nature biotechnology

Article

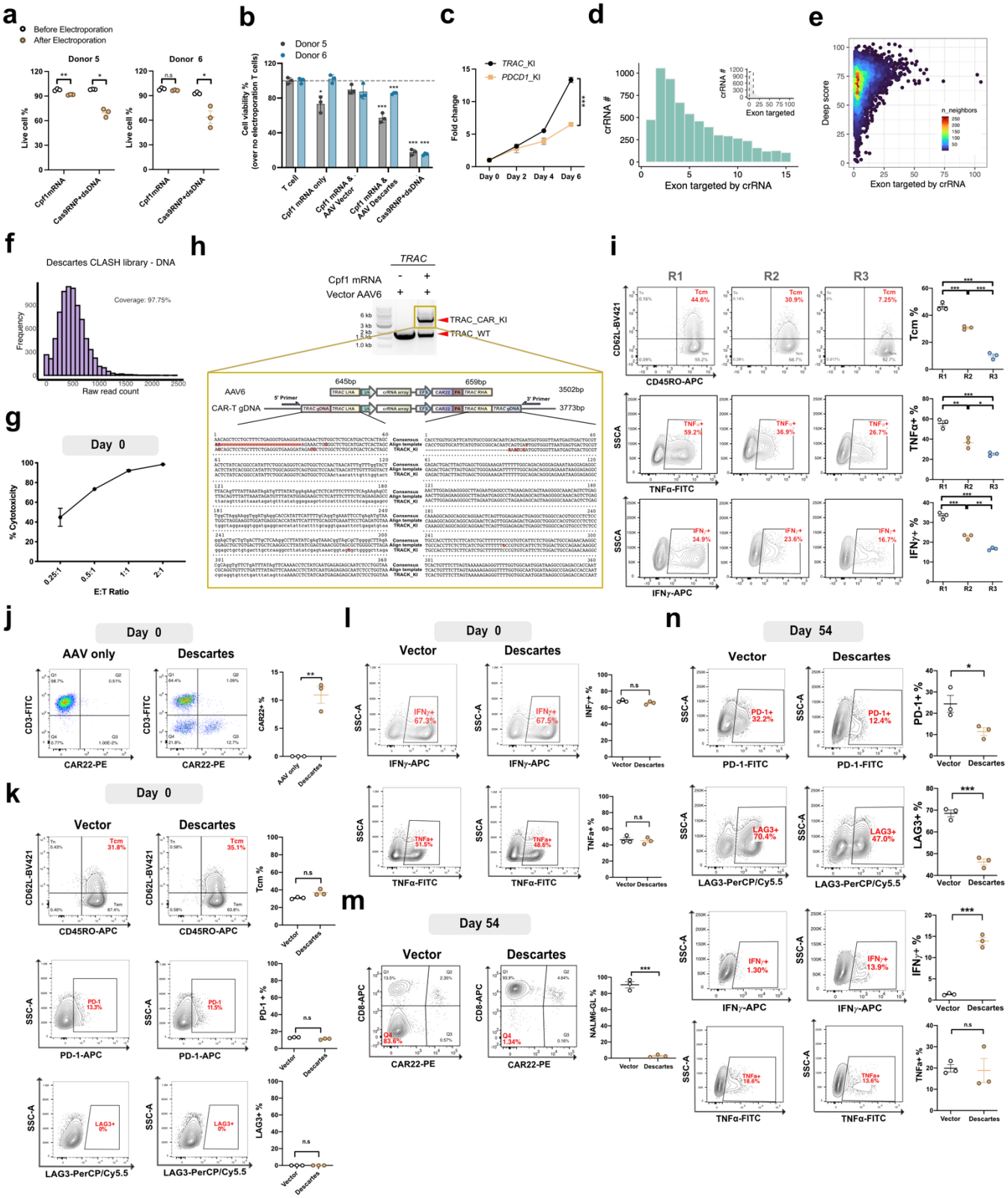
https://doi.org/10.1038/s41587-022-01639-x

Massively parallel knock-in engineering of human T cells

In the format provided by the authors and unedited

Contents

Supplementary Figure 1



Supplementary Figure 1: Establishment of long-term co-culture experimental conditions and the CLASH system for massively parallel CAR-T engineering

a, Live cell percentages before and after electroporation for 4 hours, where viability was determined by trypan blue staining. Unpaired two-sided T test was used to assess statistical significance. * $p < 0.05$ and n.s., not significant. Data are shown as mean ± s.e.m (n = 3).

b, Estimation of cells lost after electroporation with Cpf1 mRNA, CLASH-vector, CLASH-Descartes and non-viral targeting editing method after electroporation 2 days. Unpaired two-sided T test was used to assess statistical significance. * $p < 0.05$ and *** $p < 0.001$. Data are shown as mean ± s.e.m (n = 3).

c, Proliferation of CLASH-mediated *TRAC* KI and *PDCD1* KI CAR-T cells. Cells were counted every two days after electroporation (n = 3). Two-way ANOVA was used to assess significance, *** $p < 0.001$. Data are shown as mean ± s.e.m.

d, Histogram summary of the gene exons targeted by Descartes Cpf1 crRNAs. Inset histogram shows crRNA targeting for all exons, while the main panel highlights targeting within the first 15 exons.

e, Density plot showing the exon targeting of Descartes Cpf1 crRNAs, relative to the predicted activity, based on DeepCpf1 scores. The density of points is calculated as the # of neighbor points and is represented by a heat map color-scale.

f, Histogram of crRNA representation by NGS sequencing of Descartes plasmid library.

g, Baseline cytotoxicity of vector control CAR-T cells by kill assay at day 0 in donor 2. In vitro cytotoxic activity of CAR-T cells was measured by bioluminescence assay at different E/T ratios, using NALM6-GL cells stably transduced with GFP and luciferase genes as target cells. Data are shown as mean ± s.e.m.

h, Agarose DNA gel showing representative PCR reactions using primers at the *TRAC* locus (primers correspond to sequence outside the homology arms from the AAV donor). The upper band corresponds to sequences with transgene insertion. The detailed Sanger sequencing result was shown at right (misaligned DNA highlighted in pink). T cell genomic DNA was collected 5 days after electroporation. Represented data from four experiments.

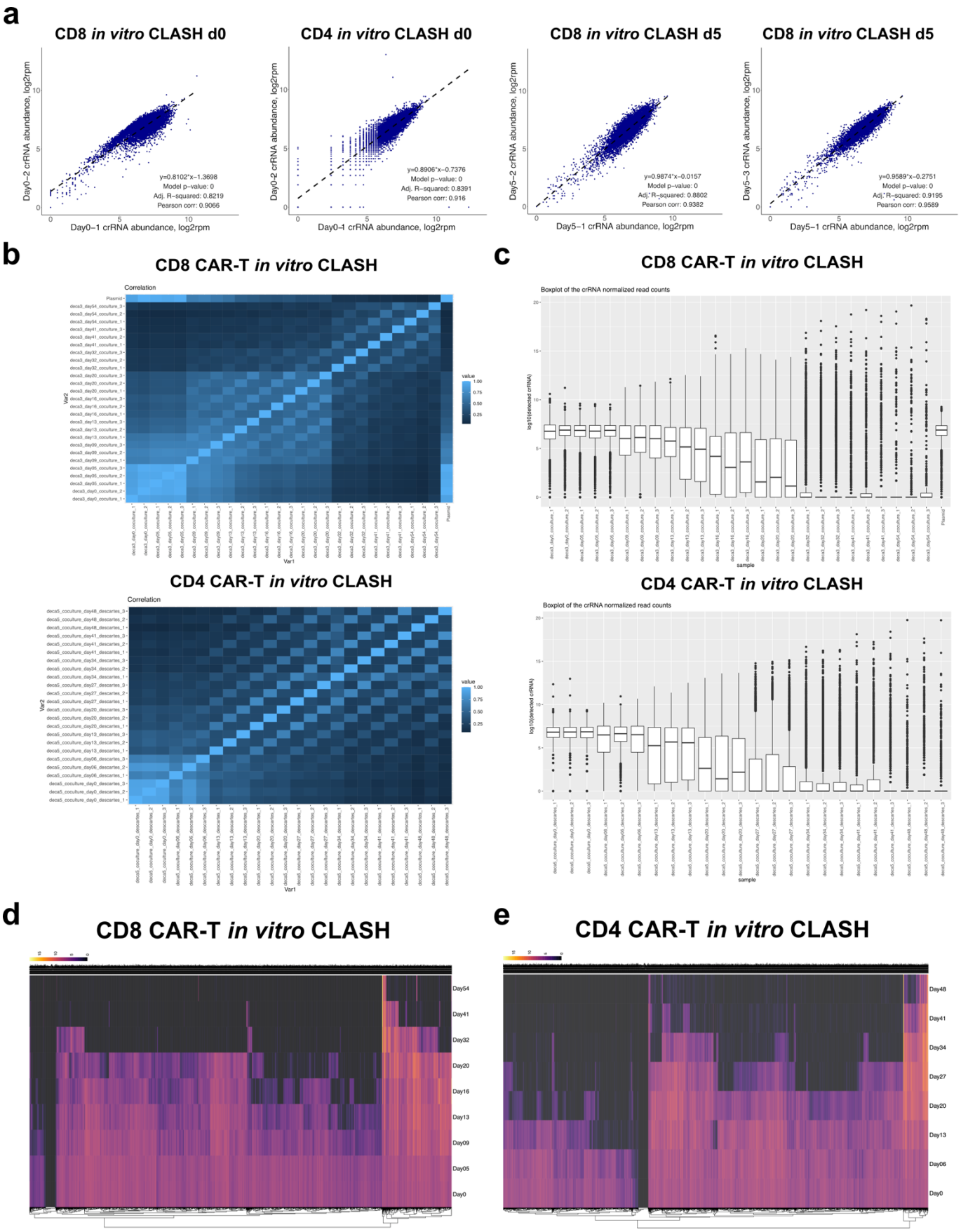
i, (Left) Representative flow cytometry analysis of memory and cellular cytotoxic marker expression on CAR-T cells after repeated co-culture with NALM6 for 3 weeks. Naive T cells (Tn) were defined as CD45RO-CD62L+, effector memory T cells (Tem) as CD45RO+CD62L- and central memory T cell (Tcm) as CD45RO+CD62L+. (Right) Quantification of Tcm, TNFα+ and IFNγ+ percentages in CAR-T cells (infection replicates, n = 3). Tukey's multiple comparison test was used to assess significance. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Data are shown as mean ± s.e.m.

