4340–4354 *Nucleic Acids Research, 2022, Vol. 50, No. 8* *Published online 7 April 2022*
https://doi.org/10.1093/nar/gkac197

Regulatory elements can be essential for maintaining broad chromatin organization and cell viability

Ying Liu[1,†], Bo Ding[2,†], Lina Zheng[3,†], Ping Xu[1,†], Zhiheng Liu[1], Zhao Chen[2], Peiyao Wu[2], Ying Zhao[2], Qian Pan[1], Yu Guo[1], Wei Wang[2,3,4,*] and Wensheng Wei[1,*]

[1]Biomedical Pioneering Innovation Center, Beijing Advanced Innovation Center for Genomics, Peking-Tsinghua Center for Life Sciences, Peking University Genome Editing Research Center, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, China, [2]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093-0359, USA, [3]Bioinformatics and Systems Biology Program, University of California, San Diego, La Jolla, CA 92093-0359, USA and [4]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093-0359, USA

Received August 04, 2021; Revised March 03, 2022; Editorial Decision March 12, 2022; Accepted March 15, 2022

ABSTRACT

Increasing evidence shows that promoters and enhancers could be related to 3D chromatin structure, thus affecting cellular functions. Except for their roles in forming canonical chromatin loops, promoters and enhancers have not been well studied regarding the maintenance of broad chromatin organization. Here, we focused on the active promoters/enhancers predicted to form many 3D contacts with other active promoters/enhancers (referred to as hotspots) and identified dozens of loci essential for cell growth and survival through CRISPR screening. We found that the deletion of an essential hotspot could lead to changes in broad chromatin organization and the expression of distal genes. We showed that the essentiality of hotspots does not result from their association with individual genes that are essential for cell viability but rather from their association with multiple dysregulated non-essential genes to synergistically impact cell fitness.

INTRODUCTION

Promoters and enhancers are regulatory elements that control gene expression in response to intra- and extracellular signals (1–4). In many cases, activated enhancers appear to engage in direct physical contact with their nearby promoters (5–7). However, there are also enhancers whose interacting promoters are distally located in the linear genome (2,8–13), and they are brought to spatial proximity by such as chromatin looping (14–17), protein oligomerization (2,18,19) or Pol II tracking along chromatin (2,20). These observations on long-range enhancer-promoter interactions highlight the important impact of 3D chromatin structure on the activities of these regulatory elements.

Recently, an increasing number of studies on chromosome spatial structures have indicated that enhancer-promoter interactions play pivotal roles in forming specific 3D structures. Imaging analyses showed that transcription factors (TFs) and polymerases are not evenly distributed in the nucleus but rather concentrated in certain regions to form spatial clusters; these regions are associated with high transcriptional activities and a more compact chromatin structure (21–23). Transcription could also affect the 3D topology, and a recent study reported that transcription elongation can be critical for chromatin organization (24). These studies suggested a mutual relationship between promoter/enhancer activity and 3D chromatin structure. How these regulatory elements positioning in such spatial clusters with active transcription could contribute to maintaining broad chromatin structures therefore has become an emerging question.

Given these observations and evidence, we hypothesized that if a promoter or enhancer is positioned in 3D space pivotal for the maintenance or stabilization of the surrounding 3D chromatin structure, perturbing such elements may impact chromatin organization beyond their specific direct enhancer-promoter interaction; namely, perturbation of such promoter or enhancer would significantly alter broad chromatin organization and disrupt regulation of multiple direct and indirect target genes simultaneously.

To investigate our hypothesis, we started with active promoters/enhancers that likely form many 3D spatial contacts with other active promoters or enhancers, referred to

*To whom correspondence should be addressed. Email: wswei@pku.edu.cn
Correspondence may also be addressed to Wei Wang. Email: wei-wang@ucsd.edu
†These authors contributed equally to this work.
Present address: Bo Ding, Thermo Fisher Scientific 5781 Van Allen Way, Carlsbad, California 92008, USA.

as hotspots hereinafter. Using our previously published algorithm EpiTensor (25), we identified hotspots at a high resolution of 200-bp based on their covariation of epigenetic marks across cell types. Interestingly, cancer-specific genetic variations (we focused on single nucleotide variations) were discovered to have a significantly higher chance of residing in hotspot regions. Through high-throughput CRISPR-Cas9 library screening of hotspots by targeted deletion, dozens of noncoding loci were identified as essential for cell growth and survival, referred to as essential hotspots. We then evaluated the impact of the 3D chromatin structure by Hi-C technology and gene expression using single cell RNA-seq upon knocking out hotspots. Importantly, we found that deleting a hotspot enhancer could alter broad chromatin organization beyond chromatin looping, which has not been reported before. Deletion of the hotspot would further impact the expression levels of multiple individually non-essential genes concurrently, which exhibited synergistic effects to affect cell fitness.

