# Supplemental Figures

# Title: Nested Epistasis Enhancer Networks for Robust Genome Regulation

**Authors:** Xueqiu Lin[1†], Yanxia Liu[1†], Shuai Liu[2†], Xiang Zhu[3,4,5], Lingling Wu[1], Yanyu Zhu[1], Dehua Zhao[1], Xiaoshu Xu[1], Augustine Chemparathy[6], Haifeng Wang[1], Yaqiang Cao[2], Muneaki Nakamura[1], Jasprina N. Noordermeer[1], Marie La Russa[1], Wing Hung Wong[3,7], Keji Zhao[2], Lei S. Qi[1,8,9]*

## Supplemental Figures

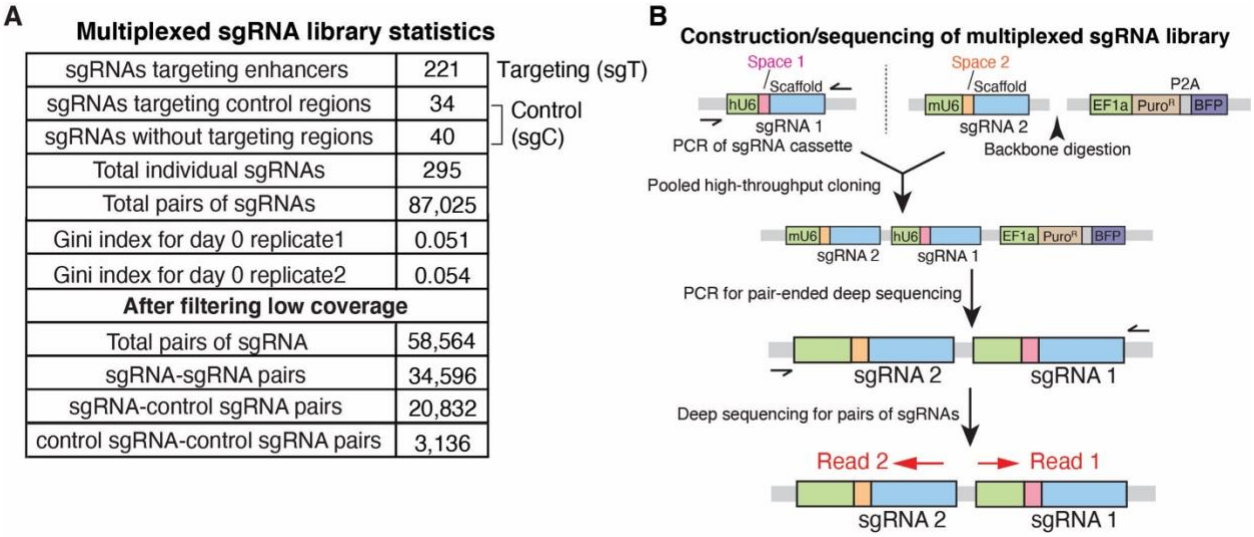



**Fig. S1. Experimental setup for the multiplexed CRISPRi screening.**

**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.

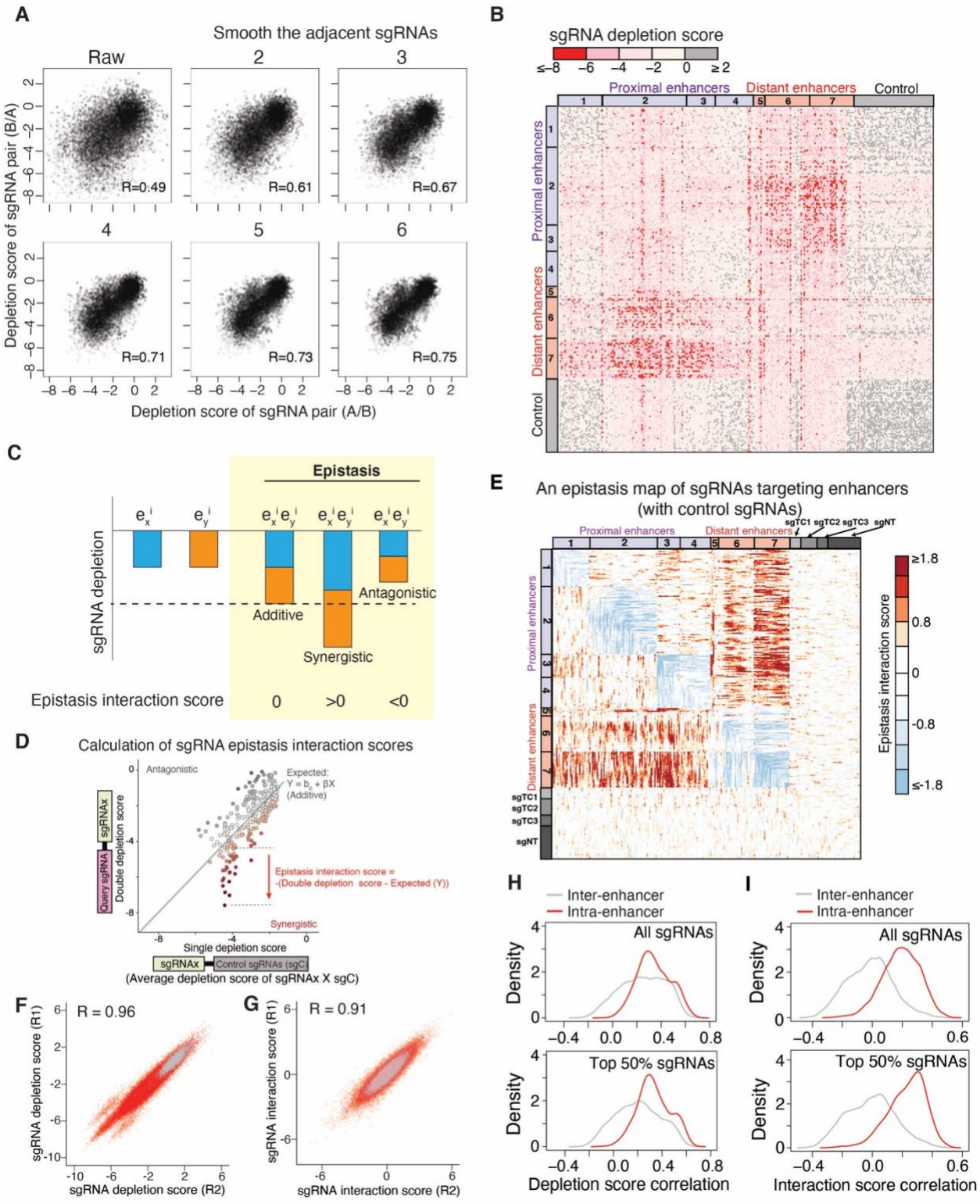



