



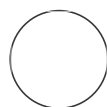
DEC 06, 2023

Methods from: Multiplex, single-cell CRISPRa screening for cell type specific regulatory elements

Troy

Florence Chardon¹, McDiarmid¹

¹Department of Genome Sciences, University of Washington



Beth Martin

ABSTRACT

CRISPR-based gene activation (CRISPRa) is a promising therapeutic approach for gene therapy, upregulating gene expression by targeting promoters or enhancers in a tissue/cell-type specific manner. Here, we describe an experimental framework that combines highly multiplexed perturbations with single-cell RNA sequencing (sc-RNA-seq) to identify cell-type-specific, CRISPRa-responsive *cis*-regulatory elements and the gene(s) they regulate. Random combinations of many gRNAs are introduced to each of many cells, which are then profiled and partitioned into test and control groups to test for effect(s) of CRISPRa perturbations of both enhancers and promoters on the expression of neighboring genes. Applying this method to candidate *cis*-regulatory elements in both K562 cells and iPSC-derived excitatory neurons, we identify gRNAs capable of specifically and potently upregulating target genes, including autism spectrum disorder (ASD) and neurodevelopmental disorder (NDD) risk genes. A consistent pattern is that the responsiveness of individual enhancers to CRISPRa is restricted by cell type, implying a dependency on either chromatin landscape and/or additional *trans*-acting factors for successful gene activation. The approach outlined here may facilitate large-scale screens for gRNAs that activate therapeutically relevant genes in a cell type-specific manner.

OPEN ACCESS



External link:

<https://www.biorxiv.org/content/10.1101/2023.03.28.534017v1.full>

Protocol Citation: Florence Chardon, Troy McDiarmid 2023. Methods from: Multiplex, single-cell CRISPRa screening for cell type specific regulatory elements.

protocols.io

<https://protocols.io/view/methods-from-multiplex-single-cell-crispra-screening-c5zv766>

MANUSCRIPT CITATION:

Multiplex, single-cell CRISPRa screening for cell type specific regulatory elements

Florence M. Chardon, Troy A. McDiarmid, Nicholas F. Page, Beth Martin, Silvia Domcke, Samuel G. Regalado, Jean-Benoît Lalanne, Diego Calderon, Lea M. Starita, Stephan J. Sanders, Nadav Ahituv, Jay Shendure

BioRxiv

doi: <https://doi.org/10.1101/2023.03.28.534017>

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: Dec 06, 2023

Last Modified: Dec 06, 2023

PROTOCOL integer ID:
91925

**Funders
Acknowledgement:**

Weill Neurohub
NHGRI
Grant ID: UM1HG011966
NIMH
Grant ID: U01MH12268
NIMH
Grant ID: R01MH116999

Cell Lines and Culture

1 *K562 cell culture*

K562s cells are a pseudotriploid ENCODE Tier I erythroleukemia cell line derived from a female (age 53) with chronic myelogenous leukemia¹⁵. All K562 cells were grown at 37°C, and cultured in RPMI 1640 + L-Glutamine (GIBCO, Cat. No. 11-875-093) supplemented with 10% fetal bovine serum (Rocky Mountain Biologicals, Cat No. FBS-BSC) and 1% penicillin-streptomycin (GIBCO/ Thermo Fisher Scientific; Cat. No. 15140122).

2 *Induced pluripotent stem cell (iPSC) culture*

Human WTC11 iPSCs equipped with a doxycycline-inducible *NGN2* transgene expressed from the *AAVS1* safe-harbor locus as well as an ecDHFR-dCas9-VPH construct (VPH consists of 12 copies of VP16, fused with a P65-HSF1 activator domain) expressed from the *CLYBL* safe-harbor locus were a gift from the Kampmann lab⁶. These iPSCs were cultured in mTeSR Plus Basal Medium (Stemcell technologies; Cat. No. 100-0276) on Greiner Cellstar plates (Sigma-Aldrich; assorted Cat. Nos.) coated with Geltrex™ LDEV-Free, hESC-Qualified, Reduced Growth Factor Basement Membrane Matrix (Gibco; Cat. No. A1413302) diluted 1:100 in Knockout DMEM (GIBCO/Thermo Fisher Scientific; Cat. No. 10829018). mTeSR Plus Basal Medium was replaced

