Supplemental Material

Supplemental Methods

Subjects and Samples

Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Hypaque density gradient centrifugation followed by blood draw in the patient and healthy donor (HD). Isolated PBMCs were cryopreserved in liquid nitrogen according to standard protocols until use.

Flow cytometry profiling of PBMCs

PBMCs were stained with antibody mixtures on ice for 30 min in RPMI 1640 (Life Technologies). Samples were subsequently acquired using a BD LSR Fortessa cytometer (BD Biosciences) and post-acquisition analysis was performed using Flowjo software (v.10.8.1; Flowjo LLC, BD Biosciences). Antibodies used for flow cytometry analyses were: anti-human CD4 in BV421 (Cat#550761, clone 581, BD Biosciences), anti- human CD8 in APC-Cy7 (Cat#555462, clone HIT2, BD Biosciences), anti-CD44 in BV711 (Cat#328108, clone 5E10, Biolegend), anti- human CD62L in PE-Cy7 (Cat#562978, clone HI10A, BD Biosciences), anti-human PD-1 in FITC (Cat#313314, clone BV10A4H2, Biolegend), and anti-human CD57 in APC (Cat#304142, clone HI100, Biolegend) and anti-mouse FMC63 scFv monoclonal antibody in PE (Cat#200106, clone R19M, CytoArt) to detect CAR-T cells.

Whole transcriptome amplification (WTA), cDNA library preparation and sequencing

PBMCs from patient and HD were barcoded using “TotalSeq C hash-tag” oligonucleotides (Biolegend): TotalSeq™-C0255 anti-human Hashtag 5 Antibody (Cat#394669, clone LNH-94; 2M2) for patient’s pre-treatment sample, TotalSeq™-C0256 anti-human Hashtag 6 Antibody (Cat#394671, clone LNH-94; 2M2) for post-treatment sample, and TotalSeq™-C0259 anti-human Hashtag 9 Antibody (Cat#394677, clone LNH-94; 2M2) for HD sample. Hashed samples were

pooled and stained with Total-seq antibodies for cell surface markers (various cell surface markers conjugated with oligonucleotides, Biolegend) per the manufacturer's protocol (https://www.biolegend.com/en-us/protocols/totalseq-b-or-c-with-10x-feature-barcoding-technology). Stained cells were encapsulated for single-cell reverse transcription using the Chromium Single-Cell V(D)J Reagent Kit v1.1 (10X Genomics, Pleasanton, CA) as per the manufacturer's protocol (https://assets.ctfassets.net/an68im79xiti/6se7DVQQ0xSCVYp4eQd4Ld/9b822ceb045ec7b20df633cddfe3021f/CG000208_ChromiumNextGEMSingleCellV_D_J_ReagentKit_v1.1_FeatureBarcodingtechnology_RevD.pdf). Single cell RNA-seq libraries (including 5' gene expression libraries, V(D)J enriched libraries, and cell surface protein libraries) were pooled and sequenced on an Illumina Novaseq System (Illumina, San Diego, CA) with a customized paired end, single indexing (26/8/0/98-bp) format according to 10X Genomics recommendations.

scRNA-seq data analysis

Preprocessing of scRNA-seq and scTCR-seq data. Raw sequencing data was processed with the Cell Ranger Software (Version 6.0.0), using the cellranger multi workflow and the hg19 references for gene expression and totalSeq antibody barcode sequences for feature barcoding data, to generate gene-cell and protein-cell matrices for each sample for further analysis. Cells with <500 genes were filtered, genes with at least one UMI count detected in at least one cell were used for further analysis. The top 1,000 most variable genes were identified based on their mean and dispersion (variance/mean). Graph-based clusters of cells were visualized by two-dimensional t-distributed Stochastic Neighbor Embedding (tSNE) using Seurat4. In each cluster, the mean expression of each gene was calculated across all cells to identify genes that were enriched in a specific cluster. Each gene from the cluster was compared to the median expression of the same gene from cells in all other clusters. Cell types were assigned to each cluster based on significance in overlapping between signature genes and cluster-specific genes. Cell type

assignment was also validated with expression of well-established cell surface markers. FindMarkers function in Seurat were used to identify differentially expressed genes between post-treatment and pre-treatment sample. Genes with *P* value < 0.05 and Log (average fold change) > 0.1 were regarded as differentially expressed genes. Heatmaps and network visualization were generated with ggplot2 and heatmap2 in the R package. Gene set enrichment analysis (http://software.broadinstitute.org/gsea) was performed to identify enriched gene sets of all genes with differential expression (based on average log fold change). Gene ontology was assessed with the R package topGO v2.26 using the algorithm elim[16], a minimum node size of 10, and genes that were expressed over 100 cells as the background gene list. *P* values derived from the gene ontology analysis were not corrected for multiple testing. We examined the biological processes GO terms[17,18] and KEGG pathways[19,20]. Pathway (only canonical pathways) enrichment analysis was performed with Genomatix Generanker (http://www.genomatix.de, Genomatix)[21]. Gene names of significantly different genes were uploaded and corresponding overrepresented pathways, related *P* values, and hit genes were saved.

TCR reads were aligned to the GRCh38 reference genome and consensus TCR annotation was performed using the cellranger vdj program (10x Genomics, version 2.1.0).

