequencing data were analyzed using a custom workflow based on Tapestri DNA/DNA + Protein Analysis Pipeline (v3.0) (https://support.missionbio.com/hc/en-us/sections/17820181627415-Tapestri-DNA-DNA-Protein-Analysis-Pipeline). Any other data supporting the findings of this study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH).

Authors' contributions

JM, PJ and DS designed the article structure and write manuscript; JM and RY were responsible for performing sequencing assays; YC and YK performed flow cytometry assays; JM, YY and SW contributed to data analysis and statistical methods; HH and ZH contributed the Jurket standard reference cell line and manuscript revision; NZ, KM, PG and AD contributed to sample collection, clinical advice, and manuscript revision; JR and RPS contributed to data interpretation and manuscript revision; LS and SH contributed manuscript revision.

NIH disclaimer:

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH author(s) are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

NIST disclaimers

Certain commercial equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification does not imply recommendation or endorsement of any product or service by the National Institute of Standards and Technology (NIST), nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References

1. Gross G, Waks T, Eshhar Z. Expression of immunoglobulin-T-cell receptor chimeric molecules as functional receptors with antibody-type specificity. Proc Natl Acad Sci U S A. 1989;86(24):10024–8.
2. Grupp SA, Kalos M, Barrett D, Aplenc R, Porter DL, Rheingold SR, et al. Chimeric antigen receptor-modified T cells for acute lymphoid leukemia. N Engl J Med. 2013;368(16):1509–18.
3. Kochenderfer JN, Dudley ME, Kassim SH, Somerville RP, Carpenter RO, Stetler-Stevenson M, et al. Chemotherapy-refractory diffuse large B-cell lymphoma and indolent B-cell malignancies can be effectively treated with autologous T cells expressing an anti-CD19 chimeric antigen receptor. J Clin Oncol. 2015;33(6):540–9.
4. Cohen AD, Mateos MV, Cohen YC, Rodriguez-Otero P, Paiva B, van de Donk N, et al. Efficacy and safety of cilta-cel in patients with progressive multiple myeloma after exposure to other BCMA-targeting agents. Blood. 2023;141(3):219–30.
5. Porter DL, Levine BL, Kalos M, Bagg A, June CH. Chimeric antigen receptor-modified T cells in chronic lymphoid leukemia. N Engl J Med. 2011;365(8):725–33.

6. FDA. Approved Cellular and Gene Therapy Products 2025 [updated 08/15/2025. Available from: https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/approved-cellular-and-gene-therapy-products
7. Li YH, Huo Y, Yu L, Wang JZ. Quality Control and Nonclinical Research on CAR-T Cell Products: General Principles and Key Issues. Engineering. 2019;5(1):122–31.
8. Leibman RS, Richardson MW, Ellebrecht CT, Maldini CR, Glover JA, Secreto AJ, et al. Supraphysiologic control over HIV-1 replication mediated by CD8 T cells expressing a re-engineered CD4-based chimeric antigen receptor. PLoS Pathog. 2017;13(10):e1006613.
9. Montini E, Cesana D, Schmidt M, Sanvito F, Ponzoni M, Bartholomae C, et al. Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. Nat Biotechnol. 2006;24(6):687–96.
10. Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E, et al. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest. 2008;118(9):3132–42.
11. U.S. Food and Drug Administration. Considerations for the Development of Chimeric Antigen Receptor (CAR) T Cell Products: Guidance for Industry. U.S. Food and Drug Administration; 2024 January 2024. Report No.: FDA-2021-D-0404.
12. Xu W, Liang J, Wang F, Yang L. Comparative evaluation of gene copy number estimation techniques in genetically modified crops: insights from Southern blotting, qPCR, dPCR and NGS. Plant Biotechnol J. 2024;22(12):3456–8.
13. Joshi M, Keith Pittman H, Haisch C, Verbanac K. Real-time PCR to determine transgene copy number and to quantitate the biolocalization of adoptively transferred cells from EGFP-transgenic mice. Biotechniques. 2008;45(3):247–58.
14. Lu A, Liu H, Shi R, Cai Y, Ma J, Shao L, et al. Application of droplet digital PCR for the detection of vector copy number in clinical CAR/TCR T cell products. J Transl Med. 2020;18(1):191.
15. Murphy LA, Marians RC, Miller K, Brenton MD, Mallo RLV, Kohler ME, et al. Digital polymerase chain reaction strategies for accurate and precise detection of vector copy number in chimeric antigen receptor T-cell products. Cytotherapy. 2023;25(1):94–102.
16. Shao L, Shi R, Zhao Y, Liu H, Lu A, Ma J, et al. Genome-wide profiling of retroviral DNA integration and its effect on clinical pre-infusion CAR T-cell products. J Transl Med. 2022;20(1):514.
17. Demaree B, Delley CL, Vasudevan HN, Peretz CAC, Ruff D, Smith CC, et al. Joint profiling of DNA and proteins in single cells to dissect genotype-phenotype associations in leukemia. Nat Commun. 2021;12(1):1583.
18. Aertgeerts M, Meyers S, Gielen O, Lamote J, Dewaele B, Tajdar M, et al. Single-cell DNA and surface protein characterization of high hyperdiploid acute lymphoblastic leukemia at diagnosis and during treatment. Hemasphere. 2025;9(2):e70085.
19. Yang Y, Parikh S, Li CY, Murphy L, Patel K, Ayaz Q, et al. A single-cell multi-omics assay for simultaneous measurement of vector copy number and protein expression in CAR T cells. Mol Ther Adv. 2026;34(2):201738.

