

Aug 08, 2024

## Single cell CRISPRi screen to identify mechanoenhancer gene targets

DOI

<https://dx.doi.org/10.17504/protocols.io.kqdg3xd7zg25/v1>

Brian D. Cosgrove<sup>1</sup>, Lexi R. Bounds<sup>1</sup>, Carson Key Taylor<sup>1</sup>, Alan L. Su<sup>1</sup>, Anthony J. Rizzo<sup>1</sup>, Alejandro Barrera<sup>1</sup>, Andrea Daniel<sup>1</sup>, Gregory E. Crawford<sup>1</sup>, Brenton D. Hoffman<sup>1</sup>, Charles A. Gersbach<sup>1</sup>

<sup>1</sup>Duke University

Andrea Daniel: This protocol was adapted from the work of Brian Cosgrove and Lexi Bounds in the Gersbach lab at Duke University.

Gersbach Lab



Andrea R Daniel

Duke University

### Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.kqdg3xd7zg25/v1>

**Protocol Citation:** Brian D. Cosgrove, Lexi R. Bounds, Carson Key Taylor, Alan L. Su, Anthony J. Rizzo, Alejandro Barrera, Andrea Daniel, Gregory E. Crawford, Brenton D. Hoffman, Charles A. Gersbach 2024. Single cell CRISPRi screen to identify mechanoenhancer gene targets. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.kqdg3xd7zg25/v1>

**Manuscript citation:**

**Mechanosensitive genomic enhancers potentiate the cellular response to matrix stiffness**

Brian D. Cosgrove, Lexi R. Bounds, Carson Key Taylor, Alan L. Su, Anthony J. Rizzo, Alejandro Barrera, Gregory E. Crawford, Brenton D. Hoffman, Charles A. Gersbach

bioRxiv 2024.01.10.574997; doi: <https://doi.org/10.1101/2024.01.10.574997>

**License:** This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Protocol status:** Working

**We use this protocol and it's working**

**Created:** January 16, 2024

**Last Modified:** August 08, 2024

**Protocol Integer ID:** 93620

**Keywords:** CRISPRi, enhancer, extracellular matrix, mechanoenhancer, cell growth and migration, ECM, matrix stiffness, single cell RNA-seq, single cell crispr, mechanoenhancer gene targets this protocol, crispr interference screen, mechanoenhancer gene target, crispr, single cell crispri screen, gene target, gene, cell

**Funders Acknowledgements:**

NIH

Grant ID: HG012053

## Abstract

This protocols describes methods for a single cell CRISPR interference screen to identify gene targets regulated by mechanoenhancers that drive cellular growth and migration.

## Troubleshooting

## gRNA library design and cloning

- 1 Following hit identification from combined migration and growth screens, a library is designed that includes the top 10 gRNA by pZ value across either screen for the 87 hit pRE (870 gRNA total).
- 2 100 non-targeting control gRNAs with similar sequence composition to the targeting gRNAs are included in the library, and 25 gRNAs targeting the promoters of contractile genes including *MYH9*, *RANGAP1*, and *CRIM1* are included, as well as the top gRNA from the MYH9 intron 3 enhancer as positive controls.
- 3 In total the library contains 1005 gRNA sequences, which are synthesized as an oligo pool by Twist Biosciences with common overhangs for cloning into the lentiviral backbone.
- 4 This oligo pool is PCR amplified, and a hU6-driven lentiviral gRNA CROP-seq vector (pLRB104) is then digested with Esp3I, gel purified, and then ligated along with the amplified oligo pool by Gibson assembly.
- 5 Following a 1x SPRI cleaning, the Gibson assembly is transformed into Endura competent cells (Lucigen) according to the manufacturer's protocol, and cultured overnight before maxi-prepping the gRNA-library plasmid.
- 6 A PCR amplicon across the gRNA region of the resulting plasmid is sequenced to a depth of ~100k-1M read pairs on an Illumina miSeq in order to verify coverage across the entire gRNA library

