a) Step 1 - Amplification of PCR products from LentiGuide-Puro

b) Step 2 - Assembly of CROPseq-Guide-Puro from PCR products via the ligase cycling reaction (LCR)
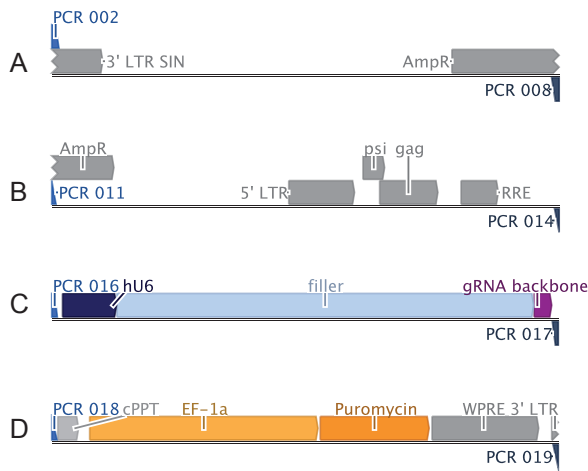
Product	PCR primers	
A	FWD 002	pCTGCTTTTTGCTTGTACTGGGTCTCTCTG
	REV 008	pCCGGGAAGCTAGAGTAAGTAGTTCGCCAG
B	FWD 011	pCAACAATTAATAGACTGGATGGAGGCGG
	REV 014	pTCTGAAACGATAATGGTGAATATCCCTGC
C	FWD 016	pTTCACCATTATCGTTTCAGACCCACC
	REV 017	pTGTCTCAAGATCTAGTTACGCCAAGC
D	FWD 018	pATCTTGAGACAAATGGCAGTATTCATCCAC
	REV 019	pATCTTGTCTTCGTTGGGAGTGAATTAGCC
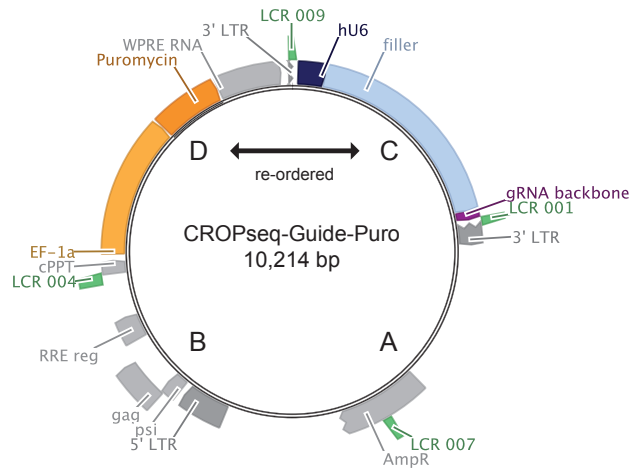
Bridge oligos directing the LCR assembly (5' > 3')

LCR 001
GGTGCTTTTTTAAGCTTGGCGTAACTAGATCTTGAGACACTGCTTTTTGCTTGTACTGGGTCTCTCTGGTTAGAC

LCR 004
CTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGAATCTTGAGACAAATGGCAGTATTC
ATCCACAATTTTAAAAGAAAAGGGG

LCR 007
AAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAG

LCR 009
GGAAGGGCTAATTCACTCCCAACGAAGACAAGATTTCACCATTATCGTTTCAGACCCACCTCCCAA

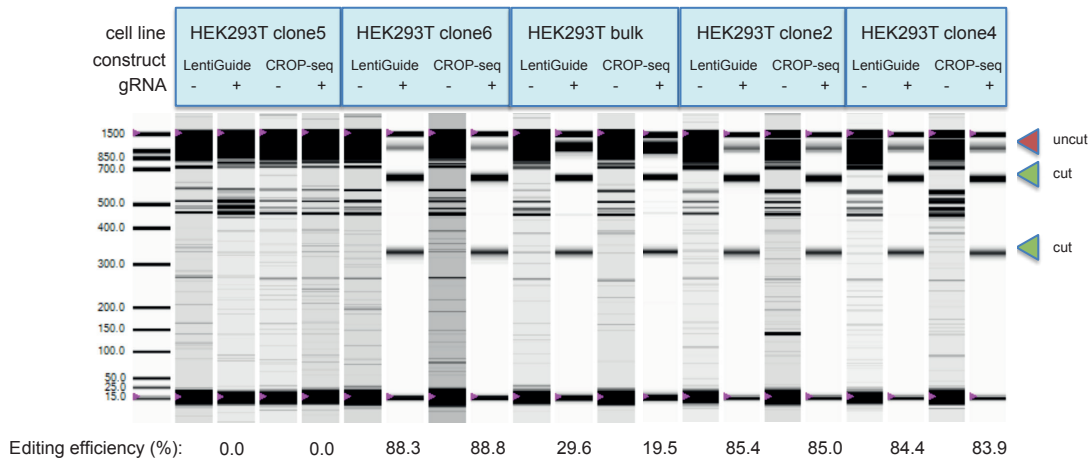**c)** Validation of gRNA cassette duplication after viral integration
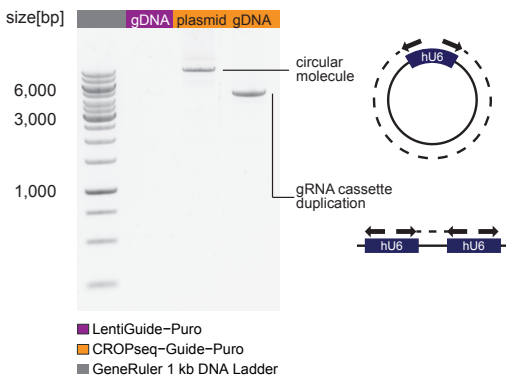
Supplementary Figure 1