j, (Left) Flow cytometry plots showing representative CAR22 knock-in into *TRAC* locus after CLASH Descartes library transduction for 5 days (AAV6 packaging, g-MOI = 5×10^3). (Right) Quantification of $CD3^-CAR22^+$ percentages in each group (infection replicates, n = 3). Unpaired two-sided t test was used to assess significance. ** $p < 0.01$. Data are shown as mean ± s.e.m.

k-l, Representative flow cytometry analysis and quantification of Tcm (**k**), exhaustion markers (**k**) and cytotoxic markers (**l**) in vector and Descartes-Lib CD22 CAR-Ts at day 0 (infection replicates, n = 3). Unpaired two-sided Mann Whitney test was used to assess significance. n.s, not significance. Data are shown as mean ± s.e.m.

m. Representative flow cytometry analysis and quantification of NAML6 cancer cell percentages in both groups at CD8 *in vitro* CLASH screen day 54 (infection replicates, n = 3). Unpaired two-sided t test was used to assess significance. *** $p < 0.001$. Data are shown as mean ± s.e.m.

n, Representative flow cytometry analysis and quantification of exhaustion markers and cytotoxic markers percentage in vector and Descartes-Lib CD22 CAR-Ts at day 54 (infection replicates, n = 3). One-way ANOVA with Tukey's multiple comparisons test was used to assess significance. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Data are shown as mean ± s.e.m.



Supplementary Figure 2: Library representation, correlation analysis and overall patterns of

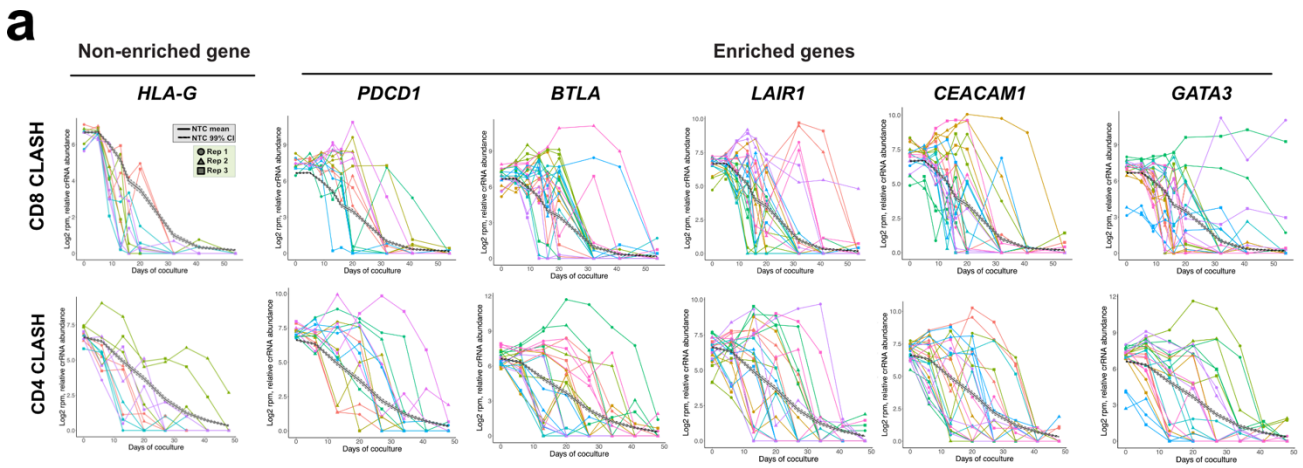
CLASH time course selection in CD4 and CD8 CAR-T cells

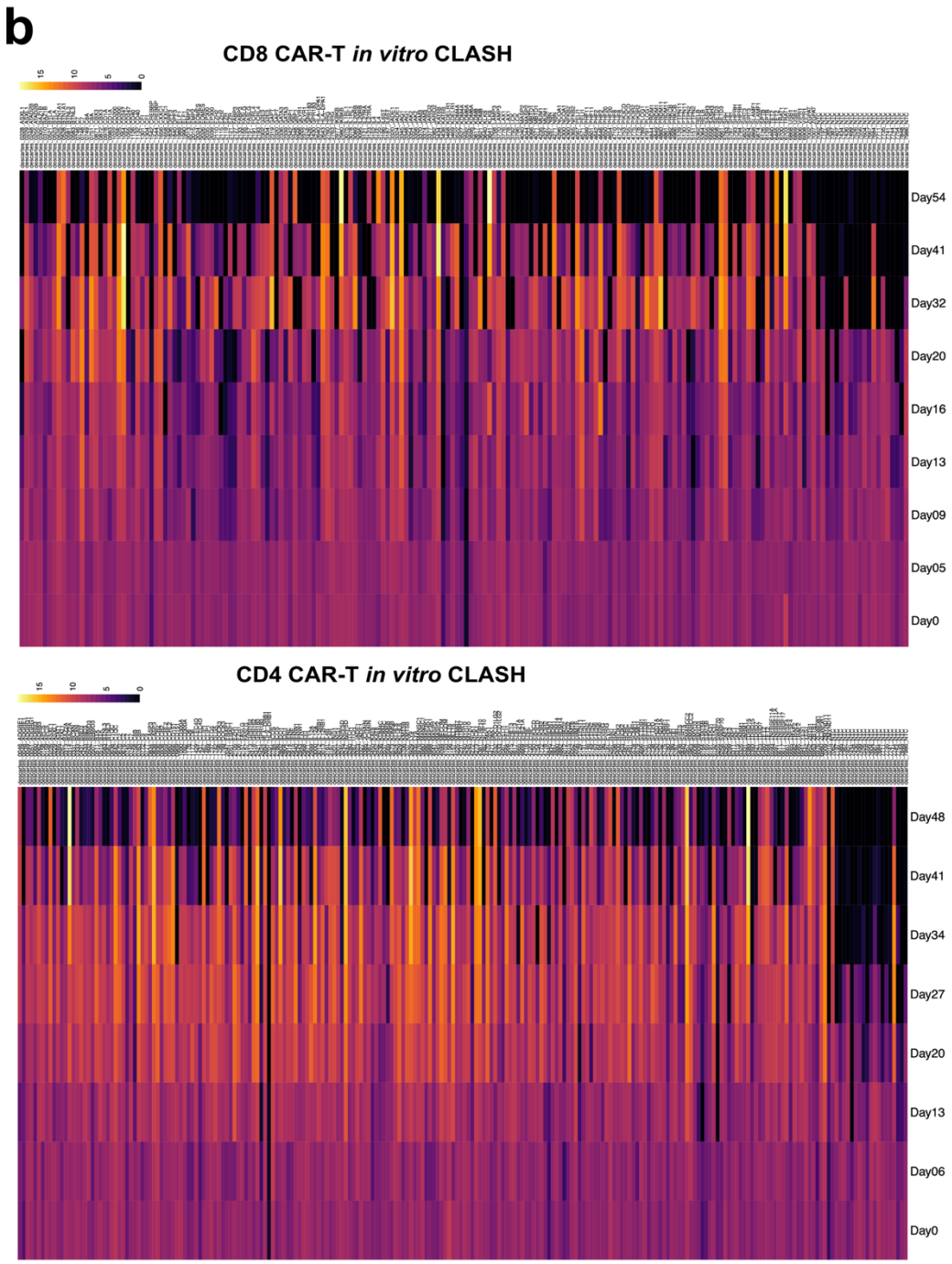
a, Correlation plots of representative transduction replicates (day 0) and early time point (day 5) replicates. After total read normalization (to reads per million reads per sample; or rpm) and log2 transformation, abundances for all crRNAs were compared between technical replicates 1 and 2 within each time point (in the following order from left to right: CD8 day0; CD4 day0; CD8 day5; CD4 day5). Using all points, a regression line was drawn, the p value of the model was calculated, and the Pearson correlation was determined.

b, Overall correlation analysis heatmaps of the crRNA representations in the genomic readouts of CD8 or CD4 *in vitro* CLASH knock-in CAR-T pools. Pearson correlations of crRNA library representation across time points were calculated based counts that were total read normalized (rpm) and log2 transformed.

c, Box plots of the normalized read counts for all crRNAs at each time point in CD8 and CD4 *in vitro* CLASH experiments. Horizontal lines indicate the first quartile, median, and third quartile, with outlier crRNAs shown as black dots.

d-e, Overall landscape of CLASH time course massively parallel CD8 (**d**) or CD4 (**e**) CAR-T mutant variant selection in the long-term co-culture. Each column shows the total read normalized, log transformed abundance of a crRNA over time (y-axis), with colors indicating degree of abundance.


a
Non-enriched gene
Enriched genes
HLA-G
PDCD1
BTLA
LAIR1
CEACAM1
GATA3
CD8 CLASH
CD4 CLASH
NTC mean
NTC 99% CI
Rep 1
Rep 2
Rep 3
Log2 rpm, relative crRNA abundance
Days of coculture

