MOLECULAR BIOLOGY

Nested epistasis enhancer networks for robust genome regulation

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]*

Mammalian genomes have multiple enhancers spanning an ultralong distance (>megabases) to modulate important genes, but it is unclear how these enhancers coordinate to achieve this task. We combine multiplexed CRISPRi screening with machine learning to define quantitative enhancer-enhancer interactions. We find that the ultralong distance enhancer network has a nested multilayer architecture that confers functional robustness of gene expression. Experimental characterization reveals that enhancer epistasis is maintained by three-dimensional chromosomal interactions and BRD4 condensation. Machine learning prediction of synergistic enhancers provides an effective strategy to identify noncoding variant pairs associated with pathogenic genes in diseases beyond genome-wide association studies analysis. Our work unveils nested epistasis enhancer networks, which can better explain enhancer functions within cells and in diseases.

Disease-associated genes, including oncogenes, are frequently associated with many remote enhancers spanning across a long genomic distance [>megabases (Mb)] (*1*–*4*). Genome-wide association studies (GWAS) reveal that noncoding variants of the regulatory elements, including enhancers, account for >90% of variants in diseases and can spread over a long distance (*5*–*8*). Although individual enhancer variants may present modest clinical risks (*9*) there are examples showing that a combination of multiple variants may greatly amplify the effects in traits and diseases (*10*–*12*). Similar to gene interactions (*13*), these enhancers may interact as an epistatic network wherein the effect of an enhancer is dependent on other enhancers to regulate gene dosage and confer robustness. Aside from these observations it remains largely unknown why multiple ultralong-distance enhancers exist for important genes and how their interactions modulate gene regulation and diseases.

Enhancer interactions were previously studied within a single enhancer cluster. For example, super enhancers were defined as a dense cluster, which contains adjacent enhancers within tens of kilobases (kb) (*14*–*16*). Other enhancer clusters similar to super enhancers were also reported including stretch enhancers and enhancer clusters (*17*, *18*). A few examples by perturbing local enhancers within these enhancer clusters showed they may interact additively or synergistically for regulatory roles (*19*–*25*). However, these short-range enhancers organized in a cluster cannot explain the prevalence of ultralong-distance enhancers in the human genome.

It remains unknown how multiple enhancers interact with one another over long genomic distances to confer regulatory roles in gene expression and disease risks. We hypothesize that by using ultralong-distance enhancers (>1 Mb), disease-associated genes have evolved high robustness to disruptive effects from genetic variations. These interactions likely occur through an elaborate network on the three-dimensional (3D) genome organization level.

High-resolution multiplexed perturbation of enhancers reveals a nested two-layer epistasis network

To gain insights into the ultralong-distance enhancer network for disease-relevant genes, we adopted a high-resolution approach to quantitatively analyze enhancer interactions in gene regulation. We chose the endogenous *MYC* locus as a model system. As an important oncogene governing cancer cell proliferation, the *MYC* locus encompasses seven enhancers (e1 to e7) spanning a 1.9-Mb region in K562 erythroleukemia cells (*26*). The reported linear correlation between *MYC* expression and cellular growth supports its use as a model system to quantitatively dissect the enhancer epistatic network over ultralong distances (*26*, *27*). We conducted a multiplexed CRISPR interference (CRISPRi) screen (*28*–*30*), using a pooled library consisting of 87,025 pairs of single guide RNAs (sgRNAs) tiling all single and pairwise combinations of seven enhancers (Fig. 1A, fig. S1, A and B, and table S1). We transduced the pooled sgRNA library into K562 cells stably expressing a doxycycline-inducible nuclease-dead dCas9-KRAB fusion and cultured cells for 30 doublings.

We calculated the depletion score of each sgRNA pair by comparing the relative abundance before and after cell culture (Fig. 1A, fig. S2, A and B, and table S2; see Methods). Using the depletion scores to fit a linear additive model we calculated enhancer interaction scores to identify epistasis interactions and generated a high-density quantitative epistasis map of enhancer-targeting sgRNAs (Fig. 1B and fig. S2, C to E; see Methods). We confirmed the epistasis interaction scores were reproducible across biological replicates and different sgRNA pairs targeting the same enhancer pair (fig. S2, F to I). We observed clusters of sgRNAs targeting the same pairs of enhancers showing similar patterns of synergistic or additive interactions, suggesting an epistatic interaction relationship between enhancer pairs (Fig. 1B and figs. S2E and S3A).

We computed the epistasis interaction scores for each enhancer pair by averaging the epistasis interaction scores of the top 25% sgRNA pairs (Fig. 1C and fig. S3B; see Methods). We observed synergistic epistasis when perturbing distant enhancer pairs (>1 Mb), with all four proximal enhancers (e1 to e4) showing strong synergistic interactions with the other three distant enhancers (e5 to e7) upon perturbation. By contrast, perturbation of enhancer pairs within the proximal or distant group mostly showed additive interactions (Fig. 1C).

Our data suggested a nested two-layer architecture of the enhancer epistasis network in regulating genes with large-scale landscapes (Fig. 1D). In the first layer (layer I), enhancer pairs (<100 kb at the *MYC* locus) behave additively after perturbation, suggesting that individual enhancers contribute independently to gene expression. In the second layer (layer II), distant enhancer pairs showed nonlinear synergistic effects after perturbation, which are speculated to function as compensatory regulatory elements for one another to maintain the robustness of gene expression upon perturbation. These synergistic enhancers are distributed over long genomic distances, which likely reduces the chance of co-mutation and thus confers robustness of gene expression against mutations or chromosome perturbations. We define synergistic regulatory enhancers (SREs) as a pair of distant enhancers with synergistic effects on gene expression upon perturbation.

We experimentally validated SREs and non-SRE pairs by examining whether they can combinatorically perturb *MYC* expression and cellular growth. Using different sgRNA pairs targeting the same SREs (e3 and e7; e4 and

[1]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA. [2]Laboratory of Epigenome Biology, Systems Biology Center, National Heart, Lung and Blood Institute NIH, Bethesda, MD 20892, USA. [3]Department of Statistics, Stanford University, Stanford, CA 94305, USA. [4]Department of Statistics, The Pennsylvania State University, University Park, PA 16802, USA. [5]Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA. [6]School of Medicine, Stanford University, Stanford, CA 94305, USA. [7]Department of Biomedical Data Science, Stanford University, Stanford, CA 94305, USA. [8]Sarafan ChEM-H, Stanford University, Stanford, CA 94305, USA. [9]Chan Zuckerberg BioHub, San Francisco, CA 94158, USA.
*Corresponding author. Email: stanley.qi@stanford.edu
†These authors contributed equally to this work.

