



## Supplementary Materials for

### **Nested epistasis enhancer networks for robust genome regulation**

Xueqiu Lin *et al.*

Corresponding author: Lei S. Qi, stanley.qi@stanford.edu

*Science* **377**, 1077 (2022)  
DOI: 10.1126/science.abk3512

#### **The PDF file includes:**

Materials and Methods  
Supplementary Text  
Figs. S1 to S17  
References

#### **Other Supplementary Material for this manuscript includes the following:**

MDAR Reproducibility Checklist  
Tables S1 to S4  
Movies S1 to S4

## Materials and Methods

### Cell culture

K562 (human myelogenous leukemia cells, female, ATCC CCL-243) and Jurkat cells (human T lymphoblast, male, ATCC TIB-152) were cultured in RPMI-1640 (Thermo Fisher Scientific, Waltham, MA) with 10% Tet-system-approved FBS (Life Technologies), GlutaMAX (Thermo Fisher Scientific), and Penicillin-Streptomycin (Thermo Fisher Scientific). K562 cells with stably integrated doxycycline (Dox)-inducible dCas9-KRAB were constructed via lentiviral transduction and cultured in the same condition as in Gilbert *et al.*, 2014 (30). Jurkat cells with stably integrated EF1 $\alpha$ -dCas9-KRAB and K562 cells with stably integrated EF1 $\alpha$ -dCas9 were constructed via lentiviral transduction. HEK293T cells (human embryonic kidney epithelial, female, ATCC CRL-3216) were cultured in DMEM with GlutaMAX (Thermo Fisher Scientific) in 10% Tet-system-approved FBS (Life Technologies). All cells were cultured at 37°C and 5% CO<sub>2</sub> in a humidified incubator.

### Plasmid construction

To clone individual sgRNAs, the lentiviral vector (pSLQ1373) was linearized, and gel purified as reported in Chen *et al.*, 2013 (53). New sgRNA sequences were PCR amplified from pSLQ1373 using different forward primers and a common reverse primer, gel purified, and ligated to the linearized pSLQ1373 vector using In-Fusion cloning (Clontech). Primers used to construct individual sgRNAs are shown in **table S1**.

To clone double sgRNAs, we used a two-step cloning procedure (**fig. S1B**) (54). First, individual sgRNA sequences were PCR amplified, and inserted into the pSLQ5004 (Addgene plasmid #118594) vector linearized by BstXI and XhoI, that contains a different (hU6) promoter than pSLQ1373. The same individual sgRNA sequences were also cloned onto pSLQ1373. Second, the double sgRNA expression vectors were constructed by inserting PCR products, amplified from pSLQ5004 containing one sgRNA into NsiI-linearized pSLQ1373 vector with another sgRNA.

### Construction of the multiplexed CRISPRi sgRNA library

We designed a library of 295 individual sgRNAs including 221 sgRNAs targeting *MYC* enhancers and two types of negative control sgRNAs (total 74), 34 sgRNAs targeting control regions in the genome, and 40 sgRNAs without a targeting sequence in the genome. The *MYC* sgRNAs were designed to tile densely on all enhancers and the promoter (26). We excluded sgRNAs containing NsiI restriction sites, which were used for the double sgRNA library construction. We synthesized individual oligos encoding sgRNAs in a 96-well format (IDT) and cloned each sgRNA into pSLQ1373 individually as described above. We also cloned the same individual sgRNAs into pSLQ5004 as previously described. After sequencing validation, we manually mixed all pSLQ1373-sgRNA constructs in equal amounts, and all pSLQ5004-sgRNA constructs in equal amounts. To generate the pooled double sgRNA library, the sgRNA sequence were PCR amplified from the pooled pSLQ5004-backbone sgRNA constructs, gel purified, and inserted into the NsiI-digested pooled pSLQ1373-backbone sgRNA constructs using In-Fusion cloning (Clontech). The double sgRNA-library pools were prepared in Stellar competent cells (Takara) and purified with a Plasmid Maxi Kit (Qiagen) (54). Primers (IDT) used to construct individual sgRNAs and the primers to quantify the representation of each paired sgRNA construct by MiSeq (Illumina) sequencing were listed in **table S1**. The sgRNA sequences were shown in **table S2**.

### Lentivirus production

To produce lentivirus, HEK293T cells were transiently transfected with pHR constructs, and packaging plasmids pCMV-dR8.91 and PMD2.G. Lentivirus was collected 72 hours after transfection by filtering supernatant through 0.45mm filters. When necessary, virus supernatant was concentrated using Lenti-X concentrator at 4°C overnight and centrifuged at 1500 g for 30 minutes at 4°C to collect virus pellets. The pellets were resuspended in cold culture medium, and directly added to cells or frozen at 80°C.

### Pooled CRISPRi screens

We transduced K562 cells harboring the doxycycline-inducible dCas9-KRAB at a multiplicity of infection (MOI) of 0.3 at a coverage of 1,000 transduced cells per sgRNA as previously described (26, 27). Forty-eight hours post transduction, we selected transduced cells with 2 µg/mL puromycin for 48 hours and collected 200 million cells as a reference sample. After maintaining cells at 1,000X coverage in 1 µg/mL puromycin and 1 µg/mL doxycycline for 30 doublings, we collected 200 million cells from the final cell population. We extracted genomic DNA from both the reference and final cell populations using the QIAamp DNA Blood Maxi kit (Qiagen) according to the manufacturer's instructions. We amplified sgRNA integrations from 500 µg genomic DNA by PCR with indexed sgRNA sequencing library primers containing Illumina adaptors and sequenced them on MiSeq using custom Illumina sequencing and index primers (table S1).

### Immunostaining

Immunostaining experiments were performed on cells seeded on a plate (Ibidi) that had been coated with Poly-L-lysine at room temperature. Cells were washed twice with PBS, fixed in 4% paraformaldehyde for 15 min at room temperature, permeabilized and blocked with 0.1% Triton X-100, 5% donkey serum in PBS (blocking buffer) for 1 hour at room temperature. After three washes with PBS, cells were incubated with primary antibodies. The BRD4 primary antibody was purchased from Abcam (Cat. No. ab128874). Cells were incubated with primary antibodies at 4°C overnight, then washed three times with PBS. After staining with corresponding secondary antibodies in the blocking buffer at 4°C overnight, cells were washed three times with PBS and stained with DAPI. The secondary antibody was purchased from Thermo Fisher Scientific (Cat. # A-31573). Washed cells were examined using a Nikon spinning disk confocal microscope.

### 2D and 3D DNA-FISH

The *MYC* genomic loci was labeled by commercial DNA-FISH probes according to the manufacturer's protocols in fixed K562 cells. The DNA-FISH probes were purchased from Biocare Medical (Cat. No. PFA 7322 V).

DNA-FISH for the *MYC* loci was performed according to a described protocol (55). To prepare samples, cells were grown in a chamber slide with a removable 12 well silicone chamber (Ibidi), fixed with 4% PFA for 20 min, washed with PBS twice, and permeabilized in 70% ethanol at 20°C overnight. Afterwards, samples were treated with methanol/acetic acid (4:1, prechilled at -20°C) at room temperature for 1 hour, washed with PBS twice, treated with 0.1 mg/ml RNase A in PBS at 37°C for 1 hour, washed again with PBS twice, and dehydrated by consecutive incubations in 70%, 85% and 100% ethanol. After air-drying, samples were heated at 95°C for 10 minutes in 70% denaturing buffer (2\*SSC, 70% formamide) and washed with 2\*SSC. For DNA-

FISH, hybridization was performed with 10 nM - 200 nM fluorescent nucleotide probes in hybridization buffer (2\*SSC, 50% formamide, 10% Dextran Sulfate) at 37°C overnight. After hybridization, samples were washed three times in 50% wash buffer (2\*SSC, 50% formamide, 0.1% Triton X-100) and then washed once with 2\*SSC. After removing the silicone chamber and drying the slide, the samples were treated with Prolong Diamond Antifade Mountant with DAPI (Thermo Fisher, P36966) and covered with cover glasses for imaging. Commercial FISH probes were hybridized according to the manufacturer's protocols.

#### Microscopy and image analysis

Microscopy was performed on a Nikon TiE inverted confocal microscope equipped with an Andor iXon Ultra-897 EM-CCD camera and 405nm, 488nm, 561nm, and 642nm lasers, the 100x PLAN APO oil objective (NA = 1.49), the 60x PLAN APO oil objective (NA = 1.40) or the 60x PLAN APO IR water objective (NA = 1.27). Images were taken using NIS Elements version 4.60 software by time-lapse microscopy with Z stacks at 0.2  $\mu$ m steps.

We used Fiji (ImageJ) (56) to quantify the BRD4 intensity at the *MYC* locus. Specifically, we used the "Find Maxima" to automatically find the XY coordinates of distinct *MYC* loci and measured the BRD4 intensity at the specific XY coordinates of each *MYC* locus. The background intensity was averaged from regions with no cells and subtracted from the raw BRD4 intensity. **Fig. 4D** and **fig. S11C&D** were plotted with a GraphPad showing mean and standard deviation. **Fig. 4F** was plotted with GraphPad showing mean and individual data points. The numbers of counted loci were listed at the bottom of each figure. The *P* values were calculated by two-tailed Fisher's exact test for paired sgRNAs effect versus the expected additive effect. The *P* values in **Fig. 4F** were calculated by using a two-sided t-test for paired sgRNAs effect versus the expected additive effect.

We used the Imaris software (Bitplane) to perform 3D visualization and reconstitution of representative cells showing the BRD4 colocalization with the *MYC* locus under different perturbation conditions. *MYC* DNA-FISH punctae were built using the Spots function (estimated XY diameters: 0.5-1  $\mu$ m), and the BRD4 surface was built using the surface function (smooth, 0.01-0.1  $\mu$ m surface detail).

#### Quantitative real-time PCR (qRT-PCR)

The qRT-PCR experiments were performed to determine the fold change of the expression of *MYC* and other tested genes. For each sample, total RNA was isolated by using the RNeasy Plus Mini Kit (Qiagen Cat. No. # 74134), followed by cDNA synthesis using the iScript cDNA Synthesis Kit (BioRad Cat. No. # 1708890). qRT-PCR was performed using the PrimePCR assay with the SYBR Green Master Mix (Bio-Rad) with primers ordered from IDT on a Bio-Rad CFX384 real-time system (C1000 Touch Thermal Cycler), according to manufacturer's instructions. Cq values were used to quantify gene expression. The relative expression of the *MYC* and other tested genes was normalized to a GAPDH internal control. The *P* values were calculated by using the two-sided t-test. All combinations of two single inhibition effects were used to calculate the mean additive effects and the standard deviation.

#### Cell viability assay

The cell viability assay was performed using Alamarblue cell viability reagents (Thermo Fisher Scientific), which quantifies the cell proliferation rate by measuring the metabolic activity

of the cells. For each condition, 100  $\mu$ l of cells were added to each well and seeded at equal concentration (5,000-10,000 cells/well) in the 96-well plate. At the time of detection, 100  $\mu$ l of Alamarblue reagents were added to each well and the plates were incubated at 37°C for 1 hour. After that, the fluorescent intensity was measured in the Synergy H1 microplate reader (Biotek) using an excitation wavelength of 540 nm and emission wavelength of 585 nm. Average fluorescent intensity of wells containing only 100  $\mu$ l culture medium was used as blanks. For each well, the relative fluorescent intensity was calculated by subtracting background (average intensity of blank wells) from raw fluorescent intensity values.

#### Trac-looping

Modified Trac-looping (34) was performed to map chromatin spatial contacts and chromatin accessibility under four conditions for the SRE pair e3&e7: wildtype, perturbation of individual e3 or e7, and perturbation of both e3&e7; and non-SRE pair e1&e4: wildtype, perturbation of individual e1 or e4, and perturbation of both e1&e4. Briefly, one million cells were fixed with 1% formaldehyde in the culture medium at room temperature for 10 minutes. Cells were washed with PBS and resuspended with a reaction buffer (50 mM Tris acetate, pH 7.5, 150 mM potassium acetate, 10 mM magnesium acetate, 4 mM spermidine, 0.5% NP-40). The short adapter and biotinylated bridging linker with Tn5, and then incubated at room temperature for 20 minutes to generate Tnp bridging complex. This complex was added to cells and mixed gently. Chromatin interacting regions were bridged at this step. EDTA was added to stop the reaction and reverse crosslinked by incubating at 65°C overnight. Genomic DNA was extracted by Phenol-Chloroform and the purified DNA was digested with restriction enzymes MluCI/NlaIII. Streptavidin beads were used to enrich biotin-labeled DNA fragments, followed by adapter ligation and amplification by PCR with Illumina multiplexing indexed primers. Libraries were purified and submitted for paired-end Hi-seq sequencing (Illumina).

#### ATAC-seq

For ATAC-seq, 50,000 cells were collected and washed with PBS. After centrifugation, cell pellets were resuspended by gently pipetting up and down 10 times with 100  $\mu$ L cold Lysis Buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% NP-40, 0.1% Tween-20, 0.01% Digitonin) and then kept on ice for 3 minutes. 1 mL Wash Buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% Tween-20) was added to the tube and mix gently. The nuclei were then pelleted by centrifuging at 1000 x g for 10 minutes at 4 °C. The nuclei pellets were resuspended with 1 x Transposition Reaction Buffer (50 mM Tris-acetate, pH 7.5, 150 mM potassium acetate, 10mM magnesium acetate, 4 mM spermidine, 0.1% Tween-20, 0.01% Digitonin). 5  $\mu$ L homemade ATAC-seq Tnp complex was added, then mixed well gently and place in a 37 °C thermomixer with 1000 rpm shaking for 30 minutes. After the reaction, DNA were purified with MinElute Reaction Cleanup Kit (QIAGEN, Cat. No. 28206). DNA were eluted with 15  $\mu$ L elution buffer, followed by PCR reaction by mixing with 2.5  $\mu$ L each indexed primer (25  $\mu$ M) and 20  $\mu$ L NEBNext High-Fidelity 2X PCR Master Mix (NEB, Cat. No. M0541S). The following PCR program was used: 72°C for 5 minutes; 98°C for 30 seconds; 11 cycles of 98°C for 10 seconds, 63°C for 30 seconds, and 72°C for 1 minute; 72°C for 5 minutes. Then 200-650 bp DNA fragments were purified for sequencing.

#### ChIP-seq

Cells were fixed with 1% Formaldehyde in culture medium at room temperature for 10 minutes. The reaction was quenched by adding 250 mM Glycine, then cells were washed twice with cold PBS. Three million fixed cells were used for each H3K9me3 ChIP-seq assay. Cells were resuspended with 1 x TE buffer plus 1 x protease inhibitor cocktail and 1 mM PMSF. Chromatin were sheared by sonication on a Diagenode Bioruptor Pico device at 4 °C for 18 cycles with 30 seconds on and 30 seconds off. Chromatin solution was adjusted to 1 x RIPA buffer (1xTE, 0.1% SDS, 0.1% Sodium Deoxycholate and 1% Triton X-100) plus 500 mM NaCl. After centrifuging at 13,000 rpm for 10 min at 4 °C, chromatin supernatant was collected and 10% of it was saved at input. 2 µg H3K9me3 antibody (Abcam, Cat. No. ab8898) or H3K27ac antibody (Abcam, Cat. No. ab4729) was mixed with 20 µL Dynabeads Protein A beads (Thermo Fischer Scientific, Cat. No. 10001D). Then the beads were incubated at room temperature for 1 hour with rotation. The beads were washed once with 1 x PBS, then added with chromatin solution to beads, and incubated at 4 °C overnight with rotation. Next, the beads were washed three times with RIPA buffer plus 500 mM NaCl, then three times with RIPA buffer plus 1 M NaCl, then three times with LiCl buffer (1 x TE, 250 mM LiCl, 0.5% NP-40, 0.5% Sodium Deoxycholate), and finally three times with 1 x TE buffer. DNA were eluted and reversed crosslinking with protease K digestion and incubating at 65 °C for over 6 hours. DNA was purified with MinElute Reaction Cleanup Kit (QIAGEN, Cat. No. 28206). DNA ends were repaired with End-It DNA End-Repair Kit (Lucigen, Cat. No. ER0720). Then “A”-tailing reaction was performed with Klenow Fragment (3' → 5' exo-) (NEB, Cat. No. M0212S) provided with dATP, then followed by adapter ligation with T4 DNA ligase (NEB, Cat. No. M0202S). Finally, the library was indexed and amplified by PCR, and DNA fragments between 200 bp and 600 bp were purified and sequenced.

#### PCR, cloning and sequencing for knockout enhancer pairs

For enhancer knockout (e1&e4 or e3&e7), we created a stable K562 cell line expressing Cas9-mCherry. We transduced the lentivirus encoding sgRNA-BFP targeting e1 or e3, and/or another lentivirus encoding sgRNA-GFP targeting e4 or e7 into Cas9-mCherry-K562 cells. We used three different fluorescent markers to sort double positive cells (thus single enhancer knockout) or triple positive cells (thus double enhancer knockout) by fluorescence-activated cell sorting (FACS). The cells with one enhancer deletion (e1, e3, e4, and e7) were cultured for 1 week for qRT-PCR experiment. Due to the high cytotoxicity from multiple DNA cleavage and the low expression of *MYC* from double enhancer knockout, we waited a longer time (~3 weeks) for the cells with double enhancer knockout to grow to an enough quantity for qRT-PCR.

To tested if knockout enhancer pairs created large deletion, primers were used to amplify segments. PCR reactions were performed using Kapa Hotstart PCR kit (Fisher Scientific) on BioRad T100 Thermal cycler under the following conditions: 95°C for 3 min; 50 cycles of 98°C 20 s, 65°C 20 s, 72°C 1 min, and final extension 72°C for 10 min. The PCR products were gel purified using Nucleospin Gel and PCR clean up kit, cloned into TOPO vectors using Zero Blunt TOPO PCR cloning kit (Thermo Fisher), and transfected into Stellar Competent cells (Takara/Clontech). For each PCR amplicon, ten colonies were picked for plasmid extraction and sequencing. sgRNA sequences are listed in **table S1**.

#### JQ1 treatments

K562 cells with stably integrated dCas9-KRAB were seeded at a density of  $1 \times 10^5$  cells/well in 24-well plates. The next day, individual or paired sgRNAs targeting enhancer were transduced into cells. One day later, 1 µg/ml of puromycin was added for 2 days. JQ1 was then added to the

cells at different concentrations (0.001  $\mu$ M, 0.01  $\mu$ M, 0.25  $\mu$ M, and 0.5  $\mu$ M) for 2 days. After the treatment, the cells were harvested for qRT-PCR analysis.

