nature portfolio

https://doi.org/10.1038/s41586-025-09313-3

Supplementary information

Precisely defining disease variant effects in CRISPR-edited single cells

In the format provided by the authors and unedited

Supplementary Information for:

Precisely defining disease variant effects in CRISPR-edited single cells

Yuriy Baglaenko[1-6#*], Zepeng Mu[1-4*], Michelle Curtis[1-4], Hafsa M Mire[1-4], Vidyashree Jayanthi[1-4], Majd Al Suqri[1-4], Cassidy Liu[1-4], Ryan Agnew[1-4], Aparna Nathan[1-4,7], Annelise Yoo Mah-Som[1-4], David R. Liu[8-10], Gregory A. Newby[8-12], Soumya Raychaudhuri[#1-4,7]
[1]Center for Data Sciences, Brigham and Women's Hospital, Boston, MA, USA.
[2]Division of Genetics, Department of Medicine Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.
[3]Division of Rheumatology, Inflammation, and Immunity, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.
[4]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA.
[5]Division of Human Genetics and Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA.
[6]Department of Pediatrics, University of Cincinnati, College of Medicine, Cincinnati, OH, USA
[7]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA.
[8]Merkin Institute of Transformative Technologies in Healthcare, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[9]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA
[10]Howard Hughes Medical Institute, Harvard University, Cambridge, MA, USA
[11]Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[12]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA

[#] co-corresponding, * equal contribution

Supplementary Notes:

Calling genotypes in CRAFTseq

Here, we present our genotyping and allele filtering strategy using data presented in **Fig. 1** in main text. We aim to summarize our DNA analysis steps to provide clarity. In the experiment discussed below, a mixture of three cell lines was used to determine the accuracy of our protocol. We amplified genomic regions around Exon 7 of PTEN to identify somatic mutations. Jurkats are reported to have two mutations in the amplified region, both insertions and or deletions (indels). In the analysis below, we unbiasedly filter cells and identify their genotypes without the prior assumption of dizygosity.

Generation of genomic DNA amplicons
Genomic DNA is amplified in a nested two-step reaction. The first reaction takes place during the pre-amplification step alongside cDNA amplification. These genomic primers are specific and have been rigorously tested at 10pg of bulk genomic DNA input to ensure robust amplification at single cell equivalence. From this reaction, an aliquot is taken for a subsequent nested PCR. In this step, “capture” oligos are added along with nested primers to amplify the region of interest and introduce a well barcode. The final product of this PCR is pooled per plate, robustly cleaned, treated with exonucleases and then sequenced.

A schematic of this entire process is shown in **Extended Data Fig. 1**. The final generated library is an amplicon (see below) that is sequenced from both ends on a MiSeq instrument. The generated FASTQ files are then processed as described (**Supplementary Figure 6**).

Demultiplexing and alignment to an amplicon
Since the barcode specifying single cell identity is custom and internally located on read 1 of the amplicon, we use a custom bash script to demultiplex the files. Each cell is demultiplexed into two separate fastq files. Those fastqs are then batch processed with **CRISPREsso2** and a specified reference amplicon sequence. The main output of CRISPResso used in downstream analysis is the allele frequency tables generated per cell that are then processed with custom R code.

An example of this output is shown below. Barcode_DNA specifies the i5 barcode of the plate and the Well_ID is mapped from a whitelist of barcodes to well locations (**Supplementary Figure 7**).

Genomic DNA Filtering and Analysis
First, let’s examine the total number of reads per cell shown below scaled at log10. DNA is recovered from every cell except three on plate 2 (TATCTGACCT) but this high recovery rate could be due to noise discussed later. To correct for this, we will filter on a minimum number of reads per cell that have aligned to the amplicon (**Supplementary Figure 8**).

Step 1: Filtering cells based on read distribution.

Previous iteration of this filtering used empirical cutoffs based on the distributions. For instance, a cutoff of 300 was chosen empirically based on the distribution of reads per cell. See below. The right hand plot shows the histogram of total reads in each cell. The left-hand plot plots these same values ranked. The dashed red-line is the cutoff (**Supplementary Figure 9**).

There were significant plate-to-plate differences in this absolute number. Below is the second plate from this same experiment. A cutoff of 30 reads per cell was chosen for this plate, although the relative distribution looks identical to plate 1 there was less overall sequencing depth (**Supplementary Figure 10**).

As an alternative and updated strategy, we have implemented an unbiased and automatic filtering strategy on cells ranked by total read counts. In this updated framework, we are motivated by the analysis performed by Lun et al. describing the Empty Droplets algorithm. Rather than choosing the threshold empirically per plate, we can use a fitted smooth spline and curvature function to find the inflection points. Below is the plot of log10(total # of reads per cell) vs rank for one of the plates (**Supplementary Figure 11**).

We consider the log(TotalReads) = f(Barcode Rank) with f = f″/(1+f′^2)^1.5. (the curvature function). The knee point represents the rank at which the total number of reads per cell begins to drop rapidly and is defined by the minimum of the above function.

The curvature of the fitted data as defined by the function above is shown (left). This minimum defines the knee point and a cutoff of 693 reads per cell shown by the dashed red line below (**Supplementary Figure 12**).

The second plate has a similar distribution and a cutoff of 66 reads (**Supplementary Figure 13**).

Replotting the total reads per cell across the plate, we can identify cells that are filtered. The distribution of missing cells is relatively random (**Supplementary Figure 14**).
The result is **605/768** cells remaining from the two plates or a recovery of **80%.**

Step 2: Filtering alleles per cell
Examining a few individual cells, we can plot all the alleles recovered from amplicon sequencing. Alleles refers to unique sequences identified in the amplicon and aligned to the reference provided. Presented are two examples: a Jurkat cell (A8 - left) and a Daudi cell (B3 - right). Their identities have been confirmed by both index sorting and RNA clustering (**Supplementary Figure 15**).

Many unique alleles are recovered, according to the CRISPResso2 output, in every cell but the vast majority of them are incredibly rare. This is likely due to PCR induced mutations and sequencing error. In the first example there are three dominant alleles, in the second example only 1. This suggests that the first cell has a multitude of alleles and may be polyploid while the second cell is homozygous with only one allele. From the literature, we know that Jurkat cells are near quadraploid and Daudi cells are near diploid.

To filter noisy alleles and identify accurate single cell genotypes, we perform filtering on a per cell basis by identifying the minimum of the first derivative of ranked alleles. This analysis identifies a right edge to the curve that is used to determine the total number of alleles per cell. The rationale behind this approach is that noisy alleles are relatively evenly distributed and rare (as seen above). There is a clear dropoff from true alleles and background noise.

For this filtering, we assume pre-filter on the top 10 alleles to speed up computation. In each cell, we scale total reads per allele with log10 and take the first derivative of the ranked alleles by number of reads. Below are two random cells with log scaled data (**Supplementary Figure 16**).

The first derivative of these plots is below with the minimum in red (**Supplementary Figure 17**).

In these examples, A5 has three alleles that are all contributing to the genotype of the cell and the rest are removed from downstream analysis. G8 has only one homozygous allele.

Performing this processing on all cells, we find that most cells only have 1 allele. However, some rare cells capture several alleles (up to 6) (**Supplementary Figure 18**).

In these cases, errors likely arose early and has propagated in amplification and sequencing. As an additional filtering step, we implement a threshold on the minimum percentage of reads per allele to remove lowly expressed / background alleles. This is set between 5-20%. A higher threshold is required to clean noisier data. The addition of non-targeting controls in each experiment provides a ground-truth on the above filtering steps.

Below is the data of all allele percentages (ie the percentage of reads per allele per cell) from the experiment in Figure 1. The red line represents the 5% cutoff chosen (**Supplementary Figure 19**).

Next, the alleles are re-aligned to a reference with biostrings in R. This is purely for visualization as the initial reference did not include indels. In other experiments, this is not routinely performed as the output of CRISPResso2 will provide the aligned sequence in the form of the allele (assuming no insertions or deletions).

The resulting alignments are trimmed and distributions plotted below along with a table (**Supplementary Figure 20-21**).

There are three dominant alleles as expected after filtering. The rare alleles represent sequences that have propagated past our set threshold. Next we filter to remove any alleles present in less than 5 cells.

In conclusion, these robust filtering steps culminate in the identification of three alleles shown below (**Supplementary Figure 22**).

A is reference, B is a 38bp indel, and C is a 9bp indel.

Step 3: Calling genotypes in individual cells.
Each cell now has a combination of one or more of these alleles as we have determined after filtering. If only one allele is present the cell is presumed to be homozygous. We collapse this information so each cell has a combination of all alleles present regardless of frequencies. For example, with 50% of reads from allele A, and 25% of reads from allele B and C, the cell would still be assigned genotype ABC.

The result looks like the graph below, plotted with ggplot2, combined with information from our RNA clustering (0 - Jurkat, 1- Daudi, 2 - HEK293T) (**Supplementary Figure 23**).

The majority of cells are reference genotypes as shown by the presence of only the A allele. Jurkat (Cell Type 0), however, have a single dominant genotype ABC. The lesser genotypes are likely the result of either allelic dropout or cancer cell instability.

Notably, Daudi (Cell Type 1) and HEK293T (Cell Type 2) only have the reference allele in the vast majority of cells with an accuracy of > 98% in Daudi and HEK293T cells with the described filtering.

Sources of error / noise.
The most likely causes of miscalled genotypes in our protocol are 1. Sequencing induced errors from repeat regions. 2. PCR induced errors from amplification of rare genomic products. 3. Background plate effects that result in a low-level of noise in all wells. The latter is the most likely cause of the background noise detected in our assay and the filtering steps above focused on removing this noise. We suspect that this occurs because of low-levels of evaporation or carryover of PCR products across wells.

Detailed Experimental and Computational Protocols for CRAFTseq

For more detailed steps see protocolols.io:
https://www.protocols.io/view/craftseq-dm6gpz8ojlzp/v2.

REAGENTS and EQUIPMENTS

Some key plastic labware items and consumables

Tips, 30uL, sterile filter Bravo (Agilent, cat.no. AGI-30.F)
Tips, 70uL, sterile filter Bravo (Agilent, cat.no. AGI-70.F)
Eppendorf PCR plate 384-wells (Eppendorf, cat.no. 951020711)
LS Columns (Miltenyi, cat.no. 130-042-401)
Pre-separation filters (70 µm) (Miltenyi, cat.no. 130-095-823)
Flow tubes with cell strainer snap cap (Corning, cat.no. 352235)
Qubit 1X dsDNA high sensitivity (HS) assay kit (Fisher Scientific, cat.no. Q33231)
Qubit assay tubes (ThermoFisher Scientific, cat.no. Q32856)
RNaseZap RNase Decontamination Wipes (ThermoFisher Scientific, cat.no. AM9786)
Agilent 4200 Tapestation loading tips (Agilent, cat.no. 5067-5599)
D1000 ScreenTape (Agilent, cat.no. 5067-5582) with D1000 Reagents (Agilent, cat.no. 5067-5583)
D5000 ScreenTape (Agilent, cat.no. 5067-5588) with D5000 Reagents (Agilent, cat.no. 5067-5589)