every other day. When 70–80% confluent, cells were passaged by aspirating media, washing with DPBS (GIBCO/Thermo Fisher Scientific; Cat. No. 14190144), incubating with StemPro Accutase Cell Dissociation Reagent (GIBCO/Thermo Fisher Scientific; Cat. No. A1110501) at 37 °C for 5 min, diluting Accutase 1:1 in mTeSR Plus Basal Medium, collecting cells in conical tubes, centrifuging at 800g for 3 min, aspirating supernatant, resuspending cell pellet in mTeSR Plus Basal Medium supplemented with 0.1% dihydrochloride ROCK Inhibitor (Stemcell technologies; Cat. No. Y-27632), counting and plating onto Geltrex-coated plates at the desired number.

3 *Human iPSC-derived neuronal cell culture, differentiation, and CRISPRa induction*

The iPSCs described above were used for the differentiation protocol below. On day -3, iPSCs were dissociated and centrifuged as above, and pelleted cells were resuspended in Pre-Differentiation Medium containing the following: Knockout DMEM/F-12 (GIBCO/Thermo Fisher Scientific; Cat. No. 12660012) as the base, 1X MEM Non-Essential Amino Acids (GIBCO/Thermo Fisher Scientific; Cat. No. 11140050), 1X N-2 Supplement (GIBCO/ Thermo Fisher Scientific; Cat. No. 17502048), 10 ng/mL NT-3 (PeproTech; Cat. No. 450-03), 10ng/mL BDNF (PeproTech; Cat. No. 450-02), 1 ug/mL Laminin mouse protein (Thermo Fisher Scientific; Cat. No. 23017015), 10 nM ROCK inhibitor, and 2 mg/mL doxycycline hyclate (Sigma-Aldrich; Cat. No. D9891) to induce expression of *NGN2*. iPSCs were counted and plated at 800K cells per Geltrex-coated well of a 12-well plate in 1 mL of Pre-Differentiation Medium, for three days. At day -2 and day -1, media changes were performed using pre-differentiation medium without ROCK inhibitor. On day -1, 12-well plates for differentiation were coated with 15 ug/mL Poly-L-Ornithine (Sigma-Aldrich; Cat. No. P3655) in DPBS, and incubated overnight at 37 degrees Celsius. On day 0, the Poly-L-Ornithine coated plates were washed three times using DPBS, and the plates were air dried in a 37 degree Celsius incubator until all the DPBS evaporated. Pre-differentiated cells were dissociated and centrifuged as above, and pelleted cells were resuspended in Maturation Medium containing the following: 50% Neurobasal-A medium (GIBCO/Thermo Fisher Scientific; Cat. No. 10888022) and 50% DMEM/F-12 (GIBCO/Thermo Fisher Scientific; Cat. No. 11320033) as the base, 1X MEM Non-Essential Amino Acids, 0.5X GlutaMAX Supplement (GIBCO/Thermo Fisher Scientific; Cat. No. 35050061), 0.5X N-2 Supplement, 0.5X B-27 Supplement (GIBCO/Thermo Fisher Scientific; Cat. No. 17504044), 10 ng/mL NT-3, 10 ng/mL BDNF, 1 ug/mL Laminin mouse protein, and 2 ug/mL doxycycline hyclate. Pre-differentiated cells were subsequently counted and plated at 400,000-450,000 cells per well of a 12-well plate coated with Poly-L-Ornithine in 1 mL of Maturation medium with 20 uM trimethoprim (TMP) (Sigma-Aldrich, Cat No. 92131) to activate the CRISPRa machinery in these cells (TMP stabilizes the degron-tagged CRISPRa machinery). On day 7, half of the medium was removed and an equal volume of fresh Maturation medium without doxycycline was added. On day 14, half of the medium was removed and twice that volume of fresh medium without doxycycline was added. On day 19, neurons were harvested for sc-RNA-seq.

Cell line generation and CRISPRa validation

5 *K562 cells*

K562 cells expressing dCas9-VP64 were generated in-house via lentiviral integration of a dCas9-VP64-blast construct⁷ (Addgene Plasmid #61422) into K562 cells. Cells were selected with 10 ug/mL blasticidin, and polyclonal cells were single-cell sorted into 96-well plates to grow up clonal cell lines expressing dCas9-VP64. Clonal cell lines were tested for CRISPRa activity by testing the ability of a CRISPRa gRNA to activate a minP-tdTomato construct²⁰, and the highest tdTomato expressing cell line was used for experiments. K562 cells expressing dCas9-VPR were purchased from Horizon Discovery/Perkin Elmer (catalog ID: HD dCas9-VPR-005), and these cells were tested for CRISPRa activity using the same tdTomato expression assay described above.