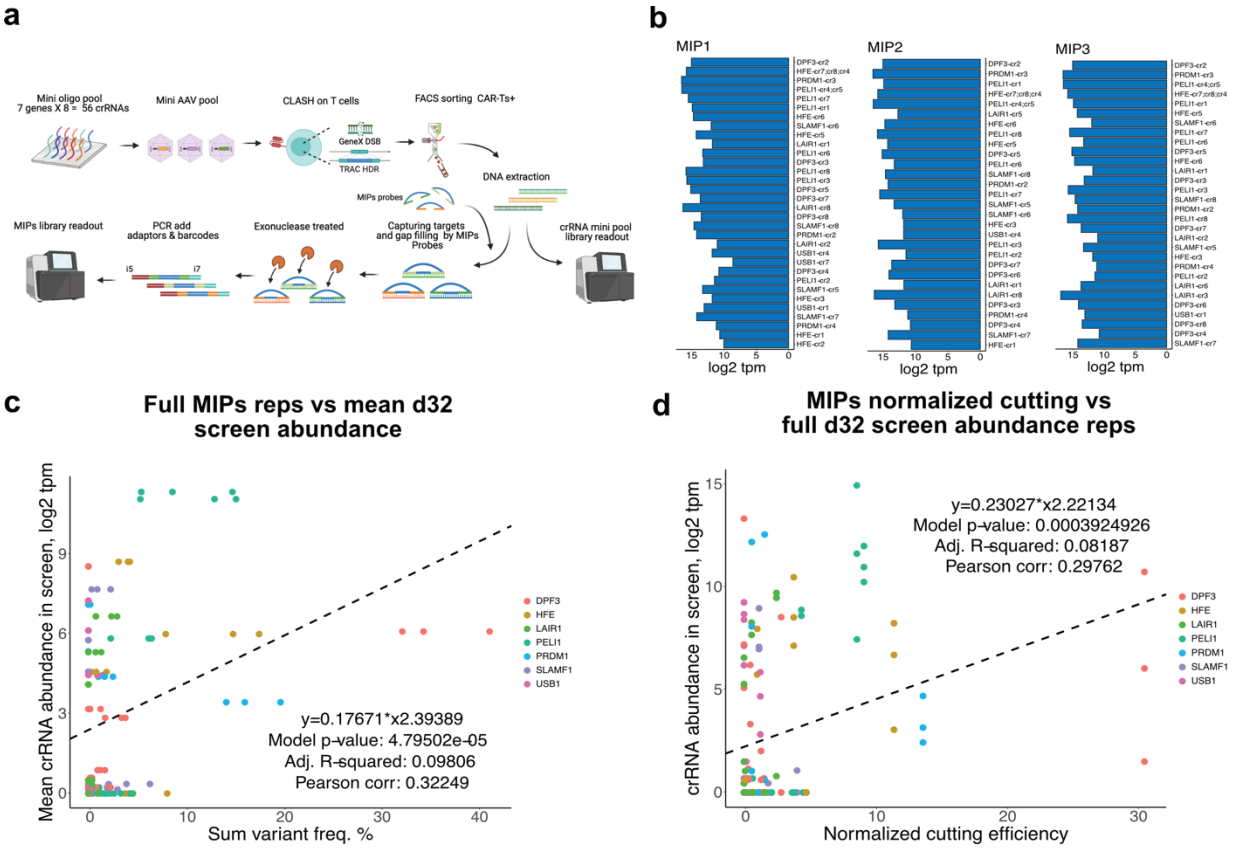

b
CD8 CAR-T in vitro CLASH
Day54
Day41
Day32
Day20
Day16
Day13
Day09
Day05
Day0
CD4 CAR-T in vitro CLASH
Day48
Day41
Day34
Day27
Day20
Day13
Day06
Day0

88

Supplementary Figure 3: Time course behaviors of representative enriched genes and involved biological pathways.

a, Time-course analysis of crRNA abundance for representative enriched genes. Different point shapes represent different replicates; different colors represent different crRNAs for the same gene. The dark solid line represents the mean trajectory for the 1,000 NTC crRNAs, and the dark dashed lines represent upper and lower 99% confidence intervals for NTCs.

b, Subset heatmap of log normalized crRNA representation over time. All the top 2 crRNAs of genes in SAMBA analysis were shown (z > 1.5 FDR= 0.01).



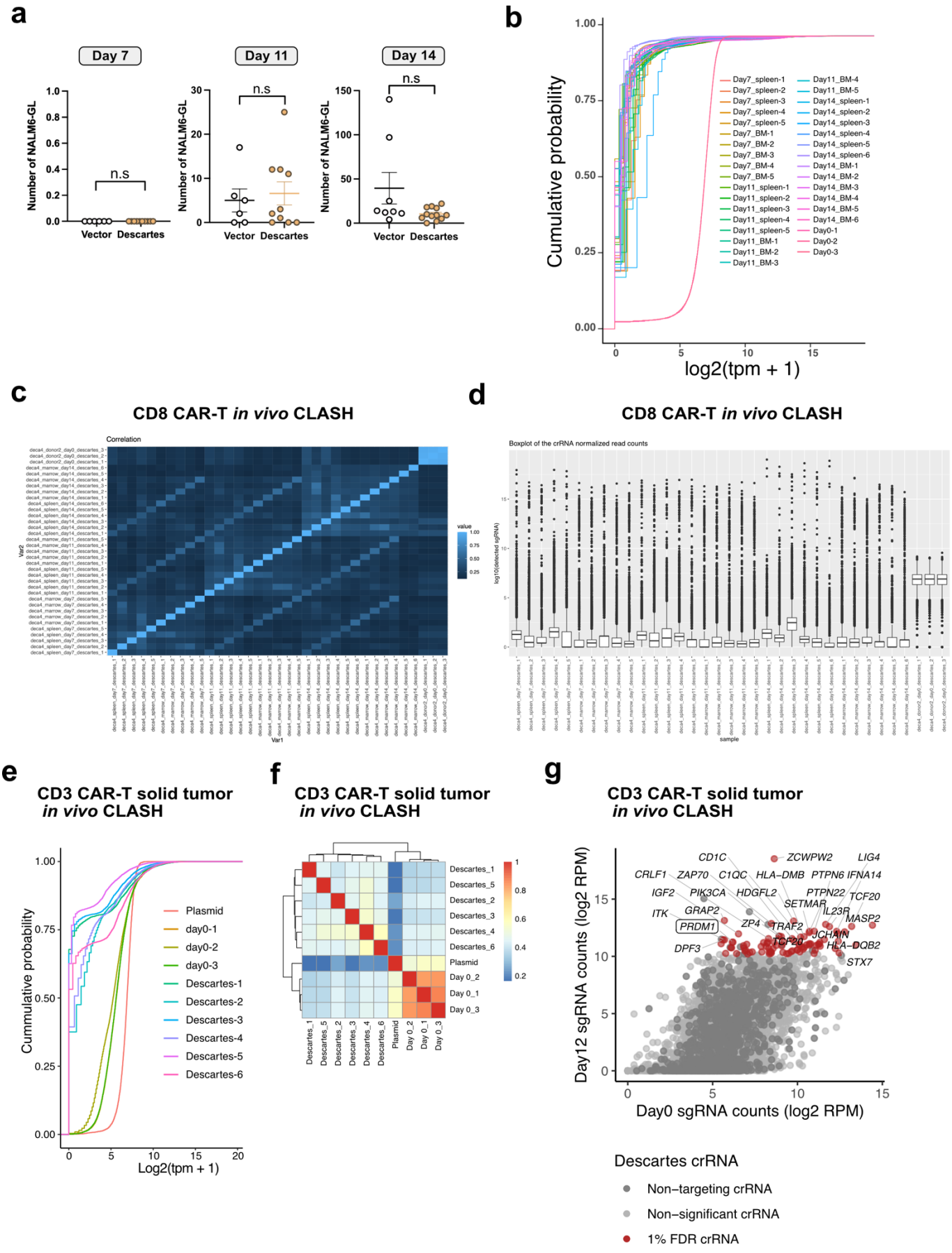
Supplementary Figure 4: MIPS-CLASH joint analysis of correlation between editing efficiency of CLASH performance in a pool setting

a, Schematic of CLASH-MIPS experiment. A 56 crRNA mini pool was used for CLASH-MIPS. A minipool of CAR-T cells was generated and subjected to FACS-based CAR-T enrichment 10 days after electroporation. Genomic DNA from pooled genome edited CAR-Ts was used for two readouts: (1) target capture / hybridization / MIPS, and (2) crRNA library readout.

b, Bar graph of crRNA library representation in MIPS samples by crRNA target windows.

c, Scatterplot of MIPS target-capture cutting efficiency for crRNA target windows by individual MIPS sample replicate compared to day 32 in vitro CLASH crRNA abundance averaged across replicates, for all genes. Linear regression model, statistics, and correlation analysis labelled.

d, Scatterplot of mean MIPS target-capture cutting efficiency for crRNA target windows normalized by mean MIPS crRNA abundance, compared to day 32 in vitro CLASH crRNA abundance by individual CLASH replicates, for all genes. Linear regression model, statistics, and correlation analysis are labelled.


a
Day 7
Day 11
Day 14
Number of NALM6-GL
n.s
Vector
Descartes
b
Cumulative probability
log2(tpm + 1)
Day7_spleen-1
Day7_spleen-2
Day7_spleen-3
Day7_spleen-4
Day7_spleen-5
Day7_BM-1
Day7_BM-2
Day7_BM-3
Day7_BM-4
Day7_BM-5
Day11_spleen-1
Day11_spleen-2
Day11_spleen-3
Day11_spleen-4
Day11_spleen-5
Day11_BM-1
Day11_BM-2
Day11_BM-3
Day11_BM-4
Day11_BM-5
Day14_spleen-1
Day14_spleen-2
Day14_spleen-3
Day14_spleen-4
Day14_spleen-5
Day14_spleen-6
Day14_BM-1
Day14_BM-2
Day14_BM-3
Day14_BM-4
Day14_BM-5
Day14_BM-6
Day0-1
Day0-2
Day0-3
c
CD8 CAR-T in vivo CLASH
Correlation
value
d
CD8 CAR-T in vivo CLASH
Boxplot of the crRNA normalized read counts
log10(detected sgRNA)
sample
e
CD3 CAR-T solid tumor in vivo CLASH
Cummulative probability
Log2(tpm + 1)
Plasmid
day0-1
day0-2
day0-3
Descartes-1
Descartes-2
Descartes-3
Descartes-4
Descartes-5
Descartes-6
f
CD3 CAR-T solid tumor in vivo CLASH
Descartes_1
Descartes_5
Descartes_2
Descartes_3
Descartes_4
Descartes_6
Plasmid
Day 0_2
Day 0_1
Day 0_3
g
CD3 CAR-T solid tumor in vivo CLASH
Day12 sgRNA counts (log2 RPM)
Day0 sgRNA counts (log2 RPM)
CD1C
ZCWPW2
LIG4
CRLF1
ZAP70
C1QC
HLA-DMB
PTPN6
IFNA14
IGF2
PIK3CA
HDGFL2
PTPN22
TCF20
SETMAR
ITK
GRAP2
IL23R
MASP2
PRDM1
ZP4
TRAF2
JCHAIN
TCF20
HLA-DQB2
DPF3
STX7
Descartes crRNA
Non−targeting crRNA
Non−significant crRNA
1% FDR crRNA

121

Supplementary Figure 5: QC analysis of CAR-Ts *in vivo* CLASH

a, Quantification of cancer cells at day 7, day 11 and day 14 (pooled spleen and bone marrow samples). Total tested live cells normalized to 1 million as inputs. Unpaired two-sided Mann Whitney test was used to assess significance. Day 7 (n = 5 mice; spleens and bone marrows, 10 samples total), day 11 (n = 5 mice, 10 samples) and day 14 (n = 6, 12 samples). n.s., not significant. Data are shown as mean ± s.e.m.

b, Empirical cumulative distribution function (CDF) of crRNA representations in the genomic readouts of CD8 *in vivo* CLASH CAR-T pool samples.