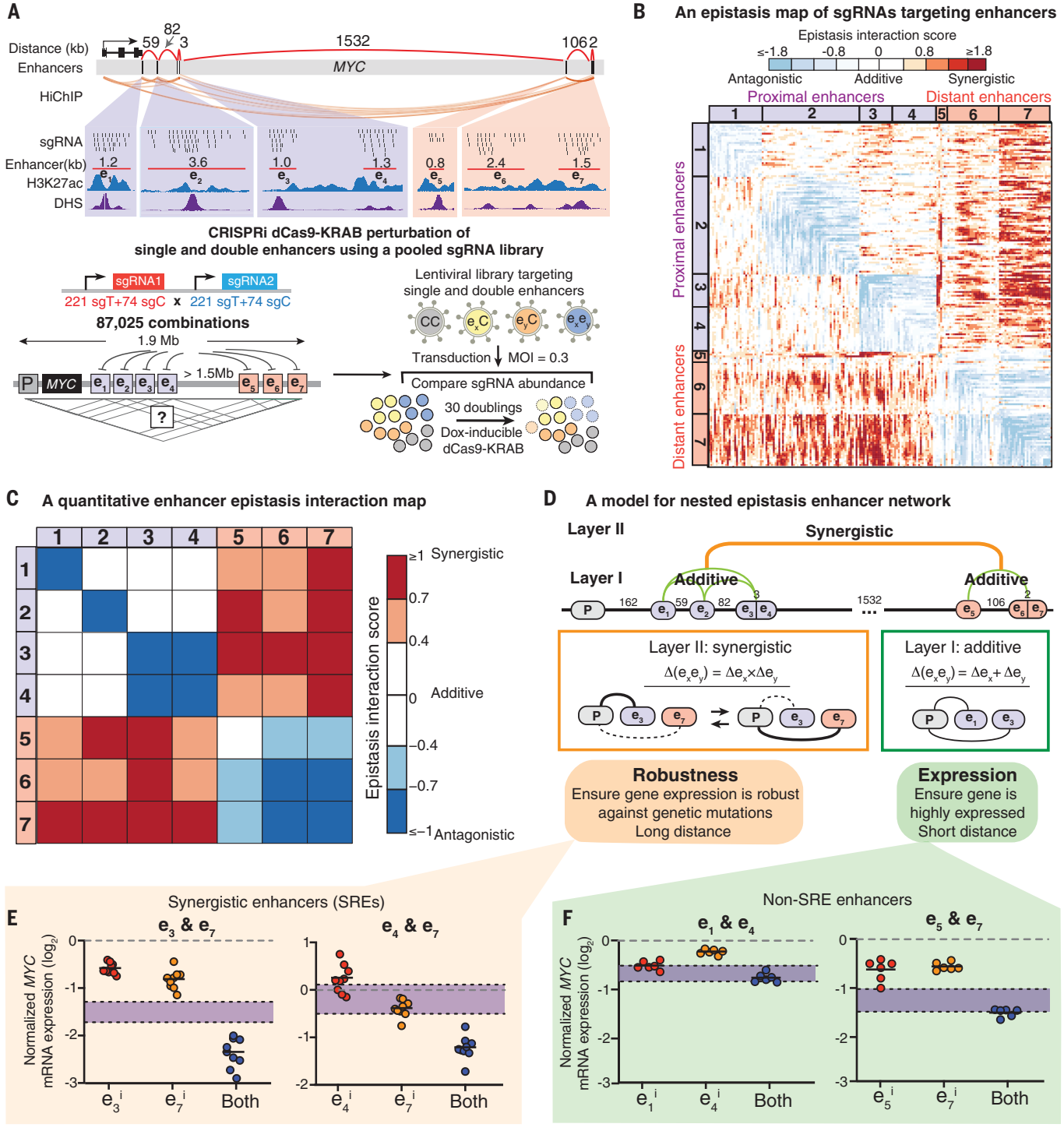


Fig. 1. High-resolution multiplexed CRISPRi perturbation of ultralong-distance enhancers at the *MYC* locus reveals a nested two-layer epistasis network. (**A**) (Top) The *MYC* locus regulated by multiple enhancers distributed over an ultralong distance (~1.9 Mb). (Bottom) Diagram showing the multiplexed CRISPRi screening for high-resolution dissection of enhancer interactions. K562 cells expressing the doxycycline (Dox)-inducible dCas9-KRAB are transduced by a pooled sgRNA library targeting single or double *MYC* enhancers. Cells are harvested to sequence the pairwise sgRNA enrichment before and after 30 doublings. sgT, targeting sgRNA; sgC, control sgRNA. (**B**) 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. (**C**) A quantitative enhancer epistasis map at the *MYC* locus. (**D**) A nested two-layer model for the enhancer epistasis network. (**E** and **F**) qRT-PCR of *MYC* mRNA expression for perturbing SREs e3 and e7 or e4 and e7 (E), or non-SREs e1 and e4 or e5 and e7 (F). P = 0.02, 1.13 × 10^{5}, 0.13, 0.61, for e3 and e7, e4 and e7, e1 and e4, e5 and e7, respectively. 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 t test.