6 *iPSC-derived neurons*

Human WTC11 iPSCs equipped with a doxycycline-inducible *NGN2* transgene expressed from the *AAVS1* safe-harbor locus as well as an ecDHFR-dCas9-VPH construct expressed from the CLYBL safe-harbor locus were a gift from the Kampmann lab⁶. These cells were tested for CRISPRa activity using the same tdTomato expression assay that was used to validate the K562 cell lines, which is described above.

gRNA selection and design

- 7 Enhancer-targeting gRNAs were selected from our CRISPRi library^{2,33}. Specifically, 50 spacer sequences (2 per candidate enhancer) were randomly selected from the list of 664 significant “hit” enhancer-gene pairs in the at-scale library. Another 50 spacer sequences targeting an additional 25 candidate enhancers (again 2 per candidate enhancer) were randomly selected from candidate enhancer non-hits (*i.e.*, gRNAs from the at-scale library targeting candidate enhancer regions with strong biochemical marks predictive of regulatory activity that did not yield significant downregulation of any neighboring genes in our previous CRISPRi study). An additional 30 TSS-positive control gRNAs were randomly sampled from the top quartile of gRNAs recommended by Horlbeck *et al.* (hCRISPRa-v2 library)¹⁶. 50 NTC negative control spacer sequences were also selected from the hCRISPRa-v2 library¹⁶. The 313 candidate promoter targeting gRNAs were either selected from the Horlbeck *et al.* library¹⁶ or designed using FlashFry³³.
- Briefly, 50 candidate promoters of 9 NDD risk genes (*TCF4*, *FOXP1*, *SCN2A*, *CHD8*, *BCL11A*, *TBR1*, *SHANK3*, *SYNGAP1*, *ANKK2*)^{23,24} were pulled from Basic GENCODE annotations³⁴ and were filtered for “type” == “transcript” and “transcript_type” == “protein coding”. Separate bed files were generated for all promoter regions defined as the 500bp upstream

of each protein coding transcript. Careful attention was paid to the strand orientation of each transcript when annotating promoter regions. Bed files were sorted and merged to combine multiple promoters with >1bp overlap into a single promoter annotation. Transcript bounds provided for each merged promoter begin +1bp from the end of the promoter and end at the position corresponding to the longest transcript mapping to that promoter. NGG-protospacer within these candidate promoters were identified using FlashFry and subsequently scored using default parameters (see FlashFry manuscript and user guide for a complete description of scoring metrics/algorithms)³³. A TSS-distance metric was then calculated for each gRNA using human fetal brain 5' Capped Analysis of Gene Expression (CAGE) data^{35, 36} obtained from FANTOM (<https://fantom.gsc.riken.jp/5/sstar/FF:10085-102B4>; CTSS, hg38). First, the strongest FANTOM annotated TSS was identified within each +/-500 bp region up and downstream of each hg38 Gencode Basic protein coding transcript TSS. For regions with a tie between the highest scoring FANTOM TSSs, the TSS position closest to Gencode annotated TSS position was prioritized. Each candidate sgRNA from FlashFry was annotated with the distance to the nearest FANTOM TSS using the command "bedtools closest -a sgRNAs_with_fantom_tss -b strongest_fantom_tss_within_gencode_promoter -D b -t first." For Gencode Basic protein coding transcripts without a human fetal brain FANTOM peak within 500 +/- bp, the distance of each sgRNA to the nearest Gencode TSS was reported instead. A distance of zero indicates that an sgRNA overlaps with the nearest annotated TSS. Multiple rounds of successively relaxing score and distance thresholds were then iterated until the top 4 gRNAs for each candidate promoter were selected (five selection rounds in total). Optimal TSS-distances were approximated using genome-wide CRISPRa design rules²⁸. gRNAs flagged for potentially problematic polythymidine tracks or GC content were excluded. The gRNA selection criteria used in each round were as follows:

■ Round

1: 1. TSS Distance between -150 and -75 BP **2.** Doench2014OnTarget >= 0.2 **3.** Dangerous_in_genome <= 1 **4.** Hsu2013 > 80.