Diversity index and Power Law curve fitting. There are many ways of defining the diversity of a population, clonal types in this study, with each method providing a different representation of the number of clones (Identical TCR chains) present (richness) and of their relative frequencies (evenness). The Shannon entropy weighs both of these aspects of diversity equally, is an intuitive measure whereby the maximum value is determined by the total size of the repertoire. Entropy values decreases with increasing inequality of frequencies as a result of clonal expansion. The Shannon entropy in a population of N clones with nucleotide frequencies pi is defined by the following equation:

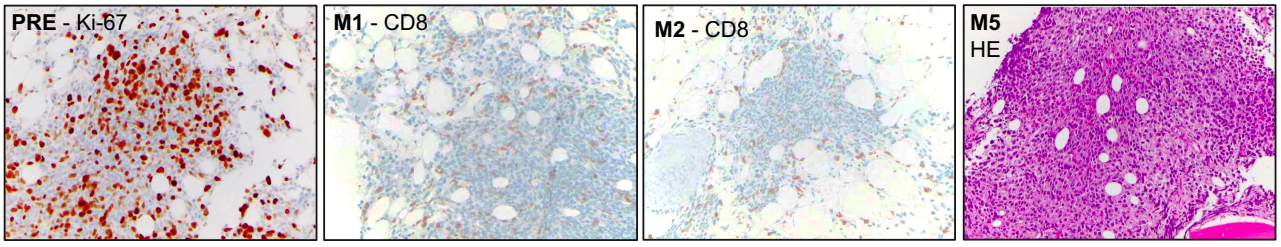
$$H(P) = -\sum_{i=1}^{n} p_i log_2 p_i$$

The Gini coefficient is a number aimed at measuring the inequality in a distribution. It is most often used in economics to measure a country's wealth distribution and has been widely used in diversity assessment of TCRs[25]. The Gini coefficient is usually defined mathematically based on the Lorenz curve or Relative mean absolute difference[26]. The Gini index and Shannon entropy for diversity and clonality analysis were calculated with the R package of tCR (https://imminfo.github.io/tcr/).

Data availability

Raw fastq files of scRNA-seq data are available in NCBI Gene Expression Omnibus with the primary accession code GSE201704.