MATERIALS AND METHODS

Predicting high-resolution regulatory element contacts by EpiTensor

Active promoters marked by H3K27ac and H3K4me3 and active enhancers marked by H3K27ac and H3K4me1 were identified in 73 normal and 5 cancer cells/tissues that were available in the NIH Roadmap Epigenomics project (26). The 3D contacts between these active promoters/enhancers in each cell/tissue were predicted by EpiTensor (25) with an EpiTensor score cut-off $\geq$ sqrt (25000). In a given cell/tissue sample, these contacts were assembled into a regulatory element interaction network (REIN) in which each node represents a promoter/enhancer and an edge represents a predicted contact.

Identification of sample-specific degree/sample-specific genetic variations (GV)

A distribution-based method was used to evaluate the cell type specificities for degree: 1) For each node, we collected the normalized degree in all samples that had epigenomic data (73 normal and 5 cancer in total); then we calculated the mean and standard deviation for each node across the normal samples, under the assumption that the normalized degrees of normal samples obey a Gaussian distribution; finally, the Z-score for each node in each sample, i.e. the sample-specific degree Z-score, was calculated using the mean and standard deviation. A node was considered to have sample-specific degree if the absolute value of the sample-specific degree Z-score was greater than 1.

We identified sample-specific GVs with a similar method: for each GV, we first calculated mean and standard deviation of B-allele frequency in all normal samples (45 in total); then, Z-score for each GV in each sample, i.e. the sample-specific allele frequency Z-score, was calculated using the mean and standard deviation. A GV was considered as sample-specific if the absolute value of the sample-specific allele frequency Z-score was greater than 1.

Cell culture

K562, H1975 and NAMALWA cells were cultured in RPMI 1640 medium (Gibco), and 293T, HeLa, A549 and Huh7.5.1 cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco). All media were supplemented with 10% fetal bovine serum (FBS, Biological Industries) and 1% penicillin/streptomycin, and cells were cultured with 5% CO_2 at 37°C.

Design and construction of the CRISPR-Cas9 paired gRNA (pgRNA) library

To explore the cellular function of hotspots, we selected 751 hotspots identified in the K562 cell line. For each hotspot, the designed sgRNAs targeted 100-bp inside regions and 1-kb outside regions flanking the two boundaries of hotspot loci. If there were not enough sgRNAs satisfying the following design rules, sgRNAs were searched among the 5-kb outside regions flanking each boundary. All the PAM motifs in the targeting regions were scanned to identify available sgRNA targeting sites. All the selected sgRNAs are located in noncoding regions and satisfy all the following conditions: (1) the targeting sequence is unique for the intended locus; (2) the sgRNA contains at least 2 mismatches to any other locus in the human genome; and (3) the GC content of the sgRNA ranges from 20% to 80%. We enumerated all possible pgRNAs from the selected sgRNAs and then retained those satisfying these conditions: (1) the two sgRNAs respectively targeted 100-bp inside regions and 1-kb (or 5-kb) outside regions flanking each hotspot boundary; (2) the deletion regions should not overlap with any promoter or exonic region of protein-coding genes; and (3) the sgRNA targeting sites are at least 30 bp away from the exon-intron boundary of protein-coding genes. The gRNA pairs were designed with one unique gRNA serving as a decoding barcode, and up to 20 pgRNAs were designed for each locus.