■ Round

2: 1. TSS Distance between -400 and -50 BP **2.** Doench2014OnTarget >= 0.2 **3.** Dangerous_in_genome <= 1 **4.** Hsu2013 > 80.

■ Round

3: 1. TSS Distance between -400 and -50 BP **2.** Doench2014OnTarget >= 0.2 **3.** Dangerous_in_genome <= 1 **4.** Hsu2013

> 50.

■ **Round**

4: 1. TSS Distance between -400 and -50 BP 2. Doench2014OnTarget ≥ 0.2 3. Dangerous_in_genome ≤ 2 4. Hsu2013 > 50.

■ **Round**

5: 1. Doench2014OnTarget ≥ 0.2 2. Dangerous_in_genome ≤ 2 3. Hsu2013 > 10 4. DoenchCFD_maxOT < 0.95

Complete oligo sequences with gRNA spacers and additional sequences for cloning into piggyFlex are listed in **Table S1**. Note all gRNAs in our library are designed/modified to start with a G followed by the 19 base pair spacer to facilitate Pol III transcription.

gRNA library cloning into piggyFlex vector

8 The 493 gRNAs with associated 10N random barcodes were ordered as an IDT oPool and PCR amplified with Q5 High-Fidelity polymerase (NEB, Cat. No. M0491S) to make double stranded DNA. The piggyFlex backbone vector was digested with Sall (NEB, Cat. No. R3138S) and BbsI (NEB, Cat. No. R0539S) in 10X NEBuffer r2 at 37 degrees Celsius overnight to ensure complete digestion of the backbone. This digestion cuts out the EF1a-puro-GFP cassette of the vector which is then added back in a later cloning step. The digestion product was run on a 1% agarose gel in TAE buffer, and the linear backbone vector (5098 base pairs in size) was gel extracted using a gel extraction kit (NEB, Cat. No. T1020S). The second product from the digestion (2878 base pairs) which contains the EF1a-puro-GFP cassette was saved for a later assembly reaction in the final cloning step (described below). The PCR amplified IDT oPool gRNAs with associated 10N random barcodes were cloned into the linear backbone using NEBuilder HiFi DNA Assembly (NEB, Cat. No. E2621S) using 0.15 pmol of the insert (gRNA library) and 0.02 pmol of the linear backbone. Assembled product was transformed into electrocompetent cells (NEB, Cat. No. C3020K) and plasmid DNA was extracted with a midiprep kit (Zymo Research, Cat. No. D4200). The resulting vector was then digested with SapI (NEB, Cat. No. R0569S), for one hour at 37 degrees Celsius. Digested product was cleaned with 0.5X AMPure beads (Beckman Coulter, Cat. No. A63880) and cleaned digested linear backbone was used for a subsequent assembly reaction to add the EF1a-puro-GFP cassette back into the final piggyFlex vector between the gRNA sequence and the 10N random barcode sequences. 0.014 pmol of the linear backbone was assembled with 0.056 pmol of the insert sequence and the assembly reaction was cleaned with a 0.5X AMPure step. The assembled product was transformed into electrocompetent cells and plasmid DNA was extracted with a midiprep kit. The final plasmid library was subsequently PCR amplified and sequenced to ensure that all 493 gRNAs were successfully cloned into the

piggyFlex vector. Note: The 10N barcode is an additional gRNA identification strategy that can be used to assign gRNAs to cells, however, we used directly sequenced gRNAs (from the 10x Genomics capture sequence) to identify gRNAs in this work as this more accurately assigns gRNA transcripts to cells³⁷.

9

Transfection of the gRNA library and selection for transfected...

10 *K562 cells*

16 million K562 cells (8 million K562-VP64 cells and 8 million K562-VP64 cells) were transfected with the gRNA library and the piggyBac transposase (System Biosciences, Cat. No. PB210PA-1) at a 20:1 molar ratio of library:transposase using a Lonza 4D nucleofector and the Lonza nucleofector protocol for K562 cells. The 16 million cells were split across 8 100 uL nucleofection cartridges, with each individual nucleofection cartridge receiving 2 million cells and 2 ug of total DNA. After nucleofection, cells were transferred to pre-warmed RPMI media in a cell culture flask and incubated at 37 degrees Celsius. One day after transfection, cells were selected with 2 ug/mL puromycin (GIBCO/Thermo Fisher Scientific; Cat. No. A1113803). After 9 days, cells were harvested for single-cell transcriptome profiling.