## Lentiviral generation and functional titering

- 7 gRNA library plasmid is co-transfected into ~18M HEK293T cells along with two lentiviral packaging plasmids using Lipofectamine 3000 (ThermoFisher).
- 8 20 hours post-transfection, the growth media is removed and fresh growth media is added. Media containing viral particles is removed at 48 hours, replaced, and removed at 72 hours post-lipofection. Media containing viral particles is stored at 4C.
- 9 Combined media containing viral particles is filtered through 0.45  $\mu$ m low-protein binding filters, and then concentrated using Lenti-X Concentrator (Takara Bio) according to the manufacturer's protocol.
- 10 Functional titering to determine MOI is performed by transducing HFF cells across a 50x-10,000x dilution range of the viral stock, and then subjecting the cells to a qPCR-based titering protocol that has been previously described in detail (1).

## Single cell CRISPRi screen

- 11 To perform screening, 775k HFF cells stably expressing dCas9-KRAB are transduced at 0.33MOI with the CROP-seq lentivirus to maintain a coverage of at least 150 cells/gRNA.
- 12 After 20 hours, viral media is removed and replaced with regular growth media, and 48 hours post-transduction the cells are subjected to selection media with puromycin (1.5 ug/mL) for 4 days.
- 13 Following puromycin selection, HFF cells were maintained until day 8, at which point cells are trypsinized and 150k cells are moved on to library prep.

## Single cell RNA-seq library preparation

- 14 Cells are washed 3x with PBS and then resuspended to a final concentration of 1000 cells/uL. Approximately 20,000 cells are loaded onto each channel of a 10X Genomics' 3' Gene Expression (GEX) v3.1 assay chip. Downstream processing is performed according to the manufacturer's protocol.
- 15 To recover the protospacer sequences (gRNA libraries), a tri-nested PCR is performed separately for each GEX library using 10% of the purified cDNA as input to reaction 1 as previously described (2).
- 16 4ng cDNA is input into a 50 uL reaction with KAPA HiFi and PCR primers prLRB470 and prLRB471. The reaction is amplified for 12 cycles and then purified using 25 uL of AMPure XP DNA beads and eluted in 25 uL H2O. 1 uL of the purified sample is input into reaction 2 using PCR primers prLRB472 and prLRB473. The reaction is amplified for 14 cycles and purified using 25 uL of AMPure XP DNA beads and eluted in 25 uL H2O. 1 uL of the purified sample is input into reaction 3 using PCR primers prLRB473 and prLRB289-302, amplifying each sample with a unique i7 sequencing index. The reaction is amplified for 7 cycles, purified using 25 uL of AMPure XP DNA beads (Beckman Coulter #A63881), and eluted in 25 uL Buffer EB (Qiagen #19086).

	A	B	C
	Primers used for CROP-seq recovery from 10X cDNA.		
	PCR number	Primer name	Primer sequence



A	B	C
1	prLRB470_U6_OUTER	TTTCCCAT GATTCCTT CATATTTGC
1	prLRB471_R1_PCR1	ACACTCTT TCCCTACA CGACG
2	prLRB472_U6_INNER_wit h_P7_adapte r	GTCTCGTG GGCTCGGA GATGTGTAT AAGAGACA GcTTGTGG AAAGGACG AAACAC
2	prLRB473_R1-P5	AATGATAC GGCGACCA CCGAGATC TACACTCT TTCCCTAC ACGACG
3	prLRB473_R1-P5	AATGATAC GGCGACCA CCGAGATC TACACTCT TTCCCTAC ACGACG
3	prLRB289_N701	CAAGCAGA AGACGGCA TACGAGAT TCGCCTTA GTCTCGTG GGCTCGG
3	prLRB290_N702	CAAGCAGA AGACGGCA TACGAGAT CTAGTACG GTCTCGTG GGCTCGG
3	prLRB291_N703	CAAGCAGA AGACGGCA TACGAGAT TTCTGCCT GTCTCGTG GGCTCGG
3	prLRB292_N704	CAAGCAGA AGACGGCA TACGAGAT GCTCAGGA GTCTCGTG GGCTCGG



