nature methods

Article

https://doi.org/10.1038/s41592-022-01705-x

Efficient combinatorial targeting of RNA transcripts in single cells with Cas13 RNA Perturb-seq

In the format provided by the authors and unedited

Supplementary Information

Efficient combinatorial targeting of RNA transcripts in single cells with Cas13 RNA Perturb-seq

Hans-Hermann Wessels[1,2,4], Alejandro Méndez-Mancilla[1,2,4], Yuhan Hao[1,2], Efthymia Papalexi[1,2], William M. Mauck III[1,2], Lu Lu[1,2], John A. Morris[1,2], Eleni P. Mimitou[1,3], Peter Smibert[1,3],
Neville E. Sanjana[1,2,†,*] & Rahul Satija[1,2,‡,*]

1. New York Genome Center, New York, NY, USA.
2. Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY, USA
3. Technology Innovation Lab, New York Genome Center, New York, NY, USA
4. These authors contributed equally

[*] These authors share correspondence
[†] E-mail: neville@sanjanalab.org
[‡] E-mail: rsatija@nygenome.org

SUPPLEMENTARY METHODS

Cell culture and monoclonal cell line generation

HEK293FT cells were acquired from ThermoFisher (R70007), NIH/3T3, and THP1 cells were obtained from ATCC (CRL-1658, TIB-202). Cell line authentication was performed by the vendor. HEK293FT and NIH/3T3 cells were maintained at 37°C with 5% CO_2 in DMEM with high glucose and stabilized L-glutamine (Caisson DML23) supplemented with 10% fetal bovine serum (Serum Plus II Sigma-Aldrich 14009C) and no antibiotics. THP1 cells were grown in RPMI plus 10% FBS at 37°C with 5% CO_2 and no antibiotics. Doxycycline-inducible *Rfx*Cas13d-NLS HEK293FT, THP1, and NIH/3T3 cells (Addgene #138149), as well as doxycycline-inducible nuclease-inactive *Rfx*dCas13d-NLS HEK293FT, have been generated as described before [11]. We sorted individual suspension THP1 cells using a flow cytometer (SONY SH800) to select single clonal lines. Each THP1-Cas13d clone was evaluated to provide homogenously strong CD46 knockdown using lentiviral integration of a single gRNA, puromycin selection, and flow cytometry.

Monoclonal Cas9-effector protein-expressing HEK293FT and THP1 cell lines were generated using dilution plating as described above [11]. For each Cas9-effector cell line, we evaluated multiple clones in their ability to provide homogenous and complete CD55 knockout or knockdown using lentiviral integration of a single sgRNA expressing cassette, puromycin selection, and flow cytometry. The cloning of KRAB-dCas9 and KRAB-dCas9-MeCP2 constructs have been described before [18]. For Cas9-nuclease, we previously cloned lentiCas9-Blast (Addgene #52962) [48]. All monoclonal CRISPR Cas effector expressing cells were maintained using 5µg/mL Blasticidin S (ThermoFisher A1113903).

Cloning of individual Cas13 and Cas9 guide RNAs

Cas13 guide RNA cloning was done as described previously [48]. Specifically, we cloned gRNA or barcode oligos into pLentiRNAGuide_001 (Addgene #138150). Guide RNA constructs with reverse transcription handles on the 3' end of the spacer sequence were synthesized and cloned together. Guide RNA constructs using CRISPR arrays used in data presented in **Fig. 2** were cloned stepwise by introducing a guide along with a direct repeat and reconstituted BsmBI restriction sites to allow for serial cloning and extension of CRISPR arrays.

pLentiRNAGuide_003 was generated by removing all LguI sites from the pLentiRNAGuide_002 plasmid, introducing the capture sequence 1 (CS1) (and *evopre*Q_1 sequence) 5' to the terminator sequence, and replacing the puromycin gene with GFP-P2A-puromycin cassette.