Cloning and validation of the CROPseq-Guide-Puro plasmid

a) As starting point for preparing the CROPseq-Guide-Puro plasmid, we amplified four PCR products (A, B, C, and D) from the LentiGuide-Puro plasmid with the indicated primer pairs. **b)** CROPseq-Guide-Puro was constructed from these four amplicons using the ligase cycling reaction (LCR). The assembly was directed by four overlapping bridge oligonucleotides to flip the order of parts C and D. This rearrangement places the hU6-gRNA cassette into the 3' LTR, downstream of the EF-1a puromycin marker. **c)** To validate the duplication of the hU6-gRNA cassette during lentiviral integration, we performed PCRs with primers that bind to the hU6 promoter but face in opposite directions. Productive amplification can occur only when amplifying from a circular plasmid or following duplication of the cassette during viral integration. As templates, we used gDNA from LentiGuide-Puro transduced cells (lane 1, resulting in no amplification), a plasmid preparation of CROPseq-Guide-Puro (lane 2), or gDNA from CROPseq-Guide-Puro transduced cells (lane 3).

a T7 endonuclease assays for bulk as well as clonally expanded HEK293T cells

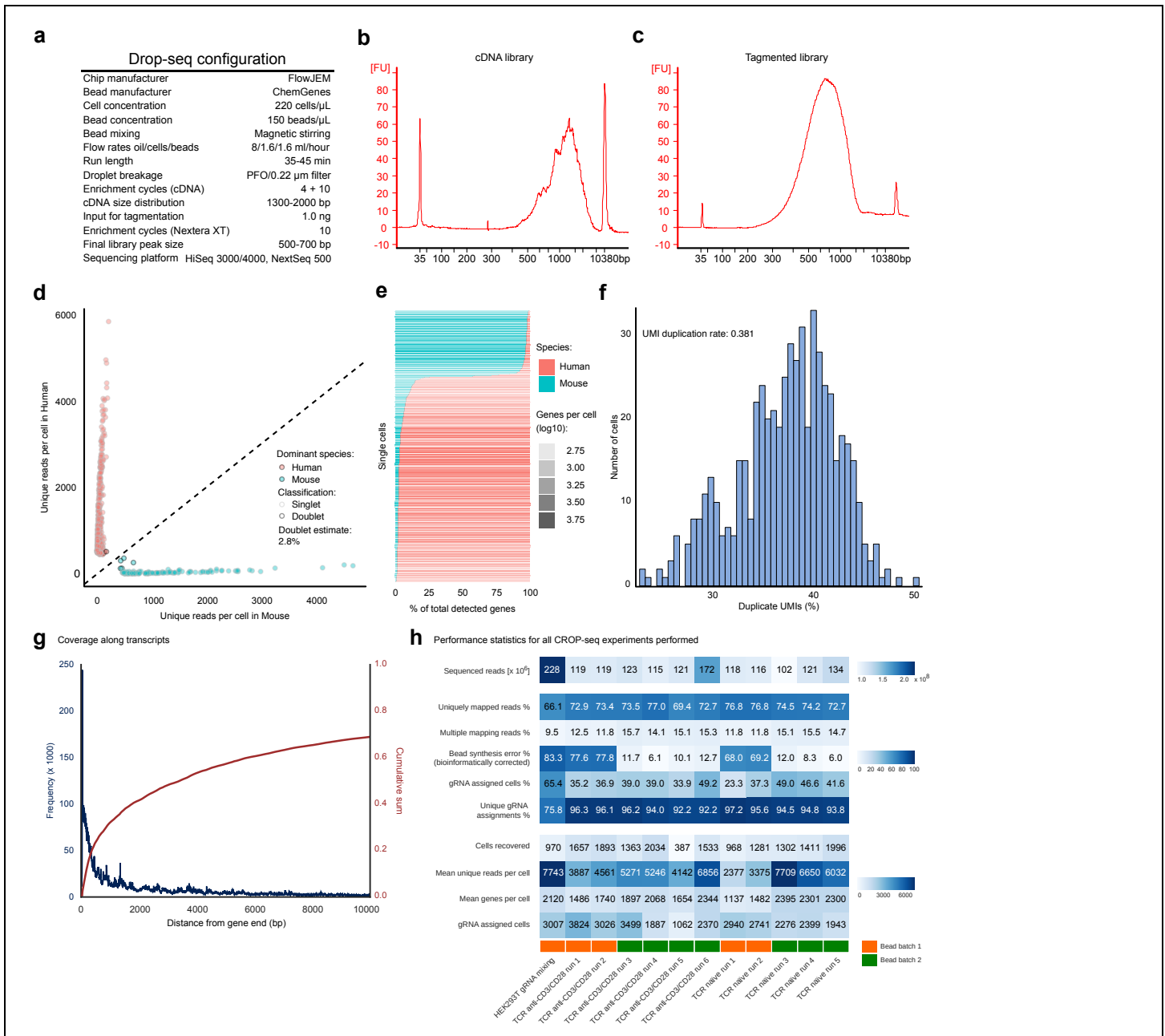| cell line	HEK293T clone5		HEK293T clone6		HEK293T bulk		HEK293T clone2		HEK293T clone4	
construct	LentiGuide	CROP-seq	LentiGuide	CROP-seq	LentiGuide	CROP-seq	LentiGuide	CROP-seq	LentiGuide	CROP-seq
gRNA	-	+	-	+	-	+	-	+	-	+

| Editing efficiency (%): | 0.0 | 0.0 | 88.3 | 88.8 | 29.6 | 19.5 | 85.4 | 85.0 | 84.4 | 83.9 |