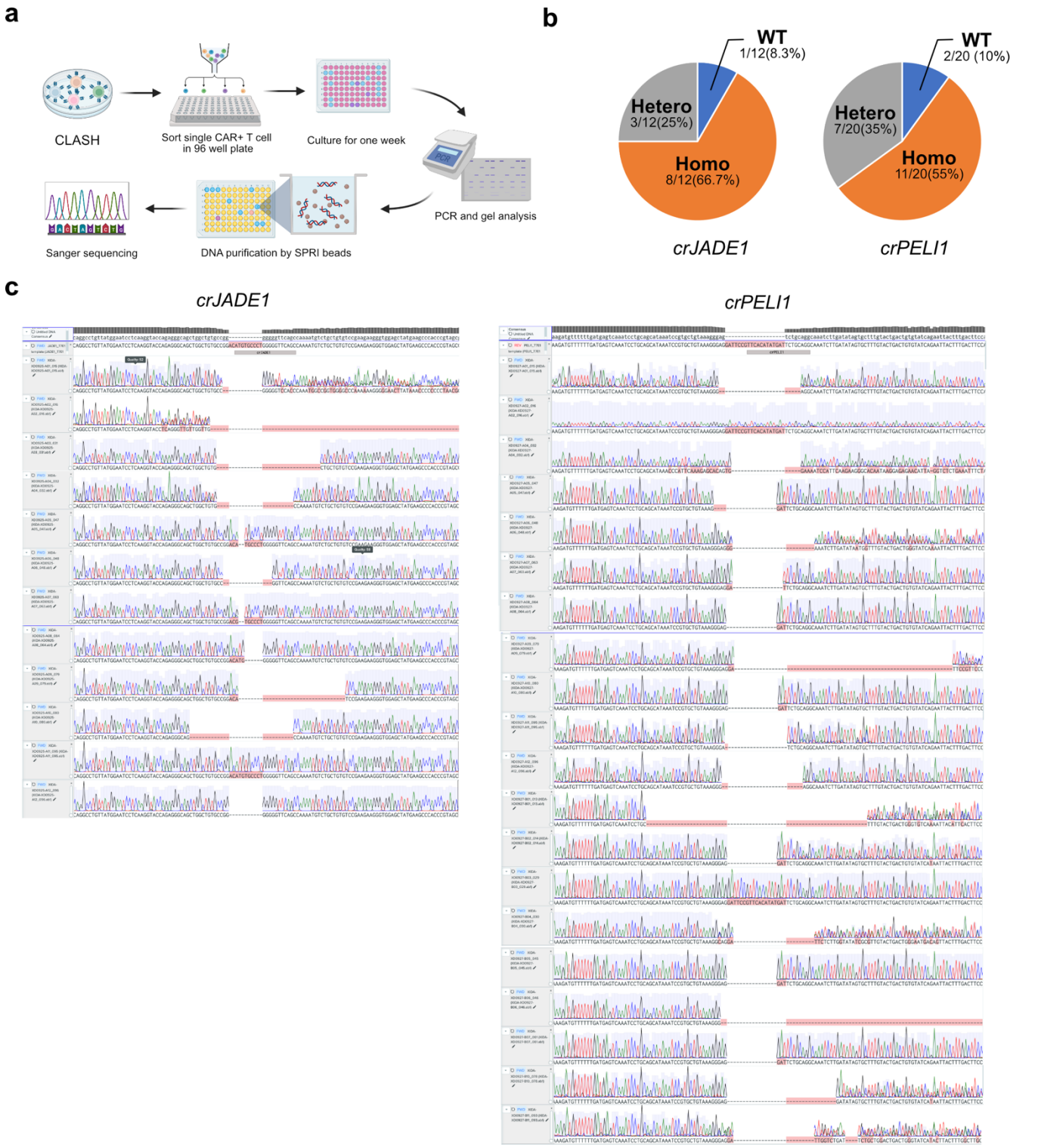
c, Correlation analysis of the crRNA representations in the genomic readouts of CD8 *in vivo* CLASH knock-in CAR-T pool. Pearson correlations of crRNA library representation across time points were calculated based counts that were total read normalized (rpm) and log2 transformed.

d, Box plots of the normalized read counts for all crRNAs at each time point in CD8 *in vivo* CLASH. Horizontal lines indicate the first quartile, median, and third quartile, with outlier crRNAs shown as black dots.

e, Empirical CDF plot of crRNA representations of Descartes library in the genomic readouts of CD3 T solid tumor CLASH CAR-T pool samples.

f, Correlation analysis of the crRNA representations in the genomic readouts of CD3 T *in vivo* solid tumor CLASH knock-in CAR-T pool. Pearson correlations of crRNA library representation across time points were calculated based counts that were total read normalized (rpm) and log2 transformed.

g, Bulk analysis of relative crRNA abundances in Descartes library in CD3 T solid tumor CLASH experiment at day 12 as compared with day 0 T cells. Gray dots are NTCs, red dots are scoring crRNAs that passed the FDR 1% cutoff, and light gray dots are remaining Descartes crRNA; top 25 crRNAs by log2 rpm of the experimental condition (y axis) are labelled with corresponding gene names.


a
CLASH
Sort single CAR+ T cell in 96 well plate
Culture for one week
PCR and gel analysis
DNA purification by SPRI beads
Sanger sequencing
b
WT
1/12(8.3%)
Hetero
3/12(25%)
Homo
8/12(66.7%)
crJADE1
WT
2/20 (10%)
Hetero
7/20(35%)
Homo
11/20(55%)
crPELI1
c
crJADE1
crPELI1
d
Before co-culture
After co-culture
Donor 5
Donor 6
TET2
NLRP10
KDM4E
EHMT1
PELI1
JADE1
PRDM1
Vector
CD69+
(Relative fold change)
e
Perforin+
Granzyme B+
TNFa+IFNg+
(Relative fold change)
Donor 5
Donor 6
Vector
PRDM1
JADE1
PELI1
EHMT1
KDM4E
NLRP10
TET2

Supplementary Figure 6: Single cell genotyping and additional individual functional analysis of immune genes scored in CLASH

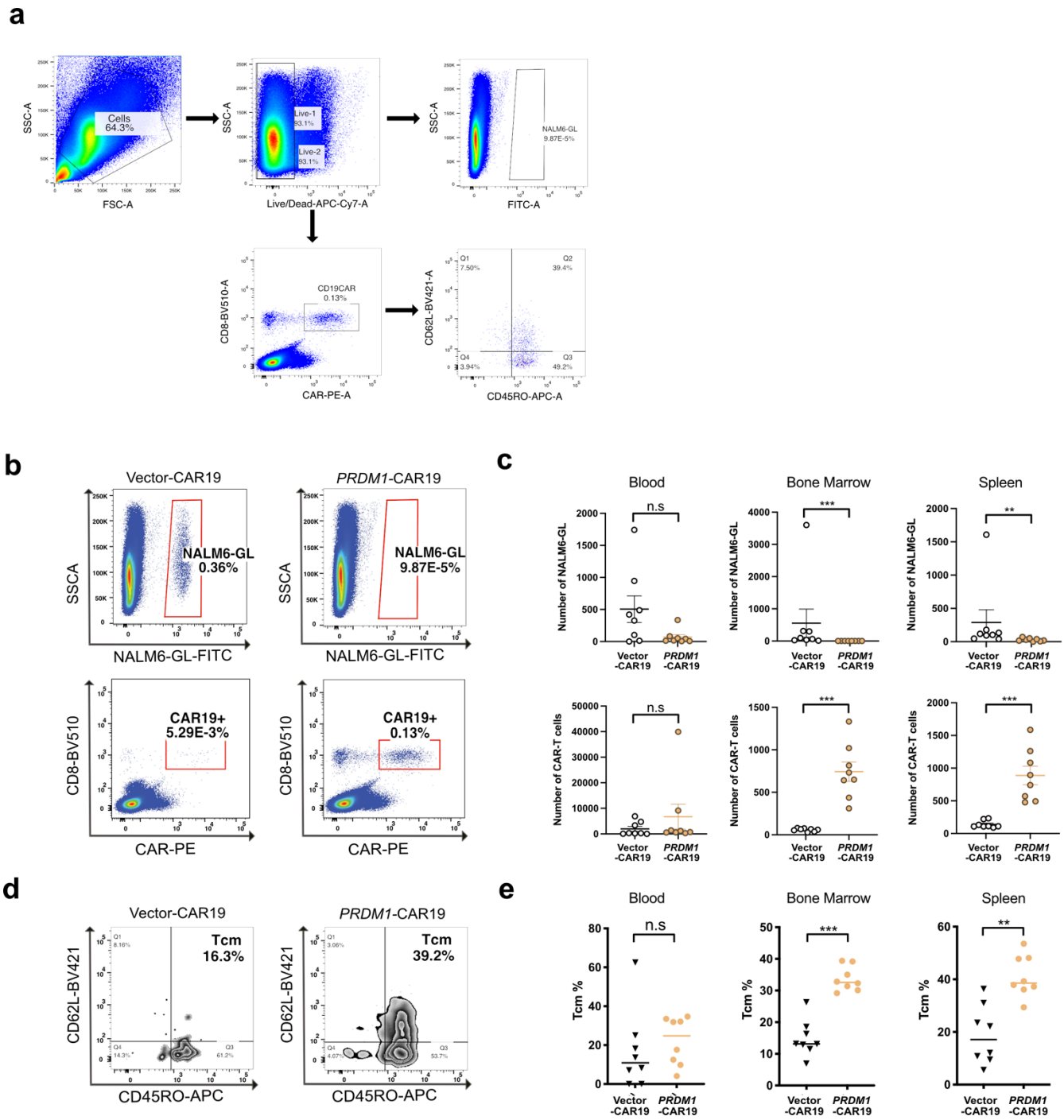
a, Schematic of CLASH single cell genotyping.

b, Pie chart showing percentage of wild type, homozygous and heterozygous/mixture single cell clone percentage (*crJADE1* n=12; *crPELI1* n=20).

c, Sanger sequencing results of *crJADE1* and *crPELI1* single clones.

d, Quantification of activation markers (CD69+) expressions in two different donors before co-culture or after co-culture with NALM6-GL for one day, normalized to vector control CAR-T cells. Unpaired two-sided t test was used to assess significance. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Data are shown as mean ± s.e.m.

e, Quantification of cytotoxicity related markers (TNFα+IFNγ+, perforin+ and Granzyme B+) expression in two different donors after co-culture with NALM6-GL for 5 hours, normalized to vector control CAR-T cells. Unpaired two-sided t test was used to assess significance. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Data are shown as mean ± s.e.m.