**Fig. S2. Data analysis for the multiplexed CRISPRi screening.**

**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 (sgRNAx, 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).

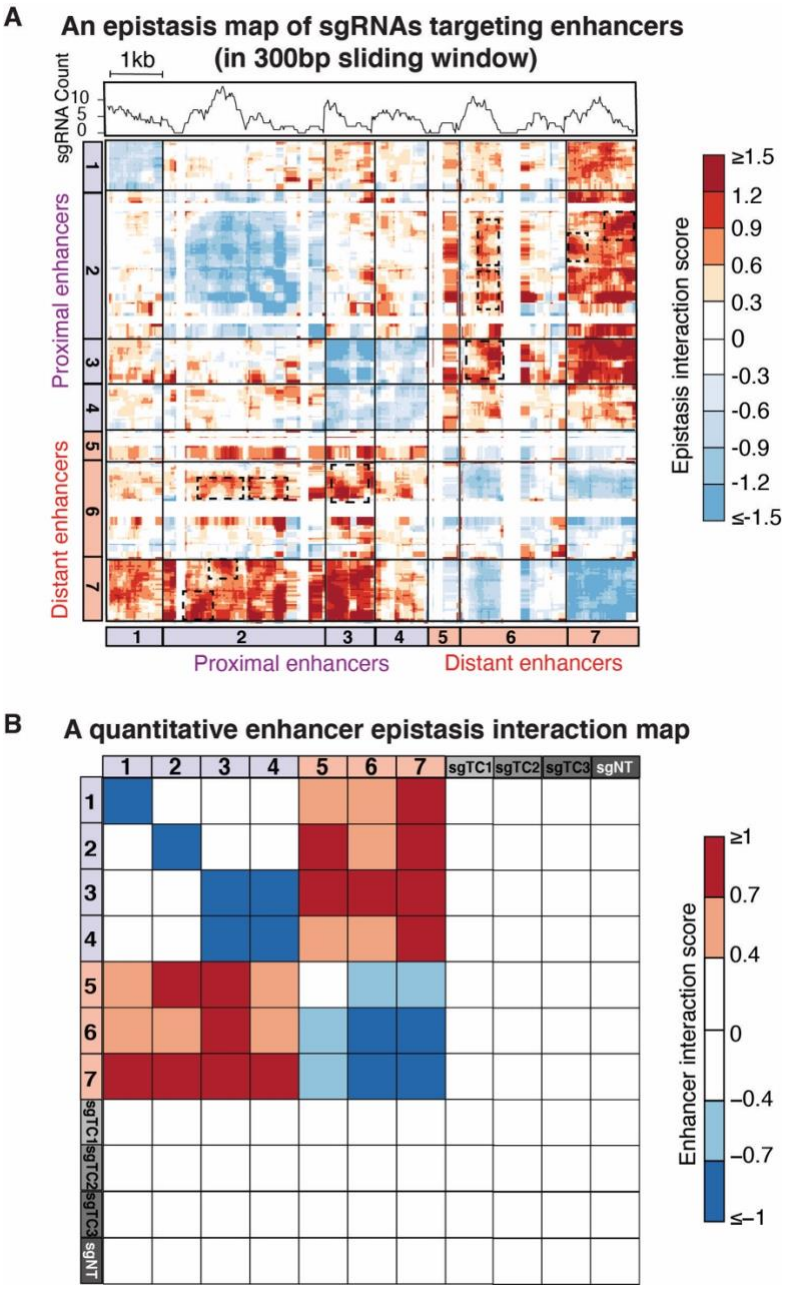



**Fig. S3. Additional data analysis for multiplexed CRISPRi screening.**

**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.

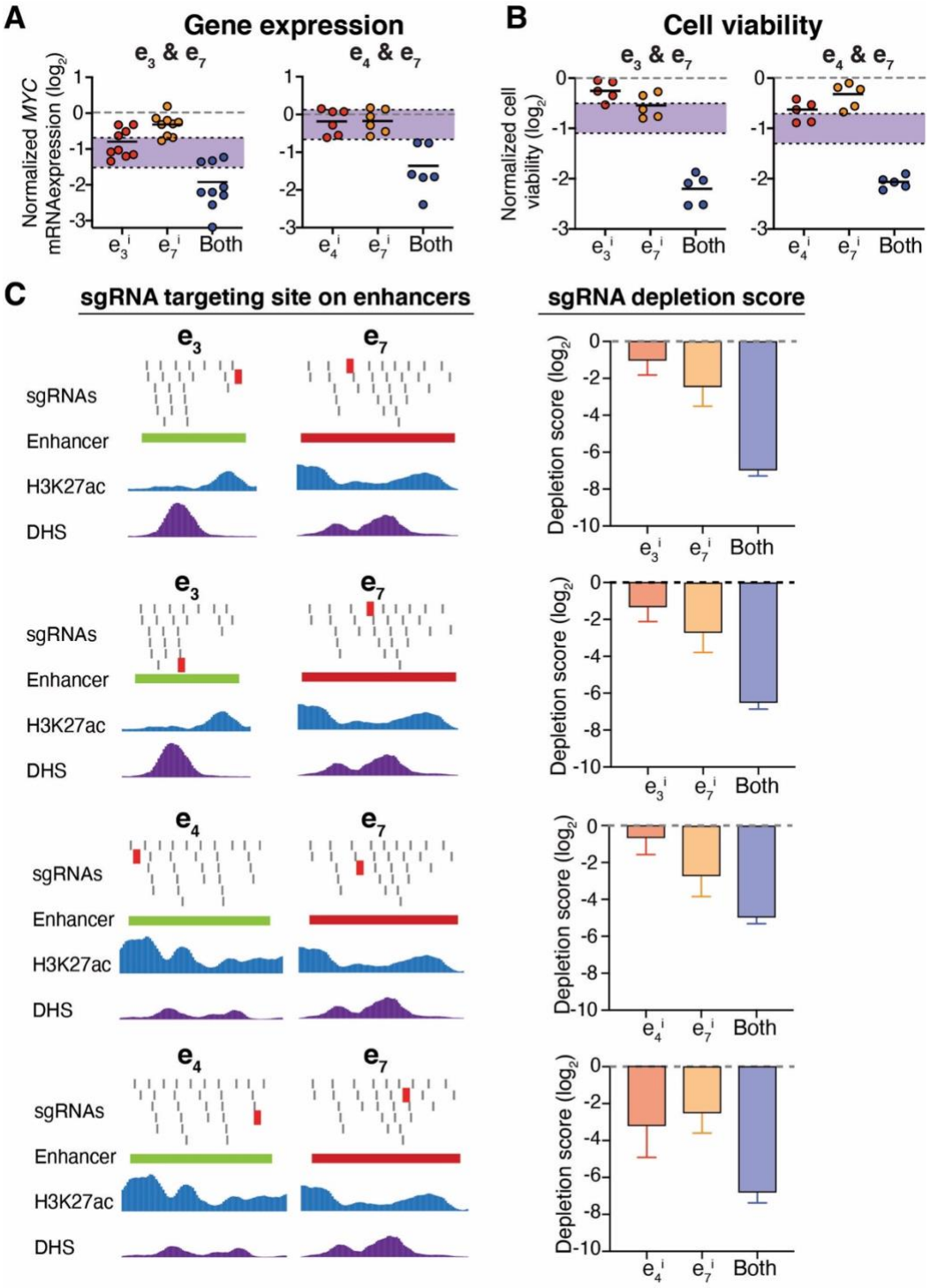



**Fig. S4. Additional experimental validation of enhancer epistasis interactions.**

**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 ± one standard derivation. *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 ± one standard derivation. *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 ± standard deviation.