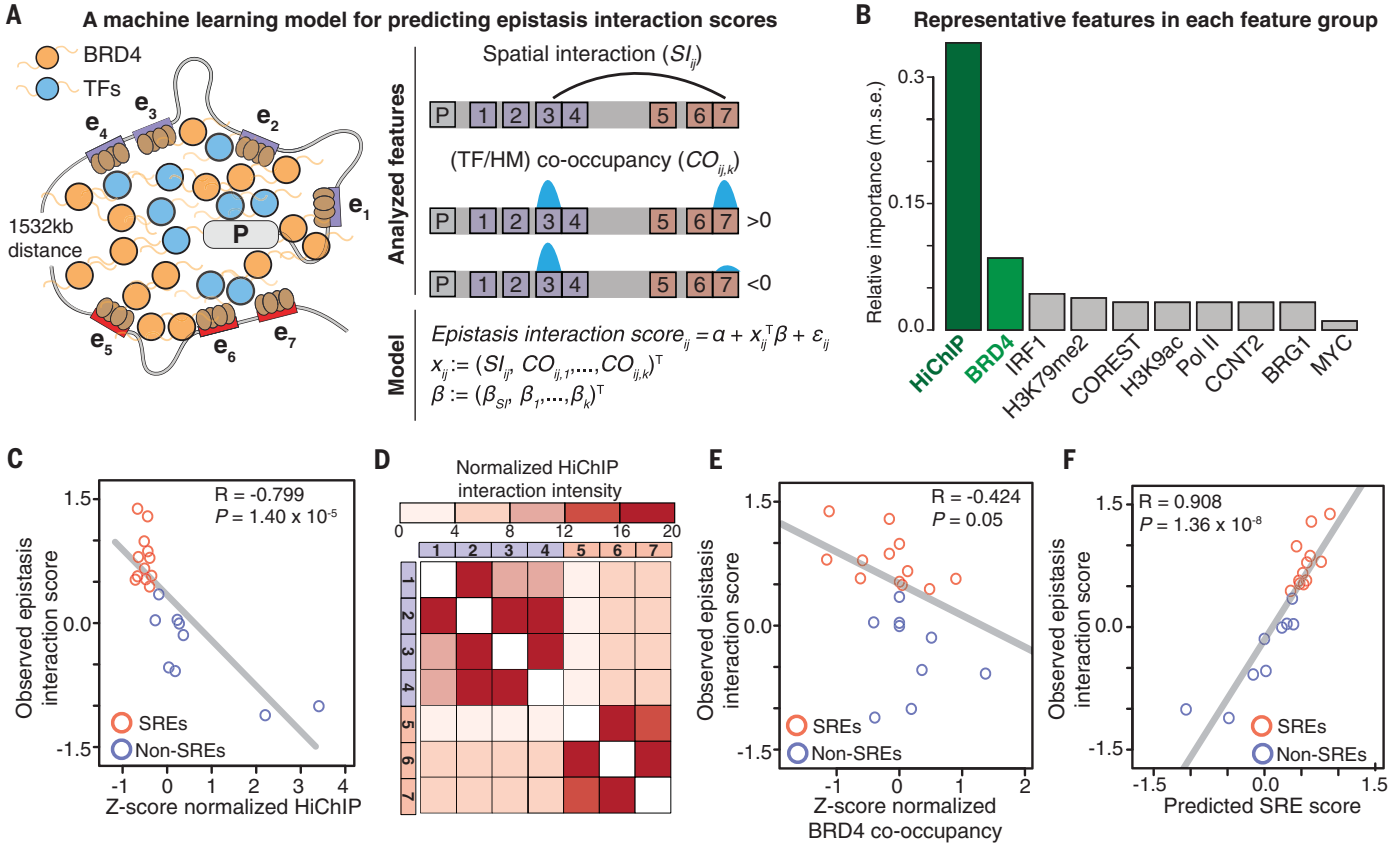


Fig. 2. A machine learning model for analyzing determinants of the SRE synergy. (**A**) An elastic net regularized linear regression model for predicting epistasis interaction scores. We selected features including the chromatin spatial interaction (SI_{ij}) and co-occupancy ($CO_{ij,k}$) of 38 TFs and 8 HM profiles. (**B**) The relative importance of each feature group for predicting epistasis interaction scores. The representative feature has the highest correlation in that group (fig. S8A). m.s.e., mean squared error. (**C** to **F**) Correlation between epistasis interaction scores and Z-scores normalized spatial contact (C) and BRD4 co-occupancy (E). (D) Heatmap of normalized HiChIP interaction intensity between enhancers. (F) Correlation between predicted SRE scores and observed epistasis interaction scores. In (C), (D), and (F), red, SREs; blue, non-SREs. The Pearson correlation coefficient (R) and *P* value are shown.

e7), we observed synergistically decreased *MYC* expression as well as cell proliferation (Fig. 1E, fig. S4, A to C, and table S1). In comparison, inhibiting enhancers within the same proximal or distant groups led to additive repression effects (Fig. 1F and table S1).

We performed H3K9me3 and H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) to characterize the resolution of using dCas9-KRAB for enhancer perturbation. We confirmed no spreading effects of KRAB on adjacent enhancers (fig. S5, A and B, and fig. S6, A and B). We also knocked out pairs of enhancers by transducing sgRNAs to K562 cells that stably expressed the nuclease Cas9 (see Methods). We confirmed consistent synergistic and additive interactions between e3 and e7 and e1 and e4, respectively (fig. S7, A and B). However, we also observed deletions of large chromatin regions when knocking out pairs of enhancers (fig. S7, C to F). This observation was consistent with reports that gene editing at multiple sites on the same DNA can induce megabase-scale chromosome deletions, which potentially confounds the study of enhancer interactions (*31*, *32*). These results together confirm that dCas9-KRAB is a high-resolution approach for studying multiple enhancer interactions without unwanted large DNA deletions.

Machine learning modeling reveals determinants of SRE synergy

We next developed a machine learning model based on an elastic-net regularized generalized linear model to analyze the determinants of SRE synergy (*33*) (Fig. 2A). We examined publicly available transcription factor (TF) binding profiles, histone modification (HM) profiles, and H3K27ac HiChIP datasets that capture DNA-DNA spatial contacts in K562 cells (table S3; see Methods) (*5*). Among all features spatial DNA contact is the most relevant feature and was inversely correlated with calculated epistasis interaction scores (Fig. 2, B and C, and fig. S8A). We found that the spatial contacts between SREs were weaker than non-SREs, which displayed an inverse pattern with the enhancer epistasis map (Fig. 2D versus Fig. 1C). In addition, the co-occupancy of bromodomain–containing protein 4 (BRD4), a key chromatin-associated coactivator, showed a strong anticorrelation with epistasis interaction scores (Fig. 2, B and E, and fig. S8, A and B).

The elastic net regression model performed better for predicting SREs compared with simple linear models using individual representative features (fig. S8, C and D). Predicted scores of all enhancer pairs were correlated with observed epistasis interaction scores assessed from the CRISPRi screen (Fig. 2F). Altogether, our machine learning model suggests that spatial DNA contacts and BRD4 co-occupancy are two major determinants for predicting SREs.

The SRE model can predict synergistic enhancer interactions at other genomic loci

We next verified whether the SRE prediction model can be generalized to study other genes that have multiple enhancers spanning an ultra-long distance in different cell types (fig. S9A;