b Summary table of all T7 endonuclease assays performed

sample_ID	cell_line	viral_construct	gene	undigested [nM]	digested 1 [nM]	digested 2 [nM]	editing_efficiency[%]
1	HEK clone5	LentiGuide-Puro	MBD1	9.26	0	0	0.0
2	HEK clone5	CROPseq-Guide-Puro	MBD1	9.52	0	0	0.0
3	HEK clone6	LentiGuide-Puro	MBD1	1.01	7.26	8.02	88.3
4	HEK clone6	CROPseq-Guide-Puro	MBD1	1.07	8.05	8.98	88.8
5	HEK bulk	LentiGuide-Puro	MBD1	7.25	2.85	3.24	29.6
6	HEK bulk	CROPseq-Guide-Puro	MBD1	7.79	1.85	1.93	19.5
7	HEK clone2	LentiGuide-Puro	MBD1	1.44	8.22	8.68	85.4
8	HEK clone2	CROPseq-Guide-Puro	MBD1	1.54	8.36	9.16	85.0
9	HEK clone4	LentiGuide-Puro	MBD1	1.58	8.3	8.76	84.4
10	HEK clone4	CROPseq-Guide-Puro	MBD1	1.69	8.55	9.09	83.9
11	HEK clone5	LentiGuide-Puro	DNMT3B	10.54	0	0	0.0
12	HEK clone5	CROPseq-Guide-Puro	DNMT3B	8.81	0	0	0.0
13	HEK clone6	LentiGuide-Puro	DNMT3B	0.92	12.37	9.71	92.3
14	HEK clone6	CROPseq-Guide-Puro	DNMT3B	0.76	11.06	8.2	92.7
15	HEK bulk	LentiGuide-Puro	DNMT3B	6.37	2.61	2.13	27.1
16	HEK bulk	CROPseq-Guide-Puro	DNMT3B	6.48	2.64	2.25	27.4
17	HEK clone2	LentiGuide-Puro	DNMT3B	1.24	14.26	11.64	91.3
18	HEK clone2	CROPseq-Guide-Puro	DNMT3B	1.17	14.25	11.2	91.6
19	HEK clone4	LentiGuide-Puro	DNMT3B	0.97	12.97	9.97	92.2
20	HEK clone4	CROPseq-Guide-Puro	DNMT3B	0.91	12.37	9.45	92.3
21	K562 bulk	LentiGuide-Puro	MBD1	0.7	6.61	6.88	90.6
22	K562 bulk	CROPseq-Guide-Puro	MBD1	0.63	6.31	6.12	90.8
23	Jurkat bulk	LentiGuide-Puro	MBD1	1.75	7.31	8.33	81.7
24	Jurkat bulk	CROPseq-Guide-Puro	MBD1	1.8	7.57	8.87	82.0
25	KBM7 bulk	LentiGuide-Puro	MBD1	0.98	6.89	7.32	87.9
26	KBM7 bulk	CROPseq-Guide-Puro	MBD1	2.55	11.7	10.72	81.5
27	K562 bulk	LentiGuide-Puro	DNMT3B	0.89	10.38	8.65	91.4
28	K562 bulk	CROPseq-Guide-Puro	DNMT3B	0.83	10.59	8.54	92.0
29	Jurkat bulk	LentiGuide-Puro	DNMT3B	1.24	11.66	9.75	89.6
30	Jurkat bulk	CROPseq-Guide-Puro	DNMT3B	1.86	10.63	9.53	84.4
31	KBM7 bulk	LentiGuide-Puro	DNMT3B	1.65	6.84	6.4	80.0
32	KBM7 bulk	CROPseq-Guide-Puro	DNMT3B	2.32	8.34	7.4	77.2

Supplementary Figure 2

Genome editing efficiencies of LentiGuide-Puro and CROPseq-Guide-Puro based on the T7 endonuclease assay

a) Clonally expanded HEK293T cell lines as well as a HEK293T bulk population were transduced with LentiGuide-Puro (LentiGuide) or CROPseq-Guide-Puro (CROP-seq) vectors containing a gRNA targeting the MBD1 locus (+) or targeting a different locus (-). Genome editing efficiencies for MBD1 were measured using the T7 endonuclease assay, which indicated highly similar performance between the two vectors. HEK293T clone 5 did not show any genome editing and was not used for further experiments. **b)** Table summarizing genome editing efficiencies for four cell lines (HEK293T, K562, Jurkat, KBM7) and two gRNAs (MBD1, DNMT3B).

a Drop-seq configuration

Chip manufacturer	FlowJEM
Bead manufacturer	ChemGenes
Cell concentration	220 cells/µL
Bead concentration	150 beads/µL
Bead mixing	Magnetic stirring
Flow rates oil/cells/beads	8/1.6/1.6 ml/hour
Run length	35-45 min
Droplet breakage	PFO/0.22 µm filter
Enrichment cycles (cDNA)	4 + 10
cDNA size distribution	1300-2000 bp
Input for tagmentation	1.0 ng
Enrichment cycles (Nextera XT)	10
Final library peak size	500-700 bp
Sequencing platform	HiSeq 3000/4000, NextSeq 500

b cDNA library

c Tagmented library

d Dominant species: Human, Mouse. Classification: Singlet, Doublet. Doublet estimate: 2.8%

e Species: Human, Mouse. Genes per cell (log10): 2.75, 3.00, 3.25, 3.50, 3.75

f UMI duplication rate: 0.381

g Coverage along transcripts

h Performance statistics for all CROP-seq experiments performed

	HEK293T gRNA mixing	TCR anti-CD3/CD28 run 1	TCR anti-CD3/CD28 run 2	TCR anti-CD3/CD28 run 3	TCR anti-CD3/CD28 run 4	TCR anti-CD3/CD28 run 5	TCR anti-CD3/CD28 run 6	TCR naïve run 1	TCR naïve run 2	TCR naïve run 3	TCR naïve run 4	TCR naïve run 5
Sequenced reads [x 10^6]	228	119	119	123	115	121	172	118	116	102	121	134
Uniquely mapped reads %	66.1	72.9	73.4	73.5	77.0	69.4	72.7	76.8	76.8	74.5	74.2	72.7
Multiple mapping reads %	9.5	12.5	11.8	15.7	14.1	15.1	15.3	11.8	11.8	15.1	15.5	14.7
Bead synthesis error % (bioinformatically corrected)	83.3	77.6	77.8	11.7	6.1	10.1	12.7	68.0	69.2	12.0	8.3	6.0
gRNA assigned cells %	65.4	35.2	36.9	39.0	39.0	33.9	49.2	23.3	37.3	49.0	46.6	41.6
Unique gRNA assignments %	75.8	96.3	96.1	96.2	94.0	92.2	92.2	97.2	95.6	94.5	94.8	93.8
Cells recovered	970	1657	1893	1363	2034	387	1533	968	1281	1302	1411	1996
Mean unique reads per cell	7743	3887	4561	5271	5246	4142	6856	2377	3375	7709	6650	6032
Mean genes per cell	2120	1486	1740	1897	2068	1654	2344	1137	1482	2395	2301	2300
gRNA assigned cells	3007	3824	3026	3499	1887	1062	2370	2940	2741	2276	2399	1943