Figure S1

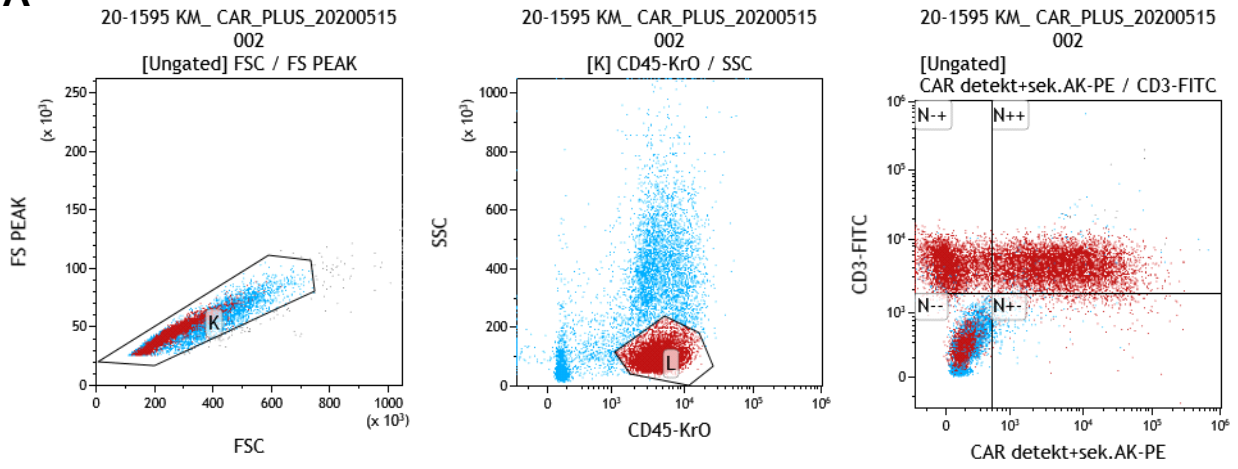


Bone marrow biopsy results

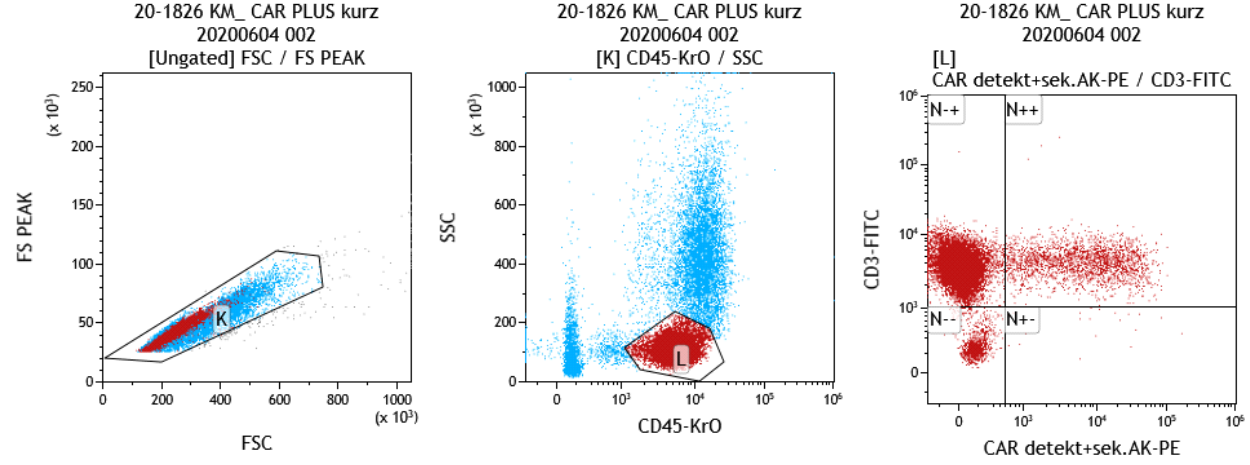
From left to right: Ki-67 staining of lymphoma cells in the bone marrow, CD8+ T-lymphocytes one month and two months after CAR transfusion, H&E staining of the bone marrow depicting pronounced BM infiltration at relapse (month 5).

Figure S2

A



B

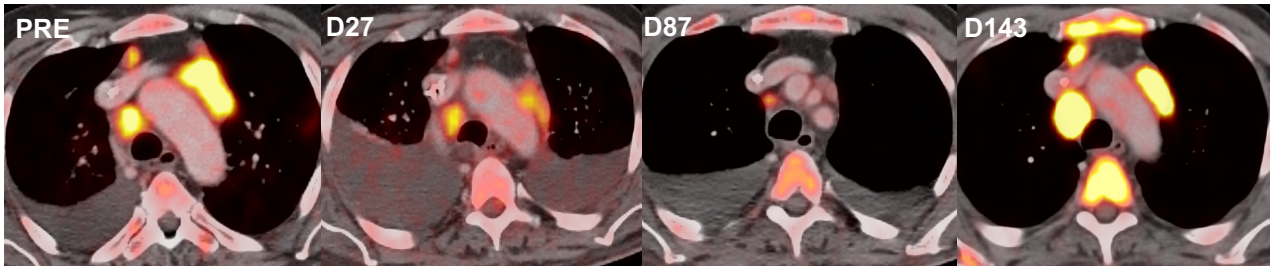


Migration of CAR T-cells to the BM by flow cytometry