20. Bejarano DA, Puertas MC, Borner K, Martinez-Picado J, Muller B, Krausslich HG. Detailed Characterization of Early HIV-1 Replication Dynamics in Primary Human Macrophages. Viruses. 2018;10(11).
21. Sadelain M, Brentjens R, Riviere I. The basic principles of chimeric antigen receptor design. Cancer Discov. 2013;3(4):388–98.
22. Rodriguez-Marquez P, Calleja-Cervantes ME, Serrano G, Oliver-Caldes A, Palacios-Berraquero ML, Martin-Mallo A, et al. CAR density influences antitumoral efficacy of BCMA CAR T cells and correlates with clinical outcome. Sci Adv. 2022;8(39):eabo0514.
23. FDA. FDA CBER Guidance Webinar. Considerations for the Development of CAR T Cell Products In: Services USDoHaH, editor. 2024.
24. Brudno JN, Kochenderfer JN. Chimeric antigen receptor T-cell therapies for lymphoma. Nat Rev Clin Oncol. 2018;15(1):31–46.
25. Gattinoni L, Speiser DE, Lichterfeld M, Bonini C. T memory stem cells in health and disease. Nat Med. 2017;23(1):18–27.
26. Nobles CL, Sherrill-Mix S, Everett JK, Reddy S, Fraietta JA, Porter DL, et al. CD19-targeting CAR T cell immunotherapy outcomes correlate with genomic modification by vector integration. J Clin Invest. 2020;130(2):673–85.
27. Wang W, Fasolino M, Cattau B, Goldman N, Kong W, Frederick MA, et al. Joint profiling of chromatin accessibility and CAR-T integration site analysis at population and single-cell levels. Proc Natl Acad Sci U S A. 2020;117(10):5442–52.
28. Niu A, Zou J, Hu X, Zhang Z, Su L, Wang J, et al. Differences in the phenotypes and transcriptomic signatures of chimeric antigen receptor T lymphocytes manufactured via electroporation or lentiviral transfection. Front Immunol. 2023;14:1068625.
29. Dabiri H, Safarzadeh Kozani P, Habibi Anbouhi M, Mirzaee Godarzee M, Haddadi MH, Basiri M, et al. Site-specific transgene integration in chimeric antigen receptor (CAR) T cell therapies. Biomark Res. 2023;11(1):67.
30. Celichowski P, Turi M, Charvatova S, Radhakrishnan D, Feizi N, Chyra Z, et al. Tuning CARs: recent advances in modulating chimeric antigen receptor (CAR) T cell activity for improved safety, efficacy, and flexibility. J Transl Med. 2023;21(1):197.
31. Alizadeh D, Wong RA, Yang X, Wang D, Pecoraro JR, Kuo CF, et al. IL15 Enhances CAR-T Cell Antitumor Activity by Reducing mTORC1 Activity and Preserving Their Stem Cell Memory Phenotype. Cancer Immunol Res. 2019;7(5):759–72.
32. Korin YD, Zack JA. Progression to the G1b phase of the cell cycle is required for completion of human immunodeficiency virus type 1 reverse transcription in T cells. J Virol. 1998;72(4):3161–8.
33. Bergantz L, Subra F, Deprez E, Delelis O, Richetta C. Interplay between Intrinsic and Innate Immunity during HIV Infection. Cells. 2019;8(8).
34. Lin P, Lin Y, Lennon DP, Correa D, Schluchter M, Caplan AI. Efficient lentiviral transduction of human mesenchymal stem cells that preserves proliferation and differentiation capabilities. Stem Cells