Supplementary Figure 7: **Flow cytometry analysis of *PRDM1* Δexon3 CAR-T cells *in vivo***

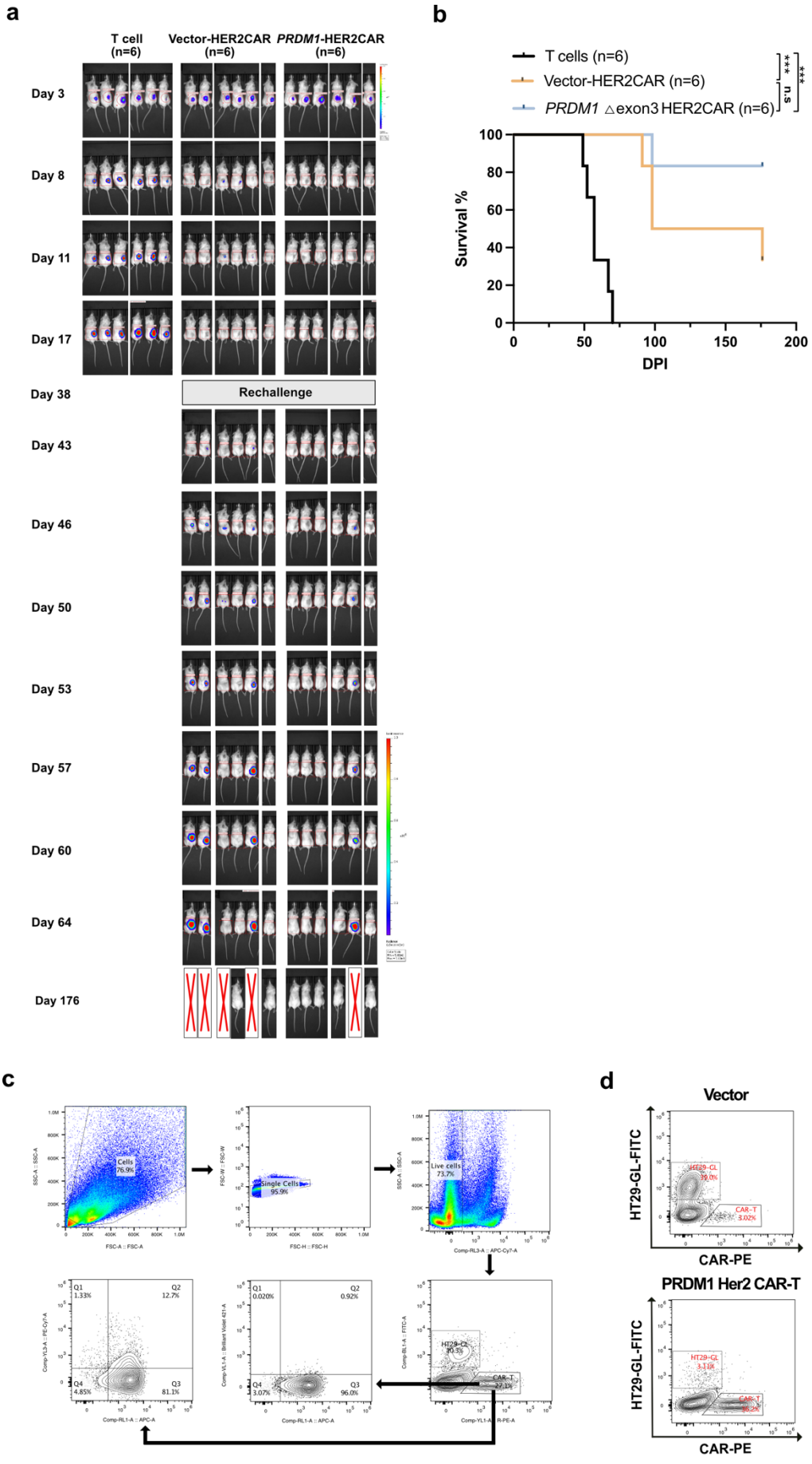
a. Gating strategy used to delineate NALM6-GL ($FITC^+$) and central memory CAR-T cells ($CD8^+CAR^+CD45RO^+CD62L^+$) in bone marrow.

b, Representative flow cytometry plots of NALM6-GL and CD19 ∆exon3 CAR-T cells in bone marrow. NSG mice were euthanized at day 18. Vector CAR19 (n = 8) and *PRDM1* ∆exon3 CAR19 (n = 8).

c, Quantification of cancer cells and CAR-T cells in blood, bone marrow and spleen. Unpaired two-sided Mann Whitney test was used to assess significance. Total tested live cells normalized to 1 million as inputs. Vector CAR19 (n = 8) and *PRDM1*∆exon3 CAR19 (n = 8). *** $p < 0.001$. Data are shown as mean ± s.e.m.

d, Representative flow cytometry plots of CAR-T memory phenotype in bone marrow. NSG mice were euthanized at day 18. Vector-CAR19 (n = 8) and *PRDM1*△exon3 CAR19 (n = 8).

e, Quantification of memory-like CAR-T ($CD45RO^+CD62L^+$) percentage in blood, bone marrow and spleen. Vector-CAR19 (n = 8) and *PRDM1*△exon3 CAR19 (n = 8). Unpaired two-sided Mann Whitney test was used to assess significance. ** $p < 0.01$, *** $p < 0.001$ and n.s., not significant. Data are shown as mean ± s.e.m.


a
T cell (n=6)
Vector-HER2CAR (n=6)
PRDM1-HER2CAR (n=6)
Day 3
Day 8
Day 11
Day 17
Day 38
Rechallenge
Day 43
Day 46
Day 50
Day 53
Day 57
Day 60
Day 64
Day 176
b
T cells (n=6)
Vector-HER2CAR (n=6)
PRDM1 △exon3 HER2CAR (n=6)
***
n.s
***
Survival %
DPI
c
Cells 76.9%
Single Cells 95.9%
Live cells 73.7%
Q1 1.33%
Q2 12.7%
Q4 4.85%
Q3 81.1%
Q1 0.020%
Q2 0.92%
Q4 3.07%
Q3 96.0%
HT29-GL 30.5%
CAR-T 97.1%
d
Vector
HT29-GL-FITC
CAR-PE
HT29-GL 32.0%
CAR-T 3.02%
PRDM1 Her2 CAR-T
HT29-GL-FITC
CAR-PE
HT29-GL 3.11%
CAR-T 96.2%

174

Supplementary Figure 8: **Bioluminescence imaging and survival analysis of HT29 rechallenge tumor model with *PRDM1* ∆exon3 CAR-T cell adoptive transfer**