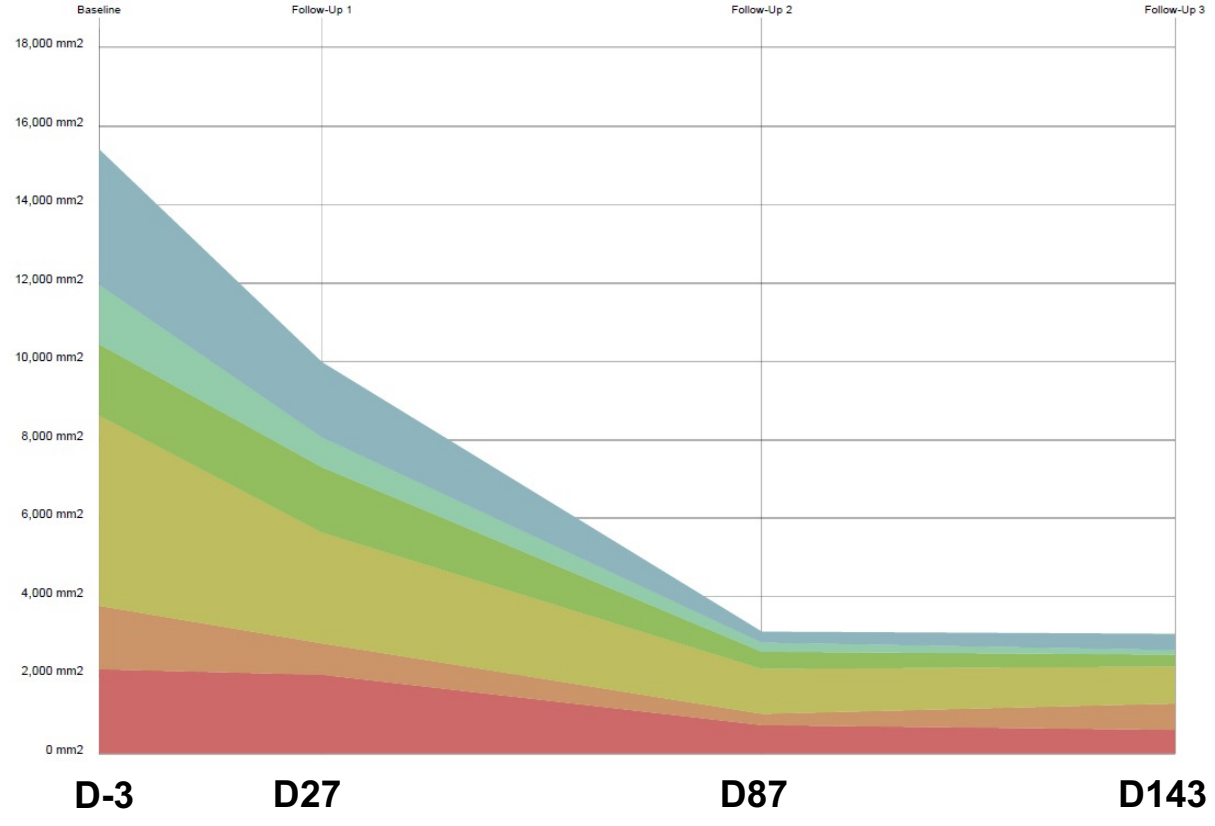
Flow Cytometric Analysis of the BM on day 27 (A) and day 47 (B): CD3+ CAR+ T-lymphocytes were detected (top right, right panel) utilizing a two-step staining method with a biotinylated CD19 protein (Blumenberg V et al. P04.01 Immunomonitoring of CD19. CAR T-cells in Large B-Cell Lymphoma- a two-center experience. *Journal for ImmunoTherapy of Cancer.* 2021;9(Suppl 1):A16-A16.).