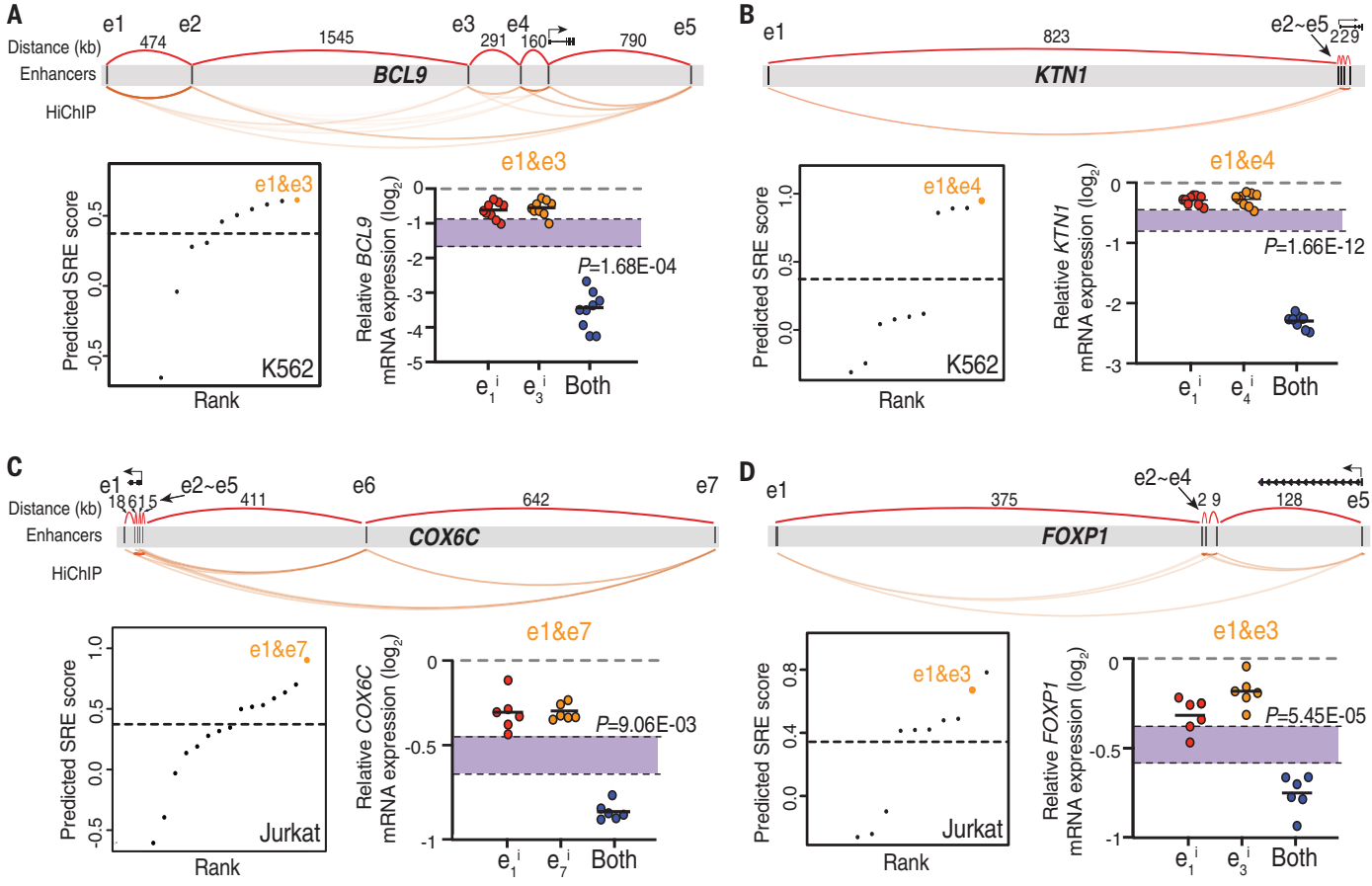


Fig. 3. Experimental validation of predicted SREs at other genomic loci in different cell types. (**A** to **D**) Prediction and validation of SREs at *BCL9* (A) and *KTN1* loci (B) in K562 cells, and *COX6C* (C) and *FOXP1* loci (D) in Jurkat cells. (Top) Diagram showing multiple enhancers spanning an ultralong distance at each genomic locus. (Bottom left) Rank of predicted SREs using the model. Dashed line represents the empirical threshold from the *MYC* locus. Orange dots indicate the validated SREs. (Bottom right) qRT-PCR of mRNA expression for each gene when perturbing the predicted SREs. 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 *t* test.

see Methods). We examined the enhancer profiles of four disease-relevant genes: *BCL9* and *KTN1* in K562 cells and *COX6C* and *FOXP1* in Jurkat cells, all of which have multiple enhancers spreading over a large genomic distance (3.3, 0.8, 1.1, and 0.5 Mb, respectively) (Fig. 3, A to D; see Methods). We used the SRE prediction model to calculate putative SREs and non-SREs and designed sgRNA pairs to target each SRE and non-SRE.

We observed synergistic changes in gene expression when targeting the predicted SRE pairs (Fig. 3, A to D and fig. S9, B to E), as well as additive effects when targeting the non-SRE pairs (fig. S9, B and C). These data suggested that our machine learning model can predict functional interaction between enhancers (SRE or non-SRE) that regulate different genes spanning an ultralong distance in different cell types. We further developed a website (http://enhancer.stanford.edu/) by exploring all 4835 putative networks of ultralong distance enhancers (≥5 enhancers; >200 kb interdistance) across six cell types (GM12878, K562, Jurkat, A549, HUVEC, and HCT116), which reports many predicted SREs and associated epistasis interaction scores.

Inhibition of SREs leads to synergistic reduction of local spatial contacts and BRD4 condensation

To experimentally examine the predicted determinants of the SRE model, we performed Trac-looping assays on CRISPRi-perturbed samples targeting an SRE pair e3 and e7 to measure both spatial contacts and chromatin accessibility (fig. S10A) (*34*). We observed that inhibition of individual enhancers decreased spatial contacts only between the targeted enhancer and other elements whereas simultaneous inhibition of e3 and e7 led to synergistic reduction of the spatial contacts at the *MYC* locus (Fig. 4, A and B), which is consistent with the observed epistatic effects on *MYC* expression and cell growth (Fig. 1E and fig. S4, A and B). By contrast simultaneous inhibition of a non-SRE pair e1 and e4 led to additive reduction of spatial contacts (fig. S10B). We also observed that inhibition of SREs showed no substantial difference from the additive effects on chromatin accessibility (fig. S10, C to E), suggesting that chromatin accessibility is less involved in synergistic interactions.

Perturbation of the distant enhancer e7 increased spatial contacts among the proximal enhancers and the promoter (e.g., e1-e3, e1-e4, e2-e3, e2-e4, e3-promoter, and e4-promoter) (Fig. 4, A and B). Similarly, perturbing e3 or e4 led to increased spatial contacts among the distant enhancers (Fig. 4, A and B, and fig. S10B). These observations imply a possible compensation mechanism on the spatial DNA contact between the SREs, which likely confers robustness of gene expression upon genome disruption (e.g., mutations or loss of DNA-TF interactions).

We next investigated the relationship between enhancer interactions and BRD4 localization. Clustered coactivator condensates mediated by BRD4 can assemble the transcription apparatus at enhancers to drive robust gene expression (*35*–*37*). Our machine