Finally, 14,399 pairs of gRNAs targeting 751 hotspots were generated for the hotspot deletion library together with 473 pgRNAs targeting the promoter regions (5 kb upstream of the transcription start site) and the first exon of 29 ribosomal genes (serving as positive controls) and 100 pgRNAs targeting the *AAVS1* locus and 100 non-targeting pgRNAs from a previous library (27) (serving as negative controls). According to the two-step cloning method (27), 128-nt oligonucleotides containing pgRNA coding sequences were synthesized (Agilent Technologies, Inc.), cloned into a lentiviral expression vector harbouring an EGFP selection marker (with a minimum representation of 150 transformed colonies per pgRNA in each cloning step) and further packaged as previously described (27).

CRISPR-Cas9 pgRNA library screening

To ensure the infection at 1,000–1,500 cells per pgRNA with an MOI of < 0.3, K562 cells stably expressing Cas9 were seeded in duplicate in T-175 flasks (Corning). Twenty-four hours later, each replicate was infected by the pgRNA library lentiviruses supplemented with 8 μg/ml polybrene. Seventy-two hours post infection, $EGFP^+$ cells were sorted

by FACS (Day 0 control group). For each replicate, the initial EGFP$^+$ pool (1500-fold coverage) was isolated for DNA extraction, and the same number of cells as the experimental group was maintained at a minimum coverage of 1,500 cells per pgRNA at each passage for 30 days. Then, cells from each condition with 1500x library coverage were respectively subjected to genomic DNA extraction, PCR amplification of sgRNA-coding sequences and high-throughput sequencing analysis (Illumina HiSeq2500 platform) as previously described (27).

Identification of functional hotspots involved in cell growth and proliferation

The raw pgRNA counts were extracted from paired-end sequencing FASTQ files by bash script based on AWK. Since the low reads in the control groups affect the analysis confidence, pgRNAs with raw reads of less than 5 were eliminated from the following analysis. The total counts were further normalized to adjust the sequence depth of each replicate in the control and experimental groups. To further filter noisy pgRNAs, we removed pgRNAs whose quantile difference of two replicates was in either 3% tail of the distribution, and 100 negative control genes were generated by randomly sampling 20 *AAVS1*-targeting pgRNAs with replacement. In each replicate, we calculated the fold change between the experimental and control group for each pgRNA, and the mean fold change of all targeting pgRNAs for each hotspot. Then, the fold changes in the two replicates were averaged for each specific hotspot. In summary, two features for each set of hotspots were calculated: 1) the mean $\log_2$FC ($\log_2$(fold change)) of all pgRNAs in the set, denoted by $\mathrm{FC}_{\mathrm{hotspot}}$; and 2) the $-\log_{10}P_{\mathrm{value}}$ of two-sided Mann-Whitney U test of all pgRNAs in the set compared with pgRNAs targeting the *AAVS1* locus, denoted by $P_{hotspot}$. To consider both the fold change and P value, we defined a screen score for the hotspots as follows:

$$\text{Screen score} = \text{sign}(\text{LFC})\left|\left|\frac{\text{LFC}-\mu_{\text{LFC}}}{\sigma_{\text{LFC}}}\right|+\text{LP}\right|$$

where LFC is the $\log_2$FC, $\mu_{LFC}$is the mean of the LFC, $\sigma_{LFC}$is the standard deviation of the LFC, and LP is the $-\log_{10}P_{\mathrm{value}}$. Hotspots with screen scores of less than -2.5 were identified as essential hotspots.

To further avoid the potential issue of cellular toxicity generated from multiple cleavages by some pgRNAs, we retrieved the GuideScan specificity score (a score reflecting the sgRNA cutting specificity) to evaluate each sgRNA (28). A specificity score was further assigned for each pgRNA, which was calculated as half of the harmonic mean of the specificity scores of the two sgRNAs. The formula is as follows:

$$\mathrm{pgRNA}_{\text{specific score}} = \frac{1}{\frac{1}{\mathrm{sgRNA1}_{\text{specific score}}}+\frac{1}{\mathrm{sgRNA2}_{\text{specific score}}}}.$$

From the identified essential hotspots through the above analysis, those targeting pgRNAs were further selected, whose specificity score is > 0.1 and $\log_2$(fold change) is < -1. To further avoid the copy number effects on dropout screening, the copy number of each hotspot locus in the K562 cell line was analysed based on ENCODE consortium copy number data (https://www.encodeproject.org/files/ENCFF486MJU/). After filtering hotspot loci with copy number amplification, the remaining hits were regarded as essential hotspots.