Bead batch 1, Bead batch 2

Supplementary Figure 3

Configuration and validation of the droplet-based assay for single-cell transcriptome profiling

a) Setup of the Drop-seq workflow used as part of CROP-seq. **b)** Bioanalyzer trace of a typical cDNA library prepared with CROP-seq. **c)** Electropherogram of a sequencing-ready CROP-seq library after tagmentation. **d)** Doublet estimates based on a HEK293T (human) / 3T3 (mouse) mixing experiment across all detected cells and transcripts (without filtering). **e)** Percent of detected genes aligning to the human and mouse transcriptomes (filtered for cells with more than 500 detected genes). **f)** PCR duplication rates based on unique molecular identifiers (UMIs) in the HEK293T (human) / 3T3 (mouse) mixing experiment. **g)** Distribution of the distance of read mapping positions to the 3'end of gene models (blue line) and their cumulative sum (red line). **h)** Detailed performance statistics for twelve CROP-seq experiments. Green and orange labels indicate different batches of Drop-seq beads, where batch 1 suffered from production problems affecting the cell barcodes, which have been bioinformatically corrected to improve the data quality of the affected samples.
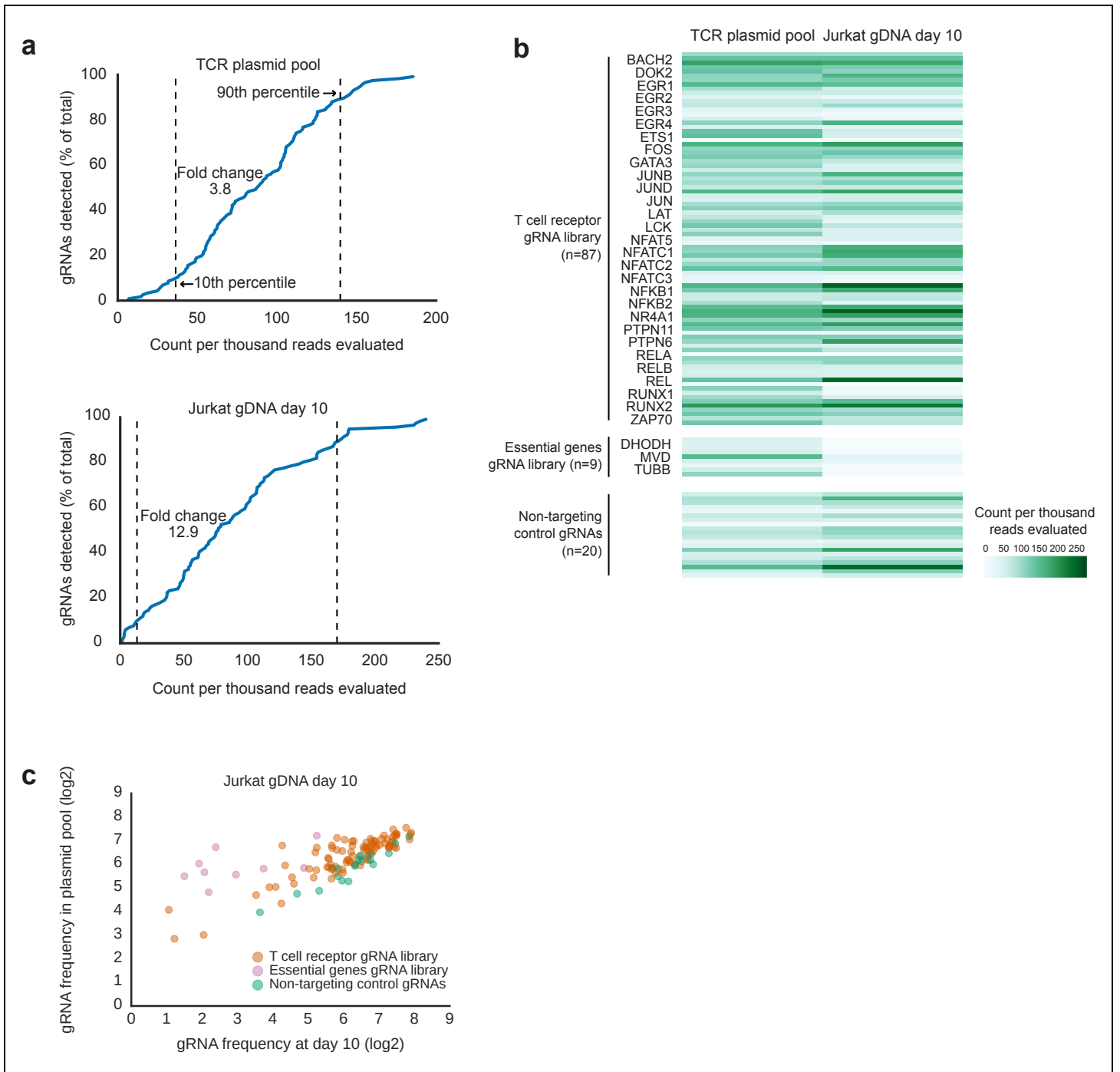
Supplementary Figure 4