#### Sequence alignment and the calculation of depletion score

We calculated the count matrix by exact match for the first 8bp of both ends of each paired end reads (**table S2**). We ensured the high reproducibility of the multiplexed CRISPRi screening by calculating counts between replicates of the same condition (Pearson correlation coefficient  $R = 0.968$  for the day 0 (D0) sample and 0.985 for the day 30 (D30) sample). To obtain sufficient representative reads for the downstream calculation, counts of sgRNA pairs were first filtered by requiring all paired constructors with at least 30 reads in D0, then added with a pseudocount of 10. The  $\log_2$  enrichment scores of all sgRNA pairs were then calculated in D30 relative to D0 by using the relative frequencies in D0 and D30. The mean and the standard deviation of enrichment scores for paired control sgRNAs (control-control; control sgRNAs: sgRNAs targeting control regions + sgRNAs without targeting regions; **fig. S1A**) were used to normalize the  $\log_2$  enrichment scores. The normalized  $\log_2$  enrichment scores were represented as the depletion score (See below equation;  $\text{Log}_2 \text{ES}$ :  $\text{Log}_2$  enrichment score). **table S2** shows the raw data of sequenced counts of sgRNA pairs at different time points for two biological replicates.

$$\text{Depletion Score} = \frac{\text{Log}_2 \text{ES} - \mu_{\text{Log}_2 \text{ES}(\text{control-control})}}{\sigma_{\text{Log}_2 \text{ES}(\text{control-control})}}$$

#### Calculation of epistasis interaction scores for sgRNA pairs and enhancer pairs

The single depletion scores were calculated for each sgRNA from the averaged depletion scores of the paired sgRNAs which contain the given sgRNA and one control sgRNA. The two permutations (A/B and B/A) of the same sgRNA pairs in our double CRISPRi tiling screening are consistent (Pearson correlation coefficient  $R = 0.49$ ;  $P < 2.2\text{e-}16$ ; The coefficient  $R$  increased up to 0.75 by smoothing the adjacent sgRNAs; **fig. S2A**). Then the double depletion score of the sgRNA pair was calculated by averaging the same sgRNAs in two permutations (A/B and B/A) to obtain the symmetric depletion score matrix. The epistasis interaction scores of the sgRNA pairs between a given sgRNA (Query sgRNA) and other sgRNAs were calculated based on their depletion scores. To do this, the depletion scores of the sgRNA pairs were first plotted against the single depletion score for all the sgRNAs paired with the Query sgRNA. Next, a linear curve was fitted to the distribution of data points (**fig. S2D**).

$$\text{Double Depletion Score}_{\text{Query sgRNA} + \text{sgRNA}_i} \sim b_0 + \beta * \text{Single Depletion Score}_{\text{sgRNA}_i}$$

$b_0$ : Single Depletion Score of Query sgRNA

The interaction score of Query sgRNA and another sgRNA (sgRNA<sub>x</sub>) was computed as the negative deviation between the observed double depletion score and the expected value of the fitted linear curve to the single depletion score of sgRNA<sub>x</sub> (**fig. S2D**). All interaction scores were normalized to the mean and the standard deviation of the interaction scores of the Query sgRNA and control sgRNA pairs.

To calculate the epistasis interaction scores of enhancer pairs, we ranked the single depletion scores of sgRNAs targeting the same enhancer and selected the top 50% sgRNAs to represent each enhancer. The interaction scores of the enhancer pairs were calculated by averaging all double

depletion scores of the top 25% sgRNA pairs (combinations of the top 50% sgRNAs targeting individual enhancers) targeting the two enhancers.

#### Prediction of SREs using the elastic net regularized linear regression model

Spatial interaction ( $SI_{ij}$ ) between paired enhancers ( $i$  &  $j$ ) were generated with Juicebox (57) using vanilla coverage square root (VC\_SQRT) normalized H3K27ac HiChIP signal at 5Kb resolution (58). To examine the co-occupancy of transcription factor (TF) binding signal and histone modification (HM) profiles at enhancer pairs, the genome-wide binding profiles of TFs in K562 cell line were downloaded from <http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeSydhTfbs> and Gene Expression Omnibus (GEO) database, and the HM profiles in K562 cell line were retrieved from ENCODE data portal (5, 59). The sample ID or individual download links for each TF/HM are listed in **table S3**. Data downloaded from GEO were processed with Bowtie (60) and MACS2 (61) to generate the sequence depth normalized bigwig file. To obtain the representative binding signals for enhancer pairs, we filtered TFs which didn't have one enrichment peak co-locating with our selected enhancers, then the binding profiles of 38 TFs were kept for TF co-occupancy measurements (**table S3**). We retrieved the enrichment signal of each TF and HM at each enhancer, and further normalized through Z-score transformation. The co-occupancy ( $CO_{ij,k}$ ) of TF binding signal or HM profile ( $k$ ) at two enhancers ( $i$  &  $j$ ) was defined as:

$$CO_{ij,k} = Z_{i,k} \times Z_{j,k}$$

We used epistasis interaction scores as predictors in a linear regression model:

$$\begin{aligned} interaction\ score_{ij} &= \alpha + X_{ij}\beta^T + \epsilon_{ij} \\ X_{ij} &:= (SI_{ij}, CO_{ij,1}, \dots, CO_{ij,k}) \\ \beta &:= (\beta_{PI}, \beta_1, \dots, \beta_k)^T \end{aligned}$$

The regression model is fitted using the glmnet R package (version 3.0-2) by elastic net regularization with the alpha set to 0.5, and the penalty parameter  $\lambda$  which gives minimum mean squared error in cross-validation curve. We kept the 9 features (H3K27ac HiChIP and another 8 profile features) with nonzero coefficients. The model was evaluated by 10-fold cross-validation.

We adapted a two-step method to evaluate the importance of features (62). First, we clustered features into feature groups by hierarchical clustering with complete linkage. The features within each group have pairwise absolute Pearson correlation coefficient of at least 0.1. Second, we evaluated the importance of each feature group by estimating the decrease of the mean squared error (m.s.e.) when that feature group was left out. Inside each group, we ranked each feature by its absolute Pearson correlation coefficient with the interaction scores, and the feature with the highest correlation was chosen as the representative feature of that group. We compared the performance of the elastic net model with a simple linear model of individual representative features in each feature group by using the Bayesian information criterion (BIC) (**fig. S8D**). Then the relative BIC was calculated by subtracting the BIC of the HiChIP simple linear model.

For SRE prediction of *BCL9* and *KTN1* in K562 cells, we used the predicted enhancers from the activity-by-contact (ABC) model (63), and applied our trained elastic net model to prioritize



the SRE pairs from the enhancer pairs. For SRE prediction of *MYC*, *CHD7*, and *CD180* in GM12878 B-lymphoblastoid cells, and *COX6C* and *FOXP1* in Jurkat cells, we downloaded the TF binding profiles and HM profiles of GM12878 cells and Jurkat cells from ENCODE data portal or GEO database (**table S3**) (5, 59), which are finally selected features in the elastic net regularized linear regression model. We used the predicted enhancers from the ABC model (63), then applied our model with these downloaded data to predict SREs for these five genes in the GM12878 and Jurkat cells.

#### Trac-looping data analysis

Raw FASTQ paired-end reads were trimmed, mapped to hg19 and background filtered by `tracPre2.py` in `cLoops2` package (64). Then paired-end tags (PETs,  $\text{MAPQ} \geq 10$ ) were converted to non-redundant BEDPE files for each sample. `cLoops2` package is available at <https://github.com/YaqiangCao/cLoops2/blob/master/scripts/tracPre2.py>.

We conducted the signal of chromatin accessibility and peak calling using the non-redundant small PETs (<1kb) for genome-wide hypersensitive sites by MACS2 (61) with parameters, including “MACS\_pvalue=0.01”, “MACS\_extsize=200” and “MACS\_shiftsize = -37” and “MACS\_broad”. And the accessibility signal was normalized to signal per million reads (SPMR). The change of chromatin accessibility was calculated with  $\log_2$  fold change of SPMR between the perturbation cells and wildtype cells at each enhancer locus.

We adapted the method to identify significant interactions in the original version of Trac-looping for robust interactions (34). Briefly, we used the model in Mango (65) to calculate the statistical significance of an “interaction” between two HSs with 3kb bin size. Only significant interactions with  $\text{PETs} \geq 3$  and  $\text{FDR} \leq 0.05$  were used in the following analyses. Then significant interactions in the perturbation samples and wild type samples were merged as an interaction union set. The dynamic changes of spatial contacts between two enhancers were calculated as the library-size normalized  $\log_2$  fold change of interacting PETs between perturbation samples and wild type in interaction union sets which overlap enhancer pairs in two ends. The overlapping was determined by bedtools “pairToPair” command (66).

#### ChIP-seq data analysis

Reads were mapped to human genome hg19 using Bowtie (60). Reads that could be mapped to multiple locations were removed. To remove PCR resulted clonal reads, only unique mapped reads were kept for subsequent analysis. Then, the remaining reads were analyzed with MACS2 (61) to call broad peak and generate the sequence depth and quantile normalized bigwig file with parameters, including “MACS\_pvalue=0.01”, “MACS\_extsize=147”, “MACS\_nomodel” and “MACS\_broad”.

#### ATAC-seq data analysis

Reads were mapped to human genome hg19 using Bowtie (60). Reads that could be mapped to multiple locations were removed. To remove the PCR resulted clonal reads, only unique mapped reads were kept for subsequent analysis. These unique reads were analyzed with MACS2 (61) to call peak and generate depth and quantile normalized bigwig with parameters, including “MACS\_pvalue=0.01”, “MACS\_extsize=200” and “MACS\_shiftsize = -37” and “MACS\_broad”. The change of chromatin accessibility was calculated with  $\log_2$  fold change of SPMR between the perturbation cells and wildtype cells at each enhancer locus.

### Identification of enhancer-linked variants

Enhancer-linked variants are variants in the LD window which also overlap with the given enhancer. Pairwise variant  $r^2$  matrices were generated by the Plink program (67). A sliding window of size 3 was shifted with 1 variant step. The consecutive windows will be kept and merged to a Linkage Disequilibrium (LD) window if their average  $r^2$  value is larger than 0.2.

### The interactive model to estimate epistatic influence of paired SRE variants on gene expression

To examine the interactive effects of SRE variants on *MYC* gene expression in Acute Myeloid Leukemia (LAML) patients (termed as interactive model), we obtained genotype data (generated from The Cancer Genome Atlas:TCGA) from dbGAP (Affymetrix SNP 6.0 array; dbGAP ID: phs000178) (68). The corresponding RNA-seq data in LAML patients were downloaded from the UCSC Xena platform (69). The gene expression value is the  $\log_2(x+1)$  transformed RSEM normalized count from the UCSC Xena platform. The imputation analysis and the selection of covariates were done following the method instruction in the PanCanQTL study (70). The covariates included in subsequent analysis include the top five principal components (PCs) from population structure analysis, the first 15 PEER factors from PEER software (71), along with age and gender. The analysis of variant-variant interactive influence was performed in a linear regression model:

$$Y_{MYC-Exp} \sim \beta_0 + \beta_1 V1 * V2 + \beta_2 V1 + \beta_3 V2 + \sum \beta_i * Covariates_i$$

$V$  represents the genotype of the variant in the equation. The  $P$  value in the variant-variant interaction term generated in the fitted linear model was used to measure the interactive effects of two non-coding variants of a SRE pair on *MYC* expression and construct the quantile-quantile (QQ) plots.

To examine the interactive influence of SRE variants in regulating *MYC* expression predicted in GM12878 cell line, we downloaded publicly available genotypes and corresponding RNA-seq data from the B lymphoblasts of 373 European individuals (39). To remove the effect of population structure on gene expression as the original study, we used SNPRelate R package (72) to perform principal component (PC) analyses for these European populations and included the first three principal components as covariates in the above linear regression model.

### Logistic regression interactive model to estimate interactive effects of SRE variants on clinical risks

To examine the interactive effects of SRE variants on clinical risks, we assessed the relapse risk in childhood acute lymphocyte leukemia (ALL) patients for SRE variants in the *MYC* locus. We obtained the genotype data and corresponding relapse records of 1,593 childhood ALL patients in (dbGAP ID: phs000638) (41, 42). For genetic association analysis, we included the treatment arms and genetic ancestry as covariates in variant-variant interaction analysis. The analysis of variant-variant interactive effects on the relapse risk for childhood ALL patients was performed in a logistic regression model:

$$\text{Logit } E(\text{relapse status}) \sim \beta_0 + \beta_1 V1 * V2 + \beta_2 V1 + \beta_3 V2 + \sum \beta_i * Covariates_i$$

*V* represents “variant” in the equation. The *P* value in the variant-variant interaction term generated in the fitted linear model was used to measure the interactive influence of SRE variant pairs in relapse risk, then construct the QQ plots.

To examine the interactive influence of SRE variants for the disease risk of Crohn’s disease, we downloaded the genotypes and the corresponding case/control records from the Wellcome Trust Case Control Consortium (WTCCC) (40). The variant-variant analysis on 4,686 Crohn’s Disease (CD) patients was performed in a similar manner as the relapse risk analysis in childhood ALL patients.

### Web development

The website <http://enhancer.stanford.edu/> uses an Express.js server to serve pre-computed results using React.js. The website provides four primary functionalities: (1) a homepage on which users can query a cell line and gene of choice to visualize and download the output of SRE prediction model among the 4,835 ultralong distance enhancer networks across 6 cell types (GM12878, K562, Jurkat, A549, HUVEC and HCT116); (2) a webpage describing the workflow of multiplexed CRISPR screening and the methods analyzing the CRISPRi screen to calculate the depletion score and the epistasis interaction score with shared code; (3) a webpage introducing how users can run the SRE prediction model with shared code; (4) a webpage providing the gene-level output of applying the SRE prediction model in GM12878 cells to interpret epistatic effects of SRE variant pairs on the relapse risk of ALL patients for 61 genes.

### **Supplementary Text**

#### High reproducibility of the multiplexed CRISPRi tiling screen across biological replicates and different sgRNAs targeting the same enhancer

Replicated screens generated highly correlated estimates for the depletion score (Pearson *R* = 0.96) and the interaction score (Pearson *R* = 0.91) of sgRNA pairs (**fig. S2F and G**). In addition, dense sgRNAs targeting the same enhancer also showed a high correlation in both the depletion score and the interaction score (**fig. S2H and I**), indicating the high reproducibility of the multiplexed CRISPRi tiling screen for perturbing pairwise enhancers for *MYC* regulation.

#### Epistatic influence of SRE variants on gene expression and clinical risk

To explore epistasis influence of the variants linked to our identified SRE from multiplexed CRISPRi screen, we used transcriptome data along with corresponding individual genotype data for 113 acute myeloid leukemia (LAML) patients from The Cancer Genome Atlas (TCGA) data portal (**See Methods**). First, we defined SRE variants by associated variants with SREs, which are contained in the same linkage disequilibrium (LD) window within a given SRE pair ( $r^2 > 0.2$ ; **fig. S12A**). Next, a linear model with the variant-variant interaction term and other covariates were fitted to estimate the epistasis influence of SRE variants. The *P* value on the interaction term was used to quantify the epistasis influence of SRE variants. We compared the *P* values calculated by SRE variants (e4&e6) to those from random permutations in the QQ plot to determine if SRE variants interacted more frequently to alter *MYC* expression than expected by chance (**Fig. 5A**).

To explore SRE prediction in other cell lines and the epistasis influence of SRE variants in gene expression and clinical risk in other cell line related tissue and diseases, we used the GM12878 B-lymphoblastoid cell line with available data (**table S3**) required in ABC model and

predicted enhancers for *MYC* and the other two genes (*CHD7* and *CD180*). Then we applied the SRE prediction model with spatial contacts, BRD4 co-occurrence and other quantitative features to prioritize SRE pairs (**Fig. 5C and H, fig. S15A**). We named the enhancers in GM1878 B cells as Be. A linear regression model was used to explore the epistasis influence of SRE variants on the expression of *MYC*, *CHD7* and *CD180* by using SNP genotype data along with corresponding B-lymphoblast transcriptome data from 373 Europeans (39) (see **Methods, Fig. 5D, fig. S12H, fig. S14B, fig. S15C**).

To examine the epistatic influence of SRE variants in clinical risk, we used two independent genotype datasets with corresponding clinical data from 1593 Childhood acute lymphoblastic leukemia (ALL) patients and 4686 Crohn's disease (CD) patients, which are GM12878 cell type-relevant tissue and diseases data (40–44). Two logistic regressions were fitted with different SRE variants to estimate the epistatic influence on altering clinical risk to characterize the epistatic effect of SRE variants on ALL relapse risk and CD disease risk (see **Methods, fig. S13 A, B and E, fig. S14A**).

#### Systematical comparison between the interactive model and the additive model

To systematically compare the interactive effects of SRE pairs with expected additive effects for e4&e6 SRE (*MYC* locus in K562 cells) variants (**Fig. 5A and B**), Be1&Be7 SRE (*MYC* locus in GM12878 cells) variants (**Fig. 5D-G**), and Be1&Be11 SRE (*CHD7* locus in GM12878 cells) variants (**Fig. 5I**), we used the ANOVA function in R to compare the interactive model and additive model in explaining *MYC* expression (in 113 LAML patients and B-lymphoblast cells of 373 Europeans) and disease risk (ALL relapse risk and CD disease risk).

The additive model did not have the interactive term in both the linear regression model for gene expression and the logistic regression model for disease risks. The resulting *P* value of ANOVA function was used to see whether the interactive model is the better fit of the data than the additive model (**fig. S12B and E, fig. S13G**).

Additive linear regression model:

$$Y_{MYC-Exp} \sim \beta_0 + \beta_1 V1 + \beta_2 V2 + \sum \beta_i * Covariates_i$$

Additive logistic regression model:

$$Logit E (relapse status) \sim \beta_0 + \beta_1 V1 + \beta_2 V2 + \sum \beta_i * Covariates_i$$

In addition, we also used AIC (Akaike information criterion) score to access whether the interactive effect can better explain the data than the additive effect for the significant variant pairs from e4&e6, Be1&Be7 and Be1&Be11 SRE pairs. The AIC score looks at both explained variation but also model complexity and parsimony, with a lower AIC score meaning that a model is closer to the 'truth'. We calculated a relative AIC score as the AIC score of the additive model minus the AIC score of the interactive model. The higher relative AIC score means the interactive model is closer to the 'truth' than the additive model (**fig. S12C and F, fig. S13G**). Both results showed that SRE variants epistatically and not additively affected gene expression and altered disease risk.

#### Genome-wide analysis of epistatic influence of SRE variants on the relapse risk for ALL patients

To evaluate the epistatic influence of SRE variants on clinical risk for important genes regulated by ultralong distance enhancer network in genome-wide scale, we used the genotype data and the corresponding relapse risk of the childhood ALL patients of 1,593 individuals. The analysis workflow (**fig. S16**) showed how we applied our trained elastic net model to prioritize the SREs for important genes in genome-wide scale and leveraged the predicted SREs to interpret the epistasis of non-coding variants on the relapse risk for ALL patients.

Briefly, we selected genes which are regulated by multiple genes spanning an ultralong distance ( $\geq 5$  enhancer across  $>200\text{kb}$ ). These genes are enriched in GSEA immune-related and oncogenic signature gene sets (**fig. S16A, fig. S17**). Then we specifically focused on the important genes ( $n=70$ ), in which 61 genes have enough data signal for subsequent analysis. To evaluate the ability of predicted SRE score in interpreting epistasis influence on disease risk, we binned the enhancer pairs into 20 groups based on the SRE score and quantified the percentage of enhancer pairs for each bin, in which the linked non-coding variant pairs showed epistatic influence on the relapse risk. Specifically, we defined the enhancer pairs with observed epistatic effects on ALL relapse risk based on the QQ plot: 1) the mean of top30% interaction  $-\log_{10}(P \text{ value})$  for enhancer variant pairs  $> \text{Expected}$ ; 2) comparing with interaction  $-\log_{10}(P \text{ value})$  of random permutation, the KS test  $P$  value is  $< 0.05$  (**fig. S16B**).