Individual validation of functional hotspots by cell proliferation assay

For each candidate hotspot without immediate overlap with the promoter or gene body of protein-coding genes, two or three pgRNAs were used for the individual validation, which were selected from the library that were consistently depleted or newly designed. To ensure the targeting specificity of all the selected pgRNAs, we required that the specificity scores are all greater than 0.15 and that the score of at least one pgRNA for each hotspot is more than 0.2. For the newly designed pgRNA, to ensure the cleavage efficiency, we further required that they don't include $\geq$ 4-bp homopolymer stretches, and their GC contents are between 0.4 and 0.7. We further ensured that each sgRNA targeting site is 400 bp inside and 1 kb outside the two boundaries of the hotspot loci. All the pgRNAs targeting each hotspot locus to be validated were individually cloned into a lentiviral expression vector containing an EGFP selection marker. The cell proliferation assay was performed as previously described (27). The experiments lasted for 15 days after the first FACS analysis, and at least 10,000 cells were analysed.

For the hotspots overlapping with the promoter or within the intron of possible essential protein-coding genes, three pgRNAs were selected for subsequent validation. The cDNA of each neighbouring coding gene was cloned into a lentiviral vector containing a puromycin selection marker and individually transduced into K562 cells. Three days after virus infection, the cells with candidate gene overexpression were enriched by puromycin treatment, and then the corresponding pgRNAs targeting the neighbouring hotspot were respectively transduced into these cells as well as into wild-type K562 control cells. The cell proliferation assay was performed as described above.

Hi-C library preparation and data analysis

Hi-C library preparation. The pgRNA Hotspot_10_25-pg2 was delivered into K562 cells via lentiviral infection at an MOI of < 1. EGFP-positive cells were then collected by FACS sorting at day 9 post infection. Before the Hi-C library preparation, the sorted cells were allowed to recover under normal cell culture conditions for 2 h. Finally, one million cells were used for Hi-C library preparation by the Arima-HiC kit (Arima Genomics, San Diego) following the manufacturer's instructions. The K562 hotspot_10_25 Hi-C library was sequenced using the Illumina NovaSeq platform.

Hi-C data processing. An in-house pipeline Juicer (29) was implemented to process the Hi-C data. Hi-C contact reads were first aligned to hg19 (GRCh37), and the reads were reserved if MAPQ greater than 30. Then, the vanilla coverage (VC) method (14) was applied to the Hi-C raw reads. Between the expected VC-normalized reads and the observed

VC-normalized reads, we conducted a Poisson distribution fitting. The normalized contacts were considered significant if the *P*-value is ≤ 0.05. HiCExplorer (30,31) and HiCPlotter (32) were utilized to visualize the processed Hi-C data.

Chromatin loops identification. The HiCCUPS software (https://github.com/aidenlab/juicer/wiki/HiCCUPS) was utilized to call the loops at 10-kb resolution in both the wild-type and hotspot-deleted K562 cells. All the other parameters in HiCCUPS were set to default.

Topological associated domain (TAD) identification. The Insulation Score method was used to call the TAD for the wild-type and hotspot_10_25-deleted cells at 10 kb resolution. TADs were visualized using HiCExplorer.

A/B compartment analysis. We performed the A/B compartment analysis on the wild-type and hotspot_10_25 deleted cells at 50-kb resolution. The eigenvectors for each individual chromosome were extracted from VC-normalized Hi-C reads using the Juicer pipeline (29). All the parameters were set to default. To determine the direction of A or B compartments in each chromosome, the K562 Pol II peak file was obtained from ENCODE (https://www.encodeproject.org/). A correlation score between the first eigenvector of each chromosome and the K562 Pol II peak density in 50 kb-sized bins was calculated.

Hi-C comparison. We used HiCRep (33,34) to calculate the Stratum-adjusted correlation coefficient (SCC) to measure the Hi-C reproducibility. We performed HiCcompare R bioconductor package (35) to detect the Hi-C contact differences across all the chromosomes. All the analyses were done at 25-kb resolution.