For direct capture Perturb-seq, we cloned all sgRNA-expressing constructs stepwise. First, we cloned dual sgRNA oligos into pJR85 and pJR89 as described before [6]. In addition, we cloned single sgRNAs into custom sgRNA expressing plasmids. We used either a human U6 (hU6) promoter driving the sgRNA scaffold described in the 10x Genomics manual CG000184 Rev C (internal CS1) (lentiGuideFB-Puro-A) or using the sgRNA cassette from pJR73 [6] driven by a bovine U6 (bU6) promoter (lentiGuideFB-Puro-B). The bU6 and sgRNA cassette from lentiGuideFB-Puro-B was then subcloned into the dual sgRNA plasmid to generate triple sgRNA expressing plasmids (see also **Extended Data Fig. 5b**). For direct capture Perturb-seq, we used lentiGuideFB-Puro-A single sgRNA constructs for pools expecting one sgRNA, pJR85/pJR89 constructs for pools expecting two sgRNAs, and pJR85/pJR89/lentiGuideFB-Puro-B constructs for pools expecting three sgRNAs. All constructs were confirmed by Sanger sequencing.

We simplified dual sgRNA cloning into pJR85/pJR89 for sgRNAs shown in **Fig. 4d** to make it a single step. Here, we provided two sets of annealed single-sgRNA oligos with the respective overhangs together with the BsmBI-digested pJR89-derived insert (sgRNA scaffold + hU6) directly to the first step of annealed oligo cloning as described before [6].

Cas13d guide vector that carry RNA stabilization sequence elements were cloned as described before [11]. All primers used for molecular cloning and guide sequences are shown in **Extended Data Table 2**. [pLentiRNAGuide_003 carrying an *evopre*Q_1 sequence, and Cas9 sgRNA vectors lentiGuideFB-Puro-A and lentiGuideFB-Puro-B] have been deposited on Addgene (#192505-#192507).

Virus production and viral transductions

For virus production of individual gRNA and sgRNA expressing plasmids, we seeded $1x10^6$ HEK293FT cells per well in 6-well format 12-18 hours before transfection. Per transfection reaction we used 7.5ul of 1 mg/mL polyethyleneimine (PEI) and 2.325 µg of total plasmid DNA (825 ng psPAX2: Addgene #12260; 550 ng pMD2.G: Addgene #12259; 1000 ng of gRNA/sgRNA expressing plasmid). Six to eight hours post-transfection, the medium was exchanged for 2 ml of DMEM + 10% FBS containing 1% bovine serum albumin (BSA). Viral supernatants were collected after additional 48 hours, spun down to remove cellular debris for 5 min at 4°C and 1000xg, and stored at -80C until use.

For pooled virus production the plasmid pools were transfected as described above in a 10 cm dish format ($1x10^7$ HEK293FT cells, 70µl PEI, 6.4 µg psPAX2, 4.4 µg pMD2.G and 9.2 µg of the plasmid pool). The pooled virus was cleared by spinning down cell debris (3min, 1000xg) and passed through a 0.45 µm filter prior to storage at -80°C. For the CaRPool-seq experiments presented in **Fig. 1, 2** and **5**, and the direct capture Perturb-seq experiments presented in **Fig. 3**, we pooled individual plasmids at equal amounts. In this way, we generated separate sgRNA pools for 1, 2, or 3 sgRNAs to a total of 6 pools (3 for Cas9-nuclease and 3 for KRAB-dCas9(-MeCP2)).

For experiments using single gRNA or CRISPR arrays, we transduced $1x10^6$ HEK293FT or THP1 cells with an MOI of 0.2-0.5. The cells were selected with 1µg/ml puromycin (ThermoFisher A1113803) starting at 24 hours post-transduction for at least 48 hours for HEK293FT cells and at least 5 days for THP1. For CaRPool-seq experiments (**Fig. 1, 2** and **5**) and direct capture Perturb-seq experiments (**Fig. 3**), we transduced $3x10^6$ cells HEK293FT or NIH/3T3 cells. For screens conducted in THP1 cells (Extended Data **Fig. 6** and **7,** and **Fig. 4**) we transduced at least $12x10^6$ cells per condition. HEK293FT and NIH/3T3 cells were selected for at least two days, and THP1 cells for at least 5 days. For single-cell screens we selected conditions with < 10% survival 48 hours post puromycin selection or fraction of GFP-positive THP1 cells below 10% (MOI < 0.1), assuring high coverage (>1000x representation) and a single integration probability >95%. Pooled screens were conducted with MOIs between 0.13 and 0.45 always maintaining coverage of >1000x. Cas13d expression was induced using 1µg/ml doxycycline (Sigma D9891). Cells were maintained with doxycycline, blasticidin, and puromycin until the pooled or single-cell experiment. During this period, the cells were passaged every 2-3 days into fresh media supplemented with doxycycline, blasticidin, and puromycin.