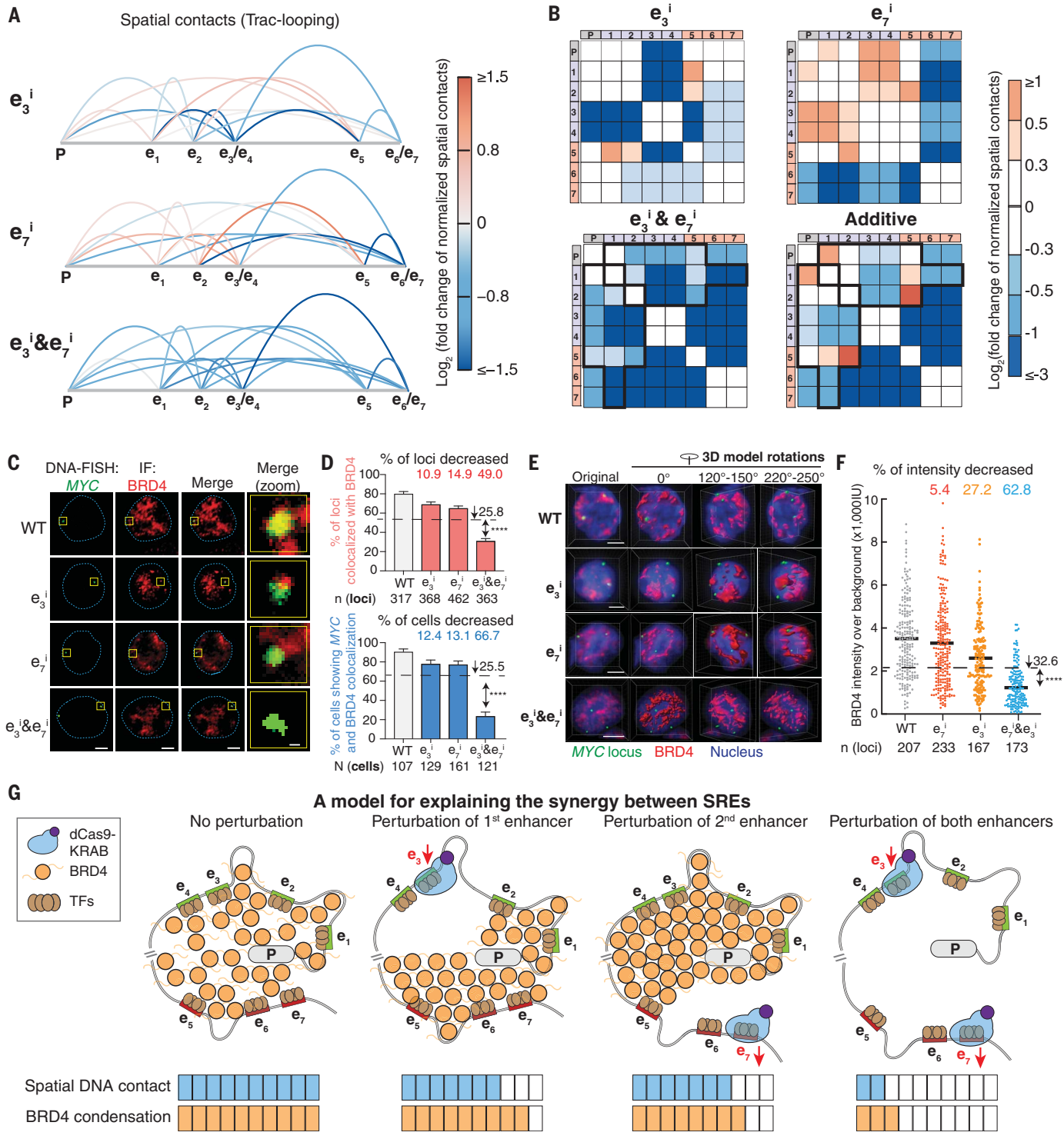
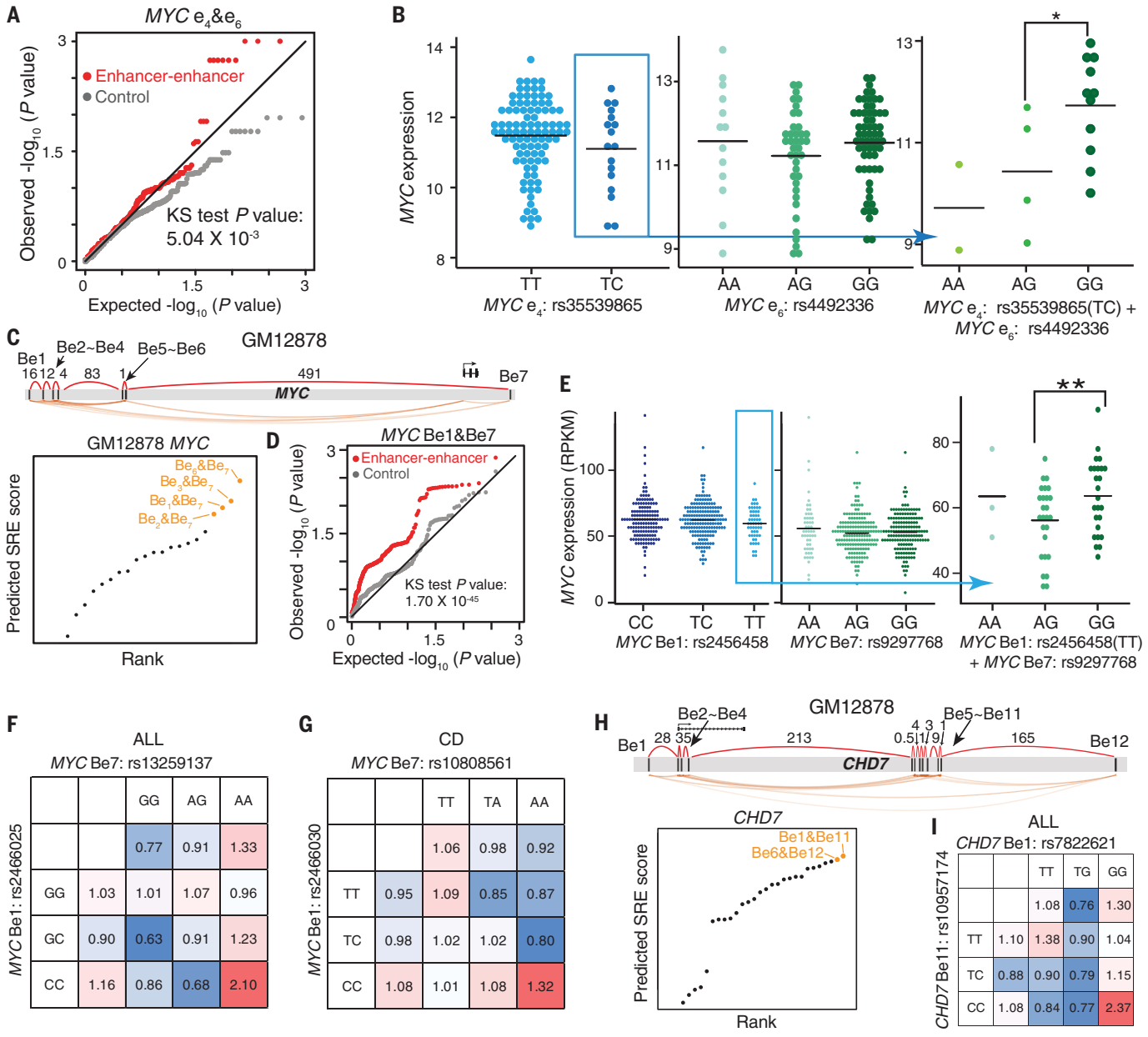


Fig. 4. Perturbation of SREs leads to synergistic reduction of spatial contacts and BRD4 condensation at the genomic locus. (**A** and **B**) Spatial contacts between the promoter and enhancers measured by Trac-looping for the *MYC* locus upon perturbation of e3, e7, and e3 and e7. Colors represent the $\log_2$ fold change of spatial contacts normalized to the wildtype cells. Black boxes in (B) indicate synergistically decreased (more than additive) spatial contacts of e3 and e7 pair perturbation. (**C** to **F**) DNA-FISH colocalization between BRD4 and the *MYC* locus of representative K562 cells for 2D (C) and (D) and 3D image analysis (E) and (F) upon perturbation of e3, e7, and e3 and e7. In (C) and (E), red, BRD4 immunofluorescence (IF) staining; green, DNA-FISH at the *MYC* locus; blue dashed line, nuclear periphery determined by DAPI staining (not shown); scale bars, 5 μm. The rightmost column in (C) shows insets in the yellow boxes. Scale bars, 500 nm. Quantification of BRD4 and the *MYC* locus colocalization are shown for 2D (D) and 3D image analysis (F). In (D), percentage of loci with colocalization is shown on the top and percentage of cells (≥2 colocalization loci) is shown on the bottom; data are represented as mean ± standard error of the mean. In (F), each dot represents an individual locus. *n* = total loci, *N* = total cells. ****$P < 0.0001$ in Fisher's exact test (D) or *t* test (F) versus the expected additive effect (dashed line). (**G**) A model to explain the synergy between SREs.

learning model predicted that the SREs were associated with distinct BRD4 clusters (Fig. 2E and fig. S11A). We examined this relationship by studying BRD4 colocalization at the *MYC* locus through immunostaining and fluorescence in situ hybridization (FISH) confocal imaging.