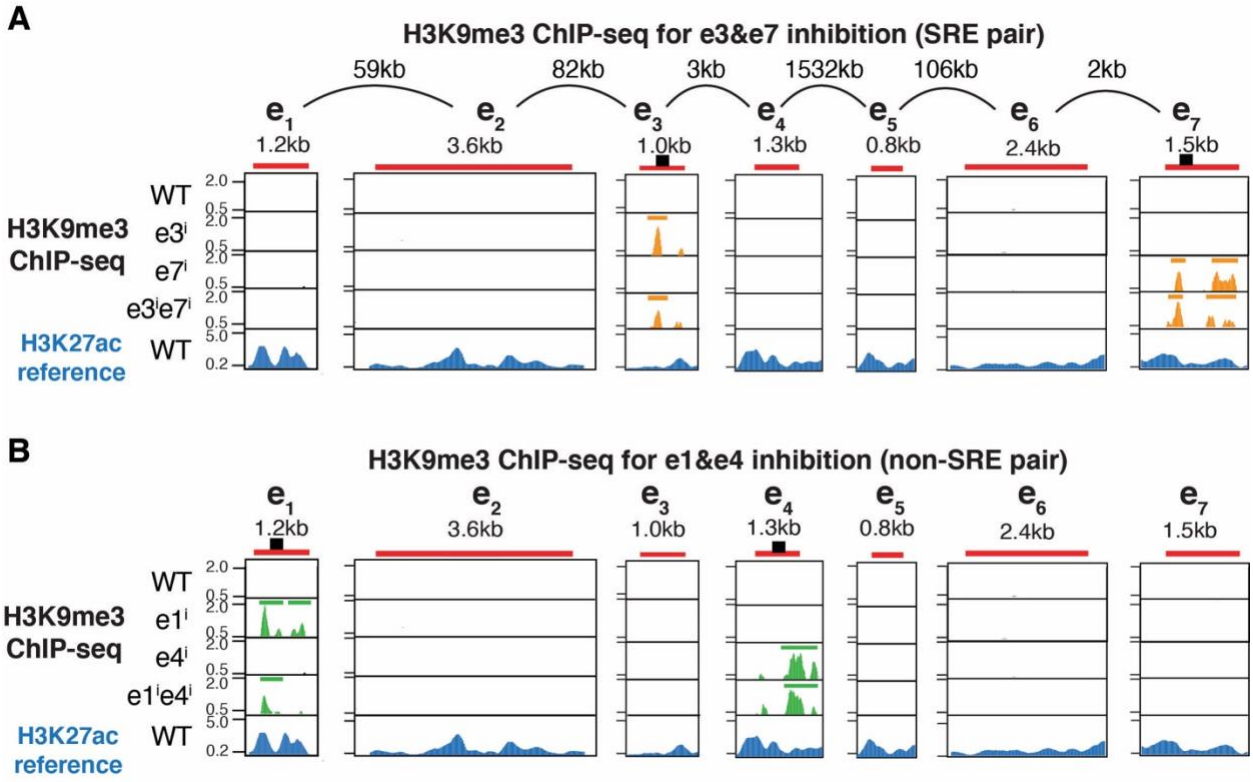



**Fig. S5. H3K9me3 ChIP-seq characterization of CRISPRi perturbation of enhancers.**

**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.

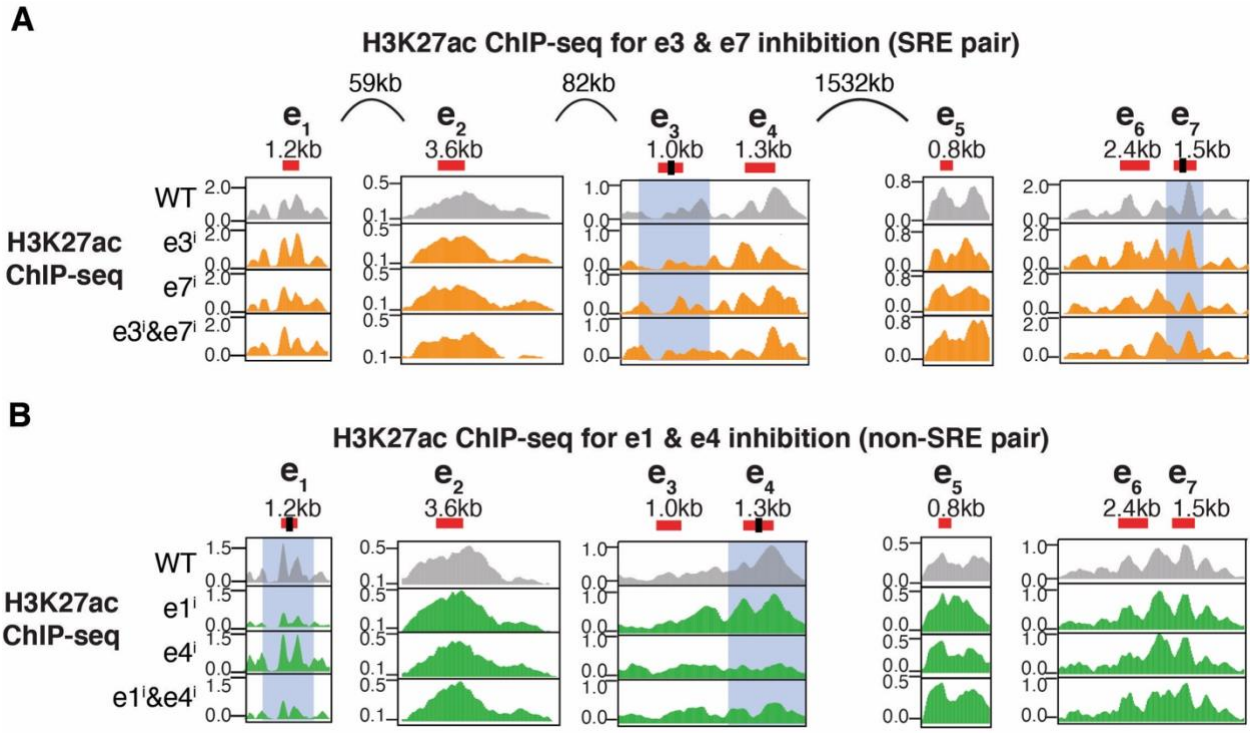



**Fig. S6. H3K27ac ChIP-seq characterization of CRISPRi perturbation of enhancers.**

**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.

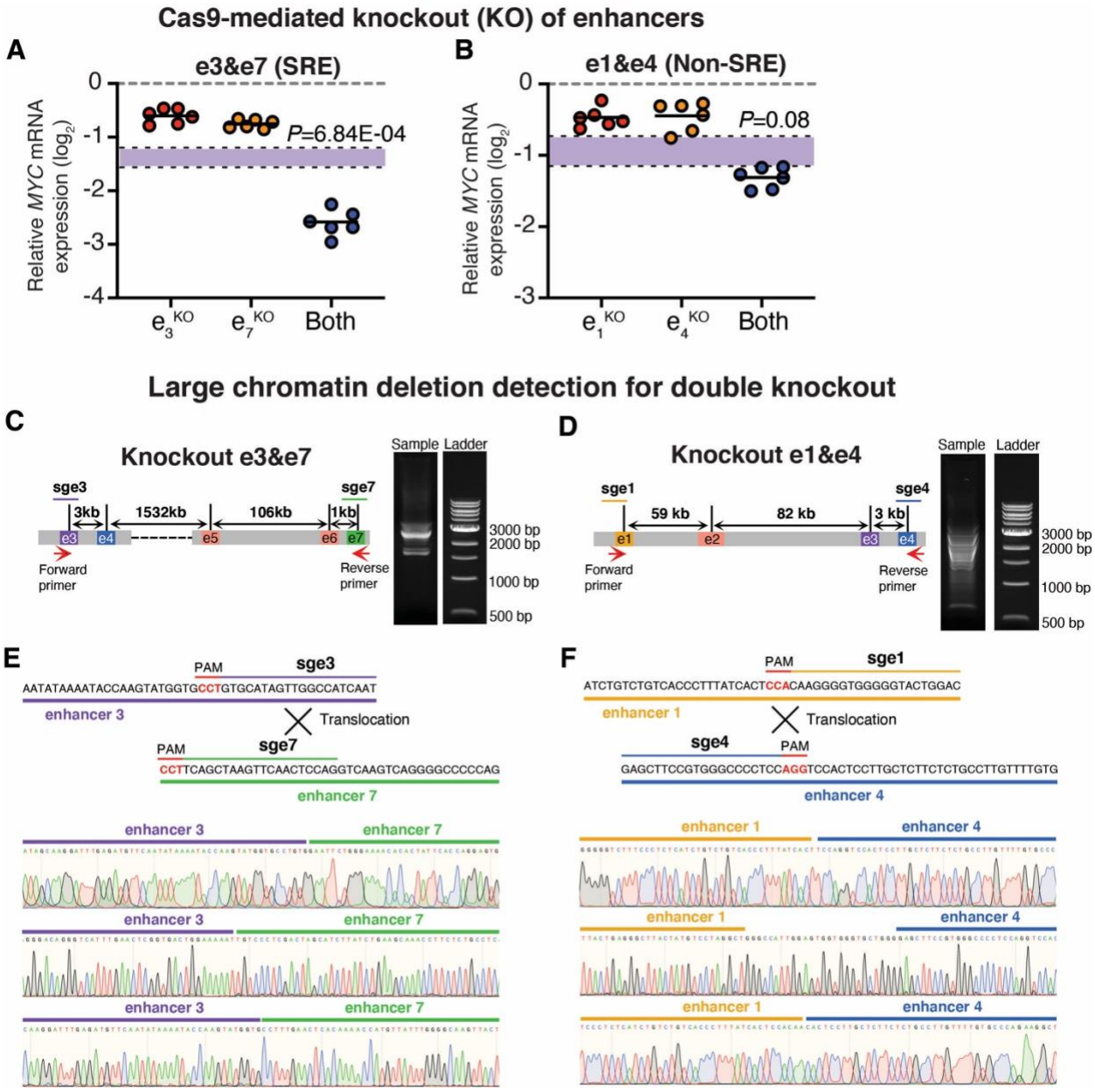



**Fig. S7. Cas9-mediated knockout (KO) of SRE and non-SRE enhancers at the *MYC* locus, and detected large chromatin deletions for double enhancer KO.**