Cell culturing, isolation, stimulation, and polarization

RPMI 1640 medium (ThermoFisher Scientific, cat.no. 11875093)
DMEM, high glucose (ThermoFisher Scientific, cat.no. 11965092)
Fetal bovine serum (FBS) (Rockland, cat.no. FBS-02-0050)
MEM non-essential amino acids solution (100X) (ThermoFisher Scientific, cat.no. 11140050)
Sodium pyruvate 100mM (ThermoFisher Scientific, cat.no. 11360070)
HEPES 1M (ThermoFisher Scientific, cat.no. 15630080)
L-Glutamine 200mM (ThermoFisher Scientific, cat.no. 25030081)
Penicillin-Streptomycin (10,000U/mL) (Fisher Scientific, cat.no. 15-140-148)
Gentamicin 10 mg/mL in distilled water, 10 mL (Gibco, cat. no 2450051)
2-Mercaptoethanol (BME) (Fisher Scientific, cat.no. 21-985-023)
PBS, pH 7.4 (ThermoFisher Scientific, cat.no. 10010049)
EDTA (0.5 M), pH 8.0, RNase-free (ThermoFisher Scientific, cat.no. AM9260G)
Naive $CD4^+$ T Cell Isolation Kit II, human (Miltenyi, cat.no. 130-094-131)
Dynabeads Human T-Activator CD3/CD28 (ThermoFisher Scientific, cat.no. 11132D)
Recombinant Human IL-2 (carrier-free) (BioLegend, cat.no. 589102)
Recombinant Human IL-7 (carrier-free) (BioLegend, cat.no. 581902)
Recombinant Human IL-12 (p70) (carrier-free) (BioLegend, cat.no. 573002)
Recombinant Human TGF-β1 (carrier-free) (BioLegend, cat.no. 781802)
Bovine Serum Albumin (Sigma-Aldrich, cat.no. A3294-50G)

CRISPR editing of cell lines and primary cells

Cas9 recombinant protein (Synthego)
sgRNA custom (Synthego, various)
HH cutaneous T-cell lines (ATCC, cat.no. CRL-2105)
Jurkat E6-1 (ATCC, cat.no. TIB-152)

Daudi cells (ATCC, cat.no. CCL-213)
HEK293T cells (ATCC, cat.no. CRL-3216)
SE cell line 4D-nucleofector X kit S (Lonza, cat.no. V4XC-1032)
SF cell line 4D-nucleofector X kit S (Lonza, cat.no. V4XC-2032)
P3 primary cell 4D-nucleofector X kit S (Lonza, cat.no. V4XP-3032)
Monarch genomic DNA purification kit (NEB, cat.no. T3010S)
Monarch total RNA miniprep kit (NEB, cat.no. T2010S)
Protector RNase inhibitor (MilliporeSigma, cat.no. 3335399001)
HiScribe T7 High-Yield RNA synthesis kit (NEB, cat.no. E2040S)
N4-Methylcytidine-5'-Triphosphate (Trilink BioTechnologies, cat.no. N-1080)
CleanCap Reagent AG (Trilink BioTechnologies, cat.no. N-7113)

Cell hashing and staining
Brilliant Violet 421 anti-human CD4 antibody (BioLegend, cat.no. 300532)
Brilliant Violet 650 anti-human CD197 (CCR7) antibody (BioLegend, cat.no. 353234)
Brilliant Violet 711 anti-human HLA-DR antibody (BioLegend, cat.no. 307644)
FITC anti-human CD27 antibody (BioLegend, cat.no. 302806)
PE anti-human CD40 antibody (BioLegend, cat. no 07958935001)
PerCP/Cyanine5.5 anti-human CD45RO antibody (BioLegend, cat.no. 304222)
PE/Cyanine7 anti-human CD127 (IL-7Rα) antibody (BioLegend, cat.no. 351320)
APC anti-human CD62L antibody (BioLegend, cat.no. 304810)
APC/Cyanine7 anti-human CD3 antibody (BioLegend, cat.no. 300426)
PE anti-human CD25 antibody (BioLegend, cat.no. 356104)
FITC anti-human CD45 antibody (BioLegend, cat.no. 368508)
Brilliant Violet 750 anti-human CD45 antibody (BioLegend, cat. no 368542)
Pacific Blue anti-human CD45 antibody (BioLegend, cat. no 368539
APC anti-human CD45 antibody (BioLegend, cat. no. 368511)
PE anti-human CD45 antibody (BioLegend, cat.no. 368510)
Brilliant Violet 421 anti-human CD45 antibody (BioLegend, cat.no. 368521)
APC anti-human CD81 (TAPA-1) antibody (BioLegend, cat.no. 349509)
TotalSeq-A human universal cocktail V1.0 (BioLegend, cat.no. 399907)
Human TruStain FcX (Fc receptor blocking solution) (BioLegend, cat.no. 422302)

CRAFTseq
All primers, adapters, and oligonucleotides (IDT or Genewiz, various)
DNase/RNase-free water (ThermoFisher Scientific, cat.no. 10977015)
Triton X-100 solution (Sigma-Aldrich, cat.no. 9036-19-5)
dCTP solution 100mM (ThermoFisher Scientific, cat.no. R0151)
dNTP (ThermoFisher Scientific, cat.no. R0182)
KAPA HiFi HotStart plus dNTPs 100U (Roche, cat.no. 07958889001)
KAPA HiFi HotStart ReadyMix 6.25 mL (Roche, cat. no. 07958935001)
Betaine 5M solution (MilliporeSigma, cat.no. B0300-5VL)
USB dithiothreitol (DTT) 0.1M solution (ThermoFisher Scientific, cat.no. 707265ML)
$MgCl_2$ 1M (ThermoFisher Scientific, cat.no. AM9530G)
Maxima reverse transcriptase (200U/μL) (ThermoFisher Scientific, cat.no. EP0743)
Recombinant RNase inhibitor (RRI) (Takara, cat.no. 2313B)
CleanNGS 500mL (CleanNA, cat.no. CNGS-0500)
Thermolabile exonuclease I (NEB, cat.no. M0568L)
Exonuclease VII (NEB, cat.no. M0379L)
Q5 HotStart high-fidelity DNA polymerase (NEB, cat.no. M0493L)
Nextera XT DNA library prep kit (24 samples) (Illumina, cat.no. FC-131-1024)

Liquid handlers and dispensers

To accurately and quickly dispense small volumes into multiple 384-well plates, we utilized advanced liquid handlers and dispensers, such as the BRAVO, I.DOT, and MANTIS. These automated systems offer precision and consistency, which are critical for minimizing variability during library preparation that requires precise reagent handling. The BRAVO platform allowed for flexible liquid handling across different formats, while the I.DOT system enabled ultra-low volume dispensing with minimal waste, making it ideal for applications requiring very small volumes. The MANTIS dispenser provided high-throughput capability with the added benefits of rapid dispensing times and reduced setup complexity. We followed the manufacturer's guidelines for configuring each system, ensuring that the workflows were fully integrated into our experimental setups.

REAGENT and EQUIPMENT SETUPS

Cell staining and sorting into lysis plates

TotalSeq panel was resuspended by equilibrating the lyophilized panel vial to room temperature for 5 minutes. The vial was then placed in an empty Eppendorf tube and spun down at 10,000xg for 30 seconds at room temperature. The lyophilized panel was rehydrated by adding 55µL of cell staining buffer (1X PBS, 0.05% BSA, 2mM EDTA), after which the cap was replaced, and the vial was vortexed for 10 seconds. The rehydrated panel was incubated at room temperature for 5 minutes. Following incubation, the vial was vortexed again and spun down at 10,000xg for 30 seconds at room temperature. The entire 55µL volume of the reconstituted cocktail was transferred to a low protein binding Eppendorf tube and centrifuged at 14,000xg for 10 minutes at 4°C.

CAUTION once the panel is rehydrated, it is recommended to use it as quickly as possible. If necessary, it can be stored temporarily at 4°C for short-term use. This is highly recommended if left with 25 µL of TotalSeq by the end of sorting

Lysis plates were set up according to the following components:
Concentration (conc) and volume (vol) of each reagent used to set up the lysis plates.

	Start Conc	Final Conc	Vol per well (uL)
Water	-	-	0.158
TritonX-100 (%)	10	0.2	0.020
DTT (mM)	100	1.2	0.012
dNTP (mM)	25	6	0.240
Betaine (M)	5	1	0.200
dCTP (mM)	100	9	0.090
RRI (U/uL)	40	1.2	0.030
OligoDT (uM)	10	2.5	0.250

Supplementary Table 16: Recipe for lysis plate.

CRAFTseq: reverse transcription (RT) and initial genomic DNA (gDNA) amplification PCR reaction

To setup the PCR reaction mixture for RT and initial genomic DNA amplification we used the following:

	Start Conc	Final Conc	Vol per well (uL)
Water	-	-	1.47
TSO (uM)	200	1.8	0.05
dNTP (mM)	10	0.3	0.15
5X KAPA High Fidelity Buffer (X) with Mg	5	1	1.00
Betaine (M)	5	0.8	0.80
DTT (mM)	100	4.8	0.24

$MgCl_2$ (mM)	1000	9.7	0.05
MaximaH Minus RT Enzyme (U/uL)	200	2	0.05
RRI (U/uL)	40	0.8	0.10
KAPA HotStart HiFi Enzyme (U/uL)	1	0.02	0.10
ADT additive primers (uM)	5	0.05	0.05
Specific genomic DNA primers (uM)	20	0.2	0.05

Supplementary Table 17: Recipe for reverse transcription and first gDNA PCR. ADT additive primers and specific genomic DNA primers are to be dispensed after reverse transcription, before first gDNA PCR.

RT step	Step	Temp	Time	
	1	50°C	1hour	
	2	85°C	5min	
	3	4°C	Hold	
PCR step				
	1	98°C	5min	
	2	98°C	20sec	21 cycles
	3	65°C	20sec	
	4	72°C	6min	
	5	72°C	5min	
	6	4°C	hold	

Supplementary Table 18: PCR cycler setup for RT and initial genomic DNA amplification.

NOTE the number of cycles for the PCR setup above is variable.

CRAFTseq: nested genomic DNA amplification PCR reaction

To setup the PCR reaction mixture for the nested genomic DNA amplification we used the following:

	Start Conc	Final Conc	Vol per well (uL)
Water	-	-	1.55
dNTP (mM)	10	0.3	0.15
5X KAPA High Fidelity Buffer (X) with Mg	5	1	1.00
KAPA HotStart HiFi Enzyme (U/uL)	1	0.02	0.10
Genomic nested capture/P7 primers (uM)	5	0.4	0.40
Unique capture barcode primer (Unblocked) (uM)	2.5	0.4	0.80

Supplementary Table 19: Nested genomic DNA amplification reaction components.

Step	Temp	Time	
1	98°C	5min	
2	98°C	20sec	20 cycles
3	65°C	20sec	
4	72°C	30sec	
5	72°C	5min	
6	4°C	hold	

Supplementary Table 20: PCR cycler setup for nested genomic DNA amplification.

CRAFTseq: Exonuclease treatment

To setup the exonuclease treatment reaction we used the following:

cDNA/ADT ExoI	Start Conc	Final Conc	Vol (uL)
Q5 buffer (X)	5	1	5
Thermolabile ExoI (U/uL)	20	1.6	2
gDNA ExoI/VII			
ExoVII buffer (X)	5	1	4
ExoVII (U/uL)	10	0.5	1
Thermolabile ExoI (U/uL)	20	4	4
Water	-	-	1

Supplementary Table 21: cDNA, ADT, and gDNA exonuclease reaction setup

CRAFTseq: P5-P7 adapter amplification

To amplify the P5-P7 adapters for all three modalities, we used the following:

cDNA P5-P7	Start Conc	Final Conc	Vol (uL)
RNA P7 unique (uM)	2	0.2	2.5
RNA P5 common (uM)	2	0.2	2.5
NPM (Nextera PCR mix)	-	-	7.5
ADT P5-P7			
ADT P7 unique (uM)	2	0.2	2.5
ADT P5 common (uM)	2	0.2	2.5
Q5 buffer (X)	5	0.8	4
dNTP (mM)	10	0.4	1
Q5 HotStart (U/uL)	2	0.04	0.5
Water	-	-	9.5
gDNA P5-P7			
DNA P7 unique (uM)	2	0.2	2.5
DNA P5 common (uM)	2	0.2	2.5
Q5 buffer (X)	5	1	5
dNTP (mM)	10	0.4	1
Q5 HotStart (U/uL)	2	0.04	0.5
Water	-	-	3.5

Supplementary Table 22: cDNA, ADT, and gDNA P5-P7 amplification reaction setup

cDNA	Step	Temp	Time	
	1	72°C	3min	
	2	95°C	30sec	
	3	95°C	10sec	16 cycles
	4	55°C	30sec	
	5	72°C	45sec	
	6	72°C	5min	
	7	4°C	hold	
ADT/gDNA				
	1	98°C	3min	
	2	98°C	15sec	16 cycles
	3	65°C	20sec	
	4	72°C	45sec	
	5	72°C	10min	
	6	4°C	hold	

Supplementary Table 23: PCR cycler setup for P5-P7 PCR

NOTE the number of cycles for the PCR setup above are variable, but 16 cycles are optimal.