Compared with wildtype K562 cells, inhibiting individual enhancers (e3 or e7) resulted in a small reduction in colocalization between BRD4 and *MYC* loci, whereas simultaneous inhibition of e3 and e7 synergistically decreased colocalization (49.0%) and the percentage of cells showing colocalization (66.7%) (Fig. 4, C and D). Similar results were observed for



		GG	AG	AA
		0.77	0.91	1.33
GG	1.03	1.01	1.07	0.96
GC	0.90	0.63	0.91	1.23
CC	1.16	0.86	0.68	2.10

		TT	TA	AA
		1.06	0.98	0.92
TT	0.95	1.09	0.85	0.87
TC	0.98	1.02	1.02	0.80
CC	1.08	1.01	1.08	1.32

		TT	TG	GG
		1.08	0.76	1.30
TT	1.10	1.38	0.90	1.04
TC	0.88	0.90	0.79	1.15
CC	1.08	0.84	0.77	2.37

Fig. 5. Synergistic interactions between predicted SRE variants influence gene expression and disease risk in an epistatic manner. (**A** and **B**) Analysis of predicted SRE variants at the *MYC* locus in K562 cells for influence on gene expression. (A) quantile-quantile (QQ) plot showing the distribution of *P* values for the epistasis influence on *MYC* expression between e4 and e6 variants (red) in LAML patients, compared with random permutations (gray); *P* value in Kolmogorov–Smirnov (KS) test. (B) *MYC* expression in LAML patients stratified by e4 and e6 SRE variants. **P* < 0.05 in Wilcoxon test. (**C** to **G**) Analysis of predicted SRE variants at the *MYC* locus in GM12878 cells for influence on gene expression and associated disease risk. (C) Diagram showing the rank of predicted SREs; orange dots show top SREs. (D) QQ plot showing the distribution of *P* values for the epistasis influence of Be1 and Be7 variants (red) on *MYC* expression in the B lymphoblasts of 373 European individuals, compared with random permutations (gray). *P* value in KS test. (E) *MYC* expression in the B lymphoblasts from individuals stratified by Be1 and Be7 variants. ***P* < 0.01 in Wilcoxon test. (F) and (G) Calculated odds ratio on the relapse risk in acute lymphoblastic leukemia (ALL) (F) and Crohn's disease (CD) (G). Odds ratios are calculated by considering the genotypes of individual variants or both SRE variants. Colors represent the odds ratios. (**H** and **I**) Analysis of predicted SRE variants at the *CHD7* locus in GM12878 cells for influence on ALL. (H) Diagram showing the rank of predicted SREs; orange dots show top SREs. (I) Calculated odds ratio on the relapse risk in ALL. Odds ratios are calculated by considering the genotypes of individual variants or both SRE variants. Colors represent the odds ratios.

another SRE pair e4 and e7 (fig. S11, B and C). We also performed 3D FISH to better quantify the fluorescent intensity of the BRD4 condensate at the *MYC* locus. Whereas individual enhancer perturbation slightly decreased the BRD4 intensity (27.2% and 5.4% for e3 and e7, respectively), simultaneous perturbation led to synergistic BRD4 reduction (62.8%) (Fig. 4, E and F, and movies S1 to S4). By contrast, simultaneous inhibition of non-SRE e1 and e4 led to additive decrease of colocalization between BRD4 and *MYC* loci (fig. S11D). We further used a BRD4 inhibitor, JQ1, to investigate whether BRD4 condensation was involved in maintaining the synergistic interaction of SRE (*38*). Consistently, with increasing JQ1 concentrations the synergistic effects from SRE perturbation decreased and then disappeared, implying the importance of BRD4 condensation for enhancer synergy (fig. S11, E and F).

These results together confirmed that SRE perturbation synergistically reduced spatial DNA contact and BRD4 condensation at the target genomic locus, which led to synergistic changes in gene expression (Fig. 4, A to F, and Fig. 1E). Based on computational and experimental analysis, we propose a speculative model (Fig. 4G): while perturbing individual enhancers modestly reduces spatial contacts and BRD4 condensation, perturbation of two distant enhancers considerably alters the 3D chromosome organization and BRD4 condensation to confer synergistic regulatory roles.

Synergistic interactions between predicted SRE variants influence gene expression and disease risk

We evaluated whether SRE genetic variants spanning the ultralong distance can alter gene regulation and disease risks in an epistatic manner (fig. S12A). We examined the effect of our validated SREs within the *MYC* locus using an acute myeloid leukemia (LAML) patient database containing genomic and transcriptomic data. In LAML patients, we observed that e4 and e6 SRE variants interacted more frequently to alter *MYC* expression than that expected by chance, additive effects, and non-SRE variants (Fig. 5A and fig. S12, B to D; see Methods and Supplementary Text). A large difference in *MYC* expression levels was observed in two patient groups stratified by the genotype combinations of e4 and e6 SRE variants, whereas there were no dynamic changes when considering the genotypes of individual SRE variants (Fig. 5B).

We further examined the epistatic effect of *MYC* SRE variants on gene regulation in B lymphoblastic cells. We named the enhancers in GM12878 B lymphoblastoid cells as Be and used the SRE model to predict the interaction network among seven enhancers and rank SREs (Fig. 5C). We examined the interactions of variants across predicted SREs in a database of B lymphoblast genomic variants and transcriptomes (*39*). Although no difference in *MYC* expression was seen when looking at the genotypes of single enhancer variants a significant difference in *MYC* expression was observed when combinatorically considering the genotypes of SRE variants at Be1 and Be7 (Fig. 5, D and E, and fig. S12, E to G; Supplementary Text), or Be6 and Be7 (fig. S12, H and I).

Next, we applied the predicted SREs to investigate the association of *MYC* SRE variants in B cell–associated diseases, acute lymphoblastic leukemia (ALL), and Crohn's disease (CD) (*40*–*44*). In the top four predicted SRE pairs, we identified two SRE instances—Be1 and Be7 and Be2 and Be7—where the SRE variant pairs can synergistically influence the clinical risk, including ALL relapse risk and CD disease risk (fig. S13, A to E; Supplementary Text). Particularly, when we stratified case and control population based on SRE variants the odds ratio was significantly higher than that of the odds ratio determined by individual SRE variants alone or additively (Fig. 5, F and G, and fig. S13, F and G; Supplementary Text).

We also predicted SREs in other gene loci in GM12878 cells and observed the epistatic influence of SRE variants in gene expression and clinical risks, including the leukemogenesis-associated *CHD7* locus and B cell antigen *CD180* locus (*45*) (Fig. 5, H and I, and figs. S13G, S14, and S15; see Methods), which both have enhancer networks spreading ultralong genomic distance (0.4 Mb and 1.2 Mb, respectively).

The SRE model better identifies epistatic influence of genome-wide noncoding variants on disease risk

Finally, we applied the SRE prediction model to the genome-wide analysis in GM12878 cells to link multiple enhancer variants to disease risk. Among more than 900 genes containing ultralong distance enhancer networks, we focused on 70 immune- or cancer-related and highly expressed genes (fig. S16, A and B, and fig. S17; Supplementary Text). Notably, the predicted SRE scores correlated well with the epistatic effects of noncoding variants on the clinical risk for ALL relapse patients (Fig. 6A; Supplementary Text). Specifically, 27.9% of predicted SREs targeting 55.7% of genes showed epistatic effects on ALL relapse risk through our SRE model, which is significantly higher than the non-SRE pairs (Fig. 6, B and C). Furthermore, the SRE model also identified significantly more ALL-associated pathogenic