**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 ± 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).

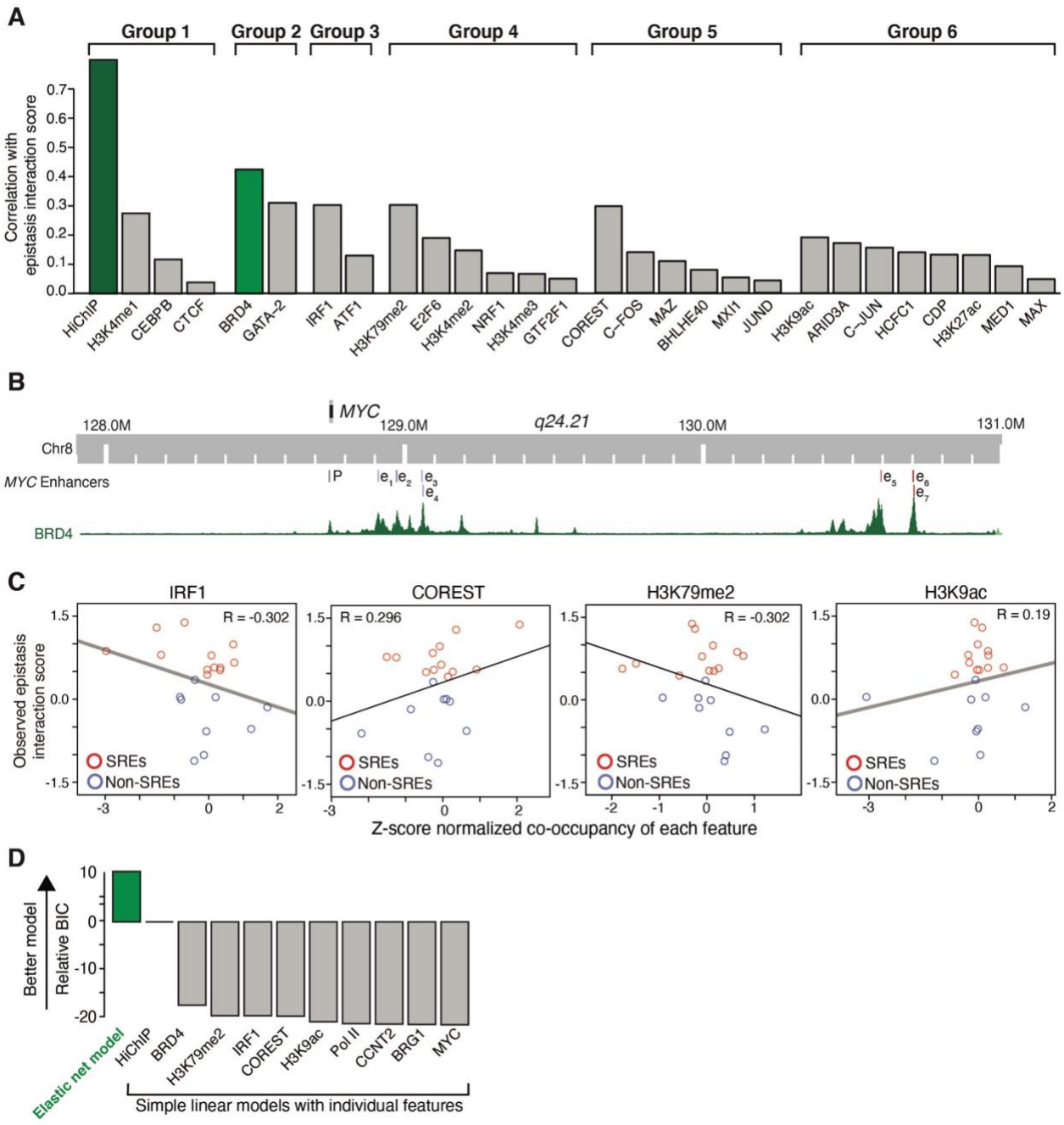



**Fig. S8. Data regression of feature groups to predict SREs using an elastic net regularized linear regression model.**

**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.

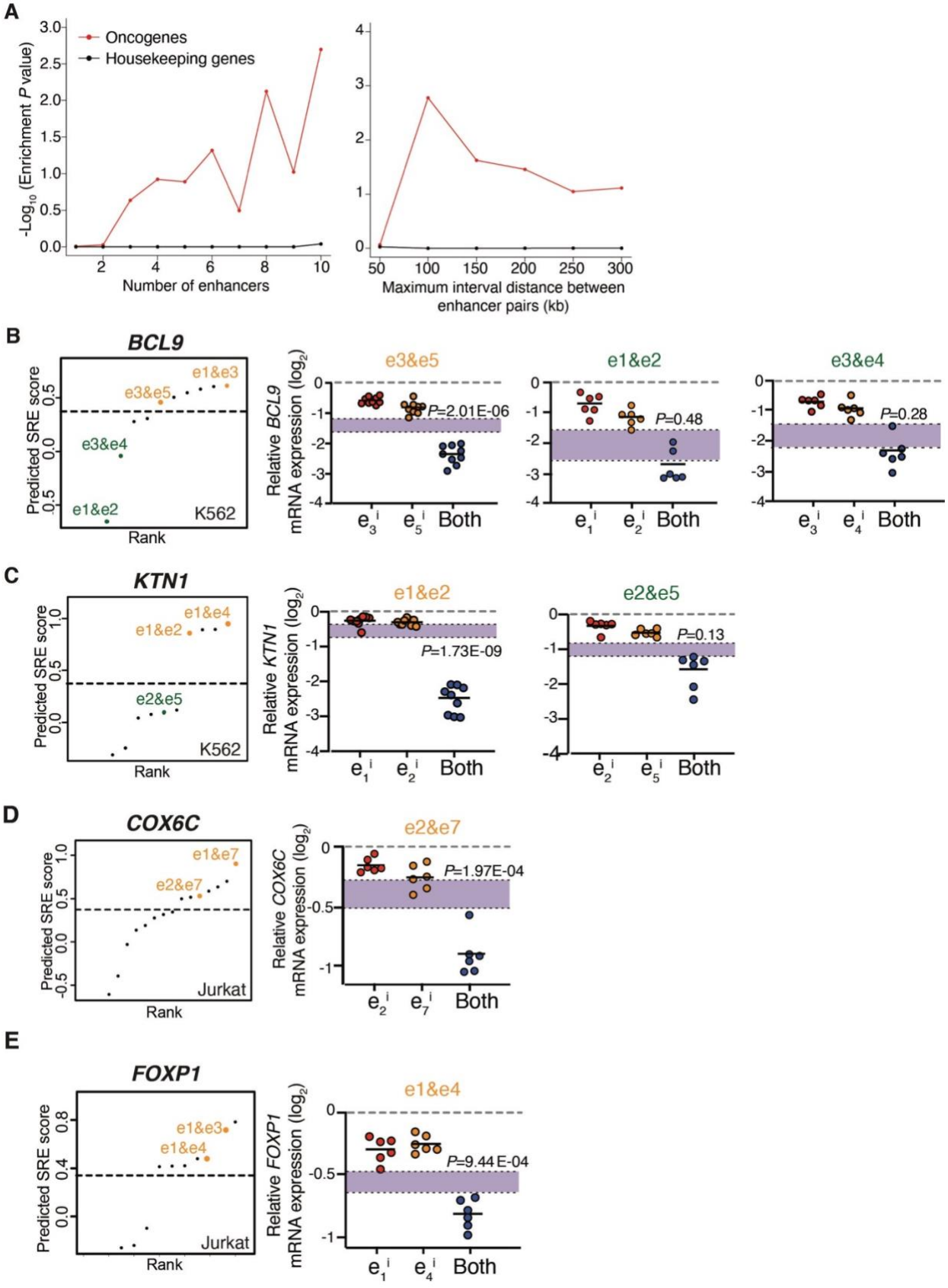

A
Oncogenes
Housekeeping genes
-Log10 (Enrichment P value)
Number of enhancers
Maximum interval distance between enhancer pairs (kb)
B
BCL9
Predicted SRE score
e1&e3
e3&e5
e3&e4
e1&e2
K562
Rank
Relative BCL9 mRNA expression (log2)
e3&e5
P=2.01E-06
e1&e2
P=0.48
e3&e4
P=0.28
Both
C
KTN1
e1&e4
e1&e2
e2&e5
K562
Rank
Relative KTN1 mRNA expression (log2)
e1&e2
P=1.73E-09
e2&e5
P=0.13
Both
D
COX6C
e1&e7
e2&e7
Jurkat
Rank
Relative COX6C mRNA expression (log2)
e2&e7
P=1.97E-04
Both
E
FOXP1
e1&e3
e1&e4
Jurkat
Rank
Relative FOXP1 mRNA expression (log2)
e1&e4
P=9.44 E-04
Both

**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 *KTN1* loci (**C**) in K562 cells. Left: Rank of predicted SREs at the *BCL9* and *KTN1* 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 *KTN1* 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 ± 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 ± one standard derivation. *P* values are calculated by two-sided Student's t-test.