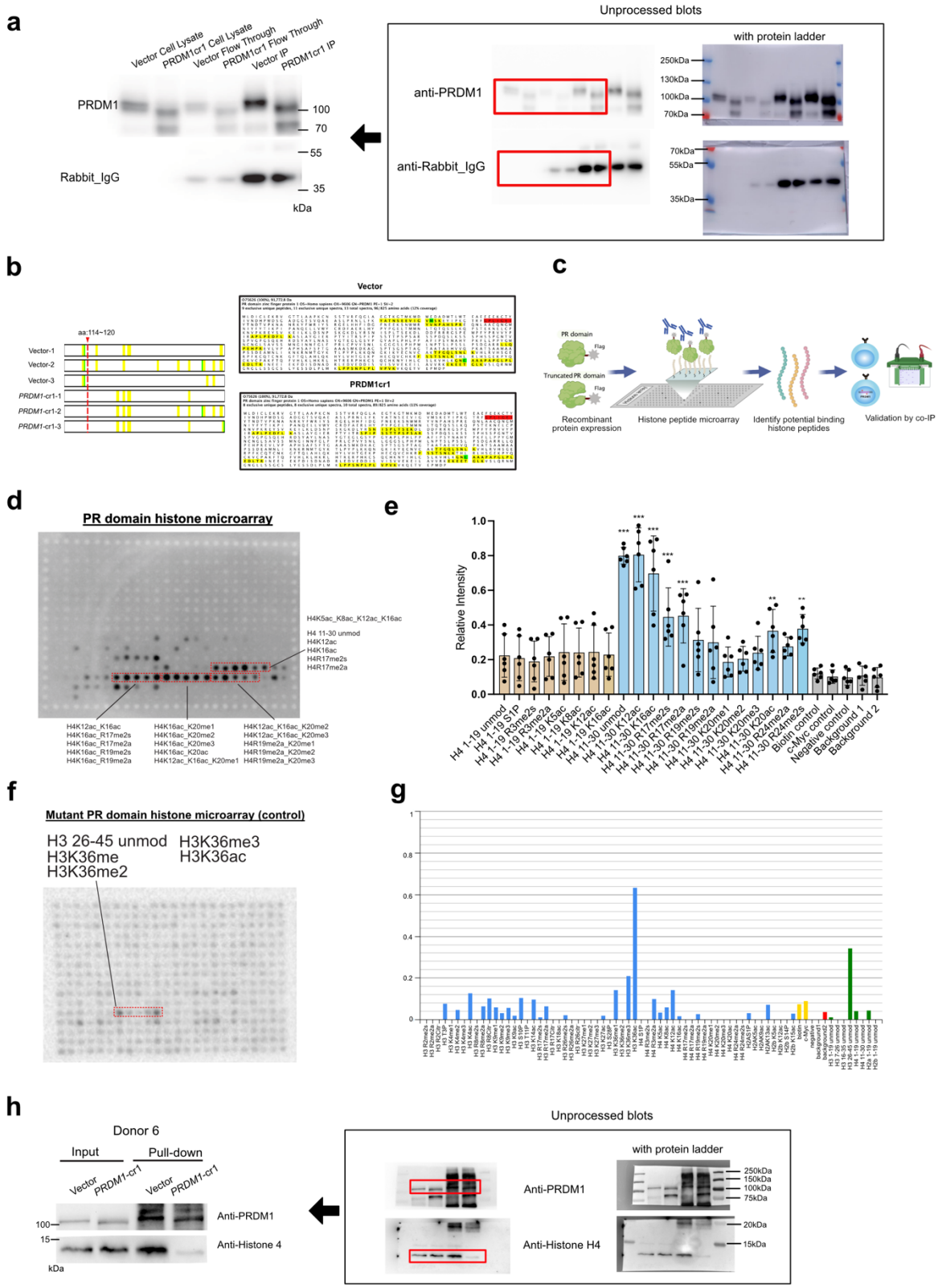
a, Bioluminescence imaging of HT29-GL-bearing NSG mice after treatment with vector or *PRDM1* ∆exon3 HER2 CAR-T cells. Cancer burden was measured as maximum photon per s per cm2 per steradian (p per s per cm2 per sr). Non-transduced $CD3^+$ T (n = 6), Vector HER2 CAR-T (n = 6) and ∆exon3 HER2 CAR-T (n = 6). Red crosses indicated absence of imaging of particular animals that passed survival endpoints.

b, Kaplan-Meier survival curves of mice in the rechallenge tumor model with different T cell treatments. Non-transduced $CD3^+$ T (n = 6), Vector HER2 CAR-T (n = 6) and ∆exon3 HER2 CAR-T (n = 6). Log-rank test was used to assess significance. *** $p < 0.001$ and n.s., not significant.

c, Gating strategy used to delineate HT29-GL (FITC+) and central memory CAR-T cells (CAR+CD45RO+CD62L+ or CAR+CD45RO+CCR7+) in solid tumor.

d, Representative flow cytometry plots of HT29-GL and ∆exon3 Her2 CAR-T cells in bone marrow. NSG mice were euthanized at day 14. Vector CAR22 (n = 4) and PRDM1 ∆exon3 CAR22 (n = 5).

Supplementary Figure 9



188

Supplementary Figure 9: Proteomic and epigenetic characterization of *PRDM1* ∆exon3 mutant CAR-T cells

a, Immunoprecipitation (IP) and western blot analysis of CAR-T cells lysate after transduction with vector or *PRDM1*-cr1 AAV6 for 5 days. The gel image is a representative of three repeated experiments.

b, (Left) The schematics of mass spectrometry (MS) protein sequence coverage of IP-enriched PRDM1 proteins in vector or *PRDM1* mutant CD22 CAR-T cells (detected peptides are highlighted in yellow, amino acids with post-translational modifications are shown in green and PRDM1 crRNA cutting sites are indicated by red dotted line). (Right) Representative MS protein sequence alignment of PRDM1 in vector or *PRDM1* mutant CD22 CAR-T cells (detected peptides are highlighted in yellow, amino acids with post-translational modifications are shown in green and *PRDM1* crRNA cutting sits are shown in red) (biological replicates, n = 3).

c, Schematic of histone peptide microarray and western blot validation by using purified PR domain and PR ∆exon3 protein.

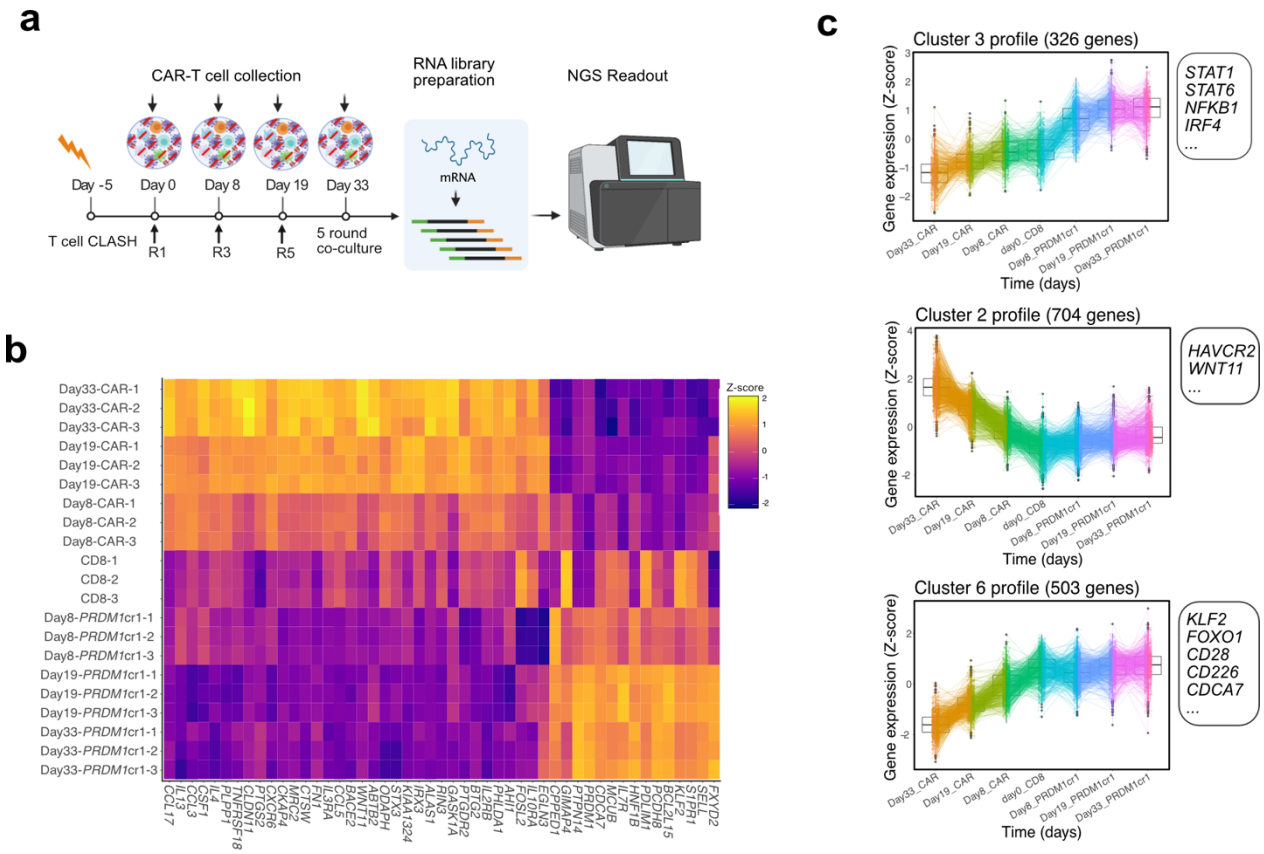
d, Representative image of histone peptide microarray. Histone peptide array probed with PR domain purified protein, followed by visualization with anti-FLAG antibody. Red boxed highlighted H4 11-30 modified and unmodified peptides, with the corresponding peptide identification annotated aside.

e. Quantification of the PR domain binding specificity by Array Analyze Software. One-way ANOVA with multiple comparisons test was used to assess significance and all the sample mean was compared with the mean of background1. ** $p < 0.01$ and *** $p < 0.001$, Data are shown as mean ± SD, n = 3 biological replicates, 6 samples.

f, Histone peptide array probed with PR ∆exon3 purified protein, followed by visualization with anti-Flag antibody (control). Red boxed highlighted H4 11-30 modified and unmodified peptides, with the corresponding peptide identification annotated aside.

g. Analysis of the PR ∆exon3 purified protein binding specificity by Array Analyze Software.