We found that the SRE scores correlated with the epistatic effects of non-coding variants on clinical risk (**Fig. 6A**). Notably, comparing with the top 40% SREs, the enhancer pairs with SRE score below 40% significantly decreased the ability in interpreting epistasis influence on disease risk. In addition, all the experimental validated SRE (**Fig. 3A-D**) also have the top 40% SRE score. Therefore, we used top40% SRE score as a threshold and access the epistasis influence of SRE variants in clinical risk. Thus, we could observe that 27.9% SREs in 55.7% genes showed the epistasis effects of SRE variants in the relapse risk for the ALL patients (**Fig. 6B and C**).

Next, we determined whether the SRE epistasis model, SRE linked non-coding variant pairs, can detect more ALL pathogenic genes compared with traditional locus-by-locus model with individual variants. In the traditional individual locus-by-locus association model, we included the treatment arms and genetic ancestry as covariates in variant-variant interaction analysis. The analysis of individual enhancer variant effects on the relapse risk for childhood ALL patients was performed in a logistic regression model:

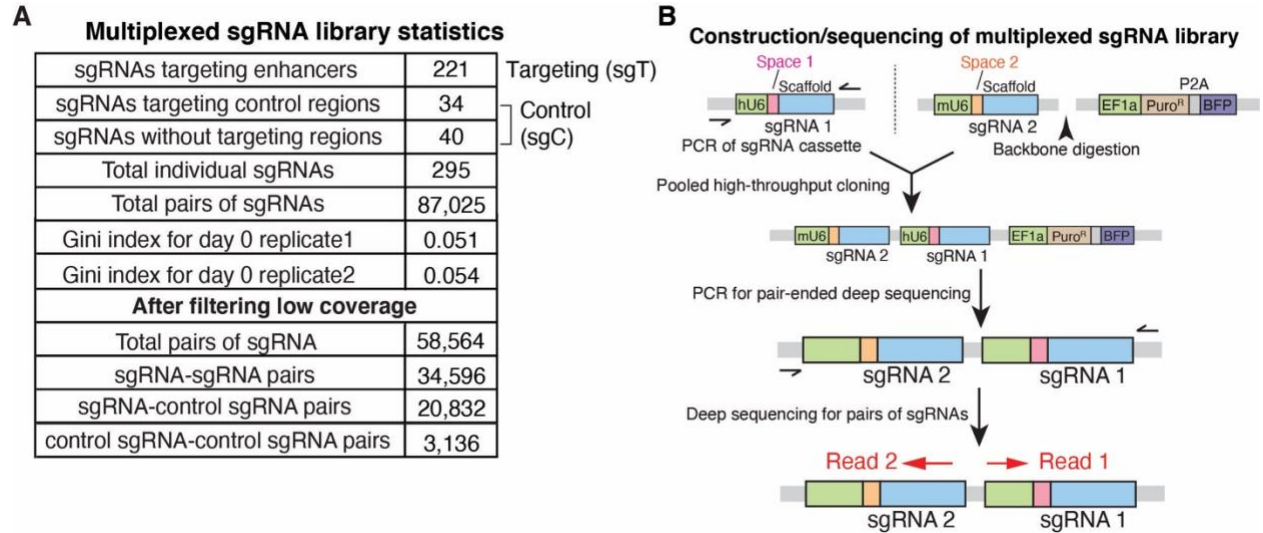
$$\text{Logit } E(\text{relapse status}) \sim \beta_0 + \beta_1 V + \sum \beta_i * \text{Covariates}_i$$

$V$  represents “variant” in the equation. The  $P$  value in the variant term generated in the fitted linear model was used to measure the influence of enhancer variants in relapse risk, then construct the QQ plots. We defined the individual enhancer with observed significant effects on ALL relapse risk based on the QQ plot: 1) the mean of top30%  $-\log_{10}(P \text{ value})$  for individual enhancer variants  $> \text{Expected}$ ; 2) comparing with interaction  $-\log_{10}(P \text{ value})$  of random permutation, the KS test  $P$  value is  $< 0.05$ .

Next, we defined the pathogenic gene based on these two different models. In SRE epistasis model, if 1) a given gene is regulated by  $\geq 1$  SREs which have observed epistasis effects on ALL relapse risk; 2) the odds ratio of one or some SRE variants are significantly  $> 1.25, 1.5, 1.75$  or  $2$ ,

we defined it as an ALL pathogenic gene. In traditional individual locus-by-locus association model, if 1) a given gene is regulated by  $\geq 1$  enhancers which have observed significant effects on ALL relapse risk; 2) the odds ratio of one or some enhancer linked variants are significantly  $>1.25$ ,  $1.5$ ,  $1.75$  or  $2$ , we defined it as an ALL pathogenic gene. In 61 important genes, 22 genes are pathogenic genes in ALL based on the literature search (Have mutation or over-/under-/do not express in ALL patients; **table S4**). Among these 22 ALL pathogenic genes, comparing with the traditional individual locus-by-locus association model, our SRE interactive model can identify significantly more ALL pathogenic genes (**Fig. 6D**). Specifically, our SRE model recovered 10 genes, whereas the locus-by-locus model only showed 2 genes at odds ratio  $>1.5$ . In addition, we included this genome wide SRE epistasis analysis associated with clinical risk on the website (<http://enhancer.stanford.edu/>).

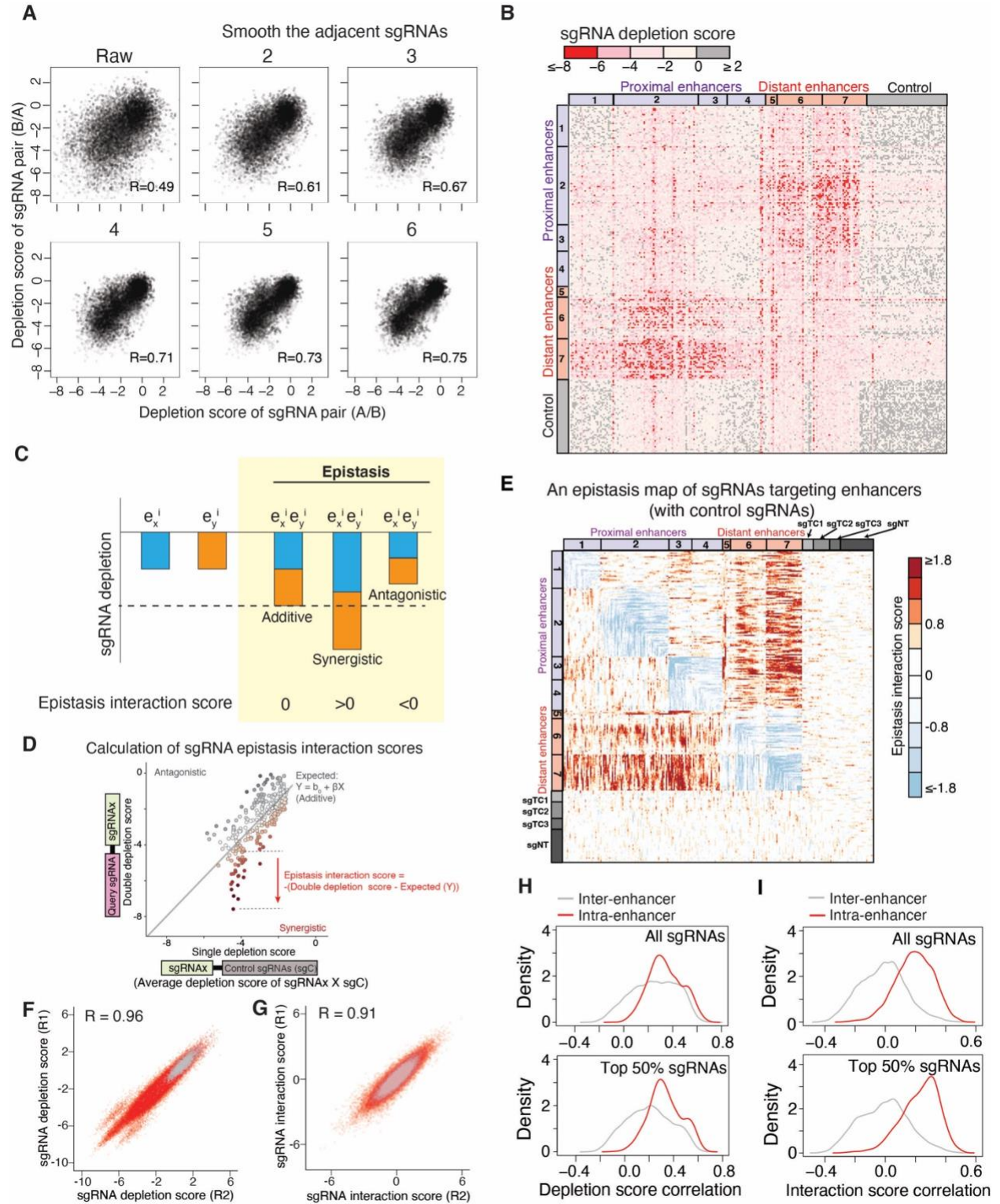
## Supplemental Figures



**Fig. S1. Experimental setup for the multiplexed CRISPRi screening.**

**A**, A summary table of the sgRNA library for multiplexed CRISPRi screening and the sgRNA pairs used in the subsequent analysis after filtering low coverage.

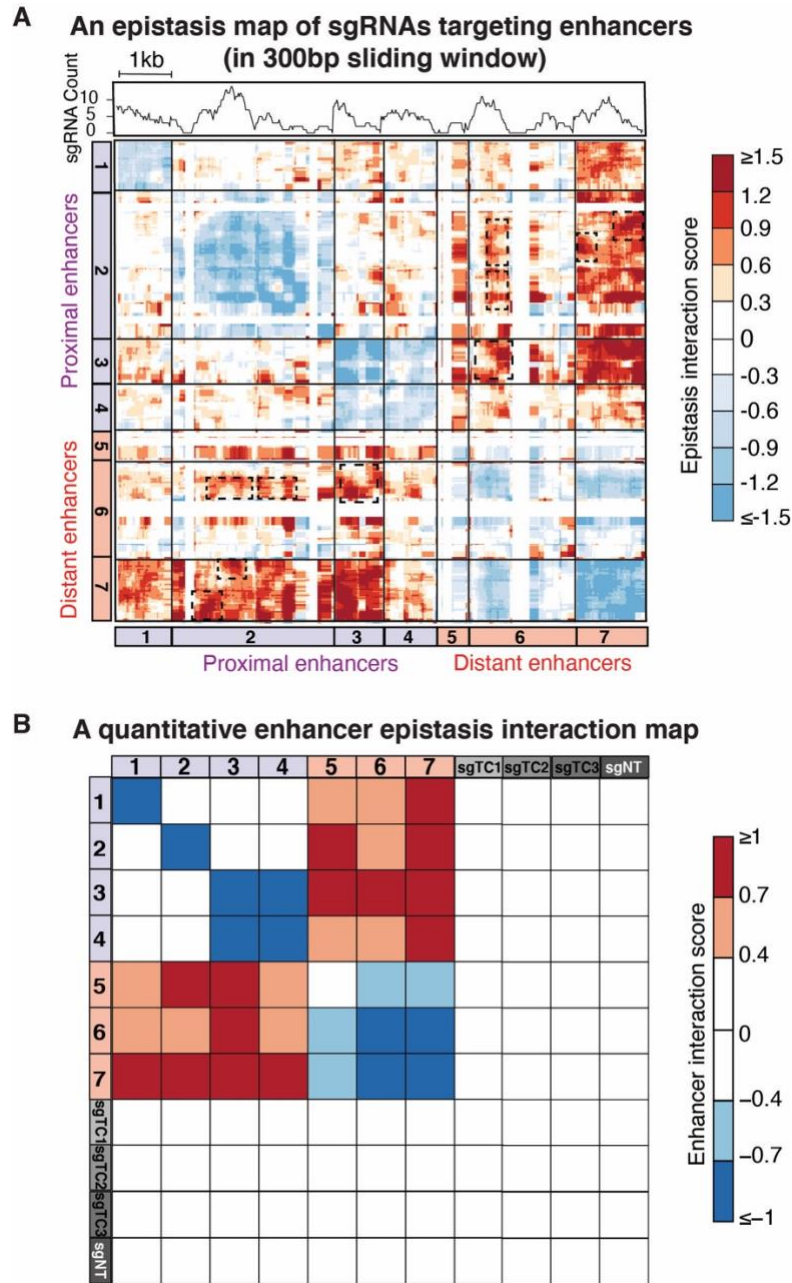
**B**, Diagram showing the construction and sequencing of the paired sgRNA library.



**Fig. S2. Data analysis for the multiplexed CRISPRi screening.**



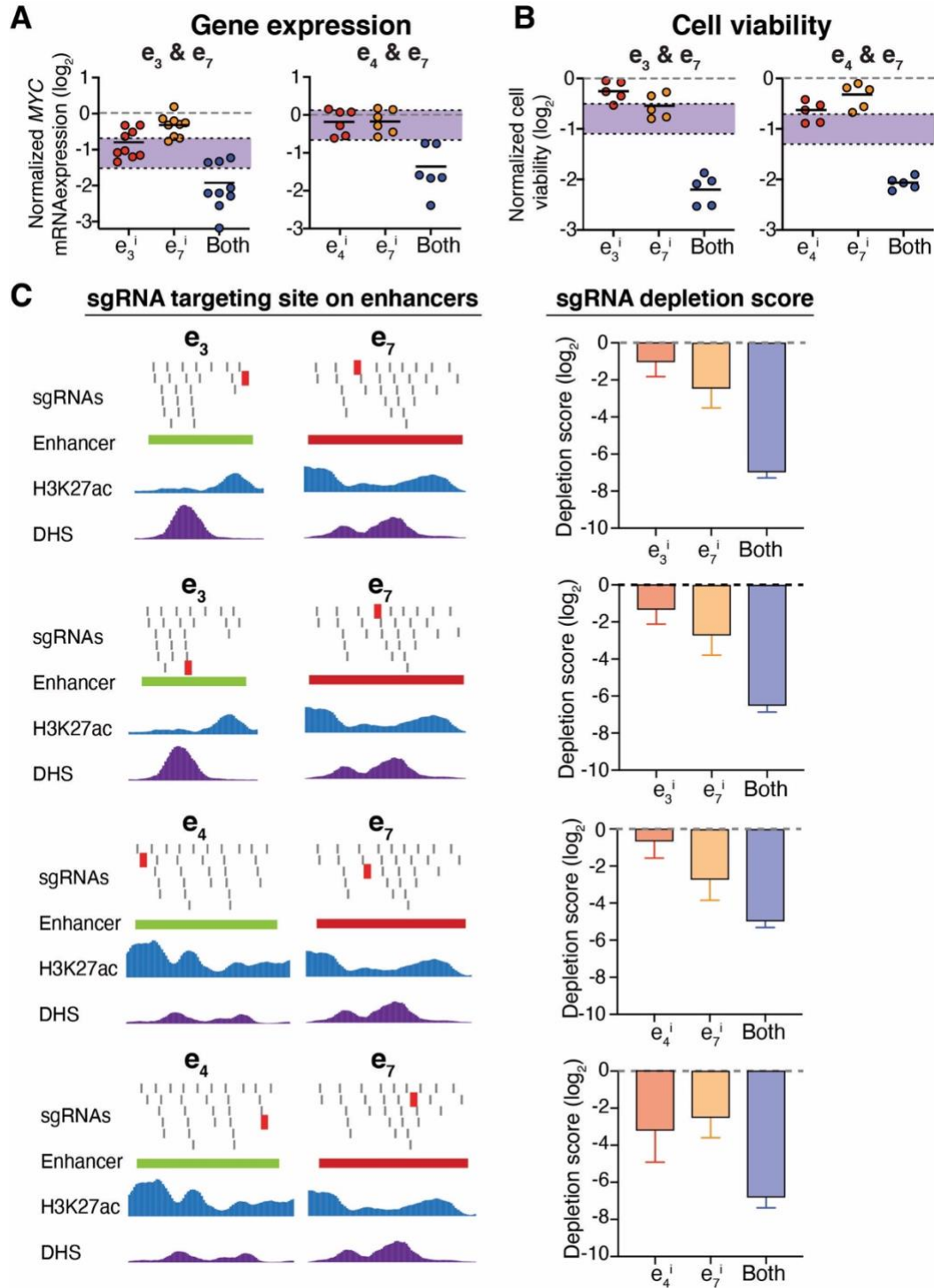
- A.** Correlation of depletion score between two permutations (A/B and B/A). Diagrams show raw depletion scores without smoothing and depletion scores after smoothing adjacent sgRNAs (2 to 6).
- B.** Heatmap showing the depletion scores ( $\log_2$  fold change) of sgRNA pairs targeting enhancers in the *MYC* locus.
- C.** Classification of enhancer-enhancer epistatic interactions based on CRISPRi-mediated sgRNA depletion scores. We define a pair of enhancers is synergistic, if perturbation of both enhancers exhibits a larger effect than the additive effects of individual enhancer perturbation.
- D.** Method for calculating the epistasis interaction scores for sgRNA pairs between a query sgRNA and all sgRNAs (sgRNA<sub>x</sub>, including 295 individual sgRNAs in the library).
- E.** A quantitative epistasis map of sgRNA pairs targeting all enhancer combinations in the *MYC* locus. Each dot represents the epistasis interaction score of a pair of sgRNAs smoothed by adjacent sgRNAs. sgTC1-3, sgRNA targeting three control regions; sgNT, non-targeting sgRNA.
- F.** Scatter plot of sgRNA depletion scores from two independent biological replicates.
- G.** Scatter plot of sgRNA interaction scores from two independent biological replicates.
- H.** Density plots of depletion score correlation between two sgRNAs targeting the same enhancer (intra-enhancer) or different enhancers (inter-enhancer). The value vector used to calculate the correlation is the depletion scores of each sgRNA paired with all sgRNAs in the library (each row in the **fig. S2B** heatmap).
- I.** Density plots of interaction score correlation between two sgRNAs targeting the same enhancer (intra-enhancer) or different enhancers (inter-enhancer). The value vector used to calculate the correlation is the interaction scores of each sgRNA paired with all sgRNAs in the library (each row in the **Fig. 1B** heatmap).



**Fig. S3. Additional data analysis for multiplexed CRISPRi screening.**

**A**, Top: sgRNA count distribution on targeting enhancers in 300bp sliding windows. Bottom: A quantitative epistasis map of sgRNA pairs targeting all enhancer combinations in the *MYC* locus in 300bp sliding window. Each dot represents the epistasis interaction score of a pair of 300bp sliding windows in two enhancers. The epistasis interaction score for each 300bp window pairs is calculated as the mean interaction score of all sgRNA pairs locating in the window pair.

**B**, A quantitative enhancer epistasis interaction map in the *MYC* locus. Each box represents the epistasis interaction score of a pair of enhancers. sgTC1-3, sgRNA targeting three control regions; sgNT, non-targeting sgRNA.