A	B	C
3	prLRB293_ N705	CAAGCAGA AGACGGCA TACGAGATA GGAGTCCG TCTCGTGG GCTCGG
3	prLRB294_ N706	CAAGCAGA AGACGGCA TACGAGAT CATGCCTA GTCTCGTG GGCTCGG
3	prLRB295_ N707	CAAGCAGA AGACGGCA TACGAGAT GTAGAGAG GTCTCGTG GGCTCGG
3	prLRB298_ N710	CAAGCAGA AGACGGCA TACGAGAT CAGCCTCG GTCTCGTG GGCTCGG
3	prLRB299_N 711	CAAGCAGA AGACGGCA TACGAGAT TGCCTCTT GTCTCGTG GGCTCGG
3	prLRB300_ N712	CAAGCAGA AGACGGCA TACGAGAT TCCTCTAC GTCTCGTG GGCTCGG
3	prLRB301_N 714	CAAGCAGA AGACGGCA TACGAGAT TCATGAGC GTCTCGTG GGCTCGG
3	prLRB302_ N715	CAAGCAGA AGACGGCA TACGAGAT CCTGAGAT GTCTCGTG GGCTCGG



- 17 Quality control of final libraries is performed prior to sequencing using the Agilent 2200 TapeStation with High Sensitivity DNA 5000 reagents, Qubit High Sensitivity dsDNA reagents, and KAPA Library Quantification Kit for Illumina platforms.

## Sequencing

- 18 GEX libraries are pooled and sequenced on a NovaSeq 6000 S4 flow cell using the parameters: 28×10×10×90.
- 19 gRNA libraries are pooled and sequenced on a NovaSeq 6000 S1 flow cell using the parameters:  
28×10×10×90.

## Data processing

- 20 Cell Ranger: All data processing steps are performed using CellRanger v6.0.1 and the human reference genome ('refdata-gex-GRCh38-2020-A') downloaded from 10X Genomics' software downloads webpage. Fastq files for each flow cell lane and sequencing run are generated from .bcl files using the cellranger mkfastq pipeline. The corresponding fastq files for each sample are then merged. The merged fastqs are then processed using the cellranger count pipeline with the number of expected cells specified (--expect-cells = 15000). The gene expression libraries are then aggregated using the cellranger aggr pipeline. The gRNA libraries are aligned to a custom bowtie index containing all protospacer sequences included in the pooled gRNA library and the UMI counts corresponding to each gRNA-cell pair are obtained.
- 21 Seurat: The gene expression and gRNA UMI count data is imported into Seurat v3.1. A gRNA is defined as 'observed in a cell' if the gRNA had at least 5 UMI counts and comprised at least 0.5% of the total gRNA UMI counts in that cell. Then calculate the total percent of mitochondrial reads per cell and filter for quality cells as follows:

```
cells[["percent.mt"]] ← PercentageFeatureSet(cells, pattern = "^MT-")  
cells ← subset(cells, subset = nCount_RNA > 10000 & percent.mt < 20)
```

## Differential expression analysis

- 22 Using the gRNA-cell assignments, differential expression testing is performed using the MAST framework (3) within Seurat v3.1 (4), comparing cells in which a given gRNA is observed versus all other cells with at least one gRNA observed excluding the given gRNA and testing all genes within +/- 1Mb of the midpoint of the pRE in which the gRNA is located.



- 23 Gene coordinates are obtained from the Ensembl Human Gene v104 reference file. P-values are then FDR-corrected on an individual gRNA-level for all tests. All genes within +/- 1Mb of any targeting gRNA are used as input features for NT gRNA tests (N=1,313). Significant gRNA-gene and corresponding pRE-gene pairs are defined as FDR < 0.01.

## Calculation of interaction distance (ep\_length)

- 24 The distance between the gRNA and the paired gene is calculated as follows: 1) the gRNA midpoint (gRNA\_mid) is defined as  $(\text{gRNA\_start} + \text{gRNA\_end})/2$ , the gene start coordinate (gene\_start) is defined as the start coordinate for genes on the '+' plus strand, and end coordinate for genes on the '-' strand. 'ep\_length' is calculated as  $\text{gRNA\_mid} - \text{gene\_start}$ .