Evaluation of the potential off-target effects by the CRISPR-Cas9 system through whole genome sequencing (WGS)

K562 cells were infected with the validated pgRNA hotspot_10_25-pg2 at an MOI of < 1. Eight days after lentiviral infection, the pgRNA-infected cells were sorted by FACS, and were further subjected to genomic DNA extraction. The whole genome sequencing (WGS) library was prepared following the manufacturer's instructions and sequenced using the Illumina HiSeq 4000 platform. Using the WGS data, we evaluated the potential off-target effects after targeted deletion of hotspot_10_25.

The K562 (wild-type) WGS data were downloaded as controls from ENCODE with accession code ENCFF313MGL, ENCFF004THU, ENCFF506TKC and ENCFF066GQD. A strict off-target evaluation was conducted according to the whole-genome sequencing approach (36,37). The putative off-target sites for hotspot_10_25 were output by Cas-OFFinder in the hg19 genome (36). To avoid missing any potential off-target locus, we considered two scenarios to detect the potential off-target loci: 1) no more than 4 base mismatches without any bulge mismatch (mismatch ≤ 4, bulge $= 0$) and 2) no more than 2 base mismatches with no more than 2 bulge mismatches 2 (mismatch ≤ 2, bulge ≤ 2). In total, we examined 746 potential off-target loci. In order to detect the candidate mutations and indels in the wild-type and hotspot_10_25-deleted K562 cells, we performed variant call according to the approaches described in GATK Best Practices (https://gatk.broadinstitute.org/hc/en-us). The sequencing reads were firstly aligned to the human reference genome (hg19) using BWA-0.7.17. Then we used the GATK4 tools MarkDuplicatesSpark (https://gatk.broadinstitute.org/hc/en-us/articles/360037224932-MarkDuplicatesSpark) to remove the duplicated reads. Finally, the reads were processed via base quality score recalibration with the GATK4 tools. Germline mutations (compared to the hg19 reference genome) were called in both wild-type and hotspot_10_25 deleted K562 cells by GTAK HaplotypeCaller (version 4.1.4.1) with default parameters. SNVs and indels in pgRNA-infected K562 cells compared to wild type K562 cells were identified via the tools GATK Mutect2 (version 4.1.4.1) with default parameters. These SNVs and indels were further compared with generated putative off-target loci.

For further confirmation, we applied the BCFTOOLS suite (version 1.9, http://www.htslib.org/doc/bcftools.html) to call variants. BCFTOOLS mpileup and call commands with default settings were used to generate raw variants. Then, variants with '%QUAL $<$ 30 || DP $<$ 30' were marked as low-quality variants by the BCFTOOLS filter command and filtered out in addition to the homozygous variants with the feature 'GT = 1/1'. We also used the BCFTOOLS isec command with parameter '-n -1 -c all' to filter the Mills and 1000G gold standard indels obtained from the GATK resource bundle (https://gatk.broadinstitute.org/hc/en-us/articles/360035890811-Resource-bundle). The putative off-target sites generated by Cas-OFFinder were checked with the variants called by the above BCFTOOLS pipelines, and no overlaps were found.

Bulk RNA-seq and data analysis

Bulk RNA-seq data processing. We downloaded the K562 bulk RNA-seq data with pgRNA targeting *AAVS1*, which was generated by our previously published research (38) (GEO accession code GSE176503). In the bulk RNA-seq library, the sequencing reads were aligned to the human reference genome (GRCh37/hg19) using HISAT2 (2.0.4) (39–41) and assembled and quantified by StringTie (1.3.5) (39,42).

Single-cell RNA-seq and data analysis

Single-cell library preparation. K562 cells were infected with the validated pgRNA hotspot_10_25-pg2. Eight days after lentiviral infection, the pgRNA-infected cells were subjected to FACS for single-cell library preparation. The single-cell library was prepared according to a previously established Drop-seq protocol (43). PolyA + RNA was reverse transcribed through tailed oligo-dT priming directly in whole-cell lysates (single droplet) using Moloney Murine Leukaemia Virus Reverse Transcriptase (MMLV RT) and temperature switch oligos. The resulting full-length cDNA contains the complete 5′ end of the mRNA as well as an anchor sequence that serves as a universal priming site for second strand synthesis. The cDNA was pre-amplified using 15 cycles with Kapa HiFi Hotstart Readymix and then

tagmented at 55°C for 5 min in a 20 μl reaction following the Illumina Nextera DNA preparation kit. 5 microliters of neutralization buffer was added to the tagmentation reaction mix to quench the reaction. The tagmented DNA was amplified by 12 cycles of standard Nextera PCR. The DNA was then purified with Ampure beads (sample to beads ratio of 1:0.6). The prepared hotspot_10_25-deleted single cell library of K562 was sequenced on an Illumina HiSeq 4000 instrument.