11 *Induced pluripotent stem cells*

6 million dCas9-VP64 iPSCs (same cells as described above) were transfected with the gRNA library and the piggyBac transposase at a 5:1 molar ratio of library:transposase using the Lonza nucleofector and the Lonza nucleofector CB-150 program. The 6 million cells were split across 6 100 uL nucleofection cartridges, with each individual nucleofection cartridge receiving 1 million cells and 17.5 ug of total DNA. After nucleofection, cells were transferred to pre-warmed mTeSR Plus basal medium with ROCK inhibitor in a cell culture flask and incubated at 37 degrees Celsius. One day after transfection, cells were selected with 20 ug/mL puromycin (note: the AAVS1-NGN2 construct has a puromycin resistance cassette on it, so a higher dose of puromycin was used to successfully select for cells that received a gRNA in the presence of an existing puromycin resistance cassette). Media changes were performed daily (mTeSR Plus basal medium with ROCK inhibitor and 10 ug/mL puromycin) for seven days prior to initiating neuron differentiation (described in "Human iPSC-derived neuron cell culture and differentiation" methods section).

10x Genomics sc-RNA-seq with associated gRNA transcript ca...

12 *K562 screen*

Cells were harvested and prepared into single-cell

suspensions following the 10x Genomics Single Cell Protocols Cell Preparation Guide (Manual part number CG00053, Rev C). Four lanes were used for the single-cell transcriptome profiling, with two lanes containing cells from the K562-VP64 cell line, and two lanes containing cells from the K562-VPR cell line. Roughly 10,000 cells were captured per lane of a 10x Chromium chip (Next GEM Chip G) using Chromium Next GEM Single Cell 3' Reagent Kits v3.1 with Feature Barcoding technology for CRISPR Screening (10x Genomics, Inc, Document number CG000205, Rev D).

13 *iPSC-derived neuron screen*

iPSC-derived neurons were harvested and prepared into

single-cell suspensions following a published protocol ³⁸. Cells were split into two batches, with one batch going through a fluorescence-activated cell sorting (FACS) step to sort on the top 40% of green fluorescent protein (GFP) expression to enrich for neurons with greater numbers of gRNAs integrated, and the second batch going directly into the 10x Genomics single-cell library preparation protocol. Sorting on the top 40% of GFP expression resulted in a two-fold increase in the mean number of gRNAs integrated in those cells as compared to unsorted cells. Four lanes were used for the single-cell transcriptome profiling, with two lanes containing GFP-positive sorted cells, and two lanes containing unsorted cells. Roughly 13,000 cells were captured per lane of a 10x Chromium high-throughput chip (Next GEM Chip M) using Chromium Next GEM Single Cell 3' HT Reagent Kits v3.1 (Dual Index) with Feature Barcode technology for CRISPR Screening (10x Genomics, Inc, Document number CG000418, Rev C).

14 *Sequencing of scRNA-seq libraries*

Final libraries were sequenced on a NextSeq 2000 P3 100 cycle kit (R1:28 I1:10, I2:10, R2:90) for each screen (K562 and iPSC-derived neuron screens). Gene expression and gRNA transcript libraries were pooled at a 4:1 ratio for sequencing.

15 *Transcriptome data processing and quality control filtering for K562 and iPSC-derived neuron screens*

CellRanger version 6.0.1 was used to perform bcl2fastq and count matrix generation. CellRanger mkfastq was run using default parameters, and CellRanger count was run using the GRCh38-3.0.0 reference transcriptome from 10x Genomics and default parameters. For the K562 screen, cells with greater than 10% mitochondrial reads and less than 4000 UMIs were filtered out. For the iPSC-derived neuron screen, cells with greater than 17% mitochondrial reads and less than 1500 unique molecular identifiers (UMIs) were filtered out. After quality control filtering, 33,944 cells were retained in the K562 screen, and 51,183 cells were retained in the iPSC-derived neuron screen. The resulting count matrix output after this filtering was used for all downstream analyses.