Figure S3

A

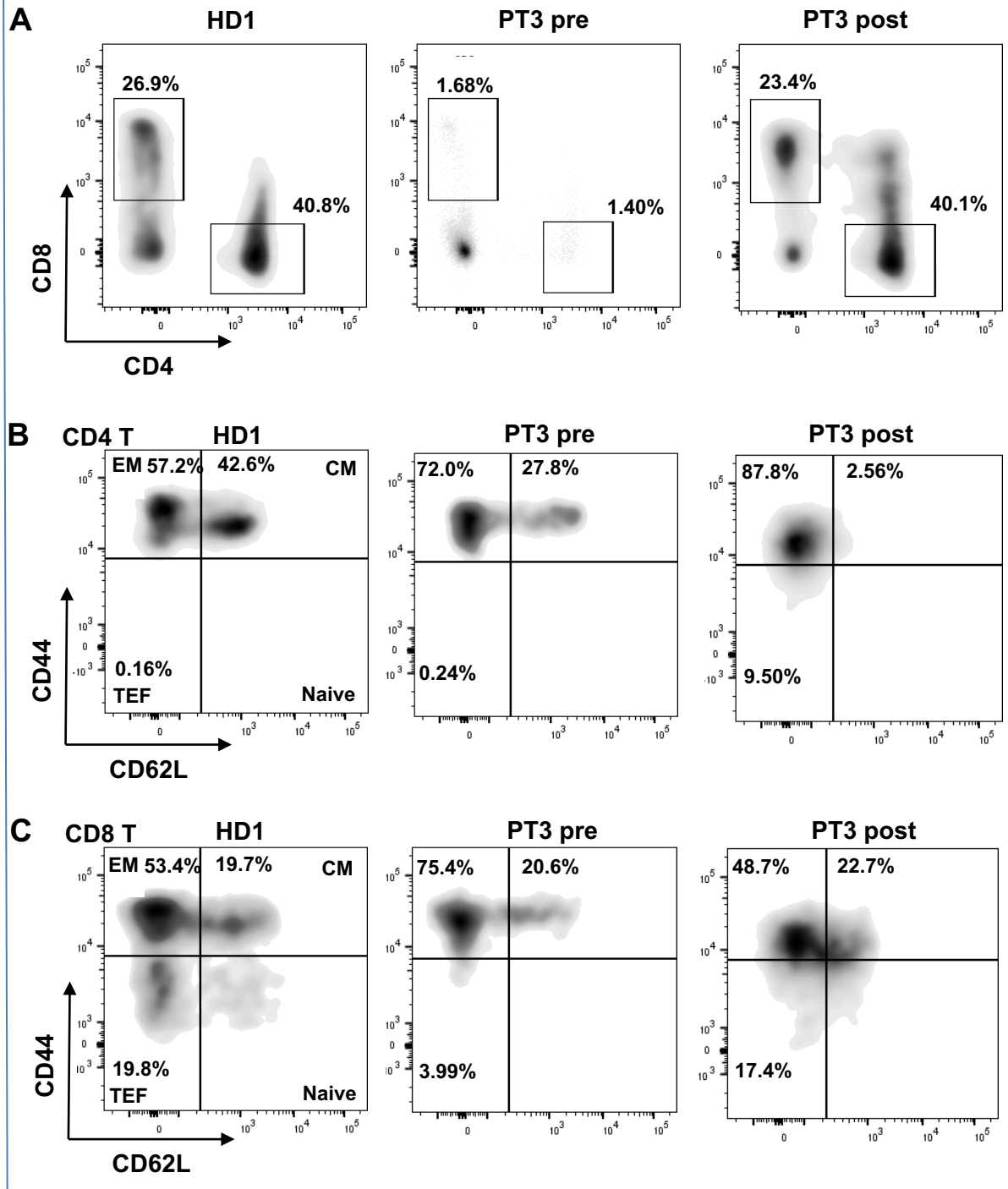


B



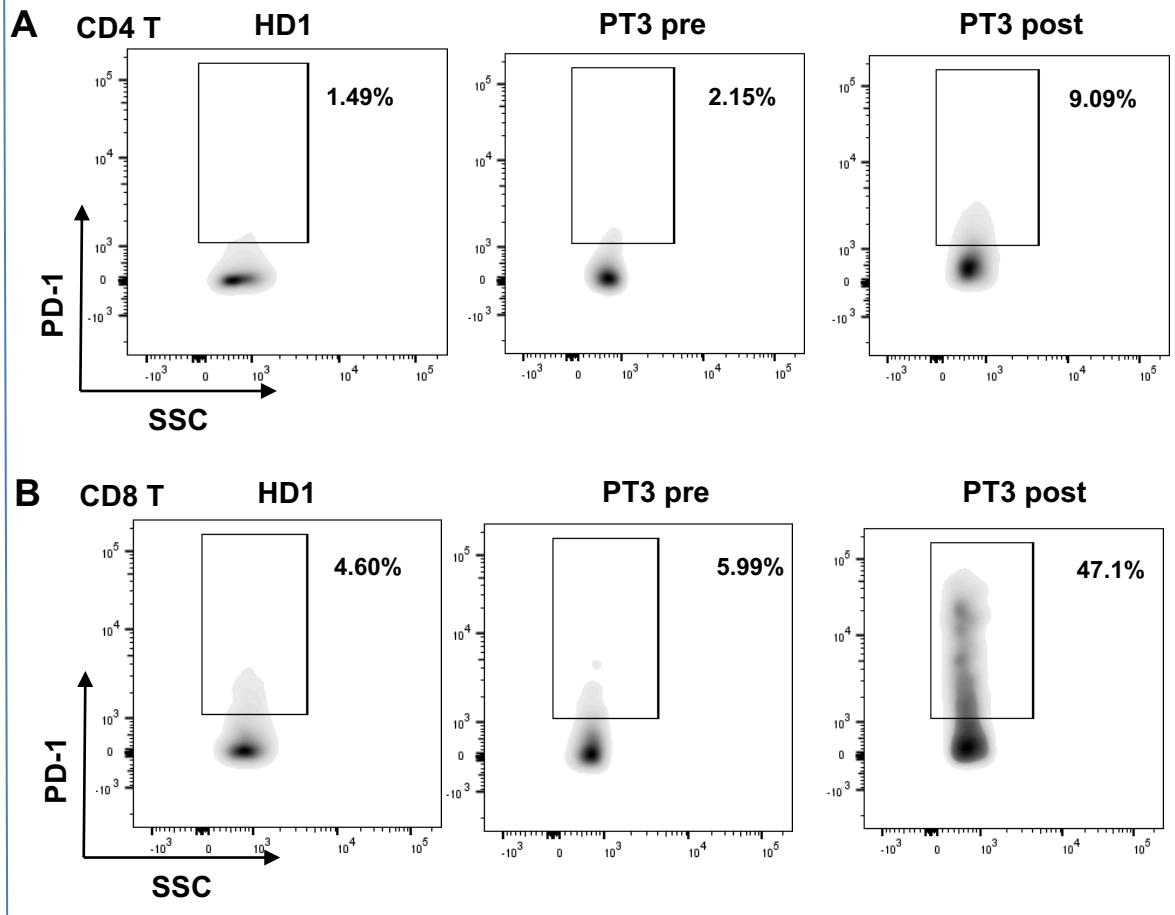
PET/CT findings

(A) Radiographic evidence of partial response and eventual relapse in ^{18}FDG PET/CT imaging across four time points. Mediastinal lymphoma masses are depicted on transverse cross-section. (B) Graphical depiction of response over time. Tumor extent was quantified as the sum of the product of perpendicular diameters according to Lugano criteria over four time points and is graphed on the y-axis.