PROCEDURE

CRISPR editing of cell lines and primary T cells

Base editor mRNAs used for editing cell lines were generated through in vitro transcription using the HiScribe T7 high-yield RNA synthesis kit following the method outlined in Chen *et al*. (2021)[1]. NEBNext polymerase was used to PCR-amplify the template plasmids, incorporating a functional T7 promoter and a 120-nucleotide polyadenine tail. Transcription reactions were conducted with full substitution of uracil by N1-methylpseudouridine and co-transcriptional 5'-capping using the CleanCap AG analog to form a 5'-Cap1 structure. The mRNAs were purified via ethanol precipitation, dissolved in nuclease-free water, and normalized to a concentration of 2µg/µL using Nanodrop RNA quantification of diluted samples. Aliquots of mRNA editors were stored at -80°C to prevent freeze-thaw cycles.

STEP 1: Editing cell lines

1.1 HH cutaneous T cells, Jurkat, and Daudi lines were cultured in complete RPMI (cRPMI) supplemented with 10% heat-inactivated FBS, 1X non-essential amino acids, 1mM sodium pyruvate, 10mM HEPES, 2mM L-glutamine, 1% penicillin-streptomycin (Pen-Strep, 10,000units/mL of penicillin and 10,000µg/mL of streptomycin), and 0.5mM BME. HEK293T cells

were cultured in complete DMEM (cDMEM) supplemented with 10% heat-inactivated FBS and 1% Pen-Strep.

1.2 For CRISPR-Cas9 mediated homology-directed repair (HDR) editing, 20μM of Cas9 protein was mixed with an equal volume of 40μM modified sgRNA (previously resuspended in TE buffer to get 40μM) and incubated at 37°C for 15 minutes to form ribonuclear protein (RNP) complexes. About $2x10^5$ to $5x10^5$ cells per nucleofection reaction were used. The cells were spun at 200xg for 10 minutes, and the media was removed. Then cells were washed once in 1X PBS and resuspended in either SE nucleofection solution (18μL nucleofector SE buffer and 4μL supplement) or SF nucleofection solution (18μL nucleofector SF buffer and 4μL supplement). Then, 18μL of cells in nucleofection solution was nucleofected with 2μL of RNPs and 1μL of the ssODN repair template using a 4D nucleofector, with a 16-well cassette, following the SE protocol (CL-120 for HH cells) or the SF protocol (CA-137 for Daudi cells).[2]

1.3 Immediately after nucleofection, 90μL of pre-warmed complete medium was added, and the cells were incubated at 37°C for 30 minutes. The cells were then transferred to 24-well culture plates containing 1mL of pre-warmed complete medium and cultured until sorting, which occurred 5 days after nucleofection.

1.4 For CRISPR-Cas9 base editing 1uL of mRNA (2ug/uL) encoding the base editor was mixed with 1uL of modified sgRNA (40μM) and kept on ice. After washing in 1X PBS cells in either 18uL of SE or SF nucleofection solution were nucleofected with 2μL of mRNA/sgRNA mixture in an 4D nucleofector using 16-well cassette (SE protocol: CL-120 for Jurkat cells, SF protocol: CA-137 for Daudi cells). Then quickly added 90uL of complete medium to cells and incubated at 37°C for 30 minutes. Cells were transferred to 24-well plates with 1mL pre-warmed complete medium and cultured until sorting day.

1.5 Editing was confirmed through flow cytometry or Sanger sequencing, verifying successful modifications at the target loci. Following the manufacturer's protocol for Monarch genomic DNA purification, we extracted genomic DNA from the edited cells, performed PCR, and conducted subsequent sequencing.

1.6 Multiplex editing was performed in cells by thawing sgRNAs corresponding to each locus of interest within the genome. After the collection and washing of cells with sterile 1X PBS, the cells were mixed with the contents of the SF nucleofection kit, along with cas9 mRNA. After this, 19 μL aliquots of the cell/media/cas9 mRNA mixture were added to the bottom of each of the wells within a 16-well cassette. Then 1 μL of sgRNA from each point of interest was added to the bottom of each well full of cells; the wells with cells intended to be multiplexed received sgRNAs from both loci of interest. Then the cells were nucleofected using the 4D nucleofector (SF protocol: CA-137 for Daudi cells). Post-nucleofection, each well received 90 μL of complete RPMI (that was supplemented with 1% gentamicin in place of Penicillin-Streptomycin). The cassette was incubated for 30 minutes at 37ºC in an incubator with 5% carbon dioxide. When incubation was complete, the cells were transferred to 24-well plates with 1 mL of pre-warmed complete RPMI and cultured for six days until sorting.

STEP 2: Stimulating primary T cells

2.1 We recruited healthy individuals and processed 40-50mL of peripheral blood. Peripheral blood mononuclear cells were isolated by density centrifugation, layering Ficoll-Paque beneath a mixture of equal amounts of blood and 1X PBS before centrifugation.The buffy coat layers were extracted, washed with 1X PBS, and resuspended in cRPMI medium. For storage, an equal

volume of RPMI supplemented with 10% DMSO and FBS was added to the cells, which were then frozen at about 10^7 cells per cryovial in liquid nitrogen until further use.

2.2 PBMCs were quickly thawed into cRPMI, spun at 500xg for 5 minutes, and washed twice in cRPMI. Following this, the cells were spun down again and resuspended in MACS buffer (1X PBS, 2% FBS, 2mM EDTA). The cell pellet was resuspended in 4µL of buffer for every 10^6 total cells. We followed the manufacturer's protocol for naïve CD4+ T cell isolation, including antibody and microbead incubation steps, as well as magnetic cell separation.

CRITICAL Make sure to thaw the tubes of PBMCs for no more than 10 seconds in order to keep the viability high. This is because the cells have been frozen in Dimethyl Sulfoxide (DMSO), which is unfavorable for cells at room temperature, but it protects cells during the freezing and thawing process. Hence it is used in a mixture with RPMI and FBS for freezing.

2.3 The appropriate number of stimulation beads (anti-CD3/CD28) ($4x10^7$ beads/mL) was washed on a magnet using 1X PBS. For every 10^6 $CD4^+$ T cells, 25µL of beads was used. After washing, the isolated naïve $CD4^+$ T cells were stimulated with a 1:1 ratio of Dynabeads in the presence of 1X stimulated complete media. The 1X stimulated complete media was prepared by first creating 2X stimulated complete media (10ng/mL IL-2, 20ng/mL IL-7). This 2X stimulated complete media was then mixed 1:1 with 125µL of the isolated cells from the previous step that was already in complete medium, resulting in the final 1X stimulated complete media (5ng/mL IL-2, 10ng/mL IL-7). The stimulation was carried out in 96-well U-bottom plates, using 250µL of the 1X stimulated complete media for every 250,000 $CD4^+$ T cells. The cells were divided evenly across the wells, rounding up or down to the nearest appropriate amount to ensure consistency. The cells were monitored to ensure robust proliferation before proceeding with editing, which took place two days after stimulation. If proceeding to polarization, stimulation was conducted for one day prior to editing.

STEP 3: Editing primary T cells

3.1 This step involves CRISPR-Cas9 base editing of activated $CD4^+$ T cells, which is performed without subsequent polarization. For the base editing 1uL of mRNA (2ug/uL) encoding the base editor was mixed with 1uL modified sgRNA (40µM) and kept on ice. Cells were pooled per donor (nucleofecting about $2.5x10^5$ to $5x10^5$ cells) and washed in 1X PBS and then resuspended in 20µL of P3 nucleofection solution (18µL nucleofector P3 buffer and 4µL supplement). Then, 18µL of cells in the nucleofection solution was nucleofected with 2µL mRNA/sgRNA mixture in an 4D nucleofector using a 16-well cassette (P3 protocol: EH-115 program for $CD4^+$ T cells). Then, 90 µL of complete media was quickly added to each well of cells before the cassette was incubated at 37º, 5% carbon dioxide for 30 minutes.Cells were transferred to 96-well U-bottom plates to the final volume of 250µL of 1X stimulated complete medium (5ng/mL IL-2). On day 3 after nucleofection, the media was replenished by carefully removing 125µL without spinning and adding back 125µL of 2X stimulated complete media (10ng/mL IL-2), and the cells were cultured until sorting day (i.e. 5 days post-nucleofection).

NOTE if proceeding to polarization the editing step above slightly changes, see step 3.2 below. The stimulation beads were not removed during the nucleofection process.

3.2 When proceeding to polarization, activated $CD4^+$ T cells were nucleofected as described in step 3.1 . However, immediately after nucleofection, 100µL of pre-warmed complete media was added to each well of the 16-well nucleofection cassette, and the cells were incubated at 37°C for 30 minutes. The cells were then split into two conditions, taking 50µL for each condition: Th1 and Treg. At a density of 1.25x105 to 2.5x105 cells per well, the two 50 µL aliquots were placed into

wells within a 96-well U-bottom plate. The cells were incubated for 4-5 hours at 37ºC, 5% CO2 before proceeding to polarization.

STEP 4: Polarizing primary T cells
4.1 We polarized about $1.5x10^5$ to $2.5x10^5$ pf T-cells in 250µL of complete medium in 96-well U-bottom plates by adding 50µL of a 5X stimulated complete medium, prepared separately for each condition Th1 or Treg. The 1X final composition for the Th1 condition, following the methods of, included 2µg/mL anti-IL-4 antibody, 10ng/mL IL-12, and 5ng/mL IL-2. The 1X final composition of the Treg condition consisted of 10ng/mL IL-2 and 5ng/mL TGF-β.

4.2 On day 4 post-polarization, media was replenished by carefully pipetting out 125µL of media from all conditions without spinning and adding back 125µL of 2X stimulated complete medium for each Th1 and Treg condition. The final 1X composition after this addition remained the same as previously stated. Cells were monitored to ensure proliferation and sorted 5 days post-polarization (i.e., 10 days post-nucleofection).

<u>Cell hashing, staining, and sorting into lysis plates</u>
STEP 5: Hashing and staining
5.1 For staining cells from editing and non-targeting control conditions, we transferred the samples to a 96-well V-bottom plate. Single-color controls and experimental conditions were centrifuged at 500xg for 5 minutes and washed once with 100µL of cell staining buffer (CSB) (1X PBS, 2 mM EDTA, 0.5% BSA). The cells were resuspended in 100µL of CSB, and 1µL of the appropriate staining antibodies was added per experimental condition. The staining was carried out for 25-30 minutes on ice, away from light. After staining, the cells were centrifuged at 500xg for 5 minutes, washed twice with 100µL of CSB, and finally resuspended in 100µL of CSB. We mixed various experimental conditions, to get 10% non-targeting control. After washing, we proceeded to TotalSeq staining.