Validation of the T cell receptor gRNA library and gRNA dynamics

a) gRNA representation in the T cell receptor (TCR) gRNA library, assessed by amplicon sequencing of the plasmid pool (top) and the gDNA of Jurkat cells at day 10 post transduction with CROPseq-Guide-Puro (bottom), both displayed as cumulative distribution plots. The fold change between the 10th and 90th percentile is highlighted as a measure of library imbalance, which expectedly increases upon transduction. **b**) Abundance of each gRNA shown as a heatmap. **c**) Scatterplots of gRNA abundance from amplicon libraries at day 10 versus the original plasmid library. Frequencies of detected gRNAs have been normalized to the evaluated reads in each experiment.
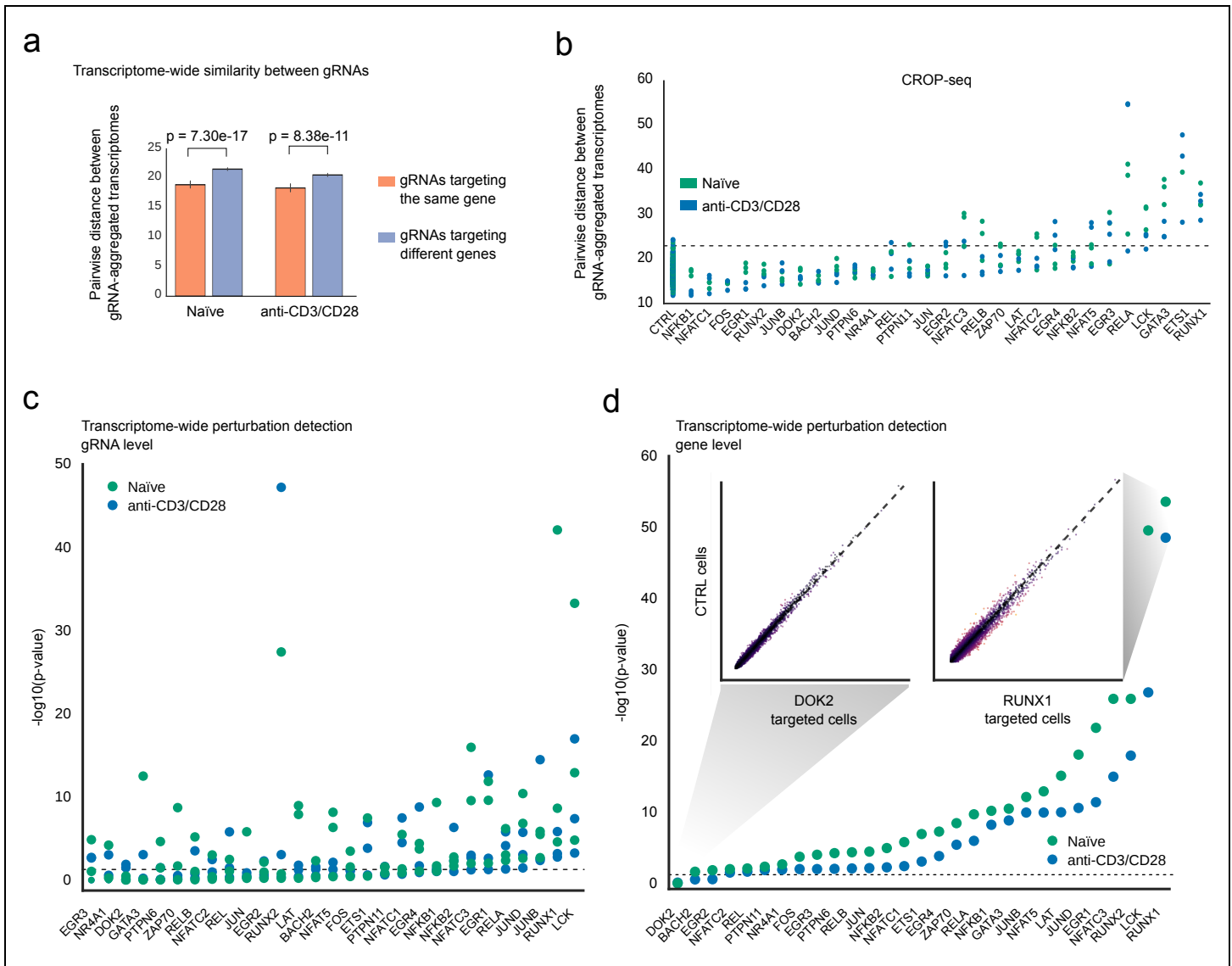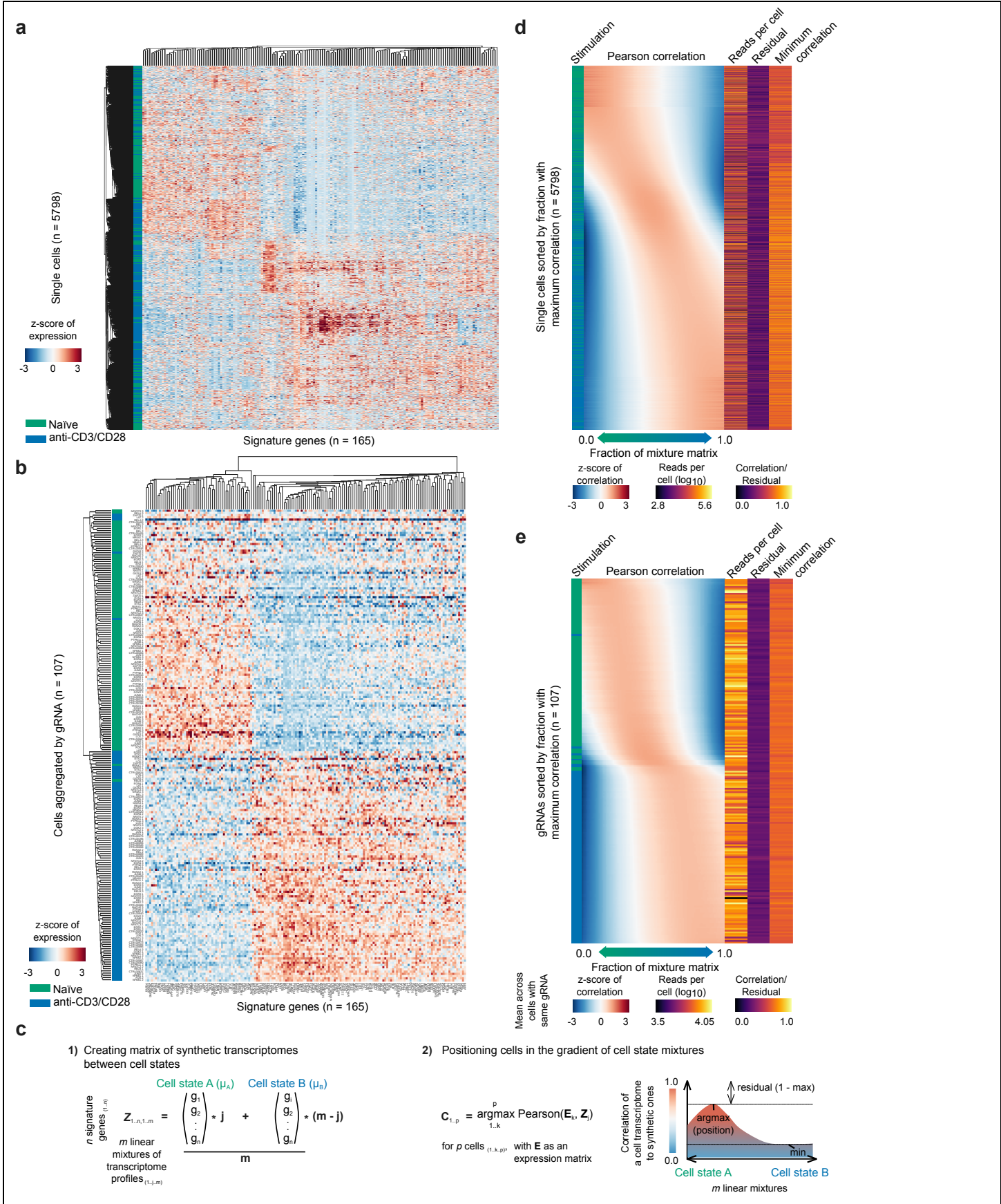
Supplementary Figure 5