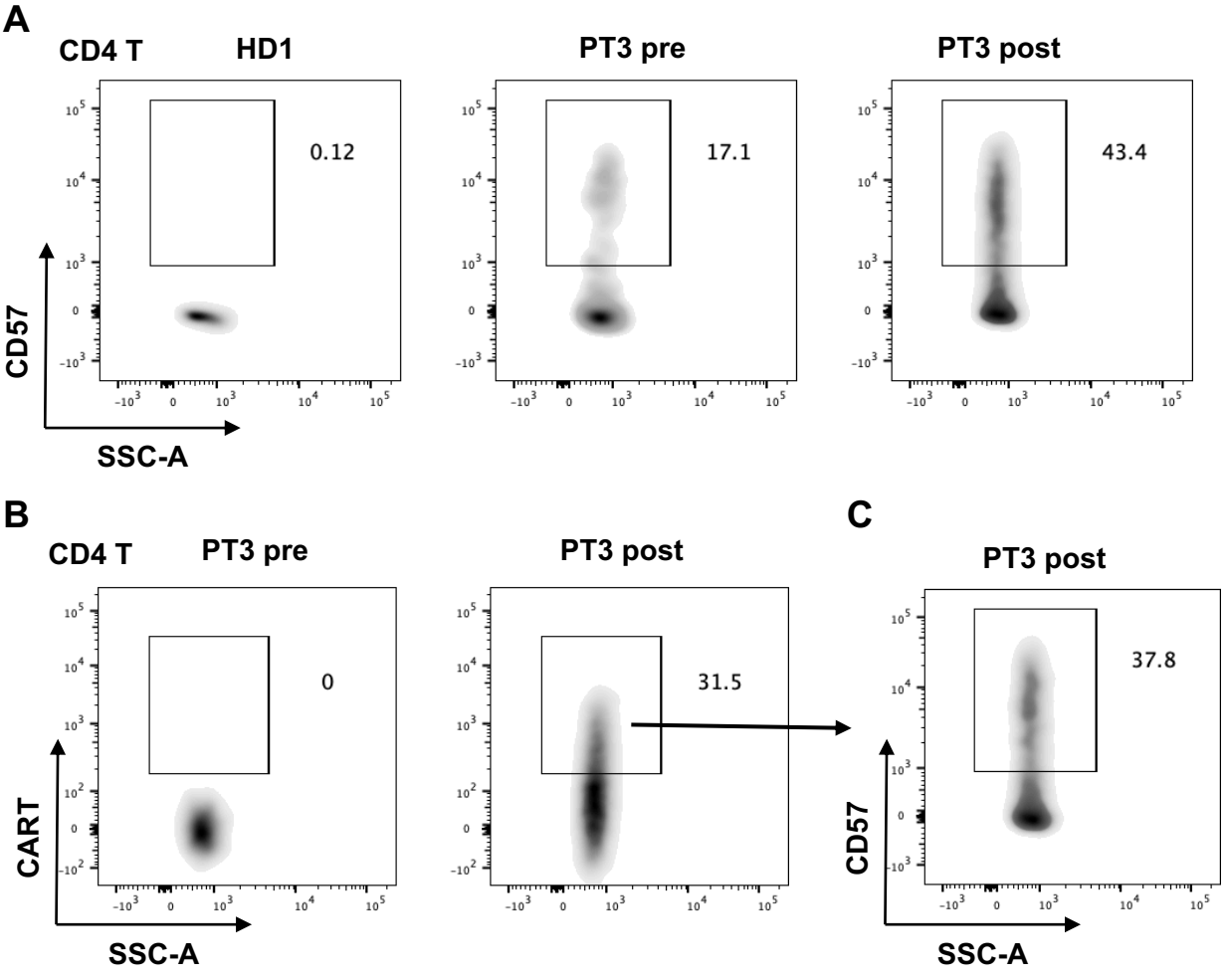
Phenotyping of T cells in pre- and post-treatment samples.

(A) CD4+ and CD8+ T cell populations in pre- and post-treatment samples. Expression of CD44 and CD62L for T cell subsets including naïve, central memory (CM), effector memory (EM), and terminal effector (TEF) CD4+ (B) and CD8+ (C) T-cells.



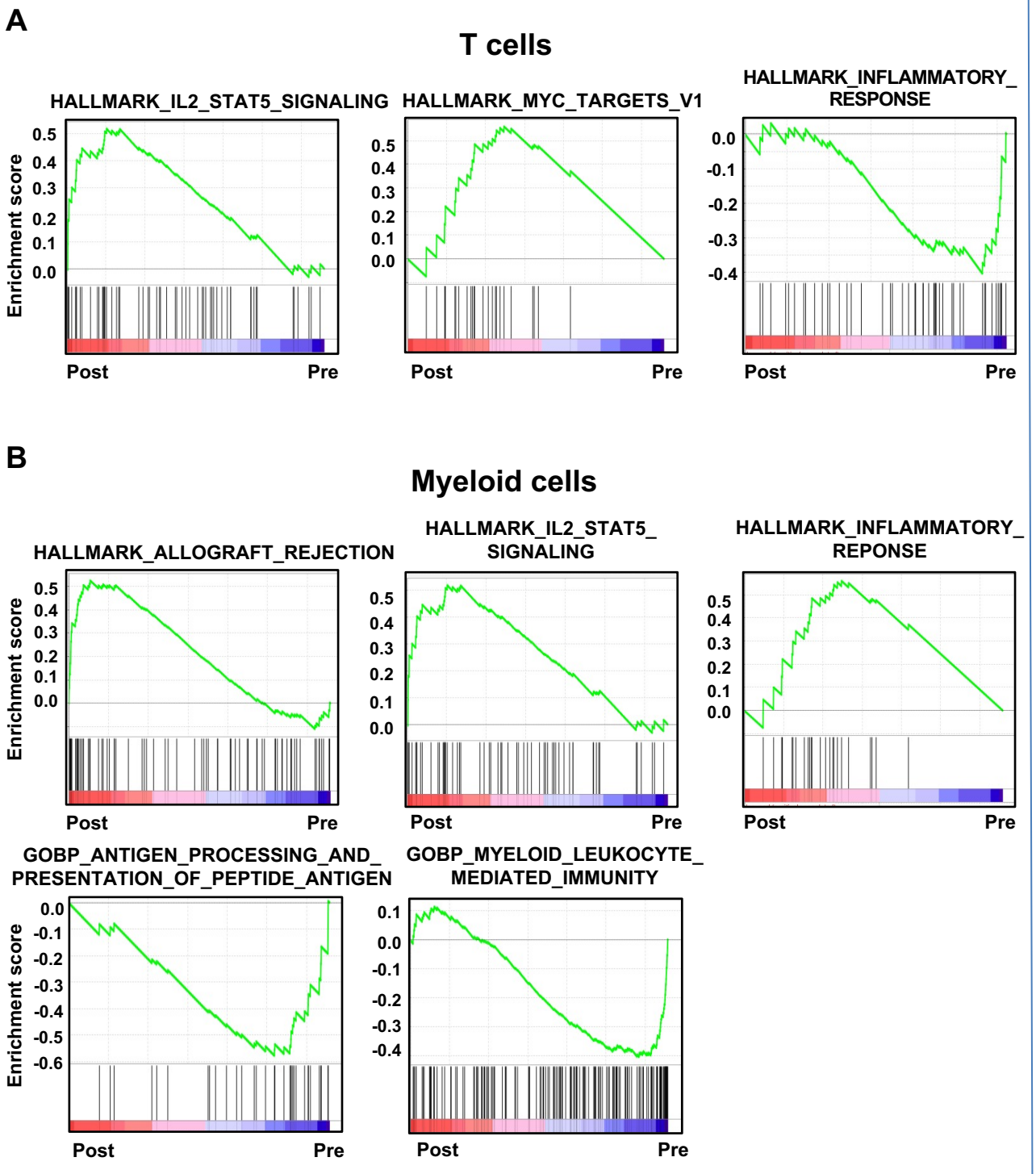
T-cell exhaustion in pre- and post-treatment samples.