(See ‘Cell Staining and sorting into lysis plates’ for how the TotalSeq cocktail is prepared)

NOTE: for the multiplexed edited cells, include a separate color stain for the multiplex edited cells, and a separate stain for the non-targeted control

5.2 Hashed cells and single-color control cells were centrifuged and blocked with 50µL of Fc block mixture in MACS (5µL Fc block in 45µL of MACS) and incubated for 10 minutes at 4°C. We added 25µL of the reconstituted TotalSeq cocktail to the tube containing 50µL of FcR-blocked hashed cells from the various experimental conditions and incubated for 30 minutes at 4°C. The cells were then washed twice with MACS buffer and resuspended at 1 million cells/mL in MACS buffer.

CRITICAL for the blocking steps in 5.2, make sure the cells are kept away from light

NOTE if the hashing antibodies and the TotalSeq panel target the same surface marker, we recommend adding the hashing antibodies and the panel at the same time.

STEP 6: Sorting
6.1 A cell lysis reaction mixture, containing water and unique OligoDT barcodes per well, was prepared using the setup in Table 1. We added 0.75µL of lysis mixture to each well of a new 384-well plate and 0.25µL of unique OligoDT to each well, such that each lysis contains a total volume of 1µL. Using the BRAVO, 0.75 µL of lysis mixture and 0.25µL of unique OligoDT was added to each well of a new 384-well plate. Cells were then sorted into lysis plates using the Bigfoot

Spectral cell sorter with standard compensation. Immediately after sorting, the plates were spun at 1000xg for 1 minute, flash-frozen on dry ice, and stored in -80°C until further library processing.

NOTE we recommend preparing lysis plates a few hours before sorting or the day before and sorting immediately afterward. After sorting, lysis plates with cells can be stored at -80°C for up to a month.

CRAFTseq

STEP 7: Pre-amplification and specific genomic DNA amplification PCR reactions

7.1 A reaction master mix without the ADT additive or genomic DNA primers was set up according to Table 2 in a clean and RNase-free environment. Lysis plates with cells are quickly incubated at 72°C for 3 minutes followed by a 4°C hold step. Plates were briefly spun at 1000xg. Then, 4µL of the master mix previously made was quickly added using MANTIS to each well of the 384-well plates with lysed cells, and the RT cycle is immediately initiated using Table 2.1. Once the RT cycle was completed 0.05µL of the ADT additive primers and 0.05µL of specific genomic DNA primers was added to each well. The final PCR cycle was then performed using Table 2.1 setup.

CRITICAL aliquots of TSO are stored in -80°C and once thawed should not be reused as TSO is susceptible to degradation. Since mRNA degrades quickly, keeping everything on ice and moving as fast as possible is highly recommended. Additional primers to perform the nested gDNA during the RT step were previously tested and resulted in gDNA loss. We highly recommend taking an aliquot of the RT-PCR product to conduct the nested gDNA amplification as indicated in the steps below.

CRITICAL the primers used in the first PCR that are the outer primers of the region of interest are the ones that the cell gDNA is the most sensitive to. This has an effect on the amounts of gDNA and cDNA post-PCR. We faced a challenge here in designing the right set of primers to be used in this first round of PCR, and through multiple rounds of testing the primers on gDNA extracted from Daudi cells and performing the outer and inner PCR cycles, we were able to identify two sets of primers that could effectively target the multiplexed Daudi gDNA.

PAUSE POINT since cDNA has been generated can pause here or proceed to the next step.

7.2 A reaction master mix was prepared according to Table 3. We aliquoted 1µL of samples generated after the RT-PCR step into new 384-well plates for the nested gDNA amplification. To these plates we added 3.2µL of the master mix and 0.8µL of the unique capture primers to each well. The plates were then subjected to nested gDNA amplification using the PCR cycle steps outlined in Table 3.1 setup.

7.3 The RT-PCR plate from step 7.1 was pooled and underwent a 0.6X SPRI clean-up. The final product is cDNA, while the supernatant, which contained the ADT, underwent a 1.4X SPRI clean-up. The gDNA plate was pooled and underwent a 1.2X SPRI clean-up.

CRITICAL SPRI beads are meant to be stored at 4ºC in the long term. Before using them for the clean up steps, make sure the SPRI beads are at room temperature and vortex to ensure they are evenly distributed throughout the tube. Also, do not use SPRI beads if they have been kept outside at room temperature for more than twenty-four hours.

CAUTION the ADT product is the supernatant in the step above and it is crucial to not discard!

7.4 To remove excess primers from the ADT, cDNA, and gDNA products we performed exonuclease treatment. A 25µL exonuclease reaction was set up according to Table 4, using 18µL of cDNA product or 18µL of diluted ADT product (2ng/µL) from the previous step. The exonuclease reaction was incubated at 37°C for 5 minutes, then deactivated at 85°C for 5 minutes, and hold at 4°C. A separate 20µL exonuclease reaction was set up with water according to Table 4, using 10µL of diluted gDNA product (2ng/µL) from the previous step. The gDNA exonuclease reaction was incubated at 37°C for 30 minutes, then deactivated at 85°C for 5 minutes, and hold at 4°C. After exonuclease treatment, a second SPRI clean-up was performed at 0.6X for cDNA and 1.2X for gDNA treated products.

NOTE an additional SPRI cleanup of the cDNA product after exonuclease treatment improves overall purity.

CAUTION we highly recommend performing a SPRI cleanup of the gDNA product after exonuclease treatment. Lack of or poor SPRI cleanup will interfere with final library amplification using P5 and P7 adapters.

CRITICAL: Measure the Qubit concentration after the second SPRI clean-up. If the concentration of the cDNA is below 0.4 ng/µL, then it is not recommended to proceed with the rest of library preparation because there is too little material to be amplified. Post-second PCR, the gDNA should be over 10 ng/µL.

7.5 cDNA concentration was measured on Qubit and 2ng was used for tagmentation with the Nextera XT kit per manufacturer's instructions in 25µL final reaction volume.

NOTE the amount of cDNA used for tagmentation here exceeds the recommended levels. If the cDNA library is overly fragmented, it may be helpful to slightly adjust the volume of the amplicon tagment mix (ATM) while still maintaining a final reaction volume of 25µL. After cDNA tagmentation we recommend immediately proceeding to the P5 and P7 PCR amplification step.

7.6 For the final amplification of pooled samples, 2ng of tagmented cDNA and 5ng of genomic DNA and ADT samples underwent a final amplification step using custom Illumina-compatible P5 and P7 primers. The PCR reaction mix was prepared according to Table 5, and the amplification program was set up as outlined in Table 5.1. All samples were subsequently cleaned using SPRI beads at 0.8X for cDNA, 1.2X for gDNA, and 1.6X for ADT prior to sequencing.

NOTE poor gDNA amplification after adding P5 and P7 adapters may result from inadequate SPRI cleanup of the gDNA product following exonuclease treatment. Make sure to measure the concentrations of the P5/P7 products post-SPRI cleanup. Measure before SPRI cleanup as well if troubleshooting.

CAUTION: If one of the exonuclease enzymes is not removed post-treatment, as would be the case with an inadequate or omitted SPRI cleanup, then the P5-P7 amplification would lead to a poor library concentration and overall quality. This is because the exonuclease enzyme is not compatible with the P5-P7 amplification mixture.

7.7 DNA products were sequenced on a MiSeq using 300 cycles, with Read 2 typically containing the critical information needed to identify CRISPR-induced mutations. A high PhiX spike-in of approximately 25% was used, as all reads were highly similar. Read 1 contained the well barcode. In contrast, RNA and ADT products were sequenced on the NextSeq or NovaSeq platforms, where Read 1 needed to be at least 26 cycles, and Read 2 was set as long as possible, preferably

exceeding 15 cycles, to maximize data acquisition. We spiked the ADT library into cDNA libraries at 1-5% percent by concentration of the final library.

ANALYSIS SECTION

PROCESSING and ALIGNMENT

I7 and I5 raw FASTQ files were merged across different lanes and runs. Transcriptomic reads were aligned to the human genome (GRCh38 2020-A) using STARsolo (version 2.7.6a)[3]. For all experiments, cells were filtered based on the following criteria: a minimum of 500 genes, 1000 unique molecular identifiers (UMIs), and less than 10% mitochondrial reads. Counts were log-transformed using the formula $ln(CP10K + 1)$. The transformed counts were then z-score scaled. Dimensionality reduction was performed using principal component analysis (PCA) on the top 3000 variable genes, which were selected using variance stabilizing transformation (VST). This was followed by batch correction with Harmony[4] (version 1.2.0) using $\theta = 1$ and otherwise default parameters. Visualization of harmonized principal components was conducted using uniform manifold approximation and projection (UMAP) from the uwot[4] package (version 0.1.16) with default parameters. For index flow cytometry analysis, raw fluorescence values were exported from index sorting as CSV files and subsequently $log10$ normalized. Kallisto-kite (version 0.27.3) was utilized to build an antibody-derived tag (ADT) reference transcriptome corresponding to the ADT panel (BioLegend TotalSeq A) used in the study and to align ADT reads[5,6]. UMI counts were centered log-ratio (CLR) normalized and z-score scaled. Visualization was conducted with PCA on CLR normalized and scaled variable ADTs, followed by plate correction using Harmony with the same parameters as used in the RNA processing. UMAP dimension-reduction was applied to the harmonized principal components.

Genomic DNA reads were demultiplexed based on cellular barcodes, and amplicons were aligned to the reference amplicon using CRISPResso2 (version 2.2.9)[7]. Allele usage statistics were calculated from CRISPResso2 outputs. For all DNA analyses, cells were filtered based on a minimum number of reads (20-40) and minimum allele frequencies (12-25%) on per plate basis, adjusted per experiment. If only one allele was recovered post-filtering, the cell was considered homozygous. Clustering of DNA editing in the DQB1 experiment was performed using supervised k-means clustering with $n = 3$.

MODELING GENE and PROTEIN EXPRESSION

Associations of gene and ADT expression to dosage and conditions was performed with negative binomial generalized linear modeling. In cell lines we filtered for genes with non-zero expression in greater than 30% of cells and a mean of 2 reads per cell. We modeled counts of gene or ADTs as function of cell-level covariates and fitted a null model 1 using:

$$(null\ model\ 1)\ \ ln\ ln\ (E_i)\ = 1 + \beta_N log_{10}(X_{i,N}) + \beta_p X_{i,p}$$

we then compared the above null model 1 to genotype or condition full models using:

$$(model\ 1.1)\ \ ln\ ln\ (E_i)\ = 1 + \beta_N log_{10}(X_{i,N}) + \beta_p X_{i,p} + \beta_G X_{i,G}$$
$$(model\ 1.2)\ \ ln\ ln\ (E_i)\ = 1 + \beta_N log_{10}(X_{i,N}) + \beta_p X_{i,p} + \beta_C X_{i,C}$$

Where E is the expected count of a gene or ADT in cell i and all other β's represent fixed effects as indicated (N = total number of expression gene counts or ADTs, p = plate identity, d = donor, G = genotype dosage of cell, C = culture condition of cell that is either CRISPR or control) for covariates in cell i. We used genomic DNA data from CRAFTseq to determine the genotype of

each cell after editing. Specifically, for the DQB1 editing experiment in cell lines we fitted a full model for the deletion size (S) that was calculated for a given gene in a cell:

$$(model\ 1.3)\ \ ln\ ln\ (E_i)\ = 1 + \beta_N log_{10}(X_{i,N}) + \beta_p X_{i,p} + \beta_S log_2(X_{i,S} + 1)$$