Bulk guide RNA detection

Puromycin-selected cells were harvested 48 hours or more after selection start. RNA was extracted from ~$1x10^6$ cells (Zymo Direct-zol RNA microPrep). Total RNA was reverse transcribed using a capture sequence-specific reverse transcription primer along with an oligo(dT)V_{30} primer (400ng RNA, 4µl 5x RT Buffer, 1µl SuperasIN, 1µl dNTPs (10mM each), 1µl Maxima H Minus RT enzyme, 1.5µl 10µM Template Switch Oligo, 1µl 10µM oligo(dT)V_{30} primer, 1µl 10µM gRNA capture sequence primer, 20µl reaction volume; 53°C/90min, 70°C/15min). The Cas13 crRNA and *GAPDH* mRNA was amplified from 1:2 diluted cDNA using partial TSO and capture sequence primers or gene-specific primers (3µl cDNA, 10µl KAPA 2x master mix, 1µl primer (10 µM each), 5µl H_2O; PCR conditions: 98°C/45s, 18x[98°C/20s, 60°C/10s, 72°C/10s], 72°C/5min). Oligonucleotides are provided in **Extended Data Table 2**.

Flow cytometry

Puromycin-selected cells were harvested 2-7 days after selection start (HEK cells) or Cas13d induction (THP1 cells). (d)Cas9 sgRNAs were evaluated 7 days post-transduction. Cells were stained for the respective cell surface protein for 30 min at 4°C in a 100µl volume and measured by fluorescence-activated cell sorting (Sony SH800) (BioLegend: CD46 clone TRA-2-10 #352405 - 3µl per $1x10^6$ cells; CD55 clone JS11 #311311 - 5µl per $1x10^6$ cells; CD71 clone CYIG4 #334105 - 4µl per $1x10^6$ cells, CD11b clone ICRF44 #301322 – 2µl per $1x10^6$ cells; CD14 clone HCD14 #325608 – 4µl per $1x10^6$ cells). For flow cytometry analysis (FlowJo v10), cells were gated by forward and side scatters and signal intensity to remove potential multiplets. If present, cells were additionally gated with live-dead staining (LIVE/DEAD Fixable Violet Dead Cell Stain Kit, Thermo Fisher L34963). For each sample, we analyzed at least 5,000

cells. If cell numbers varied, we always subsampled (randomly) all samples to the same number of cells within an experiment. For the THP1 KRAB-dCas9-MeCP2 cell experiment, we additionally gated BFP^+ cells as a marker for dual sgRNA integration.

Bulk RNA-seq

To more carefully test for the potential of Cas13 to introduce off-target effects, we performed a bulk RNA-seq experiment, where we expect to have additional power to detect differential expression for lowly expressed transcripts. For the bulk RNA-seq experiment, we performed CD55 knockdown (Cas13d cells, KRAB-dCas9-MeCP2 cells) or knockout (Cas9-nuclease cells) using three individual targeting (s)gRNAs and three individual non-targeting (s)gRNAs. Monoclonal HEK293FT cell lines were transduced with a guide expressing lentivirus (MOI 0.2-0.5) in three independent transductions. Puromycin selection was started 24 hours post-transduction (1µg/mL), and Cas13d expression was induced (1µg/ml Doxycycline). Seven days post-transduction, we confirmed efficient CD55 targeting using flow cytometry, and total RNA was extracted (Zymo Direct-zol RNA microPrep). We performed a modified version of the Smart-seq2 protocol using 100ng purified total RNA input (https://www.protocols.io/view/barcoded-plate-based-single-cell-rna-seq-nkgdctw). Bulk RNA-seq samples were processed with Drop-seq tools v1.0 [49] using a hg19 reference. On average we obtained 352,611 UMI (+/-72,067 UMI) per sample (**Extended Data Table 3**). Differential gene expression was assessed with Seurat's DESeq2 [41] implementation in *FindMarkers*.