Single cell RNA-seq processing. The single cell RNA-seq data were processed using the Drop-seq pipeline developed by the McCarroll lab (43). Low-quality reads (lower than Q10) and PCR duplicates were removed. Cells were ranked in descending order by the total number of read counts. Cells ranked before the inflection point of the cumulative distribution were selected for the following analysis. Each cell was first normalized by counts per million (CPM). The value E_i was computed as the sum of CPMs for a given gene across all the cells. E_{total} was calculated as the sum of E_i for all the genes. Then, a P_i was computed as $P_i = E_i / E_{total}$. In a given cell j, the normalized gene expression of all genes was assumed to independently and identically follow the binomial distribution $G_{ij} \sim B(N_j, P_i)$, where G_{ij} is the expected read of gene i in cell j and N_j is the total read for cell j. A P-value was computed to evaluate how each gene expression in each cell significantly deviated from the expectation based on the binomial distribution. We also calculated P-values for genes in the negative control ($\Delta AAVS1$) and wild-type bulk RNA-seq data in the same way.

Single-cell trajectory branching and pseudotime analysis. Because hotspot deletion severely hampered cell proliferation, we focused on analysing the apoptosis-related genes annotated in the KEGG database (44). The 99 apoptosis-related genes that showed differential expression upon deleting hotspot_10_25 (chr10: 74,123,469–74,124,868) in at least 10% ∼ 15% of cells (P-value < 0.05) were selected. All the normalized single cells and bulk data were clustered with trajectory branching and pseudotime analysis using Monocle (45,46). Monocle assigned a specific pseudotime value and a 'state' to each cell. Cells with the same 'state' and similar pseudotime were clustered together (46), and then the relative gene expression in each cluster was computed.

Differentially expressed genes identified from pseudotime analysis. To identify differentially expressed genes (DEGs) pairwise between different cell states, a *Wilcoxon* Rank-Sum Test (47) was used to identify genes that showed significantly up- and downregulated in the cell state pair.

Validation of synthetic lethal pairs by cell proliferation assay

Selection of the targeting sgRNA for each gene. To explore the synthetic lethal pairs among the four significantly downregulated genes located within the same TAD of hotspot_10_25 after hotspot deletion, we first determined the targeting sgRNA to ensure efficient knockdown of each gene. Three sgRNAs were selected to target the promoter region of each gene from the hCRISPRi-v2 library (48), and a non-targeting sgRNA was set as a control. These sgRNAs were further cloned into the lentiviral expression vector with an EGFP selection marker and then transduced into K562 cells stably expressing KRAB-dCas9 protein through lentiviral infection. Three days after infection, the EGFP-positive cells were sorted by FACS, and the total RNA of each sample was extracted using a RNeasy Mini Kit (QIAGEN 79254). cDNA was synthesized from 2 μg of total RNA using the Quantscript RT Kit (TIANGEN KR103-04), and real-time qPCR was performed with TB Green™ Premix Ex Taq™ II (Tli RNaseH Plus, TAKARA) to detect the expression of each indicated gene as well as of the reference gene *GAPDH*. The sgRNAs showing the most significant knockdown effect were selected for subsequent experiments to evaluate the synergistic effect. All the primers used in real-time qPCR are listed in Supplementary Table S4.

Evaluation of the growth effect of each individual gene and gene pair in K562 cells. The above four selected sgRNAs were grouped into six gRNA pairs targeting six gene pairs. The four sgRNAs and six pgRNAs were respectively cloned into the lentiviral expression vector with an EGFP selection marker and then transduced into K562 cells stably expressing KRAB-dCas9 protein at an MOI of < 1. The cell proliferation assay was performed as described above. The first time point of FACS analysis was at 6 days after lentiviral infection, and the experiment lasted for another 12 days.