In primary human cells we filtered genes for non-zero expression in greater than 15% of cells. Using a similar negative binomial generalized linear modeling we fitted models with either fixed or random effects for primary T cells data. Specifically, for the PTPRC editing experiments in primary human cells we fitted a null model 2 using:

$$(null\ model\ 2)\ \ ln\ ln\ (E_i)\ = 1 + \beta_N log_{10}(X_{i,N}) + \beta_p X_{i,p} + \beta_d X_{i,d}$$

we then compared the above null model 2 using genotype or condition full models:

$$(model\ 2.1)\ \ ln\ ln\ (E_i)\ = 1 + \beta_N log_{10}(X_{i,N}) + \beta_p X_{i,p} + \beta_d X_{i,d} + \beta_G X_{i,G}$$
$$(model\ 2.2)\ \ ln\ ln\ (E_i)\ = 1 + \beta_N log_{10}(X_{i,N}) + \beta_p X_{i,p} + \beta_d X_{i,d} + \beta_C X_{i,C}$$

For the IL2RA editing experiments in primary cells, we used random effects to account for the correlation of cells within plates and donors. Specifically, we modeled a cell-plate random effect, $k_{i,p}$, and a cell-donor random effect, $\alpha_{i,d}$. Here we fitted a null model 3 using:

$$(null\ model\ 3)\ \ ln\ ln\ (E_i)\ = 1 + \beta_N log_{10}(X_{i,N}) + k_{i,p} + \alpha_{i,d}$$

we then compared the above null model 3 using genotype or condition full models:

$$(model\ 3.1)\ \ ln\ ln\ (E_i)\ = 1 + \beta_N log_{10}(X_{i,N}) + k_{i,p} + \alpha_{i,d} + \beta_G X_{i,G}$$
$$(model\ 3.2)\ \ ln\ ln\ (E_i)\ = 1 + \beta_N log_{10}(X_{i,N}) + k_{i,p} + \alpha_{i,d} + \beta_C X_{i,C}$$

We used the glm.nb function in MASS (R package, version 7.3-60) to model fixed effects and glmer.nb function in lme4 to model both random and fixed effects. To determine the significance of these models, we used a likelihood ratio test (LRT) comparing the null and full models and calculated a p-value for the test statistic against the Chi-squared distribution with one degree of freedom. When indicated, we corrected for multiple hypothesis testing by Benjamini-Hochberg correction (FDR).

DEMULTIPLEXING DONORS

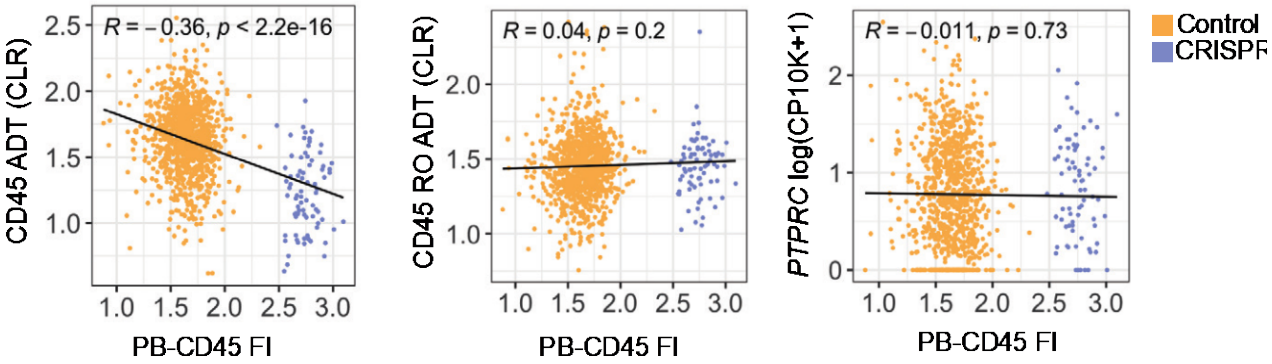
Samples were genotyped on the Illumina Multi-Ethnic Genotyping Array (MEGA) or Global Screening Array (GSA). Genotyping quality control, haplotype phasing, and whole-genome imputation were performed for MEGA as previously described[9]. Analogous steps were undertaken for GSA, with further genotyping quality control measures including Hardy-Weinberg equilibrium $p_{HWE} > 1 \times 10^{-10}$, sample relatedness (IBD) < 0.9, and a sample heterozygosity rate > 0.217. All imputed variants were filtered based on minor allele frequency (MAF) > 0.01 and imputation accuracy (Rsq) > 0.9 in both datasets. Variants were further pruned to retain only those in the intersection of filtered variants between the two genotyping arrays. Additionally, non-biallelic variants and variants within sex chromosomes or non-exonic regions were excluded. Variant coordinates were converted from GRCh37 to GRCh38 using CrossMap[8] (version 0.6.5) to match the reference assembly used for transcriptome aligned. Demuxlet (version 1.0) was run with default parameters to predict donors from genotypes[9]. Cell donor identity was assigned using

the output column “SNG.1ST,” corresponding to the most likely singlet identity. Cells with ambiguous donor assignments were excluded.

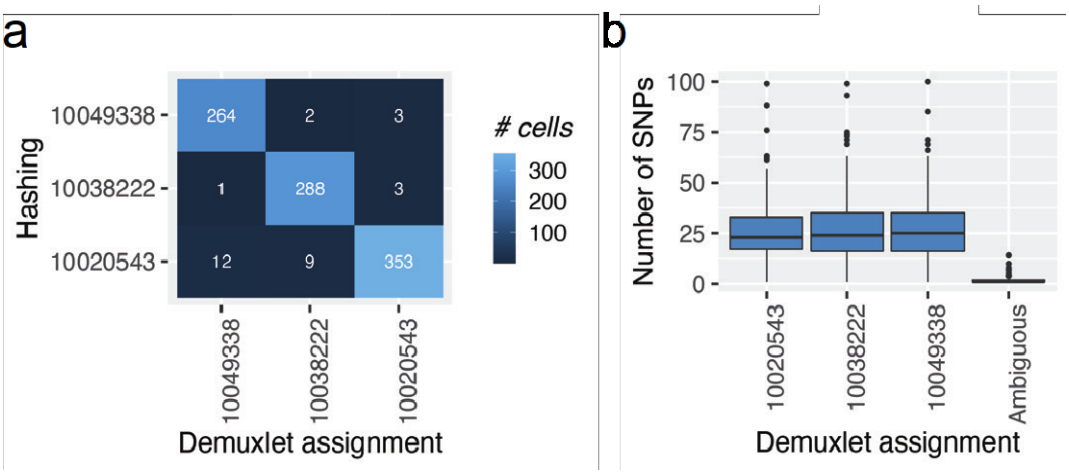
GENE SET ENRICHMENT

We performed gene set enrichment analysis using the fgsea package (version 1.20.0). As input, we utilized the results of the linear model that predicted gene expression based on genotype dosage from the CD45 experiment. In this analysis, we included all genes to ensure a comprehensive assessment of enrichment. Gene ranks were defined according to the sign of the effect size of genotype dosage, which provides directionality, multiplied by the negative $log10$ of the p-value. This method allows us to prioritize genes based on both their significance and the direction of their effects. For our pathway analysis, we incorporated the MSigDB immunological signatures (C7) pathways[10,11] that are specific to CD4 T-cells. Following this, we filtered the pathways to include only those that contained between 15-500 genes, ensuring that our analysis focused on pathways of sufficient size to be biologically relevant while avoiding pathways that were too small to yield meaningful insights.

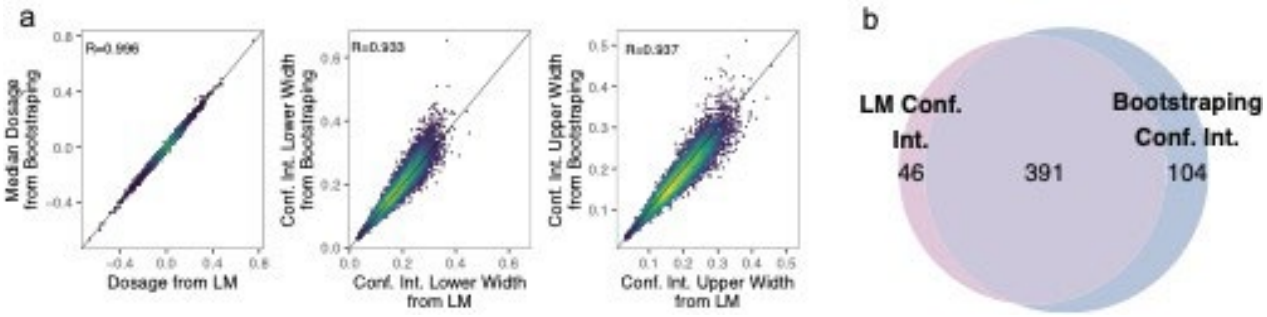
Supplementary Figures:



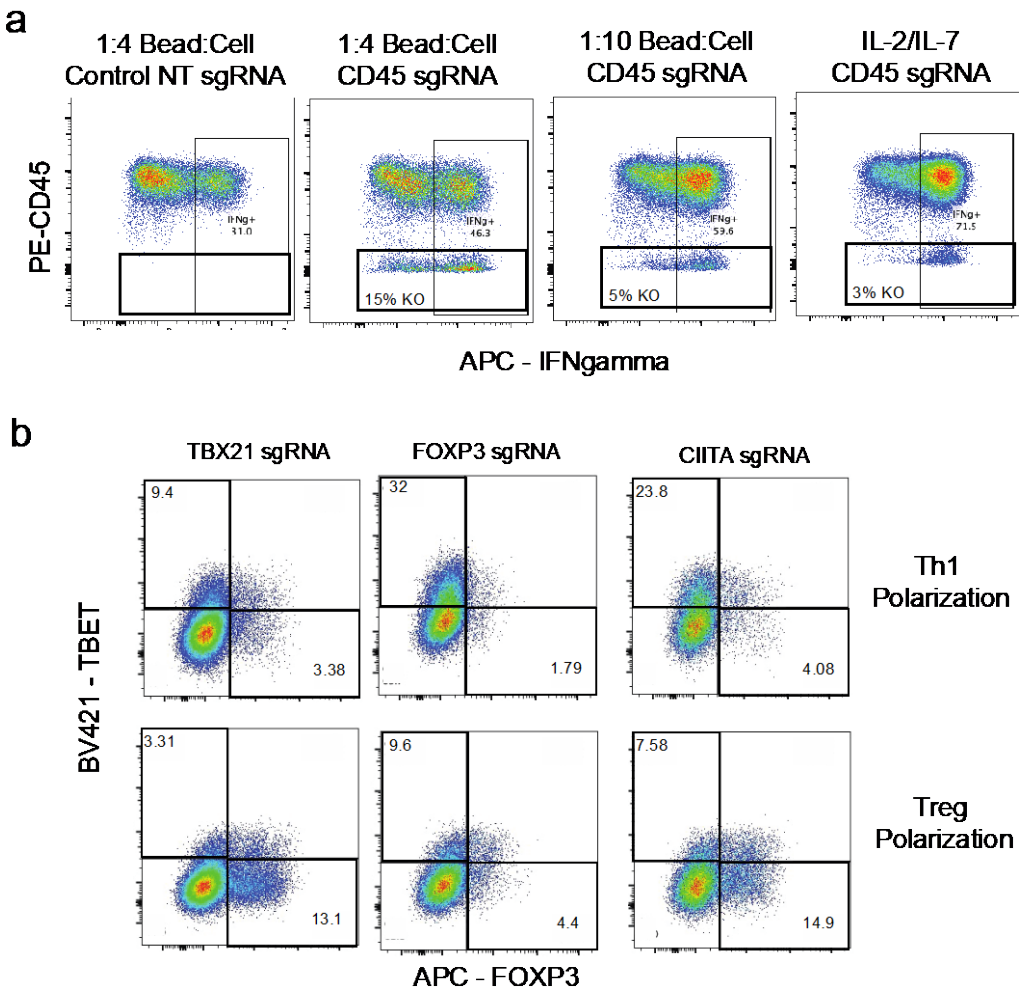
Supplementary Figure 1: Comparison of ADT normalized counts to flow cytometry intensity staining of CD45. Sequential staining with flow antibodies interferes with overall ADT counts. Data is derived from the FBXO11 experiment with no expected changes in CD45. R and p values are derived from linear regression.