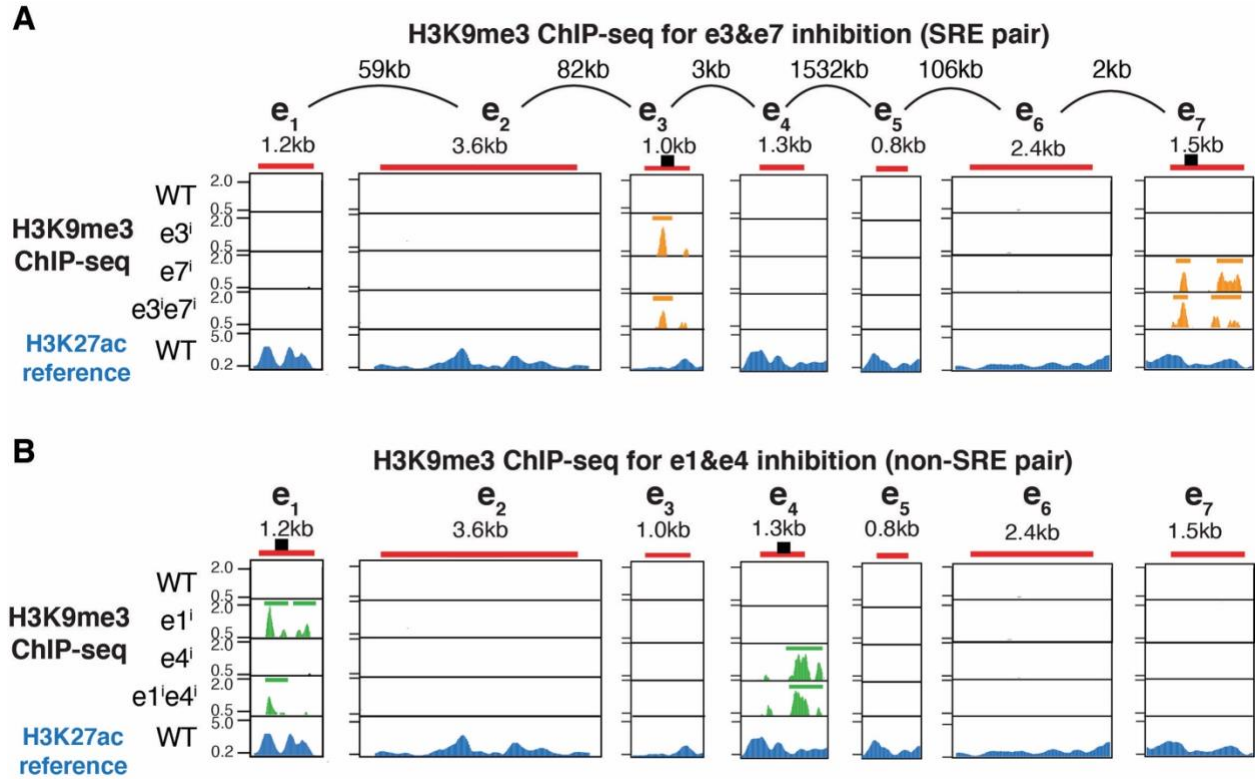


**Fig. S4. Additional experimental validation of enhancer epistasis interactions.**

**A**, qRT-PCR of *MYC* mRNA expression ( $\log_2$  fold change) by perturbing single or double enhancers. Data are represented as individual biological replicates (dots) and the mean value (black bar). The purple area indicates the expected additive effect by plotting mean  $\pm$  one standard deviation. *P* values are calculated by two-sided Student's *t*-test.

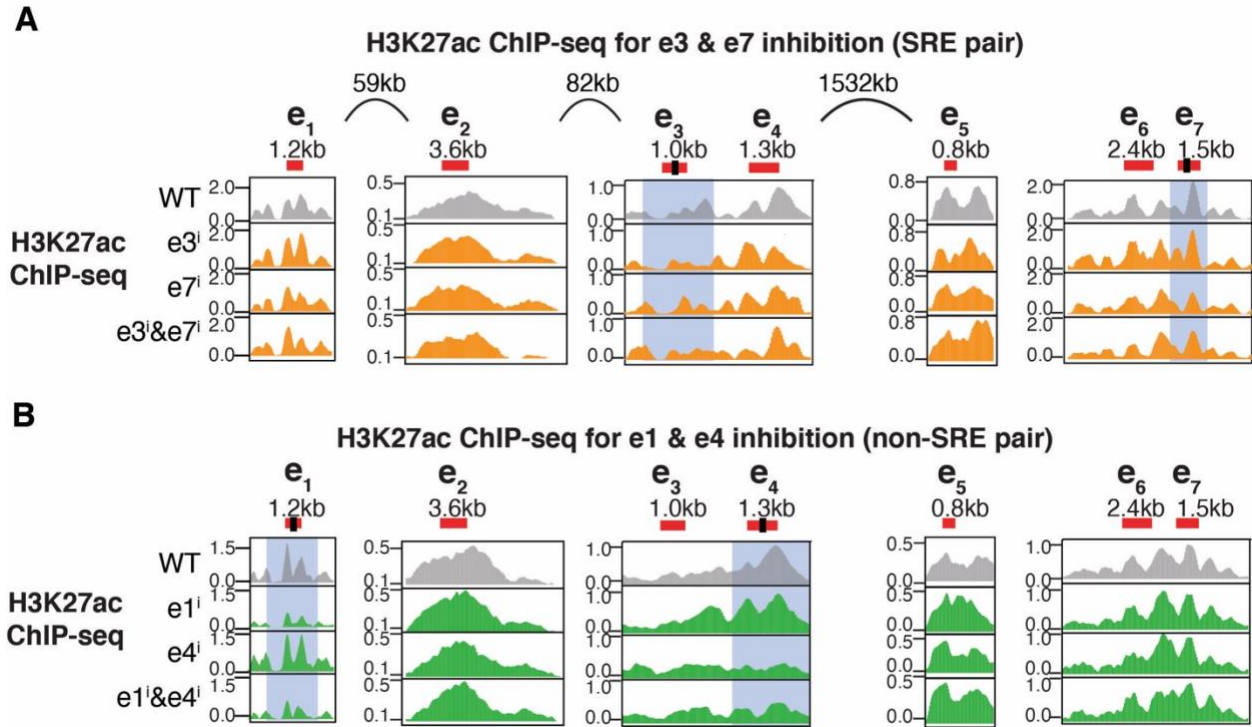
**B,** Measured cell viability normalized to wildtype cells by perturbing single or double enhancers. Data are represented as individual biological replicates (dots) and the mean value (black bar). The purple area indicates the expected additive effect by plotting mean  $\pm$  one standard deviation. *P* values are calculated by two-sided Student's t-test.

**C,** Left, diagrams showing the binding sites of sgRNAs targeting within SREs; right, diagrams showing calculated sgRNA depletion scores from multiplexed CRISPRi screening with one or two enhancers perturbed. Data are represented as mean  $\pm$  standard deviation.



**Fig. S5. H3K9me3 ChIP-seq characterization of CRISPRi perturbation of enhancers.**

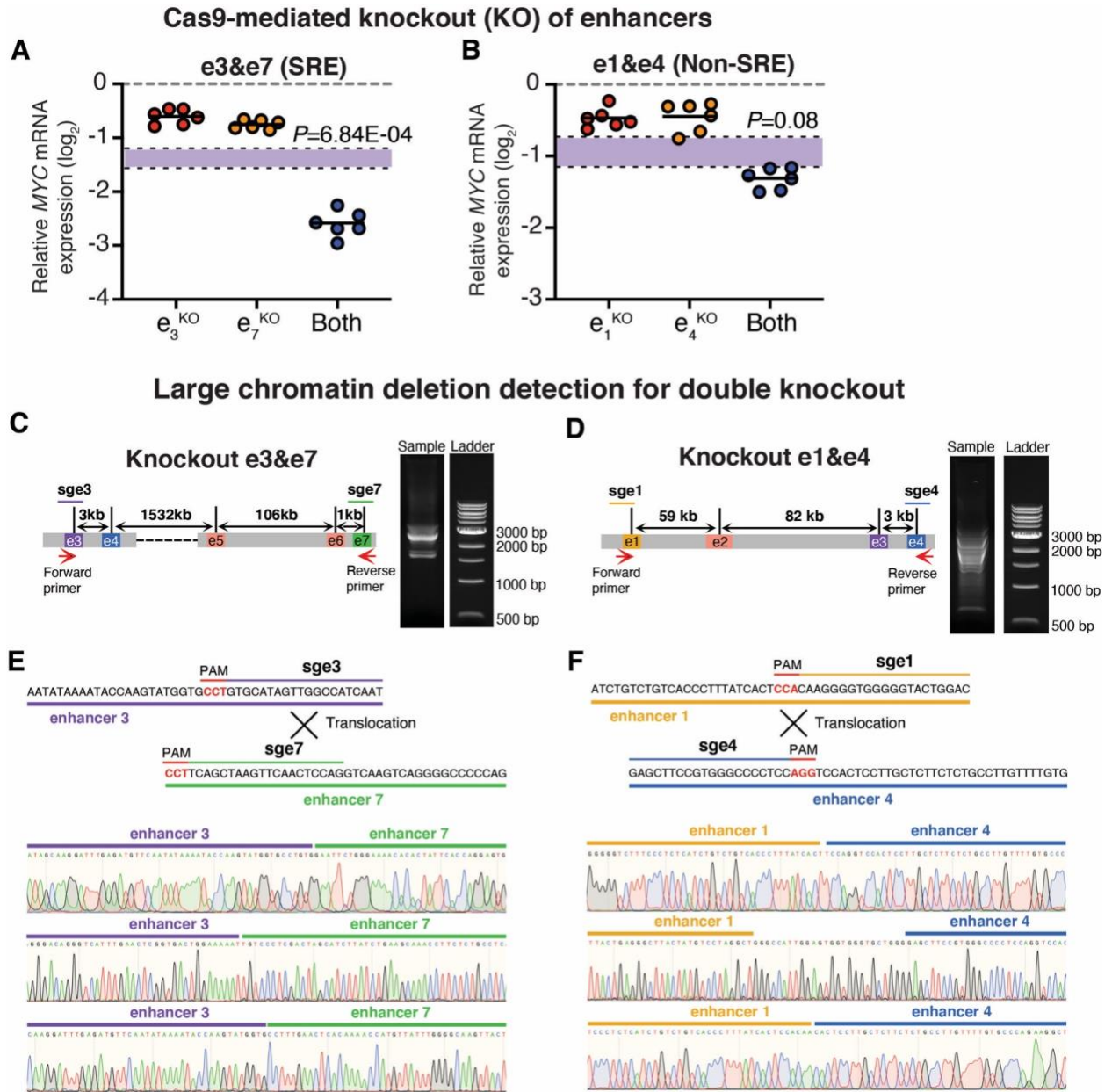
**A-B,** H3K9me3 ChIP-seq tracks in the enhancers at the *MYC* locus when perturbing SRE enhancers e3&e7 (**A**) or non-SRE enhancers e1&e4 (**B**). In **A** and **B**, from top to bottom, diagrams show wildtype cells (WT), inhibition of individual enhancers, and inhibition of both enhancers. The bars above the H3K9me3 signal represent the broad peaks of H3K9me3 called by MACS2. The bottom diagram shows the H3K27ac ChIP-seq reference tracks. The sgRNA targeting site is indicated by a black bar.



**Fig. S6. H3K27ac ChIP-seq characterization of CRISPRi perturbation of enhancers.**

**A-B**, H3K27ac ChIP-seq tracks in the enhancers at the *MYC* locus when perturbing SRE enhancers e3&e7 (**A**) or non-SRE enhancers e1&e4 (**B**). In **A** and **B**, from top to bottom, diagrams show wildtype K562 cells (WT), inhibition of individual enhancers, and inhibition of both enhancers. The sgRNA targeting site is indicated by a black bar.





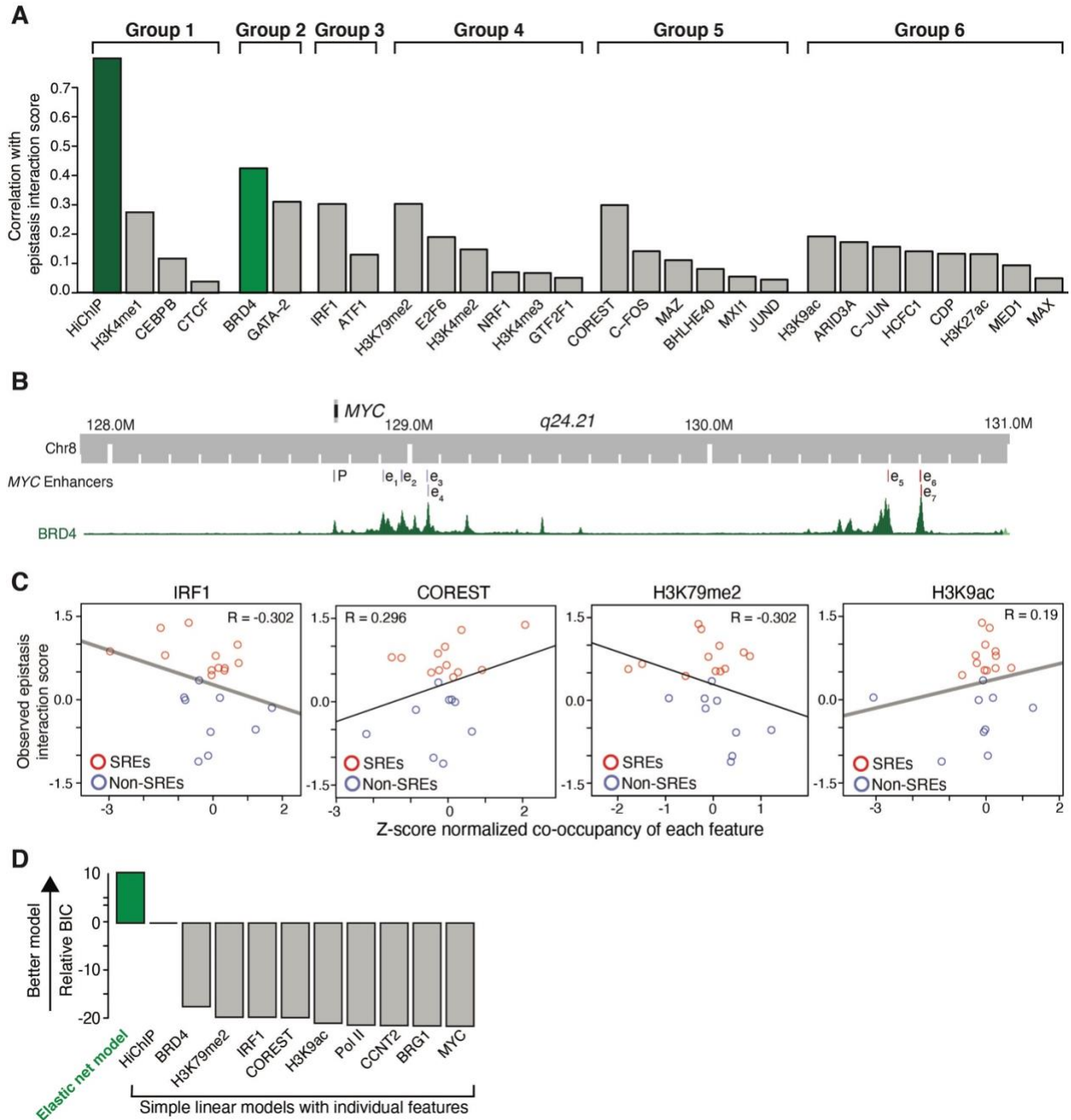
**Fig. S7. Cas9-mediated knockout (KO) of SRE and non-SRE enhancers at the *MYC* locus, and detected large chromatin deletions for double enhancer KO.**

**A-B**, qRT-PCR of *MYC* mRNA expression by knocking out individual enhancers, a pair of SRE enhancers e3&e7 (**A**), or a pair of non-SRE enhancers e1&e4 (**B**), and. Data are represented as individual biological replicates (dots) and the mean value (black bar). The purple area indicates the expected additive effect by plotting mean  $\pm$  one standard derivation. *P* values are calculated by two-sided Student's t-test.

**C-D**, Large chromatin deletions indicated by DNA gel electrophoresis showing multiple unexpected bands in double-enhancer KO cells. Left, schematic of primer binding sites on the enhancer region. Right, gel electrophoresis of PCR products showing multiple unexpected fragments; DNA ladder is shown on the right. **C**, knockout of the SRE pair e3&e7; **D**, knockout of the non-SRE pair e1&e4.

**E-F**, Top: Sequences of enhancers around sgRNA targeting sites. The sgRNA targeting sites indicated by different colors. Protospacer adjacent motifs (PAMs) are highlighted in red. Bottom: Sanger sequencing maps of PCR-amplified fragments in **C** and **D**. Six different fragments showing large chromatin deletions were detected in 20 sequenced clones (6/20).





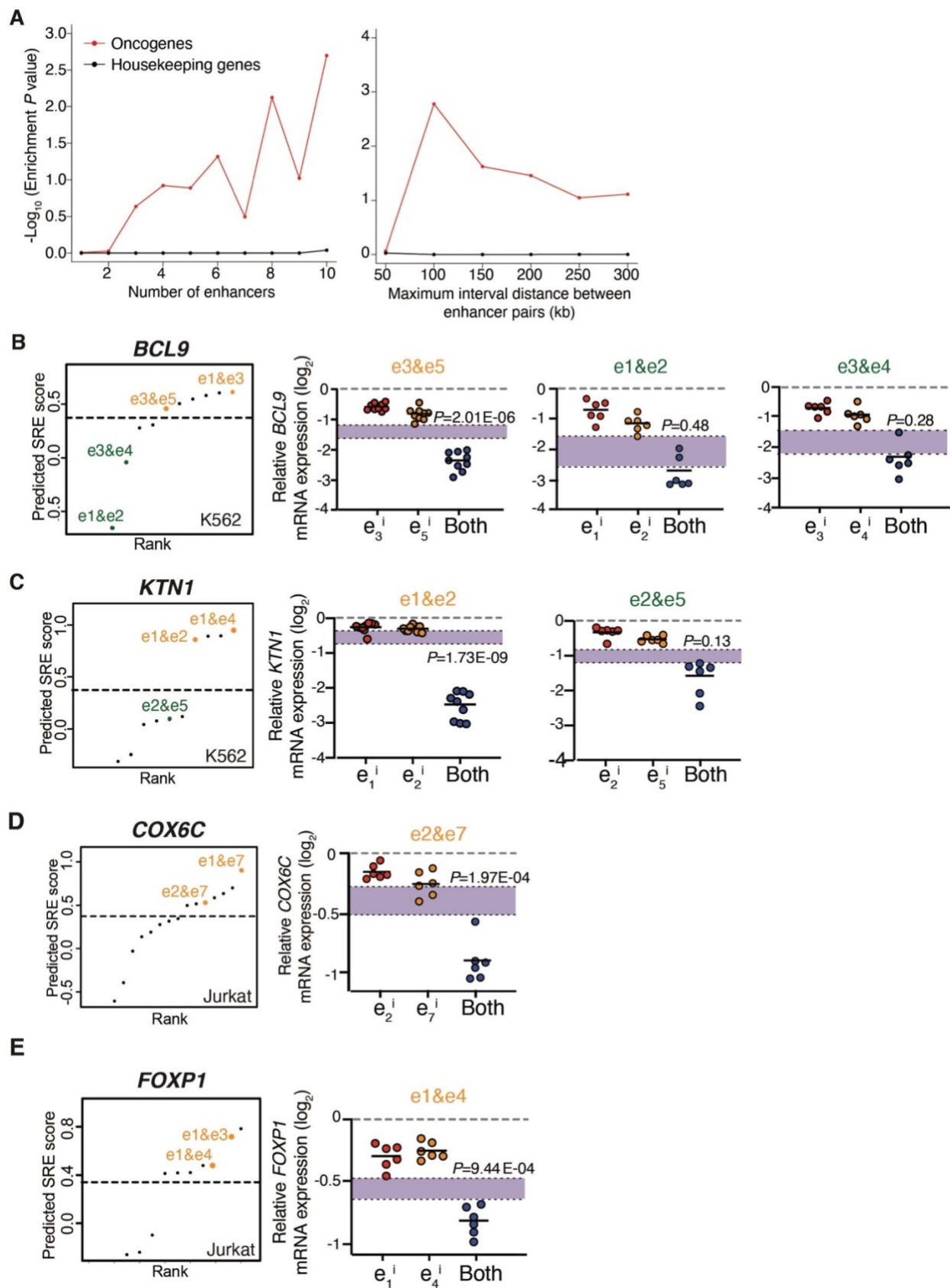
**Fig. S8. Data regression of feature groups to predict SREs using an elastic net regularized linear regression model.**

**A**, Diagram showing correlation between spatial interaction/co-occupancy and epistasis interaction scores in top6 feature groups, ranked from high to low. Green shows the top features with highest correlation with epistasis interaction scores, including HiChIP and BRD4.