## Effect size comparison between targeting and control gRNAs

- 25 For each gene with a TSS- or validated enhancer-targeting gRNA, we compare the avg\_logFC of expression for the respective gene between TSS-control, enhancer-control, pRE-targeting, and NT control gRNAs (FDR < 0.01), using a one-way ANOVA followed by Tukey's HSD with Bonferroni correction (adj. p-value). Significant differences in the change in gene expression are defined as adj. p-value < 0.05.

## Interaction distance versus effect size

- 26 Using all significant pRE-targeting gRNA gene pairs, the avg\_logFC and effect size ( $\text{avg\_logFC} * (1 - \text{FDR})$ ) are plotted versus the log10-transformed ep\_length ( $\log_{10}(\text{abs}(\text{ep\_length} + 1))$ ). Spearman correlation  $R^2$  values are calculated using the 'stat\_cor' function from the 'ggpubr' R package.

## Nearest gene prediction analysis

- 27 For each potential pRE-gene pair, we calculate the number of genes "skipped" by the element to regulate the gene as follows. First, for the significant pRE-gene connections, define the start and end coordinates for a given element and the start and end coordinates, and strand, for the paired gene. Next, count the number of genes detected in the gene expression dataset for which the entire gene body is contained within the region between the element and the connected gene. Repeat this for all significant pRE-gene connections.

## Comparison to microC looping

- 28 Chromatin contact data is intersected with all targeted pREs, TSS regions (+/- 1kb) of every gene, and all genes for which a differential expression test is performed,

separately extended by +/- 500bp with anchor 1 and anchor 2, using bedtools window -w 500.

	A	B	C
	Publicly available datasets used in this study.		
	Source	Description	Location/Accession ID
	ENCODE	SCREEN cCREs v3	<a href="https://downloads.wenglab.org/V3/GRCh38-cCREs.bed">https://downloads.wenglab.org/V3/GRCh38-cCREs.bed</a>
	4D Nucleome	HFF H3K4me1	4DNES9WN NK52
	4D Nucleome	HFF H3K4me3	4DNESWX3 2ZCG
	4D Nucleome	HFF H3K27ac	4DNESNW XHXK
	4D Nucleome	HFF ATAC	4DNESMBA 9T3L
	PMID: 32213324	HFF micro-C	4DNFI18Q7 99K
	PMID: 32832598	Single cell RNA-sequencing of primary human lung tissue	GSE135893
	PMID: 34370624	Chromatin accessibility in IPF and healthy lung tissue.	GSE180242

29 The number of pREs, TSSs, and genes with at least one chromatin contact, defined as at least one intersection with anchor 1 or anchor 2, are quantified.

30 Next, for all regions that intersect a region in the anchor 1 set, we quantify the number of pRE-gene pairs for which the corresponding contact in the anchor 2 set overlaps either

the same TSS/gene or a different TSS/gene. We repeat this for pREs intersecting the anchor 2 set with comparison of contacts for TSSs/genes in the anchor 1 set.

## Protocol references

1. M. G. Gordon, F. Inoue, B. Martin, M. Schubach, V. Agarwal, S. Whalen, S. Feng, J. Zhao, T. Ashuach, R. Ziffra, A. Kreimer, I. Georgakopoulous-Soares, N. Yosef, C. J. Ye, K. S. Pollard, J. Shendure, M. Kircher, N. Ahituv, lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements. *Nat. Protoc.* (2020), doi:10.1038/s41596-020-0333-5.
2. M. Gasperini, A. J. Hill, J. L. McFaline-Figueroa, B. Martin, S. Kim, M. D. Zhang, D. Jackson, A. Leith, J. Schreiber, W. S. Noble, C. Trapnell, N. Ahituv, J. Shendure, A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell.* 176, 377–390.e19 (2019).
3. G. Finak, A. McDavid, M. Yajima, J. Deng, V. Gersuk, A. K. Shalek, C. K. Slichter, H. W. Miller, M. J. McElrath, M. Prlic, P. S. Linsley, R. Gottardo, MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* 16, 278 (2015).
4. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck 3rd, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell.* 177, 1888–1902.e21 (2019).