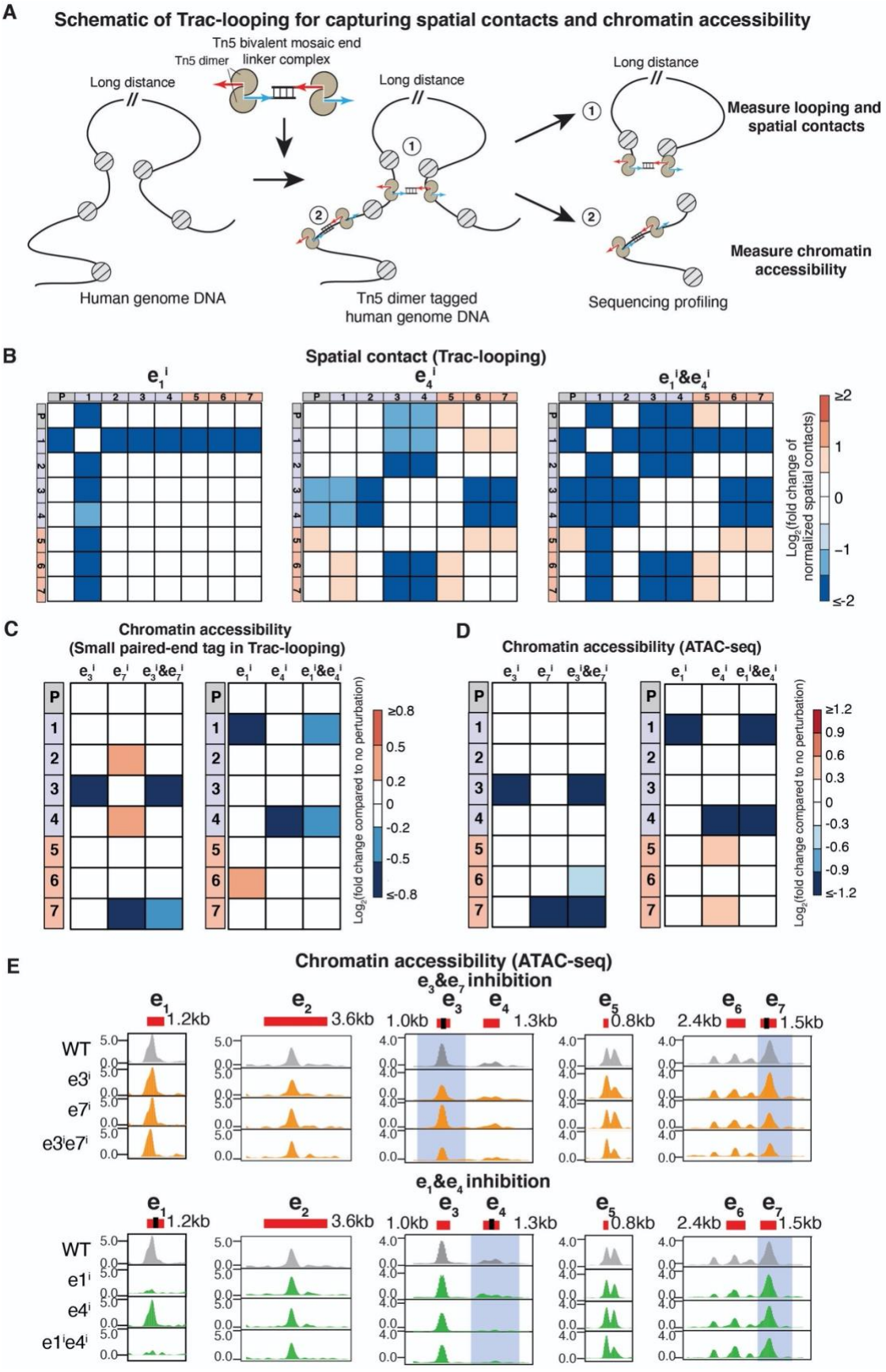

A
Schematic of Trac-looping for capturing spatial contacts and chromatin accessibility
Tn5 bivalent mosaic end linker complex
Tn5 dimer
Long distance
Long distance
Long distance
Measure looping and spatial contacts
Measure chromatin accessibility
Human genome DNA
Tn5 dimer tagged human genome DNA
Sequencing profiling
B
Spatial contact (Trac-looping)
e1i
e4i
e1i&e4i
Log2(fold change of normalized spatial contacts)
≥2
1
0
−1
≤-2
C
Chromatin accessibility (Small paired-end tag in Trac-looping)
e3i
e7i
e3i&e7i
e1i
e4i
e1i&e4i
Log2(fold change compared to no perturbation)
≥0.8
0.5
0.2
0
-0.2
-0.5
≤-0.8
D
Chromatin accessibility (ATAC-seq)
Log2(fold change compared to no perturbation)
≥1.2
0.9
0.6
0.3
0
-0.3
-0.6
-0.9
≤-1.2
E
Chromatin accessibility (ATAC-seq)
e3&e7 inhibition
e1 1.2kb
e2 3.6kb
1.0kb
e3
e4 1.3kb
e5 0.8kb
2.4kb
e6
e7 1.5kb
WT
e3i
e7i
e3ie7i
e1&e4 inhibition
WT
e1i
e4i
e1ie4i

**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 $\log_2$ fold change of spatial contacts normalized to the wildtype cells upon perturbation of e1, e4, and e1&e4.

**C,** Heatmap showing the $\log_2$ 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 $\log_2$ 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.