**B**, Genome browser diagram showing the binding profiles of BRD4 at the *MYC* locus.

**C**, Correlation between epistasis interaction scores and co-occupancy of 4 representative genomic features in **fig. S8A**, including the profiles of IRF1, COREST, H3K79me2, and H3K9ac. Red, SREs; blue, non-SREs. The Pearson correlation coefficient ( $R$ ) is shown.

**D**, Comparison of the elastic net regularized linear regression model and simple linear models fitted with individual representative features in (**Fig. 2B**) for model performance of predicting epistasis interaction scores. BIC, Bayesian information criterion. The relative BIC is normalized to the HiChIP simple linear model.

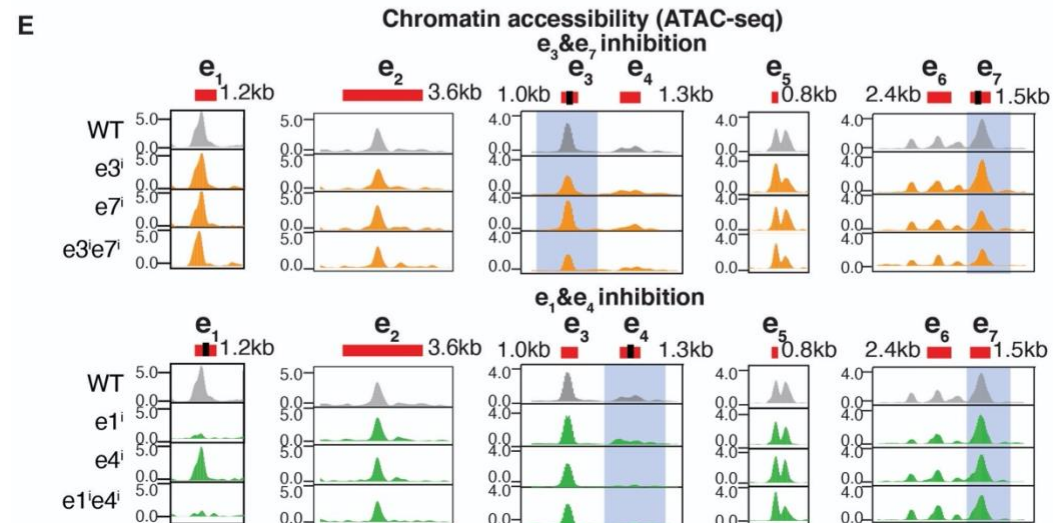
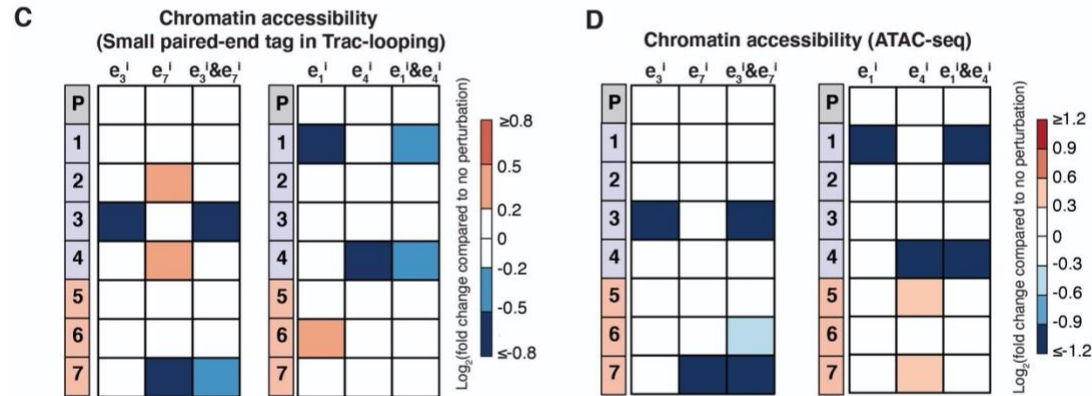
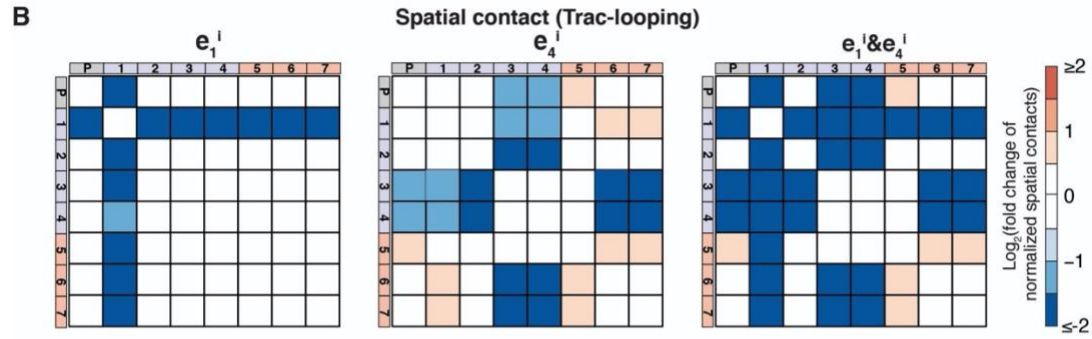
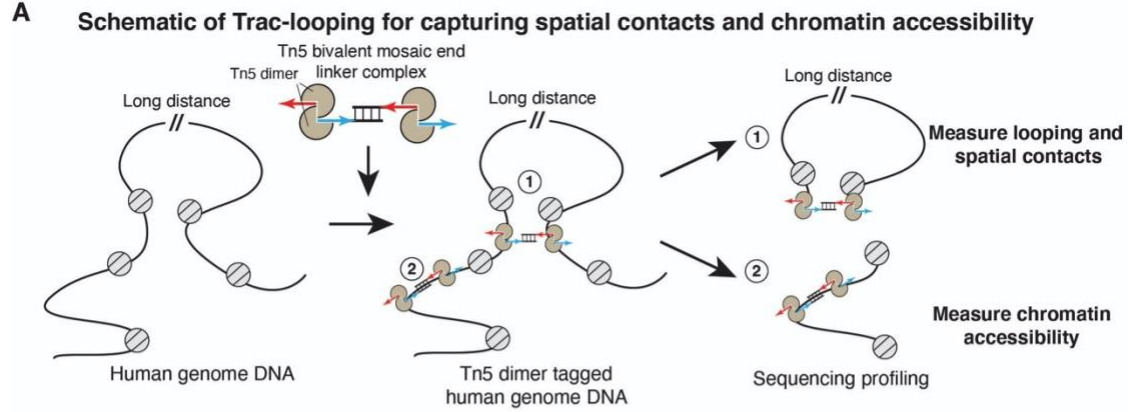


**Fig. S9. Validation of predicted SREs and non-SREs at other genomic loci in different cell types.**

**A**, Enrichment  $P$  values of oncogenes and housekeeping genes in K562 cells. Left, different numbers of enhancer; right: maximum interval distance between enhancer pairs.  $P$  values are calculated by using two-tailed Fisher's exact test.

**B-C**, Prediction and validation of SREs and non-SREs at *BCL9* (**B**) and *KTNI* loci (**C**) in K562 cells. Left: Rank of predicted SREs at the *BCL9* and *KTNI* locus in K562 cells using SRE model. Orange dots indicate the predicted SREs. Green dots indicate the predicted non-SREs. Dashed line represents the empirical threshold from *MYC* locus in K562 cells. Right: qRT-PCR of *BCL9* and *KTNI* mRNA expression when perturbing individual or a pair of enhancers. Data are represented as individual biological replicates (dots) and the mean value (black bar). The purple area indicates the expected additive effect by plotting mean  $\pm$  one standard derivation.  $P$  values are calculated by two-sided Student's t-test.

**D-E**, Prediction and validation of SREs at *COX6C* (**D**) and *FOXP1* loci (**E**) in Jurkat cells. Left: Rank of predicted SREs at the *COX6C* and *FOXP1* locus in Jurkat cells using SRE model. Orange dots indicate the predicted SREs. Dashed line represents the empirical threshold from *MYC* locus in K562 cells. Bottom right: qRT-PCR of *COX6C* and *FOXP1* mRNA expression when perturbing individual or a pair of enhancers. Data are represented as individual biological replicates (dots) and the mean value (black bar). The purple area indicates the expected additive effect by plotting mean  $\pm$  one standard derivation.  $P$  values are calculated by two-sided Student's t-test.



**Fig. S10. Workflow for Trac-looping analyses, Trac-looping data for chromatin accessibility and spatial contacts, and ATAC-seq data for chromatin accessibility.**

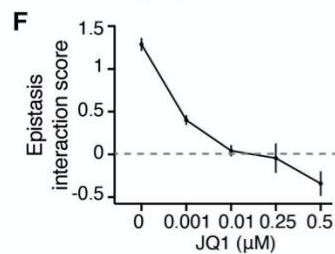
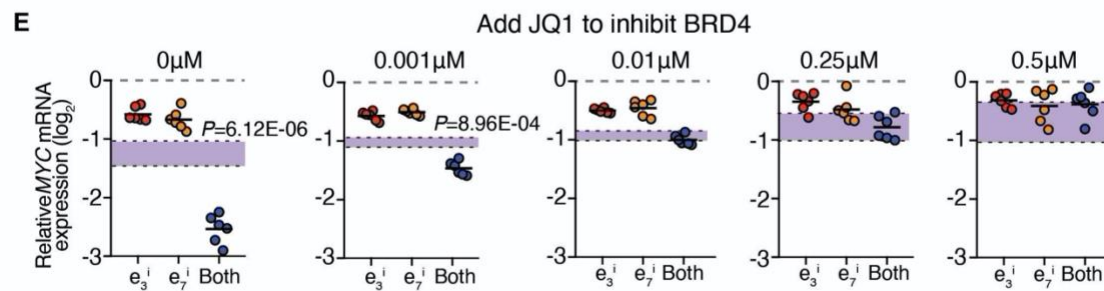
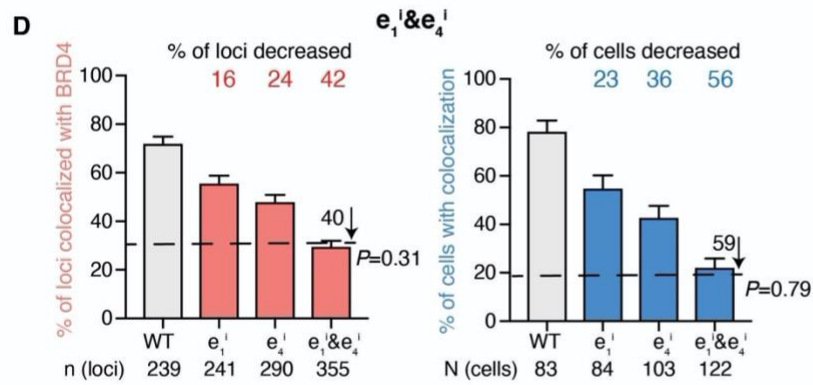
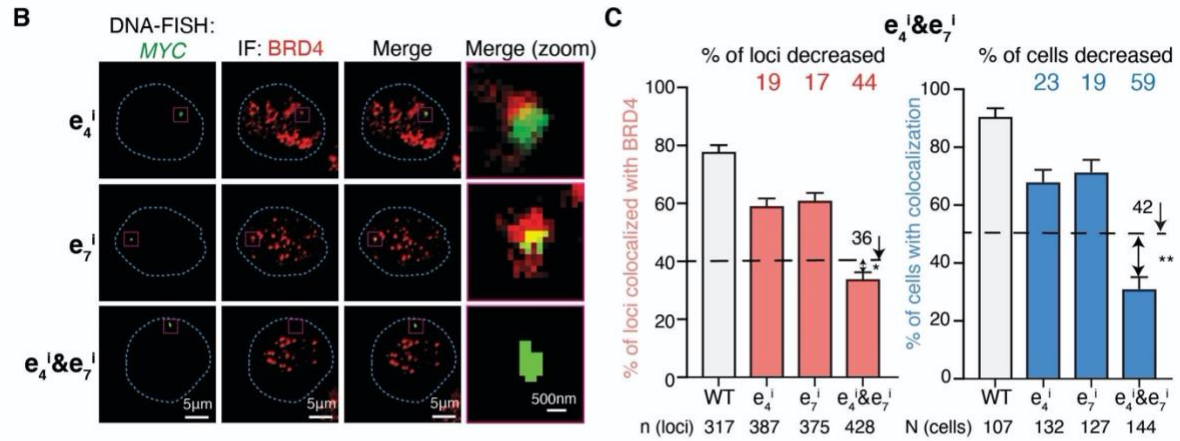
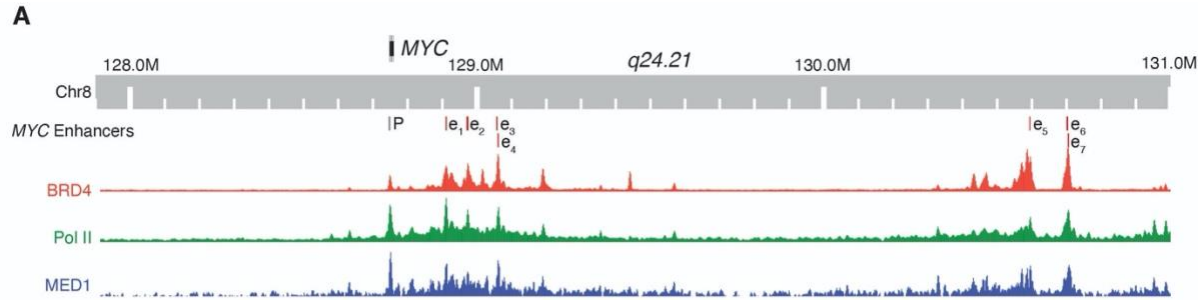
**A**, Schematic of the workflow for Trac-looping, adapted from the Lai *et al.* 2018 (34). In Trac-looping, a Tn5 bivalent linker complex containing a pair of mosaic ends with a 30-bp oligonucleotide spacer is used for transposase-mediated analysis of spatial contacts and chromatin accessibility.

**B**, Heatmap showing the changes of spatial contacts among the promoter and enhancers. Colors represent the log<sub>2</sub> fold change of spatial contacts normalized to the wildtype cells upon perturbation of e1, e4, and e1&e4.

**C**, Heatmap showing the log<sub>2</sub> fold change of chromatin accessibility normalized to the wildtype cells at the promoter and enhancers upon perturbation of e3, e7, and e3&e7 (left); e1, e4 and e1&e4 (right). The chromatin accessibility was measured by the short range paired-end tag (<1kb) in Trac-looping, which specifically measure the chromatin accessibility without the signal of spatial contacts as demonstrated in the **fig. S10A**.

**D**, Heatmap showing the log<sub>2</sub> fold change of chromatin accessibility normalized to the wildtype cells at the promoter and enhancers upon inhibition of e3, e7, and e3&e7 (left); e1, e4 and e1&e4 (right). The chromatin accessibility was measured by ATAC-seq.

**E**, ATAC-seq signal in the enhancers at the *MYC* locus when perturbing individual or a pair of SRE pair (e3&e7) and non-SRE enhancer (e1&e4). In each panel, from top to bottom, diagrams show wildtype cells (WT), inhibition of individual enhancers, and inhibition of both enhancers. The sgRNA targeting sites are indicated by black bars.





**Fig. S11. Additional data of BRD4 colocalization at the *MYC* locus.**

**A,** Genome browser diagram showing the binding profiles of BRD4, Mediator and Pol II at the *MYC* locus.

**B,** Colocalization between BRD4 and the *MYC* locus by immunofluorescence (IF) staining (red) and 2D DNA-FISH (green) in K562 cells upon perturbation of e4, e7, and e4&e7 (SRE pair). The blue dash line indicates the nuclear periphery determined by DAPI staining (not shown). The rightmost column shows insets in the yellow boxes. Scale bars, 5 $\mu$ m in the first three columns; 500 nm for the rightmost column.