Extrapolation of sgRNA detection in direct capture Perturb-seq experiments

We used published sgRNA assignment rates for single and dual sgRNA targeting using direct sgRNA capture via Feature Barcoding technology [6]. We determined the mean sgRNA assignment rate to be 80.9% by averaging the assignment rate for exactly one sgRNA (80% in single guide experiments) and taking the square root of the assignment rate of exactly two sgRNAs per cell (67% in dual guide experiments). In our simulation, we assume that a single viral particle will be taken up by a cell during a low-MOI infection. A single integration event may deliver up to three sgRNAs that are independently expressed, similar to the two sgRNA experiments described previously [6]. We assume that sgRNA-detection for each sgRNA is an independent event. These can be modeled by multiplying detection and editing probabilities p by the number of sgRNA feature n (p^n). The resulting curve shows the fraction of cells that have received exactly n sgRNAs (grey line in **Fig. 3b**).

Gene module comparison

We processed the raw Wang et al. [23] ECCITE-seq data as described above with minor modifications. Gene expression libraries were processed using CellRanger v4.0.0. Guide RNA reads were 5' adapter-trimmed (*CCAGCATAGCTCTTAAAC*), prior to quantification with Cite-seq-Count as described above.

Single-cell data was processed using Seurat, as described above, with minor modifications. Cells were filtered based on RNA features and mitochondrial read content (*nFeature_RNA >400 & <6000, percent.mt <15*) returning 4,902 filtered cells. Guide RNAs UMI counts were normalized (*normalization.method = CLR, margin = 2*) and scaled. Guide RNA identity was determined using *MultiSeqDemux(autoThresh = T)*. Cells without sgRNA assigned and cells with multiple sgRNAs assignments were removed, returning 2,384 filtered cells. The ECCITE-seq data was filtered using mixscape [9] to remove unperturbed cells (*min.de.genes = 5, logfc.threshold = 0.1*) returning 1,408 filtered cells. Differentially expressed gene were identified using *FindMarkers(logfc.threshold = 0.1, only.pos = F)* using the default Wilcoxon's rank sum test. For each target gene, we selected the 50 most up and downregulated genes ranked by p-value. Single-cell Module scores were assigned using *AddModuleScore* for each set selected DEGs of overlapping single gene perturbation (DOT1L, EP300, FLI1, GFI1, GSE1, MED24).

Reference integration and differentiation trajectory

We utilized a human bone marrow reference dataset (http://azimuth.hubmapconsortium.org) and extracted reference cells belonging to HSC, LMPP, GMP, and CD14 monocyte populations, and then integrated query cells to this monocytic differential trajectory reference by Seurat with reciprocal PCA workflow and setting k.anchor = 20. Query cells were annotated based on similarity to the myeloid reference. We next created a k-nearest graph representing cells from both CaRPool-seq data and healthy reference using the batch correct latent space learned during the mapping. This graph was used as input to the destiny package (v3.8.1), which uses diffusion maps to reduce the dimensionality of the data [50]. We did not expect or observe branching trajectories in our data, which represents cell states spanning monocyte differentiation. We used the first two diffusion map dimensions as input to monocle3 (v1.2.4) [51] in order to infer a pseudotemporal ordering.

SUPPLEMENTARY REFERENCES

48. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods 11, 783–784 (2014).
49. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161, 1202–1214 (2015).
50. Haghverdi, L., Buettner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics 31, 2989–2998 (2015).
51. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods 14, 979–982 (2017).