Expression of PD-1 in CD4+ (A) and CD8+ (B) T cells in pre- and post-treatment samples.



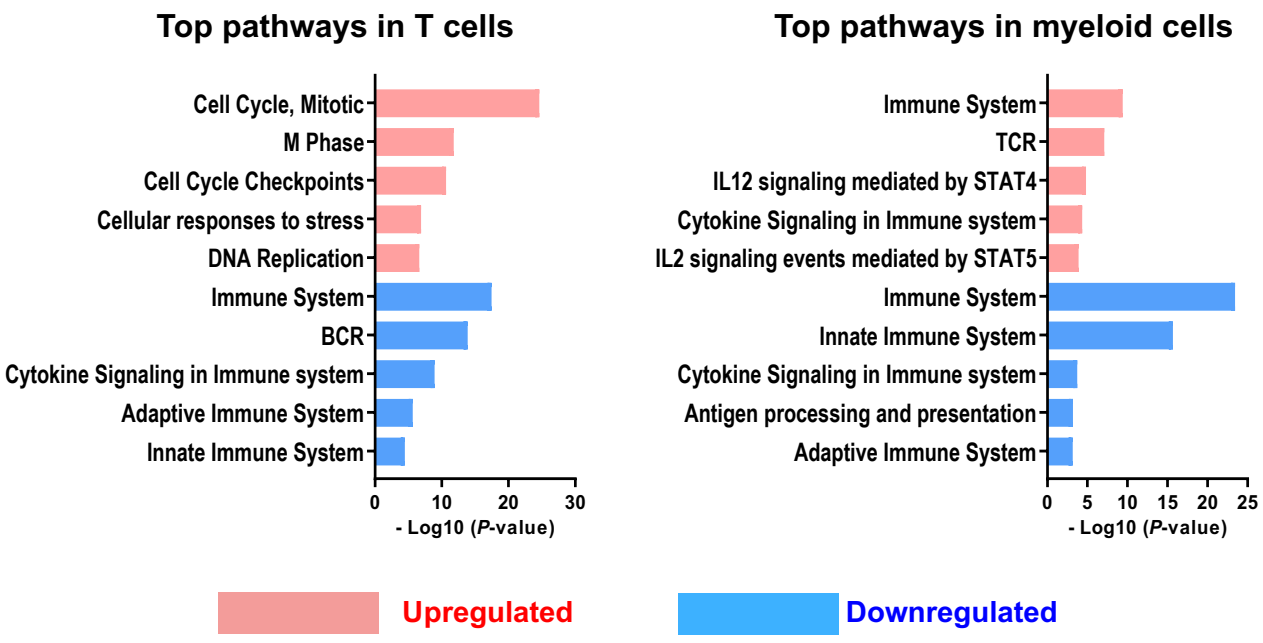
CD57 expression in pre- and post-treatment samples.

Expression of CD57 in CD4+ T-cells in pre- and post-treatment samples (A). Expression of CAR-T in CD4+ T-cells in pre- and post-treatment samples (B). Expression of CD57 in CD4+ CAR T-cells in post-treatment sample (C).



Altered gene pathways after treatment.