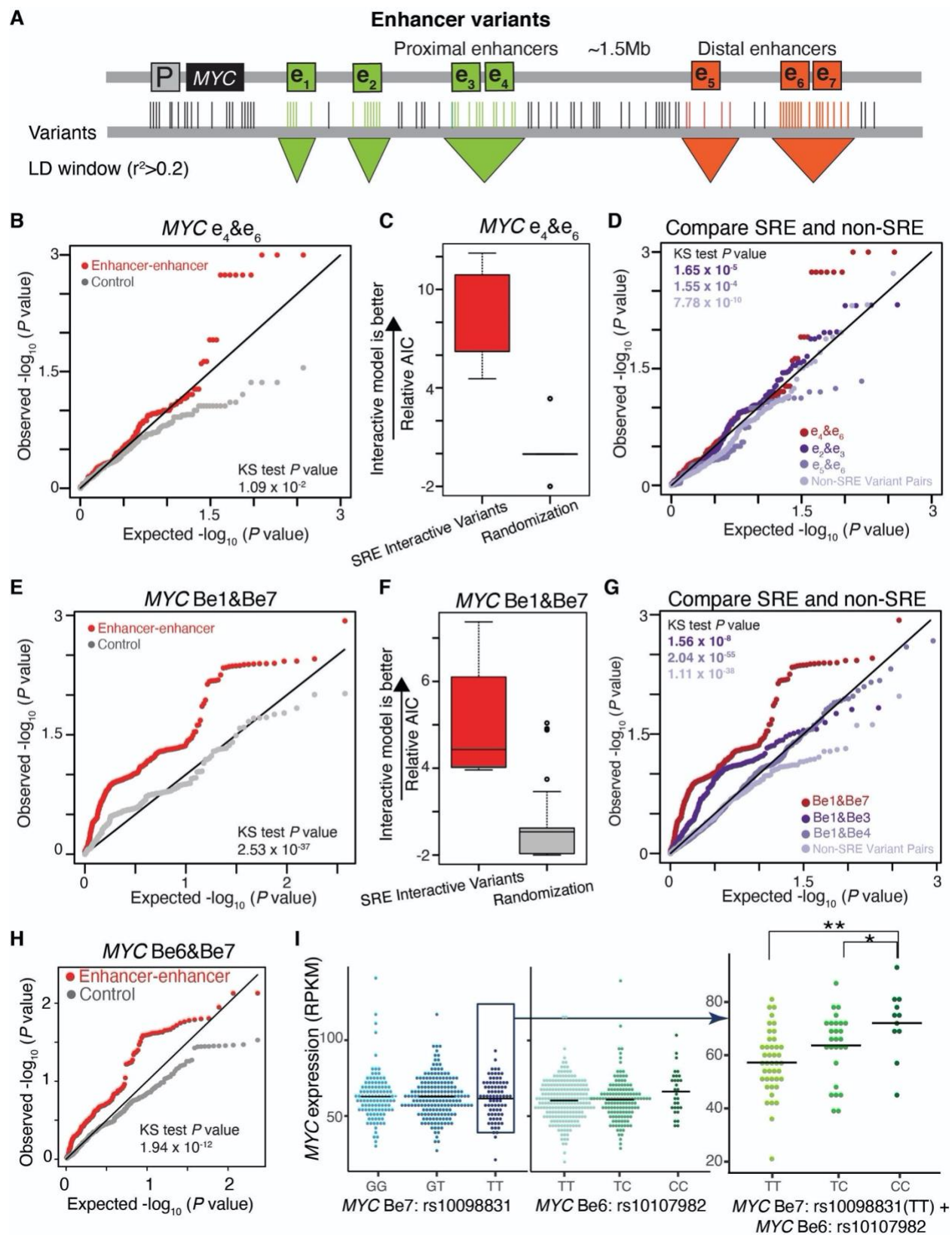
**C,** Quantification of BRD4 and 2D DNA-FISH colocalization frequency at the *MYC* locus in cells without perturbation and upon perturbation of e4, e7, and e4&e7 (SRE pair). Left, percentage of loci showing BRD4 and the *MYC* locus colocalization, n = total loci analyzed. Right, percentage of cells showing at least two loci with BRD4 and the *MYC* locus colocalization, N = total cells analyzed. Data are represented as mean  $\pm$  standard error of the mean. Dashed line represents the expected additive effect. \*: *P* value < 0.05; \*\*: *P* value < 0.01, calculated by two-tailed Fisher's exact test for paired sgRNAs effect versus the expected effect.

**D,** Quantification of BRD4 and the *MYC* locus 2D DNA-FISH colocalization frequency in control or upon perturbation of e1, e4, and e1&e4 (non-SRE pair). Top, percentage of loci showing BRD4 and the *MYC* locus colocalization, n = total loci analyzed. Bottom, percentage of cells showing at least two loci with BRD4 and the *MYC* locus colocalization, N = total cells analyzed. Data are represented as mean  $\pm$  standard error of the mean. Dashed line represents the expected additive effect. *P* value was calculated by two-tailed Fisher's exact test for paired sgRNAs effect versus the expected effect.

**E,** Effects of JQ1 on *MYC* mRNA expression measured by qRT-PCR when perturbing individual or double enhancers of e3&e7. RNA abundance is calculated by normalizing to the sample without perturbation (log<sub>2</sub> fold change). Dashed line indicates the *MYC* expression levels in samples without perturbation. Data are represented as technical replicates in three independent biological experiments (dots) and the mean value (black bar). The purple area indicates the expected additive effect by plotting mean  $\pm$  one standard derivation. *P* values are calculated by two-sided Student's t-test.

**F,** Diagram showing JQ1 effects on the computed epistasis interaction scores of e3&e7. The epistasis interaction score was calculated as -(Observed relative *MYC* expression of paired sgRNAs targeting e3&e7- Expected) in log<sub>2</sub> fold change.





**Fig. S12. Additional analysis of interactive influence of SRE variants for gene expression at the *MYC* locus in K562 cells and GM12878 cells.**

**A,** Diagram showing the method to define enhancer variants. Genomic variants are indicated by vertical lines, and those falling within the LD window for a given enhancer are highlighted and defined as enhancer variants.

**B,** QQ plot showing the distribution of *P* values for the comparison between interactive model and additive model (see **Methods**) of e4&e6 SRE variants (red) on *MYC* expression for in LAML patients, compared to random permutations of individual *MYC* expression (grey). KS test *P* value was calculated between e4&e6 SRE and the random permutations.

**C.** Model comparison for significantly interactive variants of e4&e6 ( $P < 0.05$  in **Fig. 5A**) on *MYC* expression of LAML patients and random permutations of individual *MYC* expression, measured as the relative AIC calculated by the AIC of the additive model minus the AIC for the interactive model.

**D.** QQ plot showing the distribution of *P* values for the interactive influence of e4&e6 SRE variants on *MYC* expression in LAML patients, compared to e2&e3 non-SRE variants, e5&e6 non-SRE variants, and non-SRE variant pairs sampled from all the non-SRE variants. KS test *P* values were calculated between e4&e6 SRE and the non-SRE pairs.

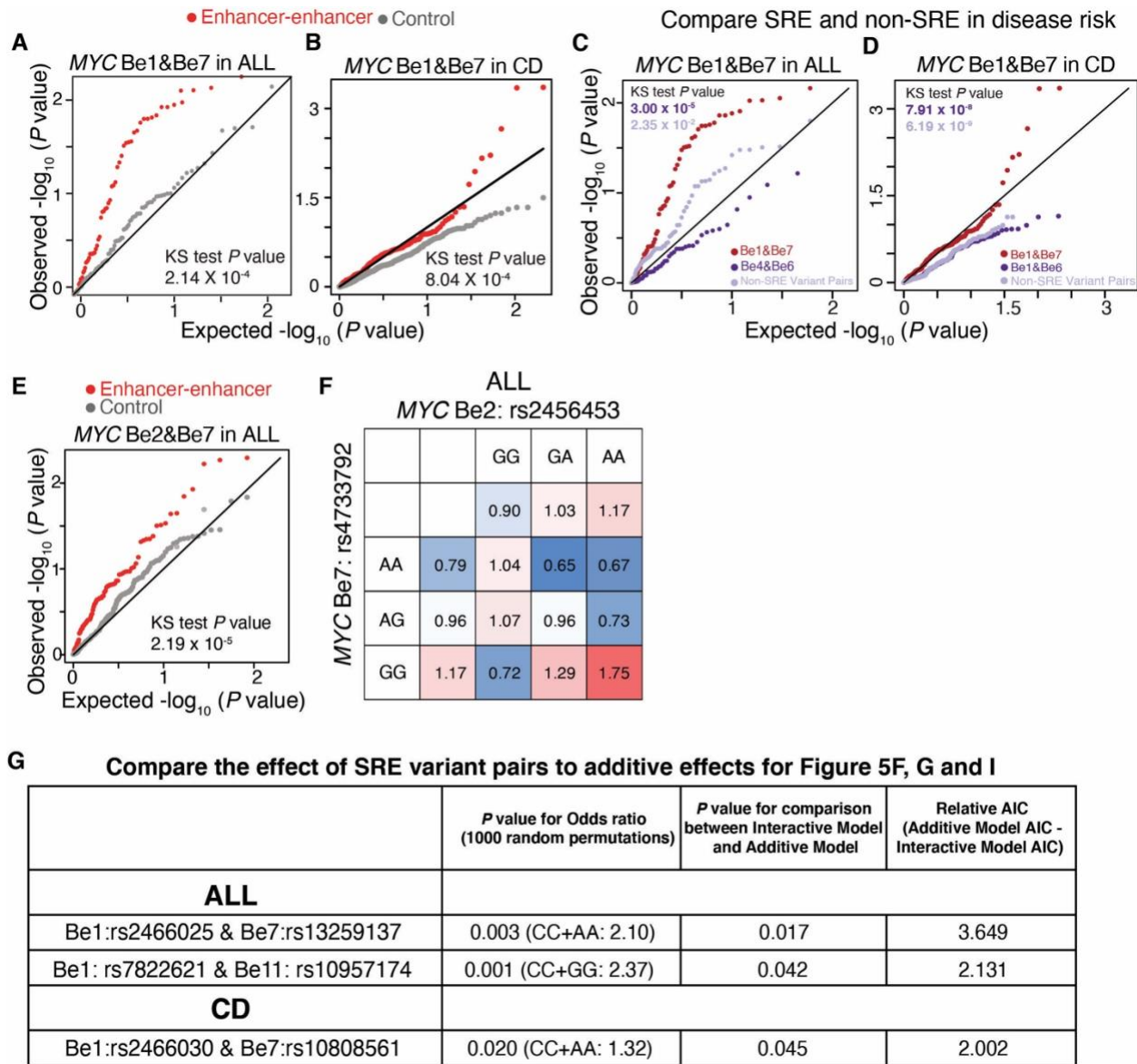
**E,** QQ plot showing the distribution of *P* values for the comparison between interactive model and additive model (see **Methods**) of Be1&Be7 variants (red) on *MYC* expression between in the B lymphoblasts of 373 European individuals, compared to random permutations of individual *MYC* expression (grey). KS test *P* value was calculated between Be1&Be7 SRE and the random permutations.

**F.** Model comparison for significantly interactive variants of Be1&Be7 ( $P < 0.05$  in **Fig. 5D**) on *MYC* expression of the B lymphoblasts of 373 European individuals and random permutations of individual *MYC* expression, measured as the relative AIC calculated by the AIC of the additive model minus the AIC for the interactive model.

**G.** QQ plot showing the distribution of *P* values for the interactive influence of Be1&Be7 SRE variants on *MYC* expression in the B lymphoblasts of 373 European individuals, compared to Be1&Be3 non-SRE variants, Be1&Be4 non-SRE variants and non-SRE variant pairs sampled from all the non-SRE variants. KS test *P* values were calculated between Be1&Be7 SRE and the non-SRE pairs.

**H,** QQ plot showing the distribution of *P* values for the interactive influence on the *MYC* expression between Be6&Be7 pairs (red) in the B lymphoblasts of 373 European individuals, compared to random permutations of individual *MYC* expression (grey). KS test *P* value =  $1.94 \times 10^{-12}$ .

**I,** *MYC* expression from RNA-seq data in the B lymphoblasts of 373 European individuals stratified by genotypes of SRE variant pairs located on Be6 (rs10107982) and Be7 (rs10098831). One-tailed Wilcoxon test was used for calculating the *P* value in different groups. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ .



**Fig. S13. Additional analysis of interactive influence for SRE variants on clinical risks of B-cell-associated diseases in the MYC locus.**

**A**, QQ plot showing the distribution of  $P$  values for the interactive influence on the relapse risks between Be1&Be7 variants (red) at MYC locus in childhood ALL patients, compared to random permutations of relapse and control population (grey). KS test  $P$  value =  $2.14 \times 10^{-4}$ .

**B**, QQ plot showing the distribution of  $P$  values for the interactive influence on the disease risks between Be1&Be7 variants (red) at MYC locus in CD patients, compared to random permutations of relapse and control population (grey). KS test  $P$  value =  $8.04 \times 10^{-4}$ .

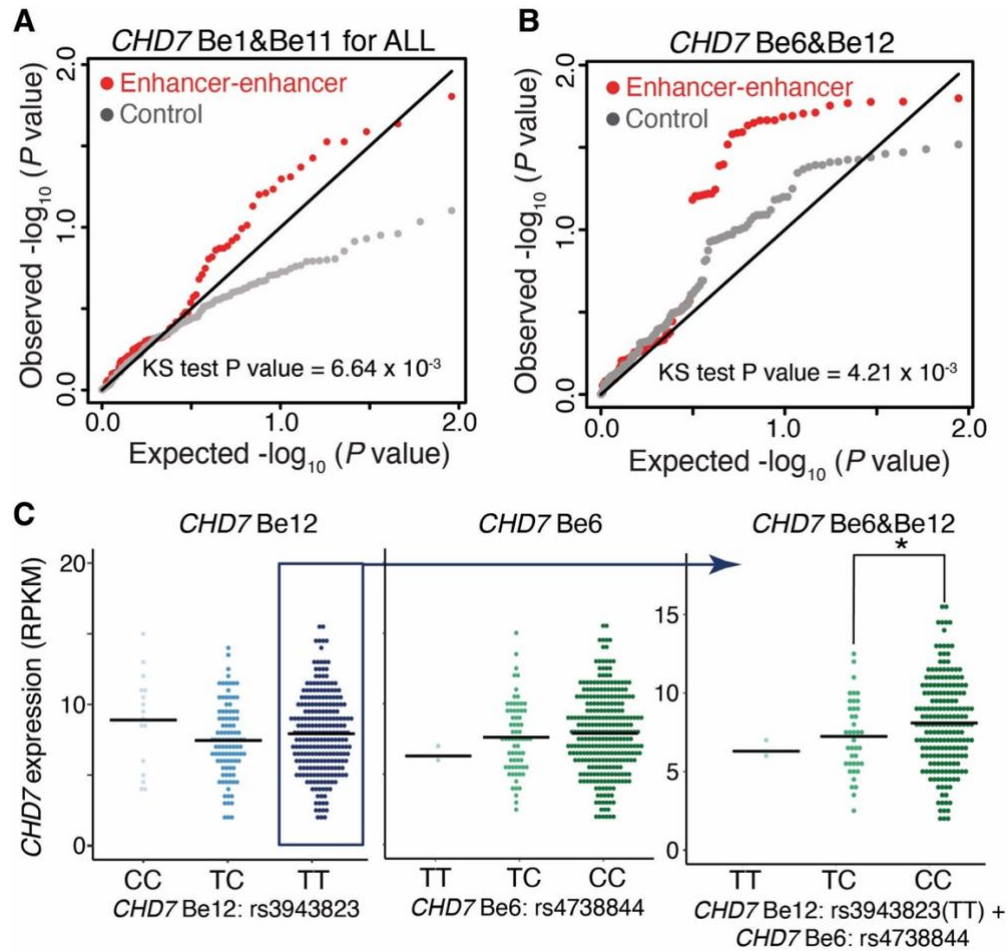
**C**, QQ plot showing the distribution of  $P$  values for the interactive influence of Be1&Be7 SRE variants on ALL relapse risk between, compared to Be4&Be6 non-SRE variants and non-SRE variant pairs sampled from all the non-SRE variants. KS test  $P$  values were calculated between Be1&Be7 SRE and the non-SRE pairs.

**D.** QQ plot showing the distribution of  $P$  values for the interactive influence of Be1&Be7 SRE variants on CD disease risk between, compared to Be1&Be6 non-SRE variants and non-SRE variant pairs sampled from all the non-SRE variants. KS test  $P$  values were calculated between Be1&Be7 SRE and the non-SRE pairs.

**E.** QQ plot showing the distribution of  $P$  values for the interactive influence on the relapse risks between Be2&Be7 variants (red) at *MYC* locus in childhood ALL patients, compared to random permutations of relapse and control population (grey). KS test  $P$  value =  $2.19 \times 10^{-5}$ .

**F.** Example of one pair of SRE variants, Be2 (rs2456453) and Be7 (rs4733792), for calculated odds ratio on the relapse risk in childhood ALL patients of 1,593 individuals. Odds ratios are calculated by considering the genotypes of individual variants, Be2(rs2456453) or Be7(rs4733792), or by the genotype combinations of both variants. Colors represent the odds ratios.

**G.** Table showing the statistical metrics for **Fig. 5F, G and I**. The first column shows the  $P$  value for the highest odds ratio calculated by considering both SRE variants. 1000 permutation analysis on randomized individuals (cases and controls) was used to define the significance of odds ratio. The second column shows the  $P$  values for the comparison between interactive model and additive model for each SRE variant pair in clinical risk. The third column shows the relative AIC calculated by the AIC of the additive model minus the AIC for the interactive model.

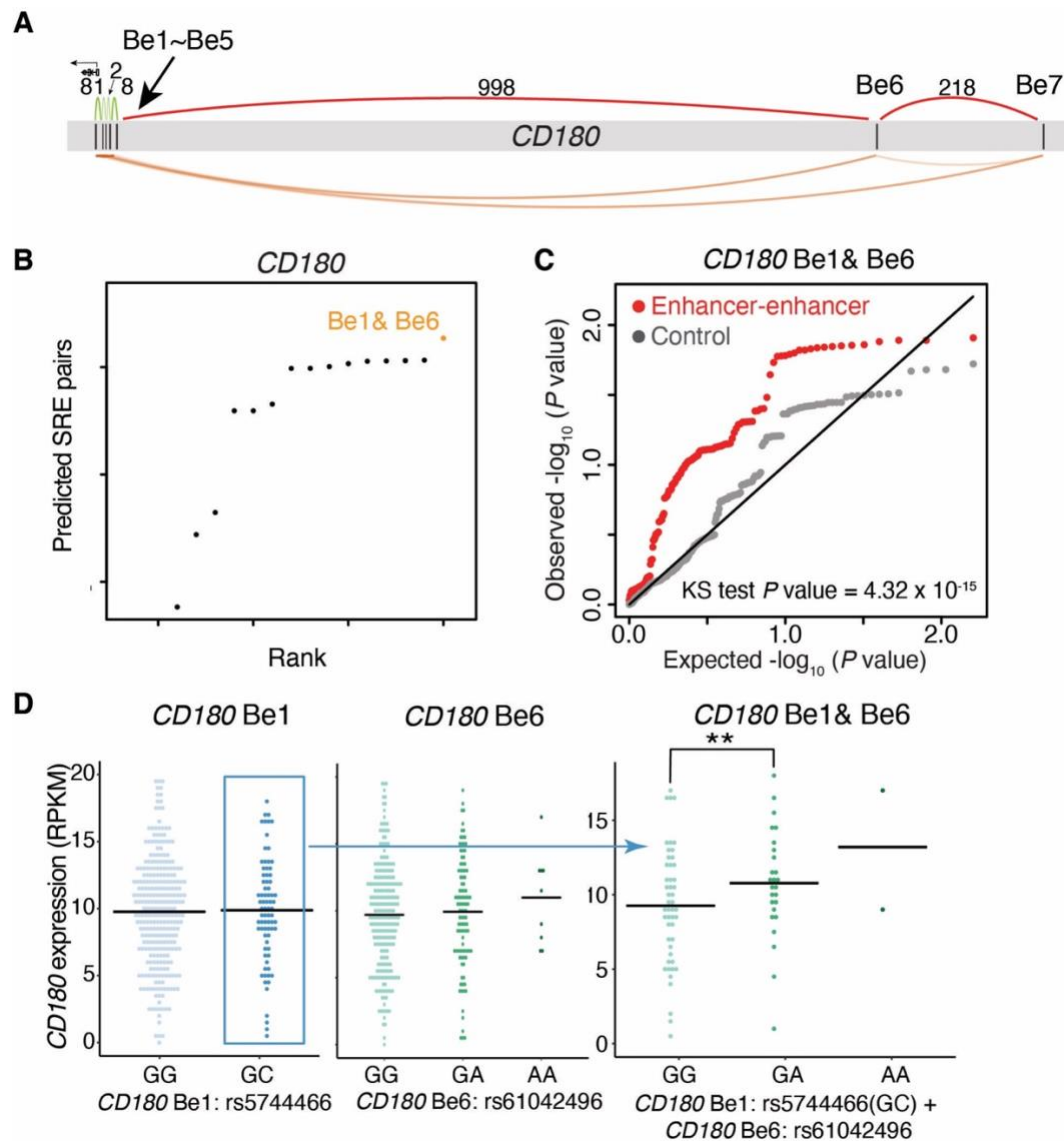


**Fig. S14. Additional analysis of interactive influence for SRE variants on gene expression and clinical risk of B-cell-associated disease in the *CHD7* locus.**

**A**, QQ plot showing the distribution of  $P$  values for the interactive influence on the relapse risks between Be1&Be11 variants (red) at *CHD7* locus in childhood ALL patients, compared to random permutations of relapse and control population (grey). KS test  $P = 6.64 \times 10^{-3}$ .