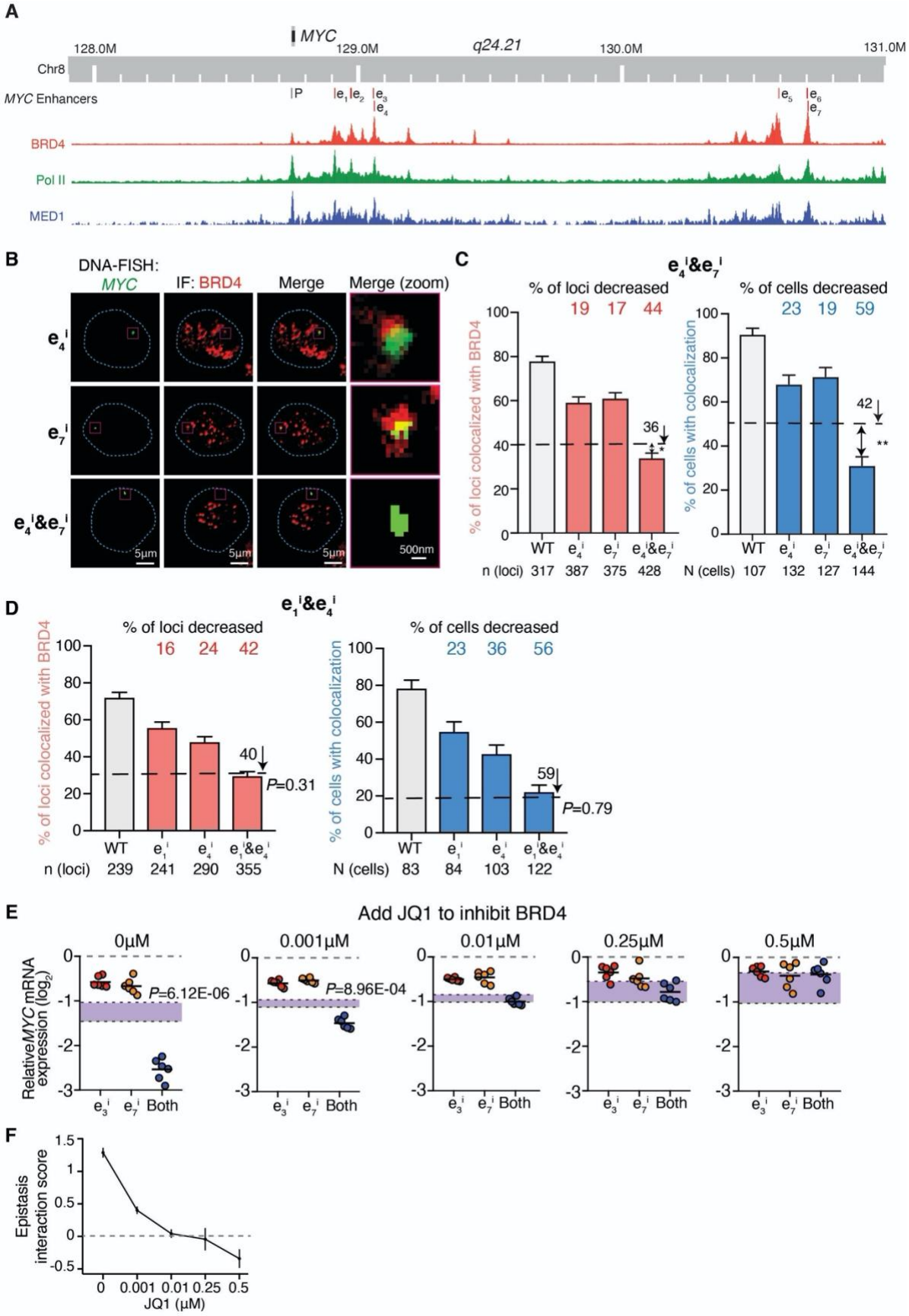

A
MYC
128.0M
129.0M
q24.21
130.0M
131.0M
Chr8
MYC Enhancers
P
e1 e2
e3 e4
e5
e6 e7
BRD4
Pol II
MED1
B
DNA-FISH:
MYC
IF: BRD4
Merge
Merge (zoom)
e4i
e7i
e4i&e7i
5μm
5μm
5μm
500nm
C
e4i&e7i
% of loci decreased
19 17 44
% of loci colocalized with BRD4
36
% of cells decreased
23 19 59
% of cells with colocalization
42
**
WT e4i e7i e4i&e7i
n (loci) 317 387 375 428
N (cells) 107 132 127 144
D
e1i&e4i
% of loci decreased
16 24 42
40
P=0.31
% of cells decreased
23 36 56
59
P=0.79
WT e1i e4i e1i&e4i
n (loci) 239 241 290 355
N (cells) 83 84 103 122
E
Add JQ1 to inhibit BRD4
0μM
0.001μM
0.01μM
0.25μM
0.5μM
Relative MYC mRNA expression (log2)
P=6.12E-06
P=8.96E-04
e3i e7i Both
F
Epistasis interaction score
JQ1 (μM)
0 0.001 0.01 0.25 0.5

**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, 5μm 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 ± 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 ± 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 ($log_2$ 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 ± 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 $log_2$ fold change.

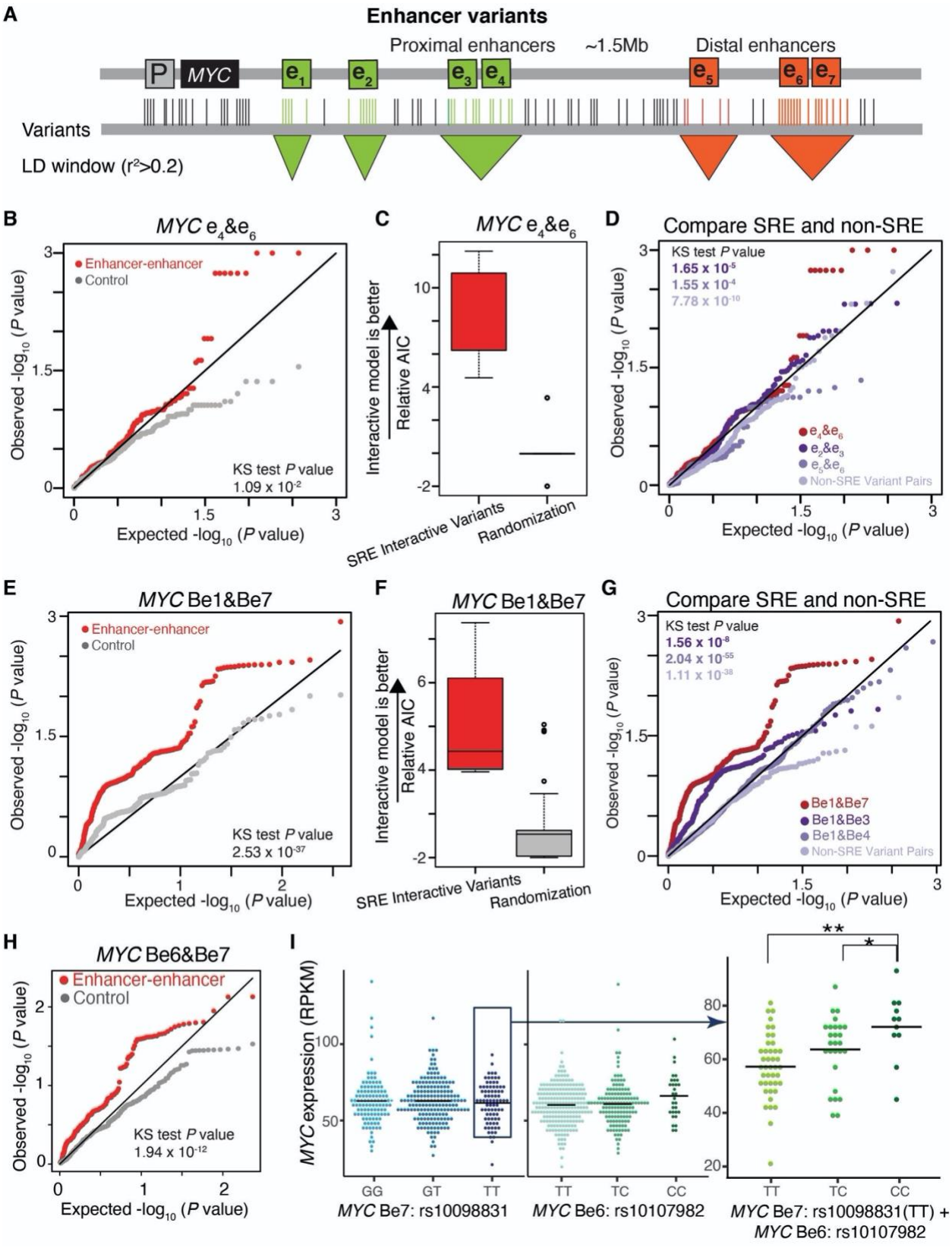
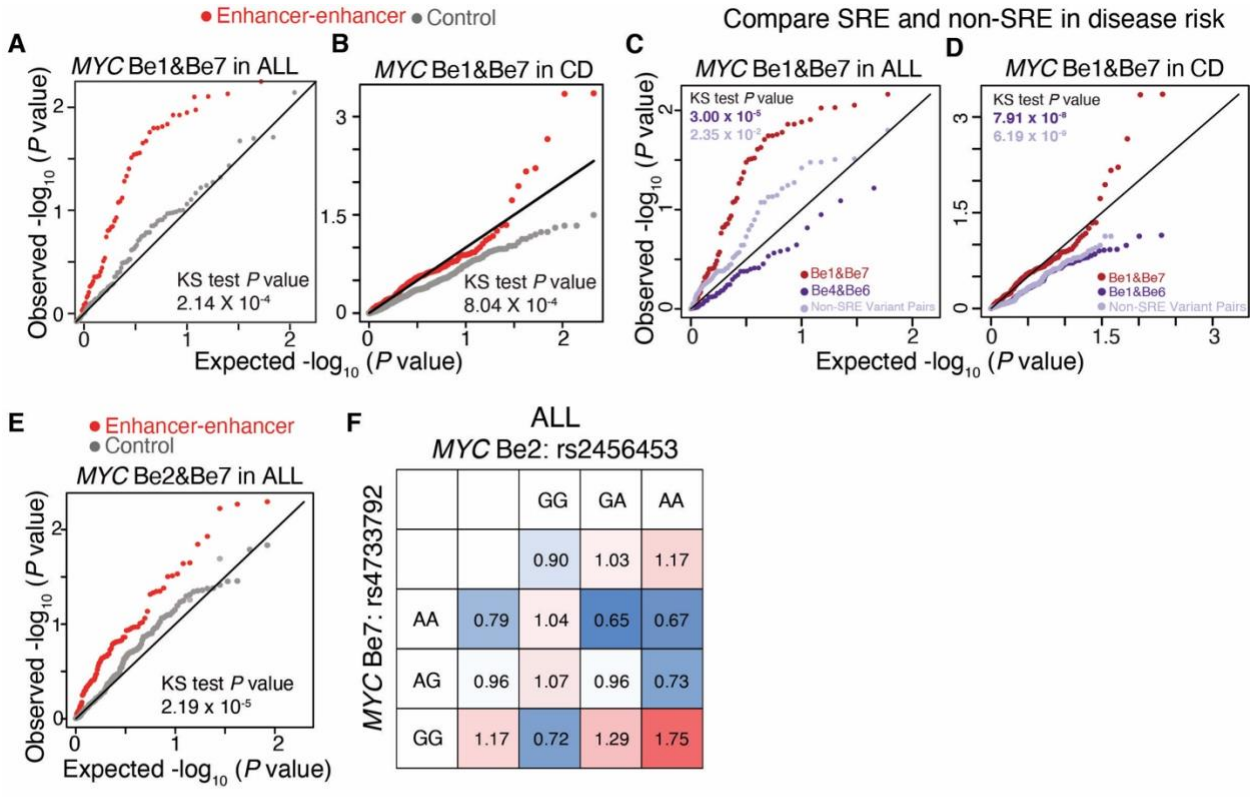