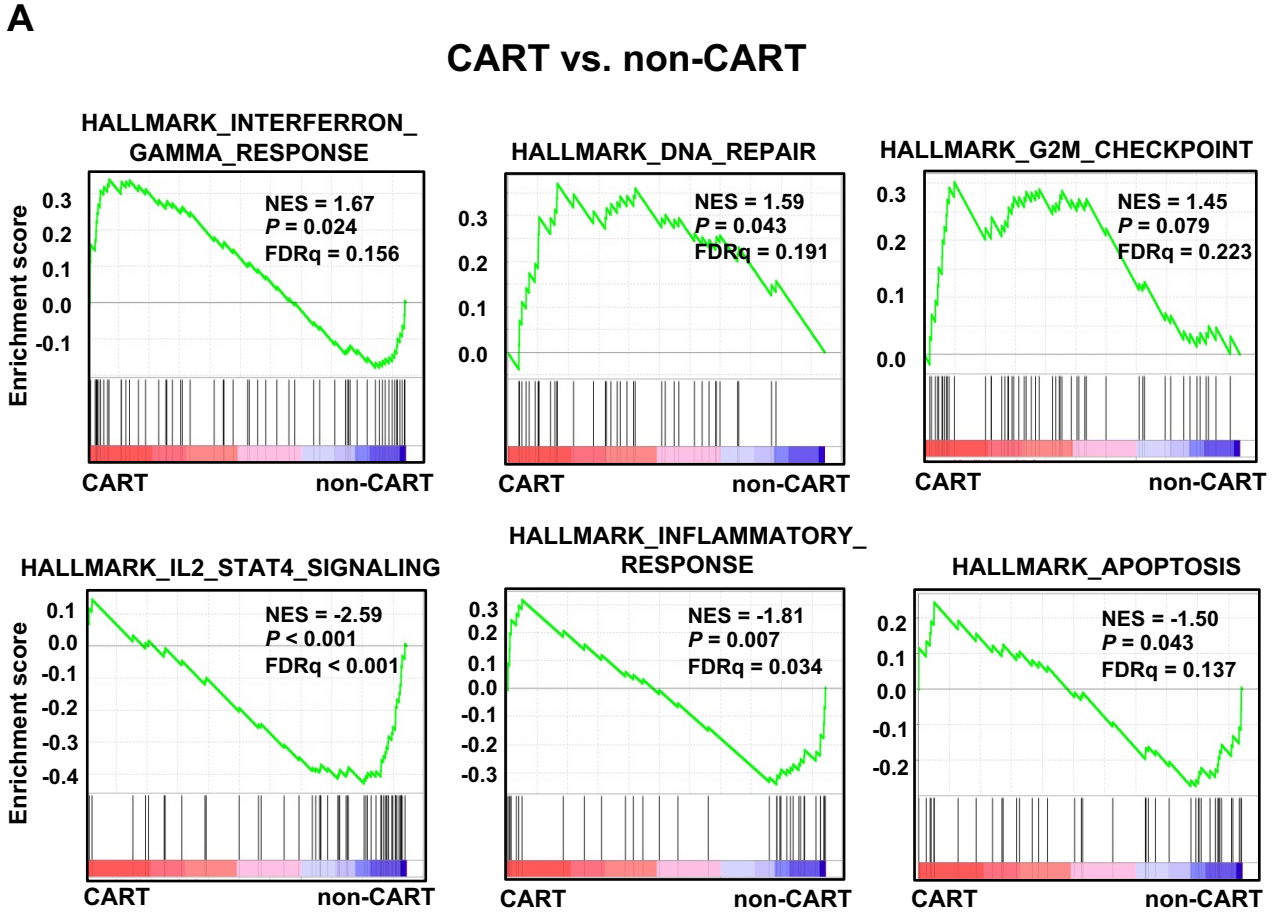
Gene Set Enrichment Analysis (GESA) plots of differentially expressed genes of T-cells (A) and myeloid cells (B) in post-treatment sample versus pre-treatment sample.



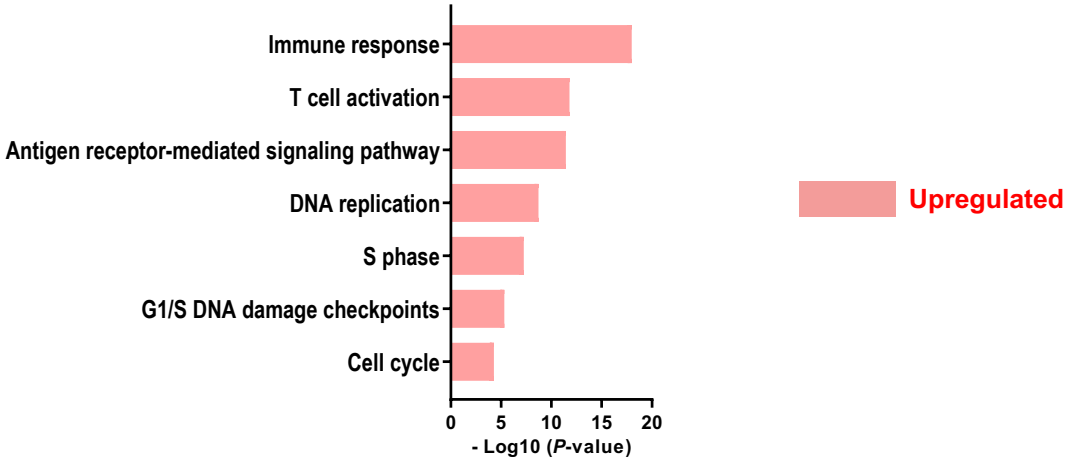
Pathway Analysis

Bar plot showing top upregulated pathways in T cells (C) and myeloid cells (D) in post-treatment sample versus pre-treatment sample, analyzed by Genomatix using top differentially expressed genes.

A



B



Altered gene pathways in CAR T-cells

(A) GSEA plots of differentially expressed genes of CART cells (top 500 T cells with highest CART expression) versus non-CART cells (bottom 500 T cells with lowest CART expression). (B) Bar plot showing top upregulated pathways in CAR T-cells right) in post-treatment sample, analyzed by Genomatix using top differentially expressed genes.