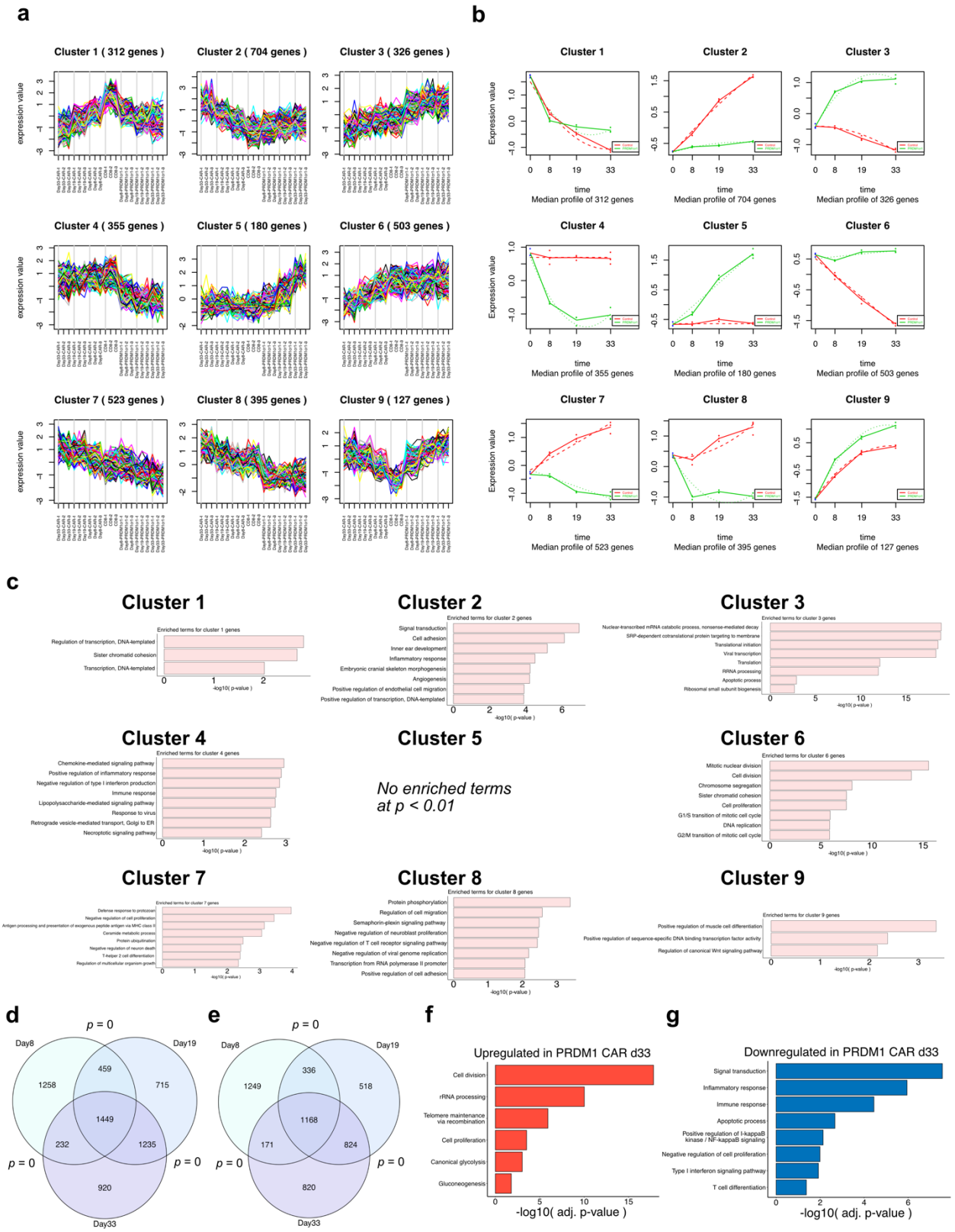
h, Co-IP Western using anti-PRDM1 and anti-histone H4 in vector control CAR-T and ∆exon3 CAR-T cells in an independent donor. Represented data from three experiments.



Supplementary Figure 10: Time-course mRNA-seq of *PRDM1* ∆exon3 mutant CAR-T cells

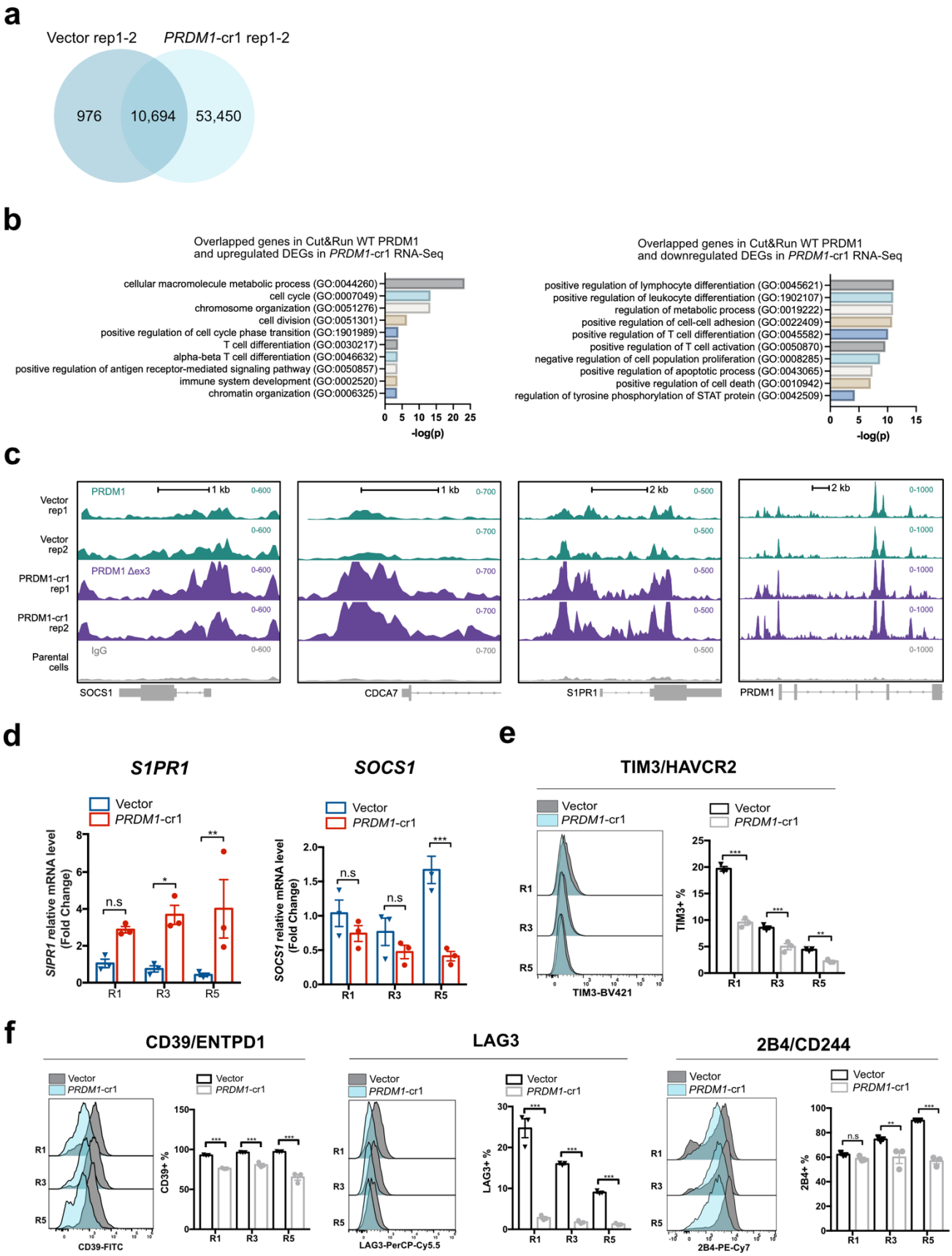
a, Schematic of time-course mRNA-seq of CLASH-generated vector and *PRDM1* ∆exon3 CD22 CAR-Ts. **b**, Time-course mRNA-seq for CLASH-generated vector and *PRDM1* ∆exon3 CD22 CAR-Ts. Heatmap of differentially expressed genes for *PRDM1* mutant vs control CAR-T cells across all 3 time points. **c**, Gene clustering analysis of *PRDM1* ∆exon3 CAR-T time-course mRNA-seq. Experiment-wide expression profiles of 3 representative clusters from maSigPro time course cluster analysis. Clusters comprise genes with similar expression patterns along time and similar behaviors between *PRDM1* ∆exon3 vs control CAR-T groups. Representative genes in each cluster were highlighted.

Supplementary Figure 11



Supplementary Figure 11: Analysis of gene expression clusters and their behaviors in *PRDM1* CAR-T time-course RNA-seq

a, Full set of experiment-wide expression profiles from maSigPro time course cluster analysis of *PRDM1* ∆exon3 CAR-T time-course RNA-seq. Each cluster comprises of genes with similar expression patterns over time and also in regard to the behaviors of vector and ∆exon3 CAR-T groups.
b, Experiment-wide gene expression profiles for each cluster represented with time-factor is in the x-axis, gene expression Z-score in the y-axis, and expression levels for a given experimental group represented by the same color. Solid lines join averages of each time group, dashed lines represent the regression curve.
c, Enriched gene ontology pathways found by DAVID analysis on gene sets of each of the 9 clusters.
d, Venn diagram of differentially upregulated genes found for ∆exon3 vs control CAR-Ts. Two-sided hypergeometric test was used to assess significance. p-value threshold < 1e-3 at three different time points.
e, Venn diagram of differentially downregulated genes found for ∆exon3 vs control CAR-Ts. Two-sided hypergeometric test was used to assess significance. p-value threshold < 1e-3 at three different time points.
f, Enriched gene ontology pathways found by DAVID analysis adjusted by Benjamini-Hochberg procedure on differentially upregulated genes for ∆exon3 vs control CAR-Ts at adjusted p-value threshold < 1e-3 on day 33.
g, Enriched gene ontology pathways found by DAVID analysis adjusted by Benjamini-Hochberg procedure on differentially downregulated genes for ∆exon3 vs control CAR-Ts at adjusted p-value threshold < 1e-3 on day 33.