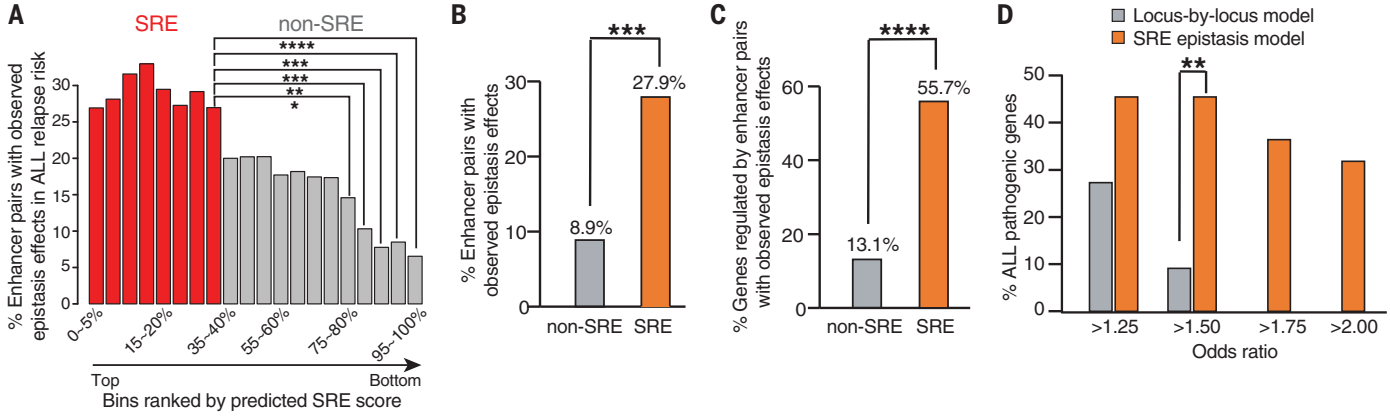


Fig. 6. Genome-wide analysis of epistatic influence of SRE variants on disease risk. (**A**) Percentage of enhancer pairs with observed epistatic effects on ALL relapse risk for predicted SREs and non-SREs. (**B** and **C**) Percentage of enhancer pairs (B) and genes (C) exhibiting interactive effects on ALL relapse risk. SRE pairs: enhancer pairs with top 40% SRE predicted score; non-SRE pairs: enhancer pairs with bottom 10% SRE predicted score. (**D**) Comparison of identified ALL pathogenic genes between the SRE model and the traditional locus-by-locus model at different odds ratio levels. In all figures, *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001 in Fisher's exact test.

genes compared with the traditional locus-by-locus model (Fig. 6D). For example, among 22 literature-reported ALL-associated pathogenic genes (table S4; Supplementary Text), our SRE model recovered 10 genes, whereas the locus-by-locus model showed only two genes (Fig. 6D; Supplementary Text). Therefore, the SRE prediction model can effectively elucidate the epistatic influence of multiple noncoding variants on associated clinical risk.

Discussion

Our work differs from previous studies on interactions (<100 kb) within enhancer clusters (e.g., super enhancers) (*19*–*25*, *46*). Although small-scale perturbations revealed additive (*21*, *22*, *46*) or synergistic (*23*, *25*) interactions within these enhancer clusters, it remains unknown whether enhancers distributed on a very large scale (>1 Mb) play interactive roles for gene regulation. Our results demonstrate that the observed nested synergistic interactions over the long distance and additive interactions in the short distance are important for an integrated function in the enhancer network; whereas the additive effects ensure a high expression level, the synergistic effects confer robustness against perturbations. Additional quantitative interaction mapping at more genomic loci in more cell types (e.g., diploid cells to rule out aneuploidy effects) should allow for the derivation of distance requirements for ultralong distance enhancer networks and a universal prediction model for enhancer networks. It should also help elucidate whether strong versus weak inhibition effects of individual enhancers determine whether they are SREs or non-SREs.

Our analysis showed that SREs are prevalent in the mammalian genome. The identification of SREs is consistent with evidence from studies in the 1000 Genomes Project, which showed that enhancer regions can be deleted without obvious phenotypic alterations (*47*, *48*). Theoretically, long-distance enhancers are less likely to be mutated at the same time, which avoids co-mutagenesis and thus provide compensation effects on important gene expression against mutations. Our website—which comprehensively explores genome-wide SREs—provides a resource to study enhancer interactions for gene regulation and multiple noncoding variants for diseases.

Because perturbations of individual enhancers may exhibit modest effects on gene expression, multiplexed perturbation of enhancers in the native chromatin context is crucial to fully elucidate their roles. We observed clusters of sgRNA pairs showing similar patterns of synergistic or additive interactions within an enhancer (fig. S3A), suggesting a high-resolution (~300 bp) subenhancer interaction mapping capability. We note that as a result of dCas9-KRAB spreading effects (500 bp~1 kb estimated by H3K9me3 peaks) (fig. S5, A and B), results from dCas9-KRAB should be validated with the Cas9 nuclease knockout for very close enhancers (<1 kb). Nevertheless, our analysis among 15 cell lines showed that >90% of enhancers have an inter-distance of >1 kb. By contrast, because the Cas9 nuclease may induce unwanted DNA deletions when perturbing multiple enhancers (fig. S7, C to F) (*31*, *32*), dCas9-KRAB offers technology for high-throughput study of enhancer interactions with high resolution and minimal side effects.

We provided a speculative model that links the 3D genome and BRD4 interaction to the ultralong distance enhancer network (Fig. 4G). In this model, large BRD4 condensates are formed by smaller distinct BRD4 clusters at individual enhancers (*49*), which connects these enhancers across ultralong distances to create weak 3D spatial contacts (*50*). This model is consistent with our quantitative mapping of enhancer networks that showed an inverse correlation between spatial contacts and synergistic interactions. Although the inverse correlation may be partly derived from the genomic distance, our experimental validation demonstrated that the 3D genome organization at SREs is casually linked to the synergistic interactions.

With more whole genome DNA sequencing data available in patients, the SRE model can be applied to infer the biological roles of SRE variants in cancer and other diseases and interpret the interactive influence of noncoding elements on disease risk to aid diagnosis and therapy.