**B**, QQ plot showing the distribution of  $P$  values for the interactive influence on *CHD7* expression between Be6&Be12 pairs (red) in the B lymphoblasts of 373 European individuals, compared to random permutations of individual *CHD7* expressions (grey). KS test  $P = 4.21 \times 10^{-3}$ .

**C**, *CHD7* expression from RNA-seq data in the B lymphoblasts of 373 European individuals stratified by the genotypes of SRE variant pairs located on Be6 (rs4738844) and Be12 (rs3943823). One-tailed Wilcoxon test was used for calculating the  $P$  value in different groups. \*,  $P < 0.05$ .



**Fig. S15. Additional analysis of interactive influence for SRE variants on gene expression at *CD180* loci in GM12878 cells.**

**A**, Diagram showing *CD180* regulated by multiple enhancers in GM12878 cells.

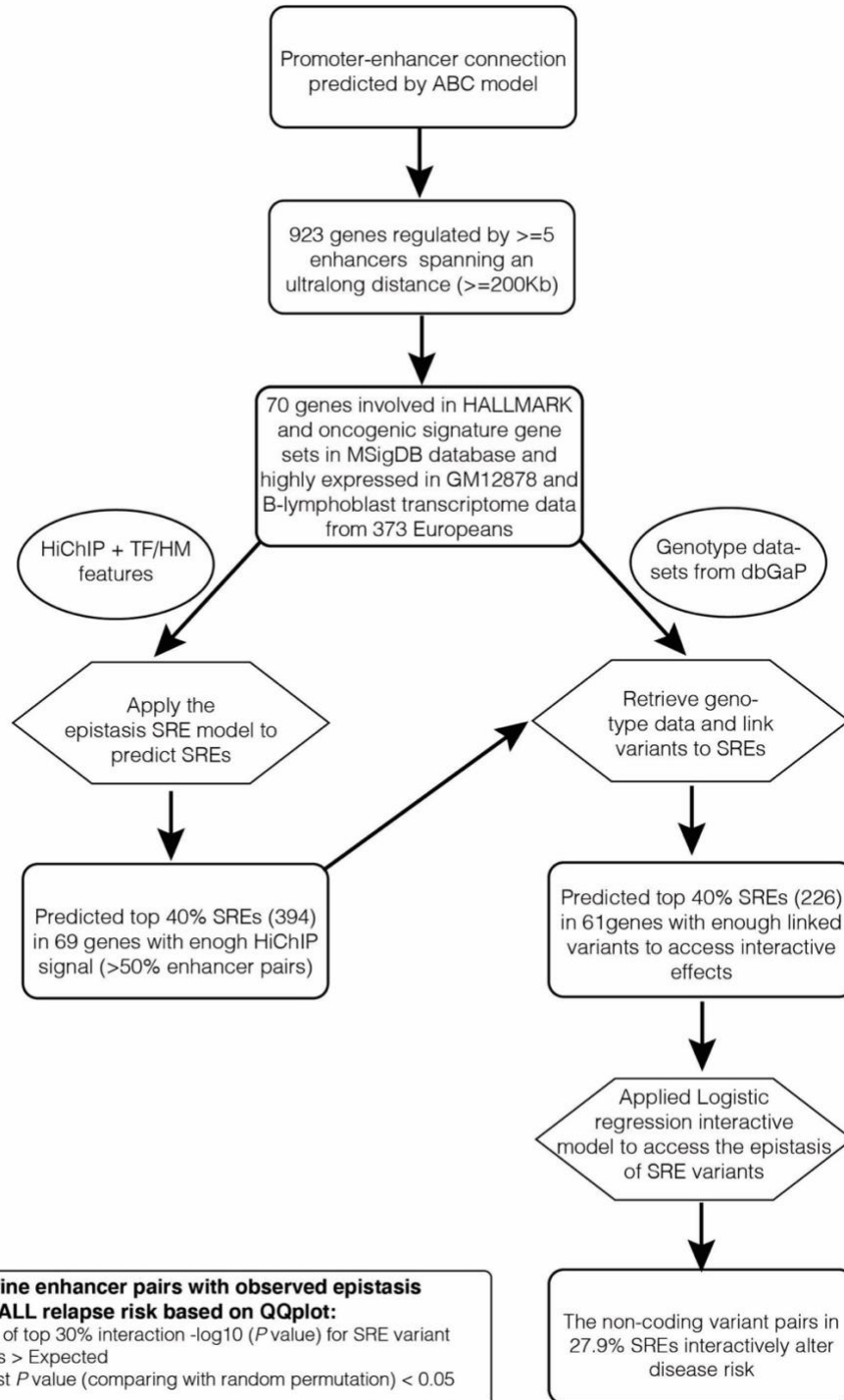
**B**, Rank of predicted epistasis interaction scores for enhancer pairs at the *CD180* locus in GM12878 cells using SRE model. Orange dot, top SRE.

**C**, QQ plot showing the distribution of  $P$  values for the interactive influence on *CD180* expression between Be1&Be6 pairs (red) in the B lymphoblasts of 373 European individuals, compared to random permutations of individual *CD180* expressions (grey). KS test  $P = 4.32 \times 10^{-15}$ .

**D**, *CD180* expression from RNA-seq data in the B lymphoblasts of 373 European individuals stratified by the genotypes of SRE variant pairs located on Be1(rs5744466) and Be6(rs61042496). One-tailed Wilcoxon test was used for calculating the  $P$  value in different groups. \*\*,  $P < 0.01$ .



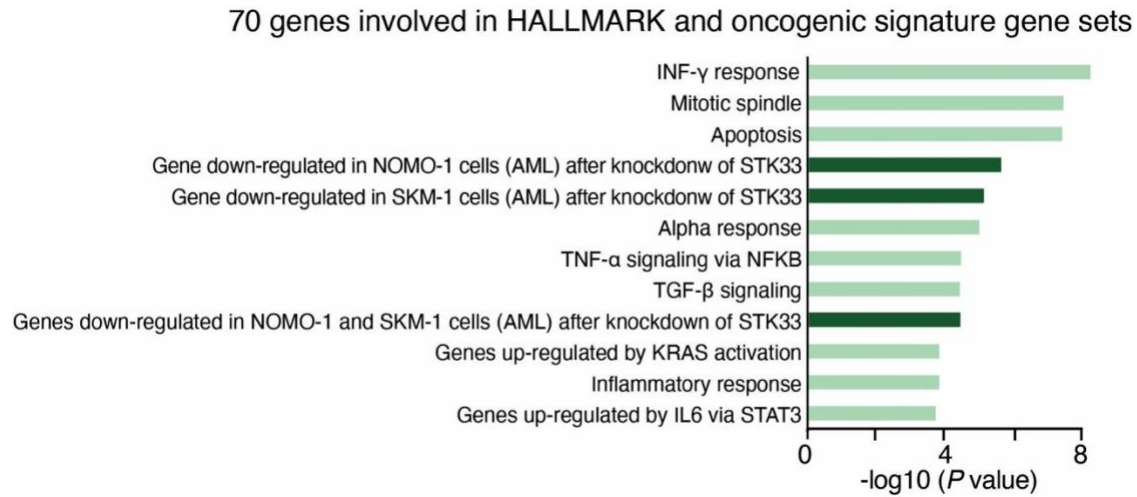
**A**



**Fig. S16. Workflow for applying the SRE prediction model to link multiple non-coding variants with complex disease on the genome-wide scale.**

**A,** The workflow of genome-wide analysis for SRE epistasis effects in ALL relapse risk.

**B,** Criteria used to define enhancer pairs with observed epistasis effects on ALL relapse risk.



**Fig. S17. Genome-wide analysis of the SRE model in GM12878 cells.** Gene Set Enrichment Analysis (GSEA) of genes regulated by ultralong distance enhancer interaction network in GM12878 cells. Light green represents GSEA immune-related gene sets, and dark green represents oncogenic signature gene sets.



## References and Notes

1. P. A. Northcott, C. Lee, T. Zichner, A. M. Stütz, S. Erkek, D. Kawauchi, D. J. H. Shih, V. Hovestadt, M. Zapatka, D. Sturm, D. T. W. Jones, M. Kool, M. Remke, F. M. G. Cavalli, S. Zuyderduyn, G. D. Bader, S. VandenBerg, L. A. Esparza, M. Ryzhova, W. Wang, A. Wittmann, S. Stark, L. Sieber, H. Seker-Cin, L. Linke, F. Kratochwil, N. Jäger, I. Buchhalter, C. D. Imbusch, G. Zipprich, B. Raeder, S. Schmidt, N. Diessl, S. Wolf, S. Wiemann, B. Brors, C. Lawerenz, J. Eils, H.-J. Warnatz, T. Risch, M.-L. Yaspo, U. D. Weber, C. C. Bartholomae, C. von Kalle, E. Turányi, P. Hauser, E. Sanden, A. Darabi, P. Siesjö, J. Sterba, K. Zitterbart, D. Sumerauer, P. van Sluis, R. Versteeg, R. Volckmann, J. Koster, M. U. Schuhmann, M. Ebinger, H. L. Grimes, G. W. Robinson, A. Gajjar, M. Mynarek, K. von Hoff, S. Rutkowski, T. Pietsch, W. Scheurlen, J. Felsberg, G. Reifenberger, A. E. Kulozik, A. von Deimling, O. Witt, R. Eils, R. J. Gilbertson, A. Korshunov, M. D. Taylor, P. Lichter, J. O. Korbel, R. J. Wechsler-Reya, S. M. Pfister, Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014). [doi:10.1038/nature13379](https://doi.org/10.1038/nature13379) [Medline](#)
2. M. J. Fullwood, M. H. Liu, Y. F. Pan, J. Liu, H. Xu, Y. B. Mohamed, Y. L. Orlov, S. Velkov, A. Ho, P. H. Mei, E. G. Y. Chew, P. Y. H. Huang, W.-J. Welboren, Y. Han, H. S. Ooi, P. N. Ariyaratne, V. B. Vega, Y. Luo, P. Y. Tan, P. Y. Choy, K. D. S. A. Wansa, B. Zhao, K. S. Lim, S. C. Leow, J. S. Yow, R. Joseph, H. Li, K. V. Desai, J. S. Thomsen, Y. K. Lee, R. K. M. Karuturi, T. Herve, G. Bourque, H. G. Stunnenberg, X. Ruan, V. Cacheux-Rataboul, W.-K. Sung, E. T. Liu, C.-L. Wei, E. Cheung, Y. Ruan, An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58–64 (2009). [doi:10.1038/nature08497](https://doi.org/10.1038/nature08497) [Medline](#)
3. X. Han, S. Chen, E. Flynn, S. Wu, D. Wintner, Y. Shen, Distinct epigenomic patterns are associated with haploinsufficiency and predict risk genes of developmental disorders. *Nat. Commun.* **9**, 2138 (2018). [doi:10.1038/s41467-018-04552-7](https://doi.org/10.1038/s41467-018-04552-7) [Medline](#)
4. X. Wang, D. B. Goldstein, Enhancer Domains Predict Gene Pathogenicity and Inform Gene Discovery in Complex Disease. *Am. J. Hum. Genet.* **106**, 215–233 (2020). [doi:10.1016/j.ajhg.2020.01.012](https://doi.org/10.1016/j.ajhg.2020.01.012) [Medline](#)
5. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). [doi:10.1038/nature11247](https://doi.org/10.1038/nature11247) [Medline](#)
6. Roadmap Epigenomics Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shores, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S. Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F.

- Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
7. M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kuttyavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, J. A. Stamatoyannopoulos, Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).  
[doi:10.1126/science.1222794](https://doi.org/10.1126/science.1222794) [Medline](#)
  8. M. Kellis, B. Wold, M. P. Snyder, B. E. Bernstein, A. Kundaje, G. K. Marinov, L. D. Ward, E. Birney, G. E. Crawford, J. Dekker, I. Dunham, L. L. Elnitski, P. J. Farnham, E. A. Feingold, M. Gerstein, M. C. Giddings, D. M. Gilbert, T. R. Gingeras, E. D. Green, R. Guigo, T. Hubbard, J. Kent, J. D. Lieb, R. M. Myers, M. J. Pazin, B. Ren, J. A. Stamatoyannopoulos, Z. Weng, K. P. White, R. C. Hardison, Defining functional DNA elements in the human genome. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6131–6138 (2014).  
[doi:10.1073/pnas.1318948111](https://doi.org/10.1073/pnas.1318948111) [Medline](#)
  9. T. A. Manolio, F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. C. Mackay, S. A. McCarroll, P. M. Visscher, Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009). [doi:10.1038/nature08494](https://doi.org/10.1038/nature08494) [Medline](#)
  10. S. Chatterjee, A. Kapoor, J. A. Akiyama, D. R. Auer, D. Lee, S. Gabriel, C. Berrios, L. A. Pennacchio, A. Chakravarti, Enhancer Variants Synergistically Drive Dysfunction of a Gene Regulatory Network In Hirschsprung Disease. *Cell* **167**, 355–368.e10 (2016).  
[doi:10.1016/j.cell.2016.09.005](https://doi.org/10.1016/j.cell.2016.09.005) [Medline](#)
  11. O. Corradin, A. J. Cohen, J. M. Luppino, I. M. Bayles, F. R. Schumacher, P. C. Scacheri, Modeling disease risk through analysis of physical interactions between genetic variants within chromatin regulatory circuitry. *Nat. Genet.* **48**, 1313–1320 (2016).  
[doi:10.1038/ng.3674](https://doi.org/10.1038/ng.3674) [Medline](#)
  12. D. C. Factor, A. M. Barbeau, K. C. Allan, L. R. Hu, M. Madhavan, A. T. Hoang, K. E. A. Hazel, P. A. Hall, S. Nisraiyya, F. J. Najm, T. E. Miller, Z. S. Nevin, R. T. Karl, B. R. Lima, Y. Song, A. G. Sibert, G. K. Dhillon, C. Volsko, C. F. Bartels, D. J. Adams, R. Dutta, M. D. Gallagher, W. Phu, A. Kozlenkov, S. Dracheva, P. C. Scacheri, P. J. Tesar, O. Corradin, Cell Type-Specific Intralocus Interactions Reveal Oligodendrocyte Mechanisms in MS. *Cell* **181**, 382–395.e21 (2020). [doi:10.1016/j.cell.2020.03.002](https://doi.org/10.1016/j.cell.2020.03.002)  
[Medline](#)
  13. P. C. Phillips, Epistasis—The essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **9**, 855–867 (2008). [doi:10.1038/nrg2452](https://doi.org/10.1038/nrg2452) [Medline](#)
  14. W. A. Whyte, D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, R. A. Young, Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).  
[doi:10.1016/j.cell.2013.03.035](https://doi.org/10.1016/j.cell.2013.03.035) [Medline](#)

15. J. Lovén, H. A. Hoke, C. Y. Lin, A. Lau, D. A. Orlando, C. R. Vakoc, J. E. Bradner, T. I. Lee, R. A. Young, Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013). [doi:10.1016/j.cell.2013.03.036](https://doi.org/10.1016/j.cell.2013.03.036) [Medline](#)
16. D. Hnisz, B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke, R. A. Young, Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013). [doi:10.1016/j.cell.2013.09.053](https://doi.org/10.1016/j.cell.2013.09.053) [Medline](#)
17. S. C. J. Parker, M. L. Stitzel, D. L. Taylor, J. M. Orozco, M. R. Erdos, J. A. Akiyama, K. L. van Bueren, P. S. Chines, N. Narisu, B. L. Black, A. Visel, L. A. Pennacchio, F. S. Collins, J. Becker, B. Benjamin, R. Blakesley, G. Bouffard, S. Brooks, H. Coleman, M. Dekhtyar, M. Gregory, X. Guan, J. Gupta, J. Han, A. Hargrove, S. Ho, T. Johnson, R. Legaspi, S. Lovett, Q. Maduro, C. Masiello, B. Maskeri, J. McDowell, C. Montemayor, J. Mullikin, M. Park, N. Riebow, K. Schandler, B. Schmidt, C. Sison, M. Stantripop, J. Thomas, P. Thomas, M. Vemulapalli, A. Young; NISC Comparative Sequencing Program; National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Authors; NISC Comparative Sequencing Program Authors, Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17921–17926 (2013). [doi:10.1073/pnas.1317023110](https://doi.org/10.1073/pnas.1317023110) [Medline](#)
18. L. Pasquali, K. J. Gaulton, S. A. Rodríguez-Seguí, L. Mularoni, I. Miguel-Escalada, Í. Akerman, J. J. Tena, I. Morán, C. Gómez-Marín, M. van de Bunt, J. Ponsa-Cobas, N. Castro, T. Nammo, I. Cebola, J. García-Hurtado, M. A. Maestro, F. Pattou, L. Piemonti, T. Berney, A. L. Gloyn, P. Ravassard, J. L. G. Skarmeta, F. Müller, M. I. McCarthy, J. Ferrer, Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014). [doi:10.1038/ng.2870](https://doi.org/10.1038/ng.2870) [Medline](#)
19. D. Hnisz, J. Schuijers, C. Y. Lin, A. S. Weintraub, B. J. Abraham, T. I. Lee, J. E. Bradner, R. A. Young, Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol. Cell* **58**, 362–370 (2015). [doi:10.1016/j.molcel.2015.02.014](https://doi.org/10.1016/j.molcel.2015.02.014) [Medline](#)
20. J. Huang, X. Liu, D. Li, Z. Shao, H. Cao, Y. Zhang, E. Trompouki, T. V. Bowman, L. I. Zon, G.-C. Yuan, S. H. Orkin, J. Xu, Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Dev. Cell* **36**, 9–23 (2016). [doi:10.1016/j.devcel.2015.12.014](https://doi.org/10.1016/j.devcel.2015.12.014) [Medline](#)
21. D. Hay, J. R. Hughes, C. Babbs, J. O. J. Davies, B. J. Graham, L. Hanssen, M. T. Kassouf, A. M. Marieke Oudelaar, J. A. Sharpe, M. C. Suci, J. Telenius, R. Williams, C. Rode, P.-S. Li, L. A. Pennacchio, J. A. Sloane-Stanley, H. Ayyub, S. Butler, T. Sauka-Spengler, R. J. Gibbons, A. J. H. Smith, W. G. Wood, D. R. Higgs, Genetic dissection of the  $\alpha$ -globin super-enhancer in vivo. *Nat. Genet.* **48**, 895–903 (2016). [doi:10.1038/ng.3605](https://doi.org/10.1038/ng.3605) [Medline](#)
22. C. Bahr, L. von Paleske, V. V. Uslu, S. Remeseiro, N. Takayama, S. W. Ng, A. Murison, K. Langenfeld, M. Petretich, R. Scognamiglio, P. Zeisberger, A. S. Benk, I. Amit, P. W. Zandstra, M. Lupien, J. E. Dick, A. Trumpp, F. Spitz, A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* **553**, 515–520 (2018). [doi:10.1038/nature25193](https://doi.org/10.1038/nature25193) [Medline](#)
23. H. Y. Shin, M. Willi, K. HyunYoo, X. Zeng, C. Wang, G. Metser, L. Hennighausen, Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat. Genet.* **48**,