Similarities and differences in the transcriptome response at the gRNA and gene level

a) Mean and standard deviation of pairwise distances (L2-norm) between CROP-seq transcriptomes for pairs of gRNAs that target the same gene (orange) or different genes (blue). Statistical significance was assessed with the Mann-Whitney U test. **b)** Pairwise distances as in panel a, shown separately for gRNAs targeting specific genes and for naïve as well as anti-CD3/CD28 stimulated cells. The dotted line indicates the 99th percentile of the distribution of distances between non-targeting gRNAs. **c)** Statistical significance of the transcriptome-wide effect induced by gRNAs targeting specific genes relative to cells with non-targeting gRNAs (based on the Mann-Whitney U test). The dotted line indicates a p-value of 0.01. **d)** As in panel c, but aggregated at the gene level by combining p-values using Fisher's method. Insets show scatterplots for the expression of two example genes with low (left) and high (right) systematic effects on the transcriptome compared to cells expressing non-targeting control gRNAs in the same stimulation condition (y-axis).

Supplementary Figure 6

Unsupervised analysis of the transcriptome response to T cell receptor stimulation

a) Principal component analysis of CROP-seq transcriptomes for cells with an assigned gRNA. **b)** Principal component analysis of median gene expression aggregated across cells expressing the same gRNA. **c)** Principal component analysis of median gene expression aggregated across cells expressing gRNAs targeting the same gene. The first principal component provided best separation between naïve and anti-CD3/CD28 stimulated cells, and the genes on the 99th percentile of loading contributions were selected as the CROP-seq derived TCR activation signature (n = 165 genes). **d)** Genes of the TCR activation signature and their absolute loading values for the first principal component in panel c. **e)** Principal component analysis as in panel a, with cells colored by the expression of three marker genes selected as part of the TCR activation signature. **f)** Enriched pathways and biological processes of genes in panel d, as identified by Enrichr. The combined enrichment score (p-value * z-score) is displayed for the top 8 terms of each gene set library.

a

Single cells (n = 5798)

z-score of expression

-3   0   3

Naïve
anti-CD3/CD28

Signature genes (n = 165)

b

Cells aggregated by gRNA (n = 107)

z-score of expression

-3   0   3

Naïve
anti-CD3/CD28

Signature genes (n = 165)

c

1) Creating matrix of synthetic transcriptomes between cell states

Cell state A (μ_A)   Cell state B (μ_B)

n signature genes $_{(1..n)}$

$$Z_{1..n,1..m} = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{pmatrix} * j + \begin{pmatrix} g_i \\ g_2 \\ \vdots \\ g_n \end{pmatrix} * (m - j)$$

m linear mixtures of transcriptome profiles $_{(1..j..m)}$

$$\underbrace{\phantom{\begin{pmatrix} g_1 \\ g_2 \\ g_n \end{pmatrix}}}_{m}$$

2) Positioning cells in the gradient of cell state mixtures

$$C_{1..p} = \underset{1..k}{\overset{p}{\mathrm{argmax}}} \, \mathrm{Pearson}(E_k, Z_j)$$

for p cells $_{(1..k..p)}$, with **E** as an expression matrix

Correlation of a cell transcriptome to synthetic ones

1.0

residual (1 - max)

argmax (position)

min

0.0

Cell state A   Cell state B

m linear mixtures

d

Stimulation   Pearson correlation   Reads per cell   Residual   Minimum correlation

Single cells sorted by fraction with maximum correlation (n = 5798)

0.0   Fraction of mixture matrix   1.0

z-score of correlation

-3   0   3

Reads per cell (log$_{10}$)

2.8   5.6

Correlation/ Residual

0.0   1.0

e

Stimulation   Pearson correlation   Reads per cell   Residual   Minimum correlation

gRNAs sorted by fraction with maximum correlation (n = 107)

0.0   Fraction of mixture matrix   1.0

Mean across cells with same gRNA

z-score of correlation

-3   0   3

Reads per cell (log$_{10}$)