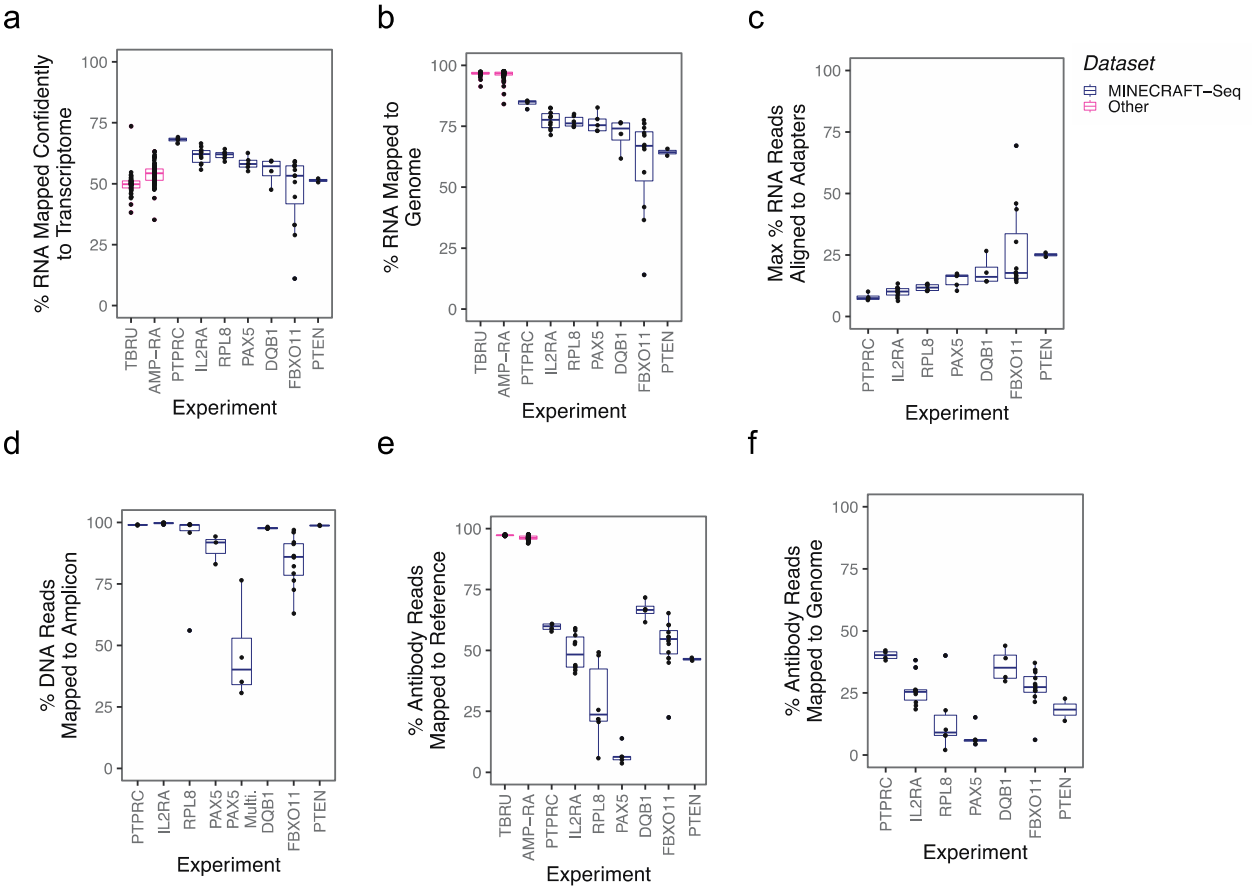
Supplementary Figure 2: Index sorting based hashing is concordant with genotype-based demultiplexing from single cell RNAseq data. (a) Concordance of genotypes and hashing identify of the three individuals from Figure 4. (b) Number of SNPs identified in RNA sequencing data in each cell per individual (N=601 cells). Boxes show the interquartile range (IQR), whiskers show the 1.5*IQR.



Supplementary Figure 3. Comparison of negative-binomial regression and bootstrap-derived summary statistics. (A) Left - Scatterplot showing concordance between effect size predictions of differentially expressed genes from the linear model and bootstrapping. Solid line is the identity line. Middle - Scatterplot showing concordance between the lower width of the confidence intervals (median to lower limit) from the linear model and bootstrapping. Right - Scatterplot showing concordance between the upper width of the confidence intervals (upper limit to median) from the linear model and bootstrapping. **(B)** Overlap of genes passing nominal significance through the linear model and bootstrapping.



Supplementary Figure 4. Titration of stimulation conditions and confirmation of polarization in naive CD4 T cells. a) Alternative bead:cell ratios of anti-CD3/CD28 microbeads were tested to determine the effects of stimulation on polarization and CRISPR-Cas editing of CD45. Naive CD4 T cells were isolated, stimulated for 1 day, and then polarized towards Th1. Intracellular IFNgamma production was measured by flow cytometry following four hours of re-stimulation of PMA and Ionomycin in the presence of monensin on day 7. b) Polarization of naive CD4 T cells was confirmed by intracellular flow cytometry of FOXP3 or TBET. Cells were nucleofected with CRISPR-Cas machinery targeting TBX21, FOXP3, or CIITA on day 1 and cells analyzed on day 7.

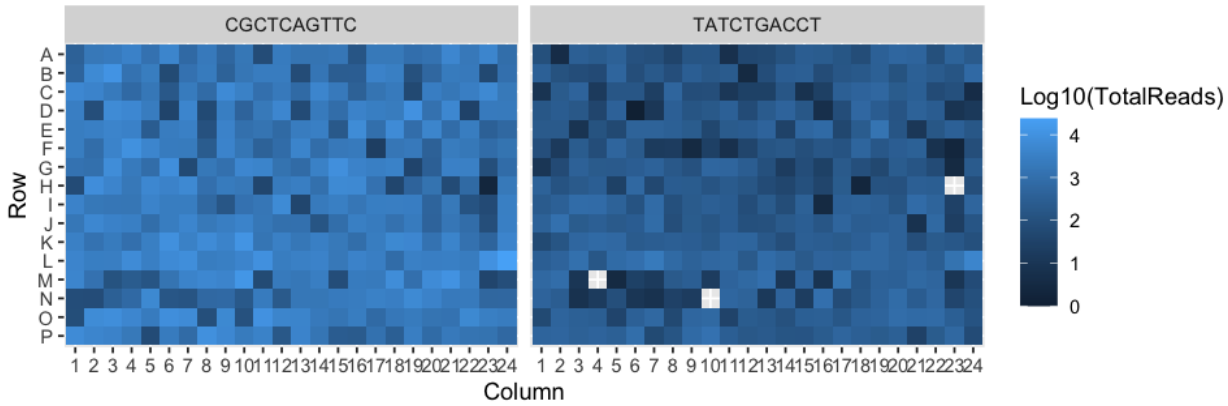


Supplementary Figure 5. QC metrics for all plates in this study and comparison with 10x data. Boxplots displaying an alignment statistic by CRAFTseq experiment type (blue) or external 10X dataset (pink). TBRU, AMP-RA, PTPRC, and IL2RA reflect experiments involving human cells. Other reported CRAFTseq experiments include cancer cell lines. Each point represents a library for external 10X datasets or a plate in CRAFTseq. Boxes show the interquartile range (IQR), whiskers show the 1.5*IQR. **(A)** Boxes show percentage of RNA reads confidently aligning to the transcriptome. **(B)** Boxes show percentage of RNA reads aligning to the genome. **(C)** For CRAFTseq experiments only, boxes show percentage of RNA reads aligning to Illumina universal adapters via FASTQC. **(D)** For CRAFTseq only, boxes show percentage of DNA reads aligning to the reference amplicon. PAX5 Multi. refers to plates where multiplexed editing occurred, whereas PAX5 refers to plates where regions of interest were edited separately (not multiplexed). **(E)** Boxes show percentage of ADT reads aligning to the antibody reference. **(F)** For CRAFTseq experiments only, boxes show percentage of ADT reads aligning to the genome.

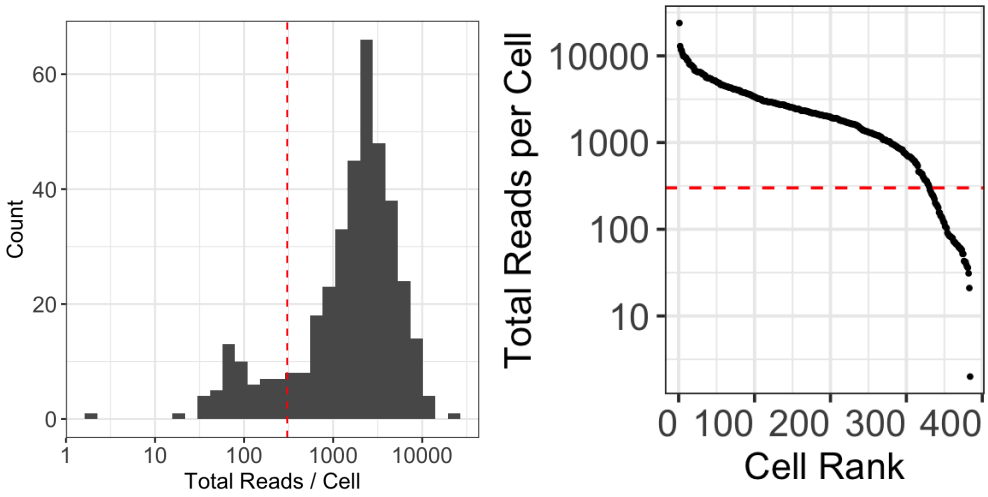
Supplementary Figure 6: Example of final constructed gDNA library.

```
# A tibble: 1,482,676 × 14
   Barcode_DNA Well_ID Aligned_Sequence        Reference_Sequence Reference_Name
   <chr>       <chr>   <chr>                   <chr>              <chr>
 1 AGAGCACTAG  A1      ----------------------… CAAAATCCCCGAGGCAT… Reference
 2 AGAGCACTAG  A1      ----------------------… CAAAATCCCCGAGGCAT… Reference
 3 AGAGCACTAG  A1      ----------------------… CAAAATCCCCGAGGCAT… Reference
 4 AGAGCACTAG  A1      ----------------------… CAAAATCCCCGAGGCAT… Reference
 5 AGAGCACTAG  A1      ----------------------… CAAAATCCCCGAGGCAT… Reference
 6 AGAGCACTAG  A1      ----------------------… CAAAATCCCCGAGGCAT… Reference
 7 AGAGCACTAG  A1      ----------------------… CAAAATCCCCGAGGCAT… Reference
 8 AGAGCACTAG  A1      ----------------------… CAAAATCCCCGAGGCAT… Reference
 9 AGAGCACTAG  A1      ----------------------… CAAAATCCCCGAGGCAT… Reference
10 AGAGCACTAG  A1      ----------------------… CAAAATCCCCGAGGCAT… Reference
# ℹ 1,482,666 more rows
# ℹ 9 more variables: Read_Status <chr>, n_deleted <dbl>, n_inserted <dbl>,
#   n_mutated <dbl>, `#Reads` <dbl>, `%Reads` <dbl>, PlateLabel <chr>,
#   MainCondition <chr>, Reference <chr>
```

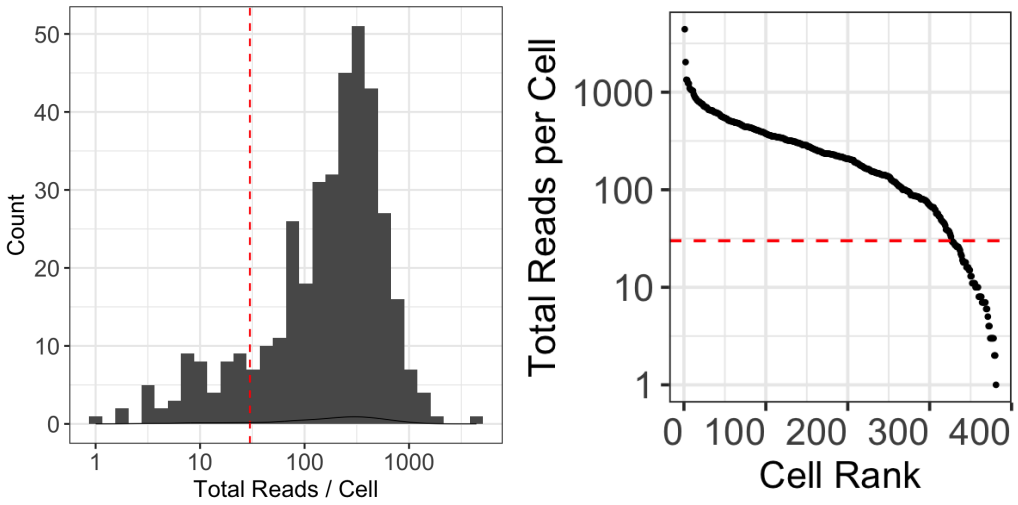
Supplementary Figure 7: CRISPResso example output for processing.