- 904–911 (2016). [doi:10.1038/ng.3606](https://doi.org/10.1038/ng.3606) [Medline](#)
24. J. Huang, K. Li, W. Cai, X. Liu, Y. Zhang, S. H. Orkin, J. Xu, G.-C. Yuan, Dissecting super-enhancer hierarchy based on chromatin interactions. *Nat. Commun.* **9**, 943 (2018). [doi:10.1038/s41467-018-03279-9](https://doi.org/10.1038/s41467-018-03279-9) [Medline](#)
25. M. W. Perry, A. N. Boettiger, M. Levine, Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13570–13575 (2011). [doi:10.1073/pnas.1109873108](https://doi.org/10.1073/pnas.1109873108) [Medline](#)
26. C. P. Fulco, M. Munschauer, R. Anyoha, G. Munson, S. R. Grossman, E. M. Perez, M. Kane, B. Cleary, E. S. Lander, J. M. Engreitz, Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016). [doi:10.1126/science.aag2445](https://doi.org/10.1126/science.aag2445) [Medline](#)
27. T. Wang, K. Birsoy, N. W. Hughes, K. M. Krupczak, Y. Post, J. J. Wei, E. S. Lander, D. M. Sabatini, Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015). [doi:10.1126/science.aac7041](https://doi.org/10.1126/science.aac7041) [Medline](#)
28. L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, W. A. Lim, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013). [doi:10.1016/j.cell.2013.02.022](https://doi.org/10.1016/j.cell.2013.02.022) [Medline](#)
29. P. I. Thakore, A. M. D’Ippolito, L. Song, A. Safi, N. K. Shivakumar, A. M. Kadi, T. E. Reddy, G. E. Crawford, C. A. Gersbach, Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* **12**, 1143–1149 (2015). [doi:10.1038/nmeth.3630](https://doi.org/10.1038/nmeth.3630) [Medline](#)
30. L. A. Gilbert, M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, E. H. Whitehead, C. Guimaraes, B. Panning, H. L. Ploegh, M. C. Bassik, L. S. Qi, M. Kampmann, J. S. Weissman, Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647–661 (2014). [doi:10.1016/j.cell.2014.09.029](https://doi.org/10.1016/j.cell.2014.09.029) [Medline](#)
31. M. Kosicki, K. Tomberg, A. Bradley, Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018). [doi:10.1038/nbt.4192](https://doi.org/10.1038/nbt.4192) [Medline](#)
32. H. Y. Shin, C. Wang, H. K. Lee, K. H. Yoo, X. Zeng, T. Kuhns, C. M. Yang, T. Mohr, C. Liu, L. Hennighausen, CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat. Commun.* **8**, 15464 (2017). [doi:10.1038/ncomms15464](https://doi.org/10.1038/ncomms15464) [Medline](#)
33. J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010). [doi:10.18637/jss.v033.i01](https://doi.org/10.18637/jss.v033.i01) [Medline](#)
34. B. Lai, Q. Tang, W. Jin, G. Hu, D. Wangsa, K. Cui, B. Z. Stanton, G. Ren, Y. Ding, M. Zhao, S. Liu, J. Song, T. Ried, K. Zhao, Trac-looping measures genome structure and chromatin accessibility. *Nat. Methods* **15**, 741–747 (2018). [doi:10.1038/s41592-018-0107-y](https://doi.org/10.1038/s41592-018-0107-y) [Medline](#)
35. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13–23 (2017). [doi:10.1016/j.cell.2017.02.007](https://doi.org/10.1016/j.cell.2017.02.007) [Medline](#)
36. W.-K. Cho, J.-H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, I. I. Cisse, Mediator and RNA

- polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018). [doi:10.1126/science.aar4199](https://doi.org/10.1126/science.aar4199) [Medline](#)
37. B. R. Sabari, A. Dall’Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, I. I. Cisse, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, R. A. Young, Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018). [doi:10.1126/science.aar3958](https://doi.org/10.1126/science.aar3958) [Medline](#)
38. P. Filippakopoulos, J. Qi, S. Picaud, Y. Shen, W. B. Smith, O. Fedorov, E. M. Morse, T. Keates, T. T. Hickman, I. Felletar, M. Philpott, S. Munro, M. R. McKeown, Y. Wang, A. L. Christie, N. West, M. J. Cameron, B. Schwartz, T. D. Heightman, N. La Thangue, C. A. French, O. Wiest, A. L. Kung, S. Knapp, J. E. Bradner, Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010). [doi:10.1038/nature09504](https://doi.org/10.1038/nature09504) [Medline](#)
39. T. Lappalainen, M. Sammeth, M. R. Friedländer, P. A. C. ’t Hoen, J. Monlong, M. A. Rivas, M. González-Porta, N. Kurbatova, T. Griebel, P. G. Ferreira, M. Barann, T. Wieland, L. Greger, M. van Iterson, J. Almlöf, P. Ribeca, I. Pulyakhina, D. Esser, T. Giger, A. Tikhonov, M. Sultan, G. Bertier, D. G. MacArthur, M. Lek, E. Lizano, H. P. J. Buermans, I. Padiou, T. Schwarzmayer, O. Karlberg, H. Ongen, H. Kilpinen, S. Beltran, M. Gut, K. Kahlem, V. Amstislavskiy, O. Stegle, M. Pirinen, S. B. Montgomery, P. Donnelly, M. I. McCarthy, P. Flicek, T. M. Strom, H. Lehrach, S. Schreiber, R. Sudbrak, A. Carracedo, S. E. Antonarakis, R. Häsler, A.-C. Syvänen, G.-J. van Ommen, A. Brazma, T. Meitinger, P. Rosenstiel, R. Guigó, I. G. Gut, X. Estivill, E. T. Dermitzakis; Geuvadis Consortium, Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013). [doi:10.1038/nature12531](https://doi.org/10.1038/nature12531) [Medline](#)
40. Wellcome Trust Case Control Consortium, Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007). [doi:10.1038/nature05911](https://doi.org/10.1038/nature05911)
41. J. J. Yang, C. Cheng, M. Devidas, X. Cao, Y. Fan, D. Campana, W. Yang, G. Neale, N. J. Cox, P. Scheet, M. J. Borowitz, N. J. Winick, P. L. Martin, C. L. Willman, W. P. Bowman, B. M. Camitta, A. Carroll, G. H. Reaman, W. L. Carroll, M. Loh, S. P. Hunger, C.-H. Pui, W. E. Evans, M. V. Relling, Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. *Nat. Genet.* **43**, 237–241 (2011). [doi:10.1038/ng.763](https://doi.org/10.1038/ng.763) [Medline](#)
42. J. J. Yang, C. Cheng, M. Devidas, X. Cao, D. Campana, W. Yang, Y. Fan, G. Neale, N. Cox, P. Scheet, M. J. Borowitz, N. J. Winick, P. L. Martin, W. P. Bowman, B. Camitta, G. H. Reaman, W. L. Carroll, C. L. Willman, S. P. Hunger, W. E. Evans, C.-H. Pui, M. Loh, M. V. Relling, Genome-wide association study identifies germline polymorphisms associated with relapse of childhood acute lymphoblastic leukemia. *Blood* **120**, 4197–4204 (2012). [doi:10.1182/blood-2012-07-440107](https://doi.org/10.1182/blood-2012-07-440107) [Medline](#)
43. J. Vijayakrishnan, R. Kumar, M. Y. R. Henrion, A. V. Moorman, P. S. Rachakonda, I. Hosen, M. I. da Silva Filho, A. Holroyd, S. E. Dobbins, R. Koehler, H. Thomsen, J. A. Irving, J. M. Allan, T. Lightfoot, E. Roman, S. E. Kinsey, E. Sheridan, P. D. Thompson, P. Hoffmann, M. M. Nöthen, S. Heilmann-Heimbach, K. H. Jöckel, M. Greaves, C. J.



- Harrison, C. R. Bartram, M. Schrappe, M. Stanulla, K. Hemminki, R. S. Houlston, A genome-wide association study identifies risk loci for childhood acute lymphoblastic leukemia at 10q26.13 and 12q23.1. *Leukemia* **31**, 573–579 (2017). [doi:10.1038/leu.2016.271](https://doi.org/10.1038/leu.2016.271) [Medline](#)
44. J. Vijayakrishnan, M. Qian, J. B. Studd, W. Yang, B. Kinnersley, P. J. Law, P. Broderick, E. A. Raetz, J. Allan, C.-H. Pui, A. Vora, W. E. Evans, A. Moorman, A. Yeoh, W. Yang, C. Li, C. R. Bartram, C. G. Mullighan, M. Zimmerman, S. P. Hunger, M. Schrappe, M. V. Relling, M. Stanulla, M. L. Loh, R. S. Houlston, J. J. Yang, Identification of four novel associations for B-cell acute lymphoblastic leukaemia risk. *Nat. Commun.* **10**, 5348 (2019). [doi:10.1038/s41467-019-13069-6](https://doi.org/10.1038/s41467-019-13069-6) [Medline](#)
  45. T. Zhen, E. M. Kwon, L. Zhao, J. Hsu, R. K. Hyde, Y. Lu, L. Alemu, N. A. Speck, P. P. Liu, *Chd7* deficiency delays leukemogenesis in mice induced by *Cbfb-MYH11*. *Blood* **130**, 2431–2442 (2017). [doi:10.1182/blood-2017-04-780106](https://doi.org/10.1182/blood-2017-04-780106) [Medline](#)
  46. D. D. Lam, F. S. J. de Souza, S. Nasif, M. Yamashita, R. López-Leal, V. Otero-Corchon, K. Meece, H. Sampath, A. J. Mercer, S. L. Wardlaw, M. Rubinstein, M. J. Low, Partially redundant enhancers cooperatively maintain Mammalian pomc expression above a critical functional threshold. *PLOS Genet.* **11**, e1004935 (2015). [doi:10.1371/journal.pgen.1004935](https://doi.org/10.1371/journal.pgen.1004935) [Medline](#)
  47. D. Xu, O. Gokcumen, E. Khurana, Loss-of-function tolerance of enhancers in the human genome. *PLOS Genet.* **16**, e1008663 (2020). [doi:10.1371/journal.pgen.1008663](https://doi.org/10.1371/journal.pgen.1008663) [Medline](#)
  48. M. Osterwalder, I. Barozzi, V. Tissières, Y. Fukuda-Yuzawa, B. J. Mannion, S. Y. Afzal, E. A. Lee, Y. Zhu, I. Plajzer-Frick, C. S. Pickle, M. Kato, T. H. Garvin, Q. T. Pham, A. N. Harrington, J. A. Akiyama, V. Afzal, J. Lopez-Rios, D. E. Dickel, A. Visel, L. A. Pennacchio, Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243 (2018). [doi:10.1038/nature25461](https://doi.org/10.1038/nature25461) [Medline](#)
  49. K. Shrinivas, B. R. Sabari, E. L. Coffey, I. A. Klein, A. Boija, A. V. Zamudio, J. Schuijers, N. M. Hannett, P. A. Sharp, R. A. Young, A. K. Chakraborty, Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol. Cell* **75**, 549–561.e7 (2019). [doi:10.1016/j.molcel.2019.07.009](https://doi.org/10.1016/j.molcel.2019.07.009) [Medline](#)
  50. N. S. Benabdallah, I. Williamson, R. S. Illingworth, L. Kane, S. Boyle, D. Sengupta, G. R. Grimes, P. Therizols, W. A. Bickmore, Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol. Cell* **76**, 473–484.e7 (2019). [doi:10.1016/j.molcel.2019.07.038](https://doi.org/10.1016/j.molcel.2019.07.038) [Medline](#)
  51. X. Lin, L. S. Qi, Code and processed data for: Nested Epistasis Enhancer Networks for Robust Genome Regulation, Zenodo (2022); <https://zenodo.org/record/6823833#.YuwiTXbMI2w>.
  52. X. Lin, L. S. Qi, SRE predictor for: Nested Epistasis Enhancer Networks for Robust Genome Regulation, Zenodo (2022); <https://zenodo.org/record/6823807#.Yuwi63bMI2w>.
  53. B. Chen, L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G.-W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi, B. Huang, Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013). [doi:10.1016/j.cell.2013.12.001](https://doi.org/10.1016/j.cell.2013.12.001) [Medline](#)
  54. Y. Liu, C. Yu, T. P. Daley, F. Wang, W. S. Cao, S. Bhate, X. Lin, C. Still 2nd, H. Liu, D.

- Zhao, H. Wang, X. S. Xie, S. Ding, W. H. Wong, M. Wernig, L. S. Qi, CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming. *Cell Stem Cell* **23**, 758–771.e8 (2018). [doi:10.1016/j.stem.2018.09.003](https://doi.org/10.1016/j.stem.2018.09.003) [Medline](#)
55. Y. Takei, S. Shah, S. Harvey, L. S. Qi, L. Cai, Multiplexed Dynamic Imaging of Genomic Loci by Combined CRISPR Imaging and DNA Sequential FISH. *Biophys. J.* **112**, 1773–1776 (2017). [doi:10.1016/j.bpj.2017.03.024](https://doi.org/10.1016/j.bpj.2017.03.024) [Medline](#)
56. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012). [doi:10.1038/nmeth.2089](https://doi.org/10.1038/nmeth.2089) [Medline](#)
57. J. T. Robinson, D. Turner, N. C. Durand, H. Thorvaldsdóttir, J. P. Mesirov, E. L. Aiden, Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Syst.* **6**, 256–258.e1 (2018). [doi:10.1016/j.cels.2018.01.001](https://doi.org/10.1016/j.cels.2018.01.001) [Medline](#)
58. M. R. Mumbach, A. T. Satpathy, E. A. Boyle, C. Dai, B. G. Gowen, S. W. Cho, M. L. Nguyen, A. J. Rubin, J. M. Granja, K. R. Kazane, Y. Wei, T. Nguyen, P. G. Greenside, M. R. Corces, J. Tycko, D. R. Simeonov, N. Suliman, R. Li, J. Xu, R. A. Flynn, A. Kundaje, P. A. Khavari, A. Marson, J. E. Corn, T. Quertermous, W. J. Greenleaf, H. Y. Chang, Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* **49**, 1602–1612 (2017). [doi:10.1038/ng.3963](https://doi.org/10.1038/ng.3963) [Medline](#)
59. T. Barrett, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C. L. Robertson, N. Serova, S. Davis, A. Soboleva, NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995 (2013). [doi:10.1093/nar/gks1193](https://doi.org/10.1093/nar/gks1193) [Medline](#)
60. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). [doi:10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923) [Medline](#)
61. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008). [doi:10.1186/gb-2008-9-9-r137](https://doi.org/10.1186/gb-2008-9-9-r137) [Medline](#)
62. J. Lyu, J. J. Li, J. Su, F. Peng, Y. E. Chen, X. Ge, W. Li, DORGE: Discovery of Oncogenes and tumor suppressor genes using Genetic and Epigenetic features. *Sci. Adv.* **6**, eaba6784 (2020). [doi:10.1126/sciadv.aba6784](https://doi.org/10.1126/sciadv.aba6784) [Medline](#)
63. C. P. Fulco, J. Nasser, T. R. Jones, G. Munson, D. T. Bergman, V. Subramanian, S. R. Grossman, R. Anyoha, B. R. Doughty, T. A. Patwardhan, T. H. Nguyen, M. Kane, E. M. Perez, N. C. Durand, C. A. Lareau, E. K. Stamenova, E. L. Aiden, E. S. Lander, J. M. Engreitz, Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019). [doi:10.1038/s41588-019-0538-0](https://doi.org/10.1038/s41588-019-0538-0) [Medline](#)
64. Y. Cao, S. Liu, G. Ren, Q. Tang, K. Zhao, cLoops2: A full-stack comprehensive analytical tool for chromatin interactions. *Nucleic Acids Res.* **50**, 57–71 (2022). [doi:10.1093/nar/gkab1233](https://doi.org/10.1093/nar/gkab1233) [Medline](#)
65. D. H. Phanstiel, A. P. Boyle, N. Heidari, M. P. Snyder, Mango: A bias-correcting ChIA-PET analysis pipeline. *Bioinformatics* **31**, 3092–3098 (2015).

[doi:10.1093/bioinformatics/btv336](https://doi.org/10.1093/bioinformatics/btv336) [Medline](#)

66. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). [doi:10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033) [Medline](#)
67. D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **81**, 559–575 (2007). [doi:10.1086/519795](https://doi.org/10.1086/519795)
68. M. D. Mailman, M. Feolo, Y. Jin, M. Kimura, K. Tryka, R. Bagoutdinov, L. Hao, A. Kiang, J. Paschall, L. Phan, N. Popova, S. Pretel, L. Ziyabari, M. Lee, Y. Shao, Z. Y. Wang, K. Sirotkin, M. Ward, M. Kholodov, K. Zbicz, J. Beck, M. Kimelman, S. Shevelev, D. Preuss, E. Yaschenko, A. Graeff, J. Ostell, S. T. Sherry, The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* **39**, 1181–1186 (2007). [doi:10.1038/ng1007-1181](https://doi.org/10.1038/ng1007-1181) [Medline](#)
69. M. Goldman, B. Craft, A. Brooks, J. Zhu, D. Haussler, The UCSC Xena Platform for cancer genomics data visualization and interpretation. *bioRxiv* 326470 [Preprint] (2018); [doi:10.1101/2021.08.29.21262792v1](https://doi.org/10.1101/2021.08.29.21262792v1)
70. J. Gong, S. Mei, C. Liu, Y. Xiang, Y. Ye, Z. Zhang, J. Feng, R. Liu, L. Diao, A.-Y. Guo, X. Miao, L. Han, PancanQTL: Systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res.* **46**, D971–D976 (2018). [doi:10.1093/nar/gkx861](https://doi.org/10.1093/nar/gkx861) [Medline](#)
71. O. Stegle, L. Parts, M. Piipari, J. Winn, R. Durbin, Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7**, 500–507 (2012). [doi:10.1038/nprot.2011.457](https://doi.org/10.1038/nprot.2011.457) [Medline](#)
72. X. Zheng, D. Levine, J. Shen, S. M. Gogarten, C. Laurie, B. S. Weir, A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012). [doi:10.1093/bioinformatics/bts606](https://doi.org/10.1093/bioinformatics/bts606) [Medline](#)