3.5   4.05

Correlation/ Residual

0.0   1.0

Supplementary Figure 7

Positioning cells and target genes on a spectrum defined by naïve and stimulated cell states

a) Hierarchical clustering of single-cell transcriptomes with unambiguously assigned gRNA (n = 5,798) based on all genes included in the TCR activation signature. Clustering for cells and genes used the Pearson correlation, and the z-score of expression is displayed along with the stimulation state for each cell (left column). **b)** Hierarchical clustering of median gene expression values aggregated across cells expressing gRNAs for the same target gene (n = 107). **c)** Analytical procedure for assigning each cell to a specific position on a spectrum defined by the CROP-seq transcriptomes of non-targeted cells in the naïve and the anti-CD3/CD28 stimulated cell state. In a first step (left), a matrix of synthetic transcriptome signatures (Z) is built by linear combination of the transcriptomes in the two defining cell states (μ_A, μ_B). In a second step (right), the position of the Z matrix that shows the maximum Pearson correlation with the transcriptome of the cells (E matrix) is taken as the cell's position along the spectrum of cell states. **d)** Correlation (row-wise z-score) of single-cell transcriptomes with a matrix comprising synthetic mixtures of transcriptome profiles between the median of non-targeted cells in both conditions (values close to one reflect similarity with anti-CD3/CD28 stimulated cells). Additional data on the transcriptome quality for each cell (unique reads per cell) and the overall correlation performance are shown as columns and reveal no relationship with the inferred position of the corresponding cells. All cells were ordered by the inferred signature value (position with maximum correlation), rather than being clustered as in panel a. **e)** Same as in panel c, but for cells grouped by gRNA target genes.

a z-score of median expression

Jurkat knockout cell lines aggregated by targeted gene (bulk RNA-seq)

-4 0 4

Signature genes (n = 165)
- Naïve
- anti-CD3/CD28

b Stimulation · Correlation · UMIs per cell · Residual · Minimum correlation

Jurkat knockout cell lines (bulk RNA-seq)

0.0 — Fraction of mixture matrix — 1.0

Log fold change over control

Mean across samples with same gene targeted

z-score of correlation
-3 0 3

UMIs per cell (log₁₀)
2.8 5.6

Correlation/ Residual
0.0 1.0

c

Pearson's r
0.675
0.660
0.645
0.630

Number of cells (log10)
2.7
2.4
2.1
1.8
1.5

Naive
Stimulated
Naive
Stimulated

Targeted gene

Pearson's r
0.68
0.64
0.60
0.56
0.52

Number of cells (log10)
1.50
1.25
1.00
0.75
0.50

gRNAs

d Predicted signature position

RNA-seq on bulk population of knockouts

CROP-seq aggregated by gRNAs

Pearson's r:0.919
Spearman's r:0.782

e Log fold change over control (aggregated by gRNAs)

RNA-seq on bulk population of knockouts

CROP-seq aggregated by gRNAs

Pearson's r:0.369
Spearman's r:0.230

Log fold change over control (aggregated by genes)

RNA-seq on bulk population of knockouts

CROP-seq aggregated by genes

Pearson's r:0.641
Spearman's r:0.555

Supplementary Figure 8

Bulk RNA-seq analysis for the arrayed validation screen

a) Hierarchical clustering of the median expression of TCR activation signature genes, based on bulk RNA-seq data aggregated across gRNAs for the same target gene. Clustering of rows and columns used the Pearson correlation, and z-scores of expression are shown. **b)** Correlation (row-wise z-score) of bulk RNA-seq transcriptomes with a matrix comprising synthetic mixtures of transcriptome profiles between the median of non-targeted cells in both conditions (values close to one reflect similarity with anti-CD3/CD28 stimulated cells). Additional metrics are shown as columns (center) and reveal no relationship with the inferred position of the corresponding samples. The effect that perturbing each target gene had on the TCR activation signature based on the bulk RNA-seq data was assessed in comparison to the control group (barplot on the right). **c)** Correlation (top) of median expression levels based on CROP-seq aggregating across target genes (left) or gRNAs (right) compared to the respective bulk RNA-seq libraries across all TCR activation signature genes. The corresponding number of cells in each group is shown at the bottom. Empty values reflect lack of matching bulk RNA-seq libraries due to failed samples in the bulk RNA-seq. **d)** Comparison of the signatures inferred from CROP-seq (x-axis) with those derived fr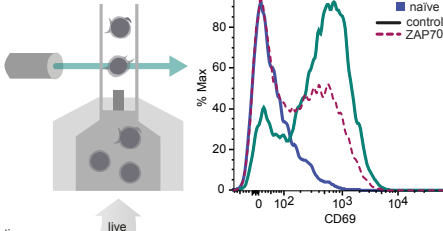om bulk RNA-seq data (y-axis) across all shared gRNAs. **e)** Comparison of the relative impact of each gRNA (left) or target gene (right) on the TCR activation signature relative to the corresponding control for CROP-seq (x-axis) and bulk RNA-seq (y-axis).

a Technical and biological validation of CROP-seq via arrayed screening with two complementary read-outs
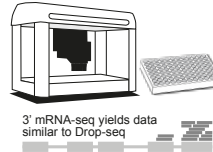
Differentially expressed surface markers from CROP-seq TCR activation signature and established T-cell activation markers (e.g. PD-1, CD38)
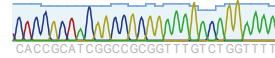
CD82
CD69
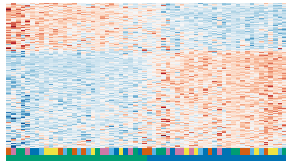
A) Validation by Flow Cytometry