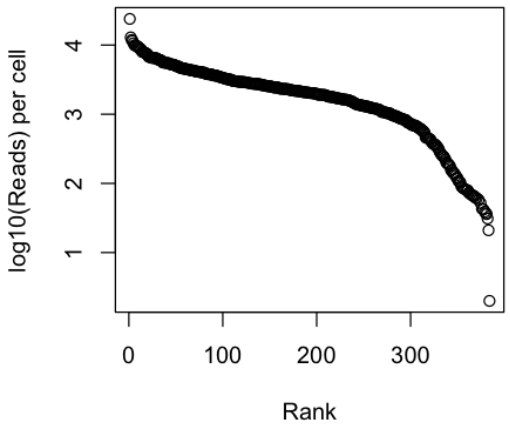
Supplementary Figure 8: Plate level reads recovered and aligned per well.



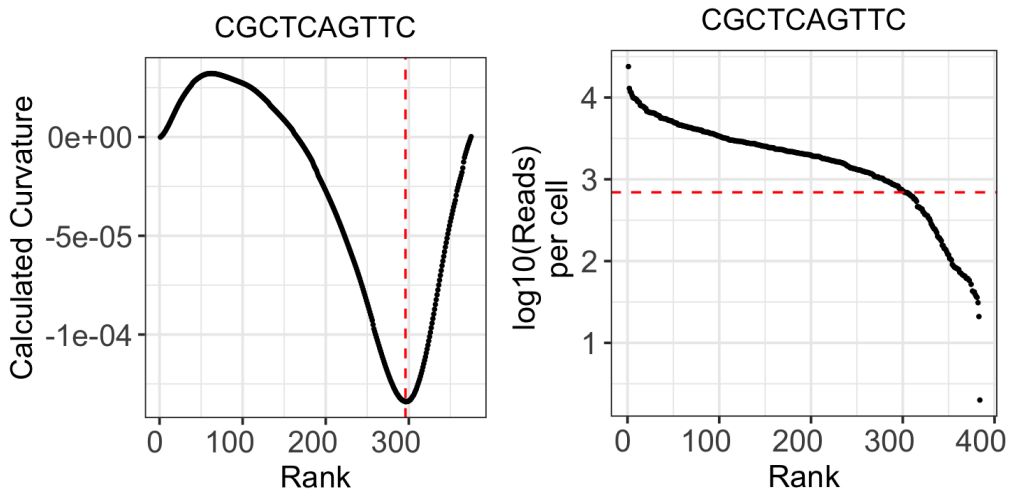
Supplementary Figure 9: Percentage of total reads per cell (left) and the same data ranked (right) for plate 1. Dotted red line represents the proposed manually chosen cutoff.



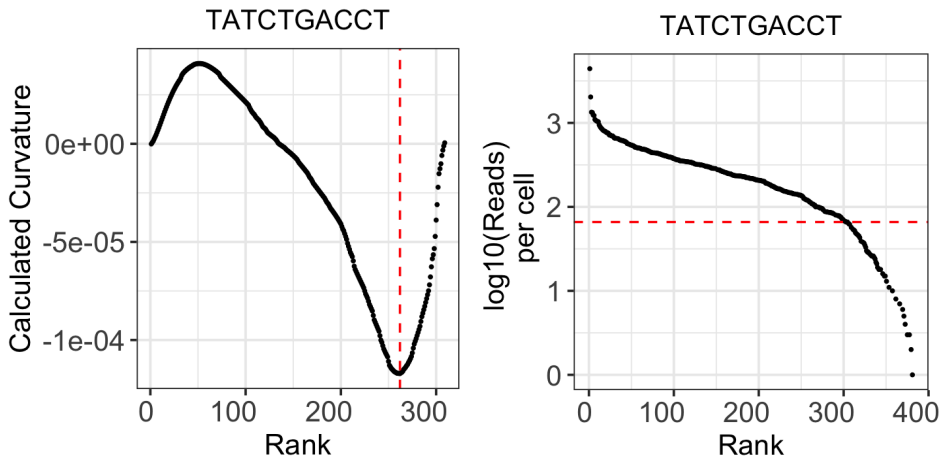
Supplementary Figure 10: Percentage of total reads per cell (left) and the same data ranked (right) for plate 2. Dotted red line represents the proposed manually chosen cutoff.



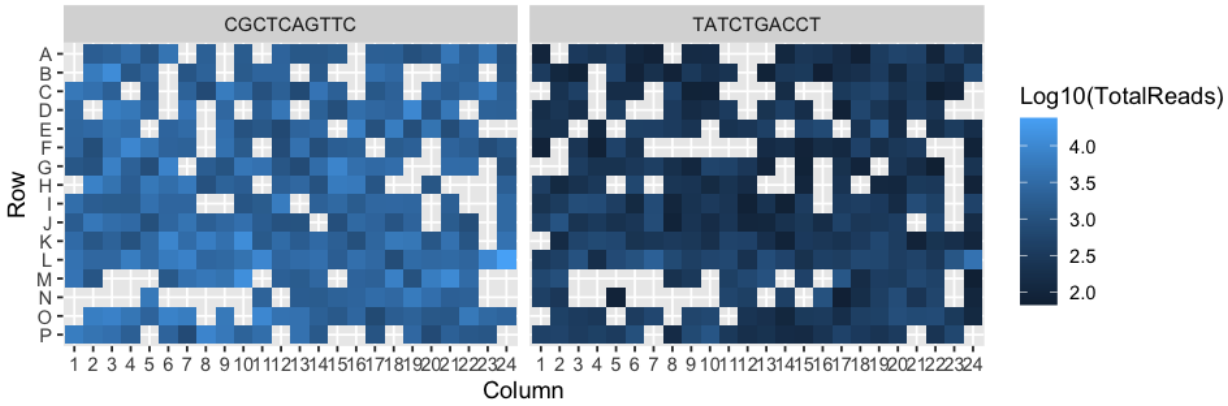
Supplementary Figure 11: Log10 total number of reads per cell for plate 1.



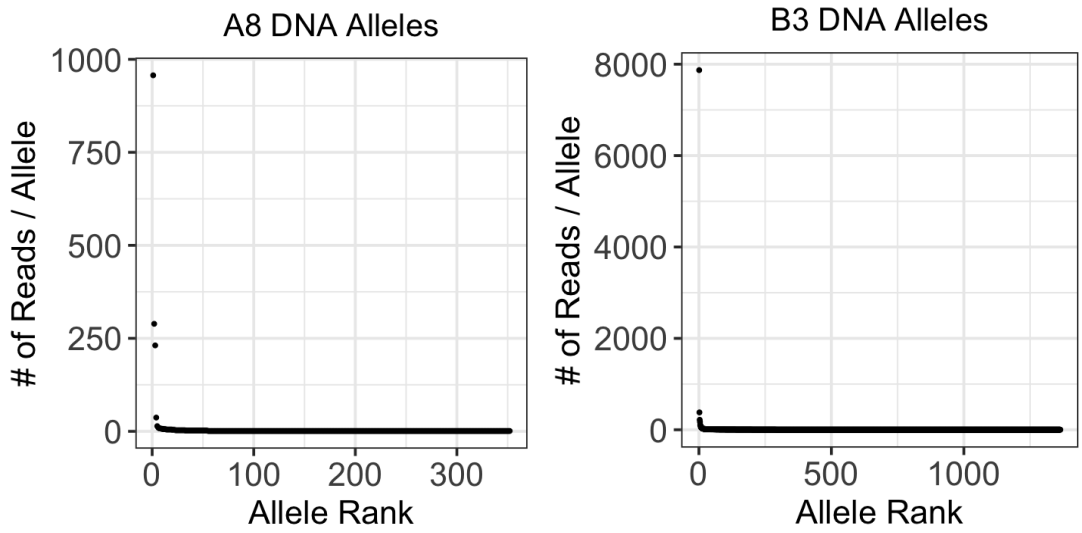
Supplementary Figure 12: Curvature plot (left) of data from plate 1(CGCTCAGTTC) and the log10 total reads per cell plotted ranked. The red line indicates the minimum calculated from the curvature and used for filtering.



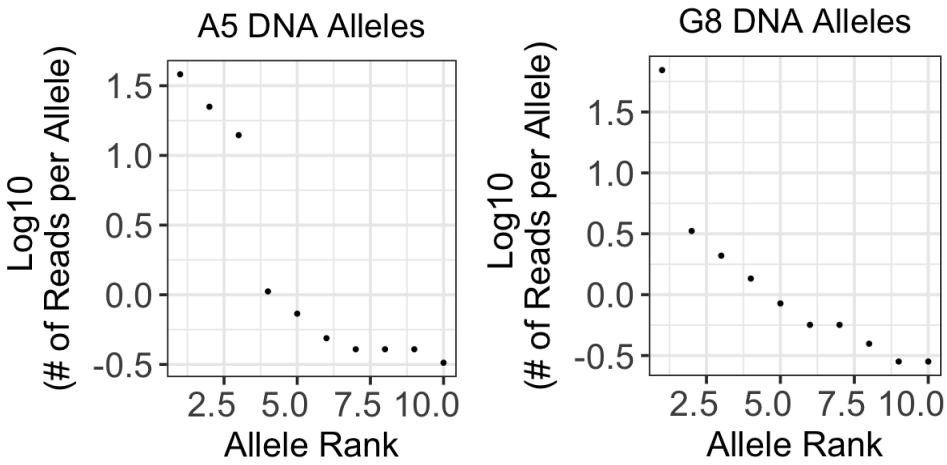
Supplementary Figure 13: Curvature plot (left) of data from plate 2(CGCTCAGTTC) and the log10 total reads per cell plotted ranked. The red line indicates the minimum calculated from the curvature and used for filtering.