**Fig. S12. Additional analysis of interactive influence of SRE variants for gene expression at the *MYC* locus in K562 cells and GM12878 cells.**

**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. *, $P < 0.05$; **, $P < 0.01$.




**G** **Compare the effect of SRE variant pairs to additive effects for Figure 5F, G and I**

| | *P* value for Odds ratio (1000 random permutations) | *P* value for comparison between Interactive Model and Additive Model | Relative AIC (Additive Model AIC - Interactive Model AIC) |
|---|---|---|---|
| **ALL** | | | |
| Be1:rs2466025 & Be7:rs13259137 | 0.003 (CC+AA: 2.10) | 0.017 | 3.649 |
| Be1: rs7822621 & Be11: rs10957174 | 0.001 (CC+GG: 2.37) | 0.042 | 2.131 |
| **CD** | | | |
| Be1:rs2466030 & Be7:rs10808561 | 0.020 (CC+AA: 1.32) | 0.045 | 2.002 |

**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 x $10^{-4}$.

**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 x $10^{-4}$.

**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 significancy 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.

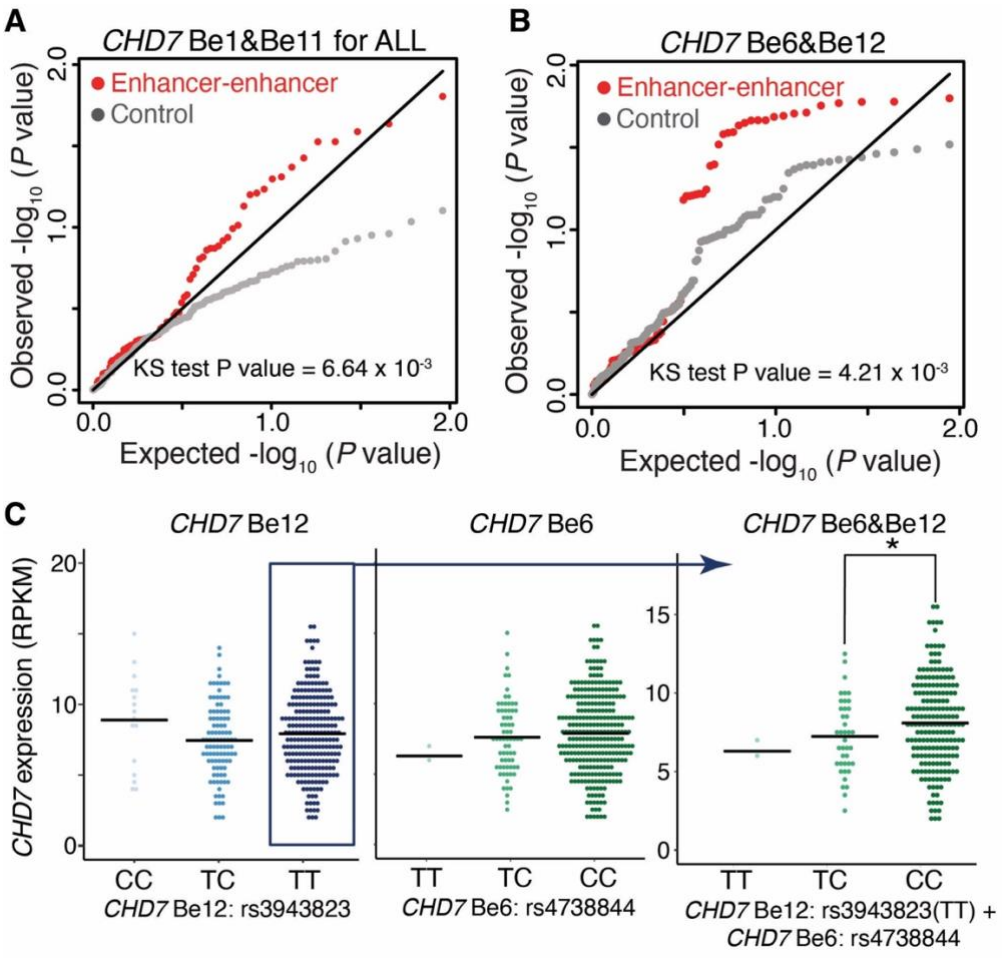



**Fig. S14. Additional analysis of interactive influence for SRE variants on gene expression and clinical risk of B-cell-associated disease in the *CHD7* locus.**