+anti-CD3/CD28
naïve
control
ZAP70

% Max
CD69

Production of single gRNA cell lines in arrayed format

TCR pathway activation or continued starvation

live cells

20 genes x 2 gRNAs
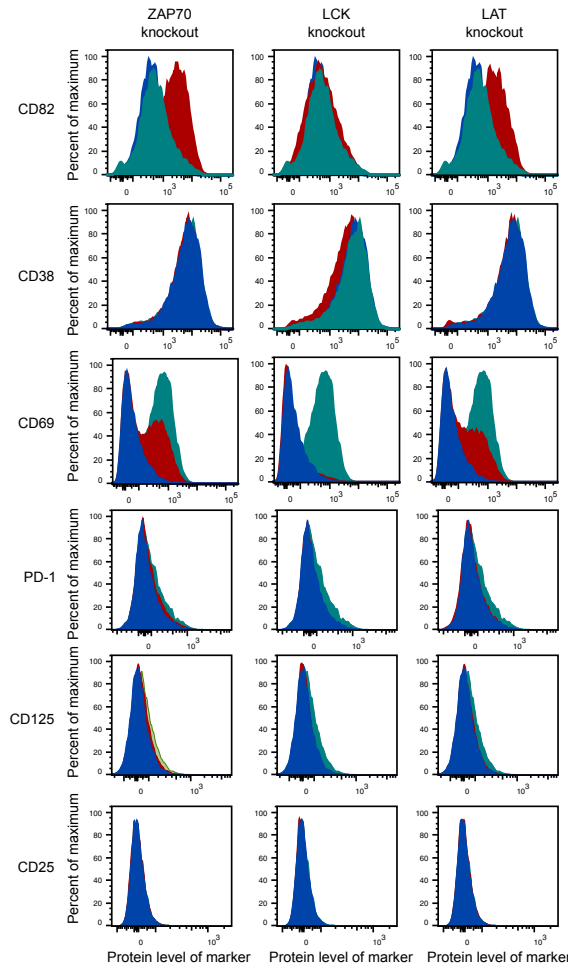8 non-targeting controls
48 single gRNA cell lines

anti-CD3
anti-CD28 stimulated
0% serum
0% serum naïve

3 hrs   7 hrs

B) Validation by bulk RNA-seq

RNA

gDNA

3' mRNA-seq yields data similar to Drop-seq

C) Guide RNA validation by Sanger sequencing

CACCGCATCGGCCGCGGTTTGTCTGGTTTT
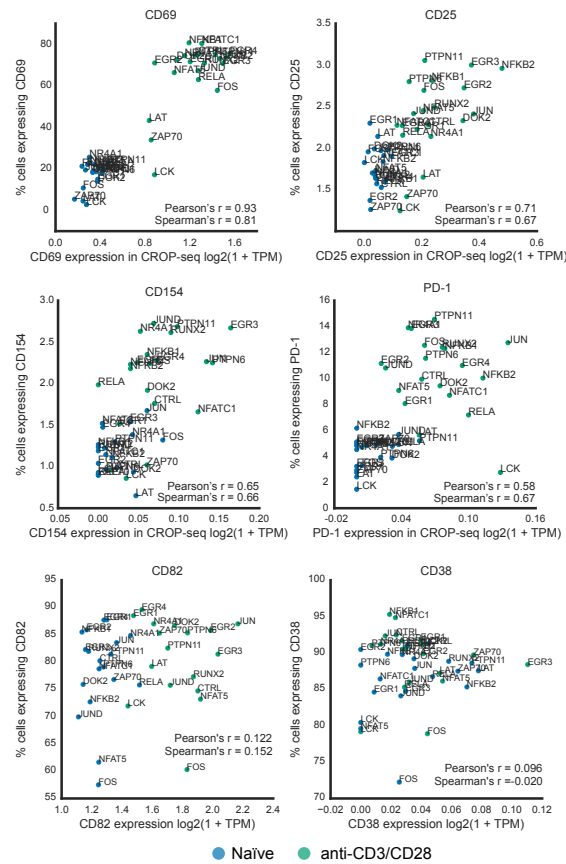
b Examples of marker gene expression changes detected by flow cytometry

naïve control
+anti-CD3/CD28 control
+anti-CD3/CD28 knockout

| ZAP70 knockout | LCK knockout | LAT knockout |

CD82, CD38, CD69, PD-1, CD125, CD25

Percent of maximum
Protein level of marker

c Correlating flow cytometry marker levels with expression values from CROP-seq

48 cell lines (20 genes x 2 gRNAs + 8 controls) and 2 conditions (n = 96)
data is aggregated by gRNA target gene

CD69
% cells expressing CD69
Pearson's r = 0.93
Spearman's r = 0.81
CD69 expression in CROP-seq log2(1 + TPM)

CD25
% cells expressing CD25
Pearson's r = 0.71
Spearman's r = 0.67
CD25 expression in CROP-seq log2(1 + TPM)

CD154
% cells expressing CD154
Pearson's r = 0.65
Spearman's r = 0.66
CD154 expression in CROP-seq log2(1 + TPM)

PD-1
% cells expressing PD-1
Pearson's r = 0.58
Spearman's r = 0.67
PD-1 expression in CROP-seq log2(1 + TPM)

CD82
% cells expressing CD82
Pearson's r = 0.122
Spearman's r = 0.152
CD82 expression log2(1 + TPM)

CD38
% cells expressing CD38
Pearson's r = 0.096
Spearman's r = -0.020
CD38 expression log2(1 + TPM)

Naïve
anti-CD3/CD28

Supplementary Figure 9

Flow cytometry analysis for the arrayed validation screen

a) Experimental strategy of the arrayed validation screen. For this screen, 48 CROPseq-Guide-Puro constructs were individually cloned, targeting 20 genes with two gRNAs each and including eight non-targeting controls. Lentivirus production and transductions were performed in 96-well plates, and cells were expanded for 10 days under puromycin and blasticidin selection. Cells were then split into two parts, serum starved for three hours, and subjected to either anti-CD3/CD28-stimulation or continuous starvation for another four hours. The resulting cell populations were validated by Sanger sequencing of the corresponding gRNAs. For validation of the CROP-seq signature, bulk RNA-seq was performed using a 3' enrichment protocol, yielding data similar to Drop-seq (n = 87 RNA-seq libraries). As a complementary single-cell and protein-based read-out, flow cytometry (n = 96 samples) was performed for surface markers enriched in the TCR induction signature derived from CROP-seq (CD69, CD82) or previously reported as markers of T cell activation (CD25, CD38, CD154, PD-1). **b)** Examples of marker expression changes for TCR pathway activators identified by CROP-seq (ZAP70, LCK, LAT). **c)** Scatterplots comparing protein levels for TCR induction markers to RNA expression values obtained by CROP-seq.