REFERENCES AND NOTES

1. P. A. Northcott *et al*., *Nature* **511**, 428–434 (2014).
2. M. J. Fullwood *et al*., *Nature* **462**, 58–64 (2009).
3. X. Han *et al*., *Nat. Commun.* **9**, 2138 (2018).
4. X. Wang, D. B. Goldstein, *Am. J. Hum. Genet.* **106**, 215–233 (2020).
5. ENCODE Project Consortium, *Nature* **489**, 57–74 (2012).
6. Roadmap Epigenomics Consortium *et al*., *Nature* **518**, 317–330 (2015).
7. M. T. Maurano *et al*., *Science* **337**, 1190–1195 (2012).
8. M. Kellis *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6131–6138 (2014).
9. T. A. Manolio *et al*., *Nature* **461**, 747–753 (2009).
10. S. Chatterjee *et al*., *Cell* **167**, 355–368.e10 (2016).
11. O. Corradin *et al*., *Nat. Genet.* **48**, 1313–1320 (2016).
12. D. C. Factor *et al*., *Cell* **181**, 382–395.e21 (2020).
13. P. C. Phillips, *Nat. Rev. Genet.* **9**, 855–867 (2008).
14. W. A. Whyte *et al*., *Cell* **153**, 307–319 (2013).
15. J. Lovén *et al*., *Cell* **153**, 320–334 (2013).
16. D. Hnisz *et al*., *Cell* **155**, 934–947 (2013).
17. S. C. J. Parker *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17921–17926 (2013).
18. L. Pasquali *et al*., *Nat. Genet.* **46**, 136–143 (2014).
19. D. Hnisz *et al*., *Mol. Cell* **58**, 362–370 (2015).
20. J. Huang *et al*., *Dev. Cell* **36**, 9–23 (2016).
21. D. Hay *et al*., *Nat. Genet.* **48**, 895–903 (2016).
22. C. Bahr *et al*., *Nature* **553**, 515–520 (2018).
23. H. Y. Shin *et al*., *Nat. Genet.* **48**, 904–911 (2016).
24. J. Huang *et al*., *Nat. Commun.* **9**, 943 (2018).
25. M. W. Perry, A. N. Boettiger, M. Levine, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13570–13575 (2011).
26. C. P. Fulco *et al*., *Science* **354**, 769–773 (2016).
27. T. Wang *et al*., *Science* **350**, 1096–1101 (2015).
28. L. S. Qi *et al*., *Cell* **152**, 1173–1183 (2013).
29. P. I. Thakore *et al*., *Nat. Methods* **12**, 1143–1149 (2015).
30. L. A. Gilbert *et al*., *Cell* **159**, 647–661 (2014).
31. M. Kosicki, K. Tomberg, A. Bradley, *Nat. Biotechnol.* **36**, 765–771 (2018).
32. H. Y. Shin *et al*., *Nat. Commun.* **8**, 15464 (2017).
33. J. Friedman, T. Hastie, R. Tibshirani, *J. Stat. Softw.* **33**, 1–22 (2010).
34. B. Lai *et al*., *Nat. Methods* **15**, 741–747 (2018).
35. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, *Cell* **169**, 13–23 (2017).
36. W.-K. Cho *et al*., *Science* **361**, 412–415 (2018).
37. B. R. Sabari *et al*., *Science* **361**, eaar3958 (2018).
38. P. Filippakopoulos *et al*., *Nature* **468**, 1067–1073 (2010).
39. T. Lappalainen *et al*., *Nature* **501**, 506–511 (2013).
40. Wellcome Trust Case Control Consortium, *Nature* **447**, 661–678 (2007).
41. J. J. Yang *et al*., *Nat. Genet.* **43**, 237–241 (2011).
42. J. J. Yang *et al*., *Blood* **120**, 4197–4204 (2012).
43. J. Vijayakrishnan *et al*., *Leukemia* **31**, 573–579 (2017).
44. J. Vijayakrishnan *et al*., *Nat. Commun.* **10**, 5348 (2019).
45. T. Zhen *et al*., *Blood* **130**, 2431–2442 (2017).
46. D. D. Lam *et al*., *PLOS Genet.* **11**, e1004935 (2015).
47. D. Xu, O. Gokcumen, E. Khurana, *PLOS Genet.* **16**, e1008663 (2020).
48. M. Osterwalder *et al*., *Nature* **554**, 239–243 (2018).
49. K. Shrinivas *et al*., *Mol. Cell* **75**, 549–561.e7 (2019).
50. N. S. Benabdallah *et al*., *Mol. Cell* **76**, 473–484.e7 (2019).
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.

ACKNOWLEDGMENTS

The authors thank all members from the Lei Stanley Qi laboratory for useful comments and help on experiments and manuscript preparation, S. Shang and X. Chen for helping with the fluorescence-activated cell sorting, M. Han for helping with the imaging, and J. Magnusson for comments. We thank Y. Ye and L. Han from UTHealth Houston for helping with the LAML linear model. We thank W. Li and J. Zhang from the University of California, Irvine; Y. Ruan and M. Kim from the Jackson laboratory; and F. Ye from Northwestern University for helpful comments. We thank Z. Duren, S. Ma, and S. Wang for helpful discussion. We thank F. Wang and C. Yu for experimental assistance. We acknowledge the data generated by the TCGA Research Network (https://www.cancer.gov/tcga), which allowed us to generate the results of variant interaction. We acknowledge the CD GWAS dataset generated by the Wellcome Trust Case Control Consortium. The ALL Relapse GWAS dataset used for the analyses described in this manuscript were obtained from dbGaP at phs000638.v1.p1. The ALL Relapse GWAS dataset was generated at St. Jude Children's Research Hospital and by the Children's Oncology Group, supported by NIH grants CA142665, CA21765, CA158568, CA156449, CA36401, CA98543, CA114766, CA140729, and U01GM92666, Jeffrey Pride Foundation, the National Childhood Cancer Foundation, and by ALSAC. **Funding:** X.Z. acknowledges supports by the Stein Fellowship from Stanford University and Institute for Computational and Data Sciences Seed Grant from the Pennsylvania State University. W.H.W. acknowledges support from NIH R01 HG010359. L.S.Q. acknowledges support from the Li Ka Shing Foundation and National Science Foundation. The project is supported by the Li Ka Shing Foundation, the National Cancer Institute of the National Institutes of Health under award no. R01CA266470, and a National Science Foundation CAREER award (L.S.Q., award 2046650). L.S.Q. is a Chan Zuckerberg Biohub investigator. **Author contributions:** X.L., Y.L., and L.S.Q. conceived of the concept. Y.L., X.L., S.L., and L.S.Q. planned and designed the experiments. Y.L. and X.L. designed the sgRNA library. Y.L. and L.W. constructed the double sgRNA library. Y.L. performed the CRISPRi screens. D.Z. cloned 192 plasmids in the library and helped with deep sequencing. X.X. cloned 96 plasmids in the library and helped with deep sequencing. X.L. analyzed the CRISPRi screen data and built the SRE model. X.L. applied the model to predict SREs of other genes and designed sgRNAs. Y.L. generated sgRNAs and performed qPCR experiments. S.L. performed Trac-looping, ATAC-seq, and ChIP-seq. X.L. and Y.C. analyzed Trac-looping, ATAC-seq, and ChIP-seq data. Y.L. and Y.Z. performed imaging experiments and 2D image analysis. H.W. performed the 3D image analysis and generated supplementary movies. Y.L. performed the JQ1 experiment. X.Z. mentored X.L. on the SRE variant analysis. A.C. and X.L. developed the enhancer website. X.L., Y.L., and L.S.Q. wrote the manuscript. M.N., H.W., M.L.R., and J.N.N. provided critical comments on the manuscript. L.S.Q. initiated the project. W.H.W., K.Z., and L.S.Q. supervised the project. **Competing interests:** L.S.Q. is a founder and scientific advisor of Epicrispr Biotechnologies, and a scientific advisor of

Laboratory of Genomics Research. The roles are unrelated to this study. **Data and materials availability:** The CRISPRi functional tiling screen, Trac-looping data, ChIP-seq data, and ATAC-seq data have been deposited in the Gene Expression Omnibus under the accession ID GSE160768. The codes for the analysis of CRISPRi screen and the SRE prediction model are publicly accessible at Zenodo (*51*, *52*). The CRISPRi double sgRNA library and key plasmids will be available on Addgene (https://www.addgene.org/Stanley_Qi/).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk3512
Materials and Methods
Supplementary Text
Figs. S1 to S17
Tables S1 to S4
References (*53*–*72*)
MDAR Reproducibility Checklist
Movies S1 to S4

View/request a protocol for this paper from *Bio-protocol*.

Submitted 7 July 2021; resubmitted 24 March 2022
Accepted 28 July 2022
Published online 11 August 2022
10.1126/science.abk3512