**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. *, $P < 0.05$.

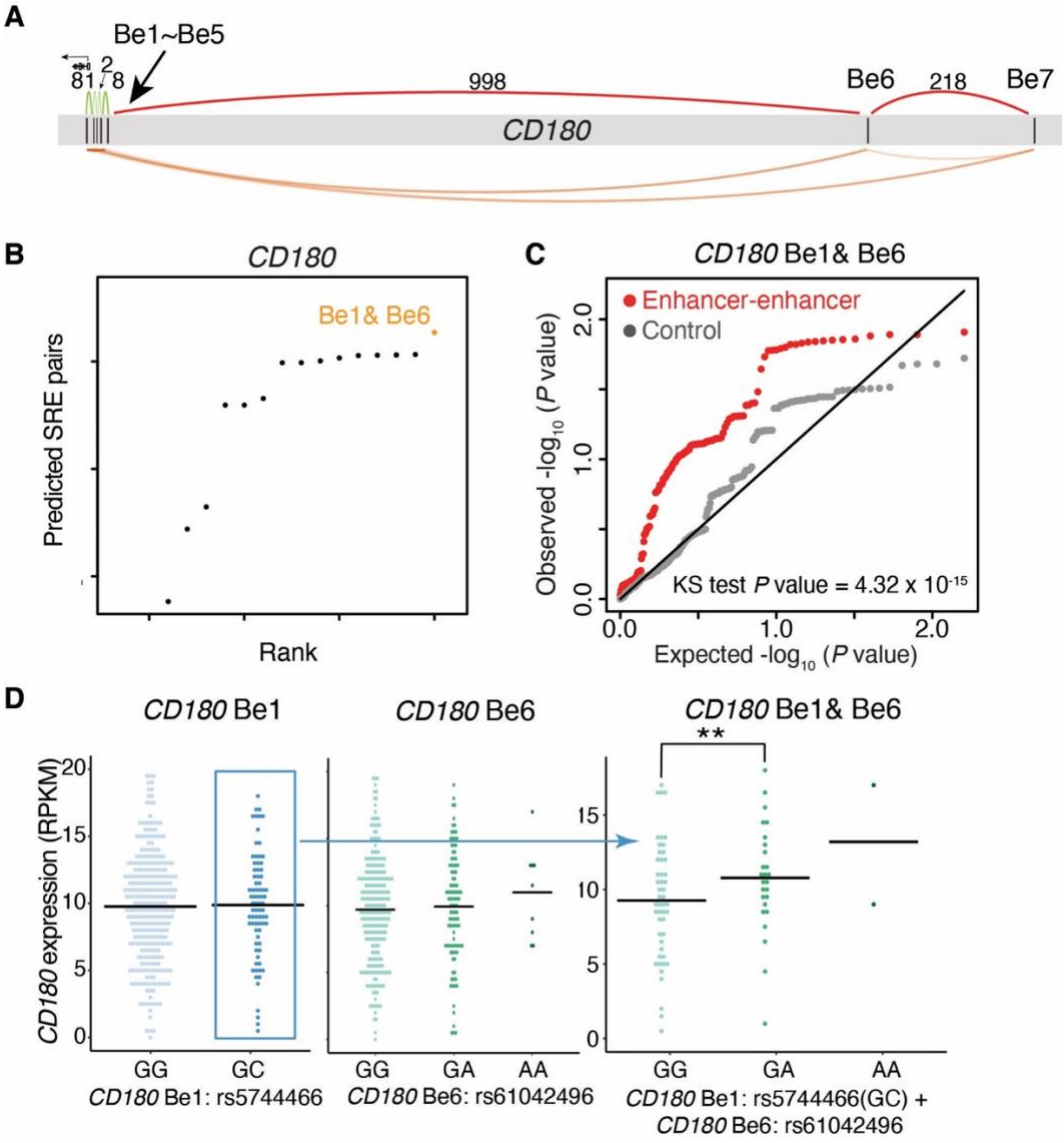



**Fig. S15. Additional analysis of interactive influence for SRE variants on gene expression at *CD180* loci in GM12878 cells.**

**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. **, $P < 0.01$.

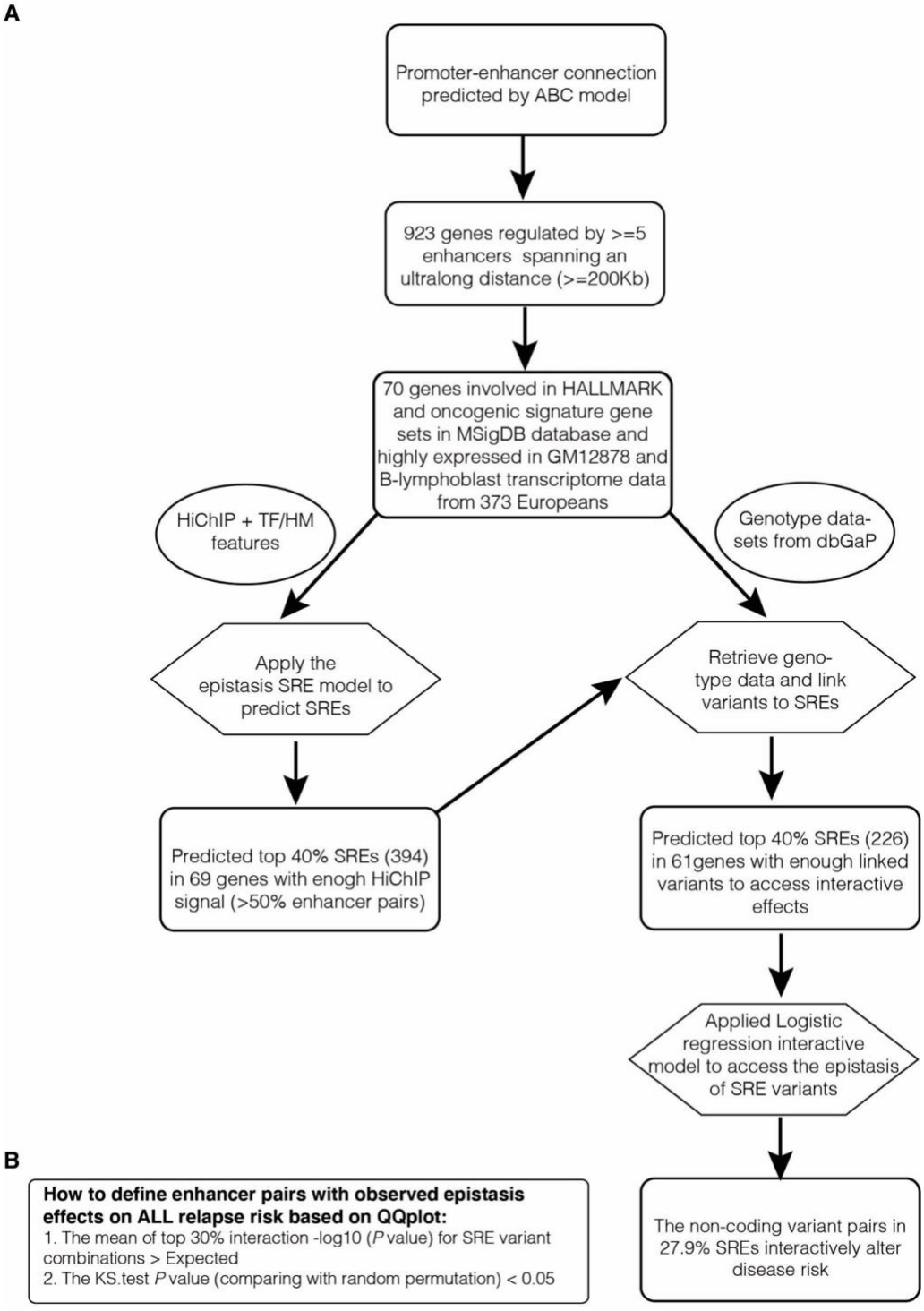



**Fig. S16. Workflow for applying the SRE prediction model to link multiple non-coding variants with complex disease on the genome-wide scale.**

**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.

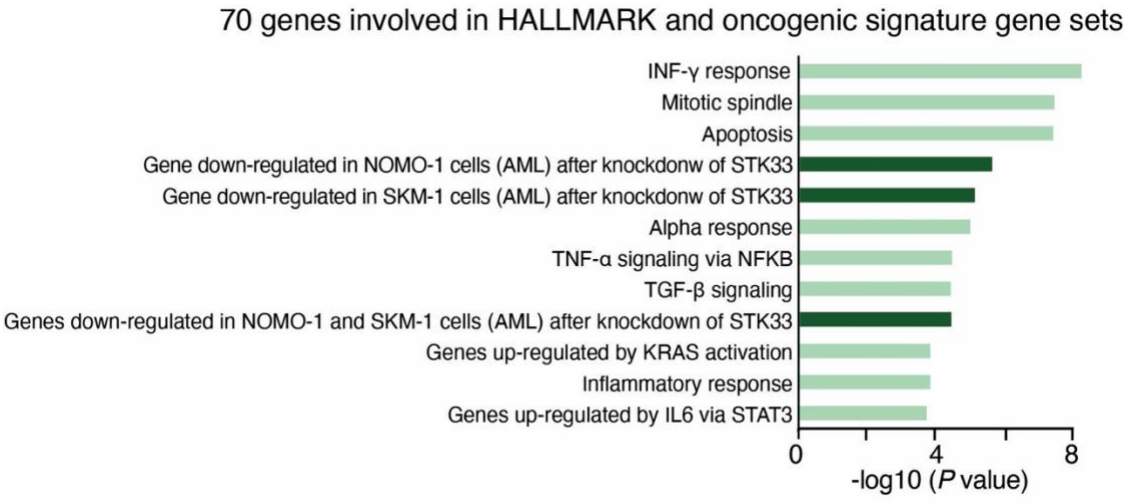



**Fig. S17. Genome-wide analysis of the SRE model in GM12878 cells.** Gene Set Enrichment Analysis (GSEA) of genes regulated by ultralong distance enhancer interaction network in GM12878 cells. Light green represents GSEA immune-related gene sets, and dark green represents oncogenic signature gene sets.