Transl Med. 2012;1(12):886–97.
35. Juillerat A, Tkach D, Busser BW, Temburni S, Valton J, Duclert A, et al. Modulation of chimeric antigen receptor surface expression by a small molecule switch. BMC Biotechnol. 2019;19(1):44.
36. Roe T, Reynolds TC, Yu G, Brown PO. Integration of murine leukemia virus DNA depends on mitosis. EMBO J. 1993;12(5):2099–108.
37. Finkelshtein D, Werman A, Novick D, Barak S, Rubinstein M. LDL receptor and its family members serve as the cellular receptors for vesicular stomatitis virus. Proc Natl Acad Sci U S A. 2013;110(18):7306–11.
38. Hrecka K, Hao C, Gierszewska M, Swanson SK, Kesik-Brodacka M, Srivastava S, et al. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. Nature. 2011;474(7353):658–61.
39. Fassati A, Goff SP. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. J Virol. 2001;75(8):3626–35.
40. Zheng Z, Chinnasamy N, Morgan RA. Protein L: a novel reagent for the detection of chimeric antigen receptor (CAR) expression by flow cytometry. J Transl Med. 2012;10:29.
41. Kaech SM, Ahmed R. Memory CD8 + T cell differentiation: initial antigen encounter triggers a developmental program in naive cells. Nat Immunol. 2001;2(5):415–22.
42. Gattinoni L, Lugli E, Ji Y, Pos Z, Paulos CM, Quigley MF, et al. A human memory T cell subset with stem cell-like properties. Nat Med. 2011;17(10):1290–7.
43. Ellis J. Silencing and variegation of gammaretrovirus and lentivirus vectors. Hum Gene Ther. 2005;16(11):1241–6.
44. Pearce EL, Poffenberger MC, Chang CH, Jones RG. Fueling immunity: insights into metabolism and lymphocyte function. Science. 2013;342(6155):1242454.
45. Kustikova O, Fehse B, Modlich U, Yang M, Dullmann J, Kamino K, et al. Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. Science. 2005;308(5725):1171–4.
46. Pipkin ME, Sacks JA, Cruz-Guilloty F, Lichtenheld MG, Bevan MJ, Rao A. Interleukin-2 and inflammation induce distinct transcriptional programs that promote the differentiation of effector cytolytic T cells. Immunity. 2010;32(1):79–90.
47. Chen D, Tang TX, Deng H, Yang XP, Tang ZH. Interleukin-7 Biology and Its Effects on Immune Cells: Mediator of Generation, Differentiation, Survival, and Homeostasis. Front Immunol. 2021;12:747324.
48. Yi JS, Du M, Zajac AJ. A vital role for interleukin-21 in the control of a chronic viral infection. Science. 2009;324(5934):1572–6.
49. Luo M, Gong W, Zhang Y, Li H, Ma D, Wu K, et al. New insights into the stemness of adoptively transferred T cells by gammac family cytokines. Cell Commun Signal. 2023;21(1):347.
50. Cole EB, Lamcaj S, Sydenstricker AV, Voss AG, Hiner CR, Hur HB, et al. IL-7/IL-15/IL-21 cytokine-fusion scaffold generates highly functional CAR T cells enriched in long-lived T memory stem cells. Sci Adv. 2026;12(11):eaec2632.

51. Cieri N, Camisa B, Cocchiarella F, Forcato M, Oliveira G, Provasi E, et al. IL-7 and IL-15 instruct the generation of human memory stem T cells from naive precursors. Blood. 2013;121(4):573–84.
52. Klebanoff CA, Crompton JG, Leonardi AJ, Yamamoto TN, Chandran SS, Eil RL, et al. Inhibition of AKT signaling uncouples T cell differentiation from expansion for receptor-engineered adoptive immunotherapy. JCI Insight. 2017;2:23.
53. Boroughs AC, Larson RC, Marjanovic ND, Gosik K, Castano AP, Porter CBM, et al. A Distinct Transcriptional Program in Human CAR T Cells Bearing the 4-1BB Signaling Domain Revealed by scRNA-Seq. Mol Ther. 2020;28(12):2577–92.
54. Toda M, Shao ZF, Yamaguchi KD, Takagi T, D'Alessandro-Gabazza CN, Taguchi O et al. Differential Gene Expression in Thrombomodulin (TM; CD141) and TM Dendritic Cell Subsets. PLoS ONE. 2013;8(8).

Tables

Table 1 to 3 are available in the Supplementary Files section.