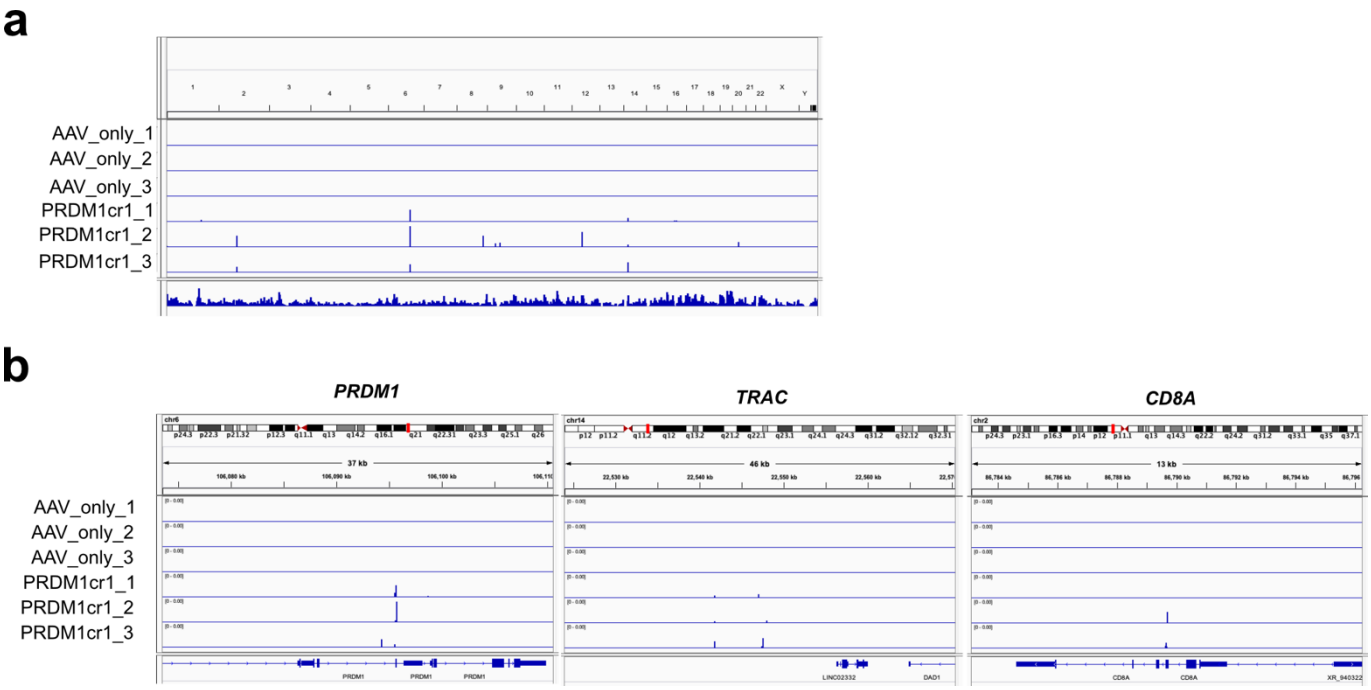


Supplementary Figure 12: Interrogation of direct targets and immunological mechanisms in CLASH generated *PRDM1* Δexon3 CAR-T cells

a, Venn diagram showing relationship between the WT PRDM1 and PRDM1 ∆exon3 Cut & Run peaks.
b, Enriched gene ontology pathways found by GO website on overlapped genes between PRDM1 WT / PRDM1 ∆exon3 mutant protein bound genes with DEGs from day 33 *PRDM1*cr1 ∆exon3 mutant vs control CAR-T with Adjusted p-value threshold < 1e-4. Mann-Whitney U test was used to assess significance.
c, Genome browser view of Cut & Run signal on a segment of *SOCS1*, *CDCA7, S1PR1* and *PRDM1*. Vector (WT PRDM1, two replicates), PRDM1 ∆exon3 (two replicates) and averaged IgG control.
d, RT-qPCR analysis of *SIPR1* and *SOCS1* mRNA expression levels on vector and *PRDM1* ∆exon3 CAR-T cells at 1st, 3rd and 5th round co-culture with NALM6 cells (infection replicates, n = 3). Two-way ANOVA with Sidak's multiple comparisons test was used to assess significance. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Data are shown as mean ± s.e.m.
e-f, Representative flow cytometry analysis of a panel of T cell exhaustion-related cell surface markers including TIM3/HAVCR2, CD39, LAG3 and 2B4/CD244 on vector and *PRDM1* ∆exon3 CAR-T cells at 1st, 3rd and 5th rounds of co-culture with NALM6 cells (infection replicates, n = 3). Two-way ANOVA with Sidak's multiple comparisons test was used to assess significance. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Data are shown as mean ± s.e.m.

Supplemental Figure Prep Note:

Supplemental Figure 4a, 6a, 9c, 10 were created with BioRender.com.



Supplementary Figure 13: Additional visualization of CLASH-*PRDM1*-cr1 CAR-T's genome-wide AAV-integration

a, IGV visualization of CLASH-*PRDM1* off-target integration events throughout the human genome. Peaks represent normalized off-target integration reads with y-axes standardized across samples and no windowing function.

b, IGV visualization of CLASH-*PRDM1* off-target integration events for specific genomic regions, including *PRDM1, TRAC* and *CD8A*. Peaks represent normalized off-target integration reads with y-axes standardized across samples and no windowing function.

Supplementary Datasets

Supplementary Dataset S1

CLASH library, *in vitro* and *in vivo* screening experiments and analyses

A zip file of folders and files, including individual txt files below:

S1.1_Descartes_Library_Lists

Descartes library gene list (initial, 954)

Descartes library gene list (final, 901)

Composition of the Descartes library (all crRNAs)

Details of Descarte library annotation (with genomic features and Deep scores)

S1.2_CD8_invitro_CLASH

Sample metadata for CD8 *in vitro* CLASH

Descartes library representation for CD8 *in vitro* CLASH (processed read counts)

Bulk analysis for later time points of CD8 *in vitro* CLASH

SAMBA full time-course analysis of CD8 *in vitro* CLASH

S1.3_CD4_invitro_CLASH

Sample metadata for CD4 *in vitro* CLASH

Descartes library representation for CD4 *in vitro* CLASH (processed read counts)

Bulk analysis for later time points of CD4 *in vitro* CLASH

SAMBA full time-course analysis of CD4 *in vitro* CLASH

S1.4_CD8_invivo_CLASH

Sample metadata for CD8 *in vivo* CLASH in a tumor model

Descartes library representation for CD8 *in vivo* CLASH (processed read counts)

Bulk analysis for later time points of CD8 *in vivo* CLASH

S1.5_CD3_invivo_CLASH

Sample metadata for CD3 *in vivo* CLASH in a tumor model

Descartes library representation for CD3 *in vivo* CLASH (processed read counts)

Bulk analysis for day 12 of CD3 *in vivo* CLASH

Supplementary Dataset S2

All CLASH - MIPS processed data and correlation analyses, with metadata.

An excel file, including individual tabs below:

Metadata of MIPS experiment

Coordinates for collapsed crRNA target region windows used for MIPS mapping.

MIPS library representation sample reads for crRNA target windows, log2 transformed.

Day 32 in vitro screen library representation for crRNA target windows, log2 transformed.

Day 54 in vitro screen library representation for crRNA target windows, log2 transformed.

Coarse filtering MIPs target capture variants for 56 crRNA pooled library sample MIPS-1.

Coarse filtering MIPs target capture variants for 56 crRNA pooled library sample MIPS-2.

Coarse filtering MIPs target capture variants for 56 crRNA pooled library sample MIPS-3.

Coarse filtering MIPs target capture variants for vector control pXD60-1.

Coarse filtering MIPs target capture variants for vector control pXD60-2.

Coarse filtering MIPs target capture variants for vector control pXD60-3.

Mean MIPs target capture variant frequencies and mean MIPs library abundance for crRNA target windows.

MIPs cutting efficiency normalized by library abundance for crRNA target windows compared to day 32 screen library abundance by each individual replicate.

MIPs cutting efficiency for crRNA target windows by each individual replicate compared to day 32 screen library abundance averaged across replicates.

MIPs cutting efficiency for crRNA target windows averaged across replicates compared to day 32 screen library abundance by each individual replicate.

Supplementary Dataset S3

All Nextera amplicon sequencing indel variant frequencies, with metadata.

An excel file, including individual tabs below:

Metadata of Nextera-NGS experiments

Vcf format indels of *PRDM1,* various crRNAs and donors

Vcf format indels of *PRDM1,* various crRNAs and donors*,* respective controls

Supplementary Dataset S4

***PRDM1* ∆exon3 CD22 CAR-T time-course mRNA-seq.**

An excel file, including individual tabs below:

Sample metadata for CD22 CAR-T time-course mRNA-seq.

Δexon3 vs. Vector gene-level differential expression analysis on day 8.

Δexon3 vs. Vector gene-level differential expression analysis on day 19.

Δexon3 vs. Vector gene-level differential expression analysis on day 33.

The maSigPro clustering analysis with Z-score normalized input.

Cluster membership gene lists from maSigPro mRNA-seq time-course analysis.

DAVID gene ontology analysis of day 33 Δexon3 upregulated genes (adj. $p < 0.001$)

DAVID gene ontology analysis of day 33 Δexon3 downregulated genes (adj. $p < 0.001$)

DAVID gene ontology analysis for cluster 1 genes from maSigPro

DAVID gene ontology analysis for cluster 2 genes from maSigPro

DAVID gene ontology analysis for cluster 3 genes from maSigPro

DAVID gene ontology analysis for cluster 4 genes from maSigPro

DAVID gene ontology analysis for cluster 5 genes from maSigPro

DAVID gene ontology analysis for cluster 6 genes from maSigPro

DAVID gene ontology analysis for cluster 7 genes from maSigPro

DAVID gene ontology analysis for cluster 8 genes from maSigPro

DAVID gene ontology analysis for cluster 9 genes from maSigPro

Supplementary Dataset S5

Genome-wide chromatin binding of PRDM1 WT and exon3-skip mutant via Cut-n-Run in human CD22 CAR-T cells.

An excel file, for PRDM1 WT and Δexon3 mutant CAR-T Cut-n-Run, including individual tabs below :

Sample metadata of Cut-n-Run

Processed read counts of Cut-n-Run

Peak identification of Cut-n-Run

List of PRDM1 WT bound genes

List of PRDM1 Δexon3 mutant bound genes

Overlap analysis of PRDM1 WT bound genes vs *PRDM1* Δexon3 differentially expressed genes

Overlap analysis of PRDM1 exon3-skip mutant bound genes vs *PRDM1* Δexon3 differentially expressed genes

Pathway analysis of PRDM1 WT bound genes and *PRDM1* Δexon3 upregulated genes

Pathway analysis of PRDM1 WT bound genes and *PRDM1* Δexon3 downregulated genes

Supplementary Dataset S6

All genome-wide AAV on-target and off-target integration processed data, with metadata.

An excel file, including individual tabs below:

Normalized reads and positions for off-target integration events for AAV only sample 1.

Normalized reads and positions for off-target integration events for AAV only sample 2.

Normalized reads and positions for off-target integration events for AAV only sample 3.

Normalized reads and positions for off-target integration events for PRDM1cr1 sample 1.

Normalized reads and positions for off-target integration events for PRDM1cr1 sample 2.

Normalized reads and positions for off-target integration events for PRDM1cr1 sample 3.

Total and gene-specific (*PRDM1, TRAC, CD8A*) off-target integration percentages.

Table S1. Supplementary DNA oligonucleotide information

Oligo sequences used in this study listed in an excel file.

Supplementary source data and statistics

Source data and statistics of non-NGS type data provided in an excel file.

Source data of uncropped blots images provided in a pdf file.