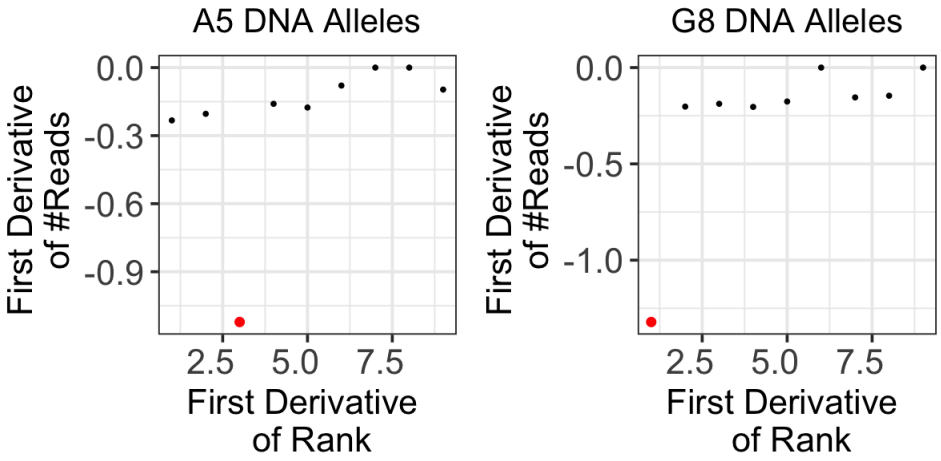
Supplementary Figure 14: Plate level reads recovered and aligned per well with filtering of wells/cells from the above analysis.



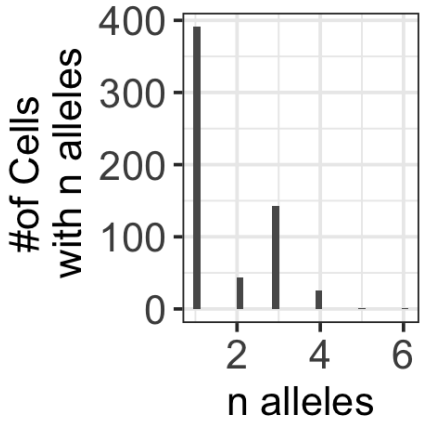
Supplementary Figure 15: Total number of reads per allele recovered plotted for two individual cells.



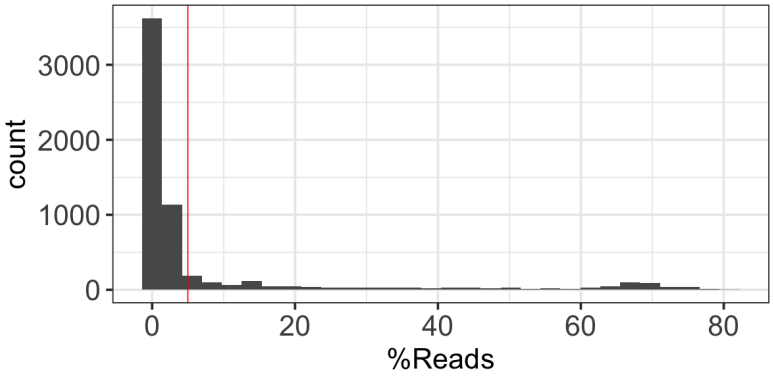
Supplementary Figure 16: Log 10 total number of reads per allele recovered plotted for two individual cells.



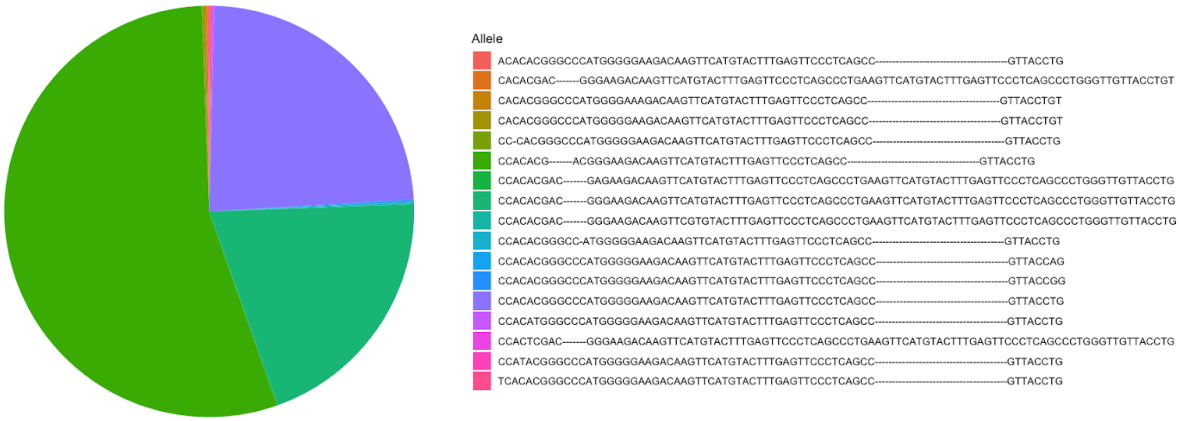
Supplementary Figure 17: First derivative of log 10 total number of reads per allele recovered plotted for two individual cells.



Supplementary Figure 18: Total number of alleles after filtering recovered per cell across both plates.



Supplementary Figure 19: Percentage of reads per cell assigned to a unique and filtered allele. This is calculated for each individual cell.

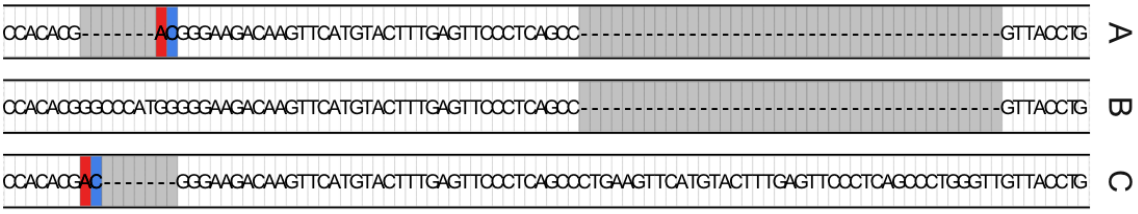


Supplementary Figure 20: Pie chart of filtered alleles recovered across both plates.

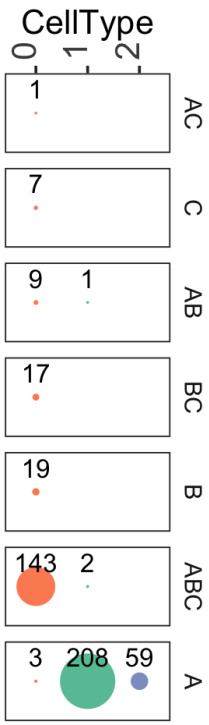
A tibble: 17 × 2

.	n
<chr>	<int>
ACACACGGGCCCATGGGGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCC----------GTTACCTG	1
CACACGAC-------GGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCCCTGAAGTTCATGTACTTTGAGTTCCCTCAGCCCTGGGTTGTTACCTGT	1
CACACGGGCCCATGGGGAAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCC----------GTTACCTGT	1
CACACGGGCCCATGGGGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCC----------GTTACCTGT	2
CC-CACGGGCCCATGGGGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCC----------GTTACCTG	1
CCACACG-------ACGGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCC----------GTTACCTG	551
CCACACGAC-------GAGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCCCTGAAGTTCATGTACTTTGAGTTCCCTCAGCCCTGGGTTGTTACCTG	1
CCACACGAC-------GGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCCCTGAAGTTCATGTACTTTGAGTTCCCTCAGCCCTGGGTTGTTACCTG	202
CCACACGAC-------GGGAAGACAAGTTCGTGTACTTTGAGTTCCCTCAGCCCTGAAGTTCATGTACTTTGAGTTCCCTCAGCCCTGGGTTGTTACCTG	1
CCACACGGGCC-ATGGGGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCC----------GTTACCTG	1
CCACACGGGCCCATGGGGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCC----------GTTACCAG	1
CCACACGGGCCCATGGGGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCC----------GTTACCGG	1
CCACACGGGCCCATGGGGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCC----------GTTACCTG	238
CCACATGGGCCCATGGGGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCC----------GTTACCTG	1
CCACTCGAC-------GGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCCCTGAAGTTCATGTACTTTGAGTTCCCTCAGCCCTGGGTTGTTACCTG	1
CCATACGGGCCCATGGGGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCC----------GTTACCTG	1
TCACACGGGCCCATGGGGGAAGACAAGTTCATGTACTTTGAGTTCCCTCAGCC----------GTTACCTG	1

Supplementary Figure 21: Table of filtered alleles recovered across both plates. N represents the number of unique occurances of that allele across the experiment.



Supplementary Figure 22: Final representative alleles recovered in this experiment after all filtering steps.



Supplementary Figure 23: Distribution of combinations of alleles in this experiment plotted by cell type as assigned with ADT clustering. See figure 1 and main text for further details.

Supplementary Table captions.

Supplementary Table 1-23 captions.

Supplementary Table 1. Antibody-derived tags (ADT) panel with unique barcodes used.
Supplementary Table 2. Capture plate with unique well barcode sequences.
Supplementary Table 3. OligoDT plate with unique well barcode sequences.
Supplementary Table 4. Specific P7 adapter sequences for ADT, DNA, and RNA applications.
Supplementary Table 5. Additional primer or adapter specific sequences.
Supplementary Table 6. CRISPR editing protospacer.
Supplementary Table 7. Primers used in the study for reverse transcription, genomic DNA amplification, and library generation.

Supplementary Table 8. Antibodies.
Supplementary Table 9. ADT DGE PTEN Experiment.
Supplementary Table 10. RNA DGE PTEN Experiment.
Supplementary Table 11. PAX5 Interaction.
Supplementary Table 12. PAX5 Interaction Single Genes.
Supplementary Table 13. Alignment statistic per experiment.
Supplementary Table 14. Number of cells passing QC per modality per plate.
Supplementary Table 15. Percentage of cells passing QC per modality per plate.
Supplementary Table 16. Concentration (conc) and volume (vol) of each reagent used to set up the lysis plates.

Supplementary Table 17. RT and initial genomic DNA amplification reaction components.
Supplementary Table 18: PCR cycler setup for RT and initial genomic DNA amplification.
Supplementary Table 19: Nested genomic DNA amplification reaction components.
Supplementary Table 20: PCR cycler setup for nested genomic DNA amplification.
Supplementary Table 21: cDNA, ADT, and gDNA exonuclease reaction setup.
Supplementary Table 22: cDNA, ADT, and gDNA P5-P7 amplification reaction setup.
Supplementary Table 23: PCR cycler setup for P5-P7 PCR.

References

1 Chen, P. J. *et al.* Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635-5652 e5629 (2021). https://doi.org/10.1016/j.cell.2021.09.018

2 Gutierrez-Arcelus, M. *et al.* Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. *Nat Genet* **52**, 247-253 (2020). https://doi.org/10.1038/s41588-020-0579-4

3 Kaminow, B., Yunusov, D. & Dobin, A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. *bioRxiv*, 2021.2005.2005.442755 (2021). https://doi.org/10.1101/2021.05.05.442755

4 Melville, J. *uwot: The Uniform Manifold Approximation and Projection (UMAP) Method for Dimensionality Reduction* (2023).

5 Booeshaghi, A. S., Min, K. H. J., Gehring, J. & Pachter, L. Quantifying orthogonal barcodes for sequence census assays. *Bioinform Adv* **4**, vbad181 (2024). https://doi.org/10.1093/bioadv/vbad181

6 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527 (2016). https://doi.org/10.1038/nbt.3519

7 Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* **37**, 224-226 (2019). https://doi.org/10.1038/s41587-019-0032-3

8 Zhao, H. *et al.* CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006-1007 (2014). https://doi.org/10.1093/bioinformatics/btt730

9 Kang, H. M. *et al.* Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotechnol* **36**, 89-94 (2018). https://doi.org/10.1038/nbt.4042

10 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740 (2011). https://doi.org/10.1093/bioinformatics/btr260

11 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005). https://doi.org/10.1073/pnas.0506580102