16 *Neuron differentiation transcriptome projection*

Single-cell transcriptome data from a time course study of

iPSC-derived neurons²⁶ was downloaded from <https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-10632> (Accession No. E-MTAB-10632, matrices_timecourse.tar.gz), and integrated with the neuron CRISPRa screening dataset described here. Seurat v4 was used for all data analyses³⁹. The CRISPRa dataset was randomly downsampled to 5,000 cells for this analysis. Count matrices from both matrices were filtered to include only shared genes from the two datasets (n=14,777 genes). SelectIntegrationFeatures() and FindIntegrationAnchors() were run using default parameters to identify anchors for integration. 20,606 anchors were found and 2,953 anchors were retained for data integration. IntegrateData() was run using the retained 2,953 anchors to integrate the two datasets. After integration, standard Seurat single-cell analysis was performed to scale the data, and run the PCA and UMAP algorithms.

gRNA assignment and differential gene expression testing

17 Genomic coordinates (hg38) for final gRNA spacers were isolated using a loop built around the matchPattern() function from the BSgenome package⁴⁰. A 2Mb window (1Mb upstream and downstream) around each gRNA was then calculated and all genes within the 2Mb window were isolated using a loop built around ENSEMBL biomaRt getBM() function^{41, 42}. These 1Mb neighboring gene sets were then filtered to unique entries (unique HGNC symbols) for compatibility with the Seurat FindMarkers() function used in DE testing. A global UMI filter of 5 gRNA UMIs/cell was used to assign gRNAs to single cell transcriptomes for both K562 and iPSC- derived neuron datasets (note this heuristic threshold was chosen based on manual inspection of the UMI count distributions for each gRNA and prior work)². gRNA UMI counts for each cell were derived from the count matrix of passing cells output by Cell Ranger (which applies an automatic total UMI threshold to cells) and which also passed QC. Expression of a given gene within 1Mb of the gRNA of interest was compared between all cells with a given gRNA and all other cells as control. log2() fold changes in expression for a given gene were calculated using the Seurat FindMarkers() function with the following arguments: ident.1 = gRNA_Cells, ident.2 = Control_Cells, min.pct = 0, min.cells.feature = 0, min.cells.group = 0, features = target_gene, logfc.threshold = 0. A Wilcoxon rank-sum test was used to generate raw differential expression *P*-values. This process was then iterated for all genes within 1Mb of all gRNAs. NTCs were tested against all genes within 1Mb of any targeting gRNA. Only tests involving genes detected in >0.2% of test gRNA and control cells were carried forward. These raw differential expression *P*-values were then used to calculate empirical *P*-values to call EFDR < 0.1 sets². Specifically, an empirical *P*-value was calculated for each gRNA-gene test as:

$$\frac{[(\text{the number of NTCs with a } P\text{-value lower than that test's raw } P\text{-value}) + 1]}{[\text{the total number of NTCs tests} + 1]}$$

Empirical P -values were then Benjamini-Hochberg corrected, and those < 0.1 were kept for 10% EFDR sets. Log2 fold changes between gRNA and control cells were visualized using the gviz package⁴³ along with tracks for RefSeq transcripts (ENSEMBL biomaRt), H3K27ac, and ATAC seq peaks. The K562 ATAC and H3K27ac data were downloaded from ENCODE⁴⁴. ATAC-seq and H3K27ac CUT&RUN data from 7-8 week old NGN2-iPSC inducible excitatory neurons was obtained from Song et al. 2019³². As previously described, ATAC-seq and CUT&RUN reads were trimmed to 50bp using TrimGalore with the command `--hardtrim5 50` before alignment (<https://github.com/FelixKrueger/TrimGalore>). ATAC-seq reads were realigned to hg38 using the standard Encode Consortium ATAC-seq and ChIP-seq pipelines respectively with default settings and pseudo replicate generation turned off. Trimmed, sorted, duplicate and chrM removed ATAC-seq bam files from multiple biological replicates were combined into a single bam file using samtools merge v1.10⁴⁵. Trimmed CUT&RUN reads were realigned to hg38 using Bowtie2 v2.3.5.1 with the following settings `--local --very-sensitive-local --no-mixed --no-discordant -I 10 -X 700` and output sam files were convert to bam format using samtools view^{45, 46}. Duplicated reads were removed from the CUT&RUN bam file using Picard MarkDuplicates v2.26.0 with the `--REMOVE_DUPLICATES =true` and `--ASSUME_SORTED=true` options (<http://broadinstitute.github.io/picard/>). Finally, bam files were converted using the bedtools genomecov followed by the UCSC bedGraphToBigWig utility.