RESULTS

Small-world network formed by 3D contacts between promoters and enhancers

To identify regulatory elements (promoters or enhancers) that are likely to be important for chromatin organization, we chose to start with those involved in many interactions with other loci in the genome. We first identified active promoters (marked by H3K27ac and H3K4me3) and enhancers (marked by H3K27ac and H3K4me1) in 73 normal and 5 cancer cell lines/tissues with all 3 marks using data from the Roadmap Epigenomics project (26). The 3D contacts between these active promoters/enhancers in each cell line/tissue predicted by EpiTensor (25) were assembled into a regulatory element interaction network (referred to as REIN hereinafter), in which nodes are promoters/enhancers and edges represent 3D contacts (see Materials and Methods). We resorted to computational prediction by EpiTensor (25) because Hi-C data with sufficient resolution to define the interactions between promoters and enhancers were rare. We have previously shown that chromatin contacts could be successfully predicted by EpiTensor (25), which detects epigenetic covariation patterns between promoter-enhancer, promoter-promoter and enhancer-enhancer pairs at 200-bp resolution via tensor analysis. Such a covariation indicates that possible 3D contacts can be formed between active regulatory elements in a cell type-specific manner. Therefore, when considering spatial contacts in a particular cell type or tissue, we only considered those formed between active promoters and/or enhancers, as marked by open chromatin or H3K27ac, because these contacts are likely to establish functional regulation. EpiTensor predictions were shown to be highly con-

cordant with the Hi-C, ChIA-PET and eQTL results in different cell types (25).

In REIN, each node represents a promoter/enhancer in the given cell line/tissue, and each edge represents a contact predicted by EpiTensor. The degree of a given node reflects its total contacts. We examined the topological properties of REIN using SNAP software (49). Through computational simulations, the cluster coefficients of REIN (the percentage of node pairs that connected when they were connected to another node) were found to be similar to an equivalent (same number of nodes and edges) regular lattice network (50), and their path length (the largest required number of steps between node pairs) was similar to that of the equivalent random network (51) (Figure 1A). These properties showed that the REINs are small-world networks. Small-world networks are characterized by robustness in that they are resistant to random attacks (random removal of nodes) but vulnerable to targeted attacks (removal of specific nodes) on high-degree nodes that have significantly more contacts than the other nodes (52). We selected the top 10% high-degree nodes in REIN as 'hotspots' for further analysis.

Mutations in hotspot enhancers and promoters could alter 3D contacts

We collected all the genomic loci identified as hotspots in at least one cell line/tissue. In total, we found 48,110 regions, the majority of which are enhancers, and 12,754 of them overlap with promoter regions (1 kb around the transcription start sites). Consistent with our previous analysis (25), these loci tend to be active (overlapping with H3K27ac signals) in more cell types than the non-hotspot loci (Figure 1B). We noticed that the number of interactions a hotspot forms varies significantly across cell types/tissues, and on average, a locus was identified as a hotspot only in 7 out of 78 cell types/tissues. Particularly, promoter hotspots are shared by more cell types/tissues (on average 17 out of 78) than enhancer hotspots (on average 4 out of 78), which is not unexpected as enhancers are known to be cell type/tissue specific.

Given the importance of high-degree nodes in a small-world network, mutations in hotspot loci may have a severe impact on the network structure. To investigate this possibility, we analysed the loci that are active in all examined cell lines but show a significant change in degree. We first identified nodes with sample-specific degrees: using the degree numbers of each node in all 73 normal samples as the background distribution, we identified nodes that are active in a specific sample and whose degree also significantly deviates from the mean. We then determined the sample-specific genetic variations (GVs). We collected 1,197,917 GVs in 45 normal and 17 cancer samples (DCC accession number ENCFF105JRY). For each GV in each sample, if its B-allele frequency significantly deviates from the mean in the 45 normal samples, we considered this GV specific to the sample (see Materials and Methods).

The nodes with a sample-specific degree containing at least one sample-specific GV, which are called degree-GV correlated nodes (Figure 1C), are suitable candidates to investigate the relationship between GV and degree. We first analysed 4 normal (GM12878, H1, HEK293 and IMR90) and 4 cancer cell lines (HeLaS3, HepG2, K562 and MCF-7) and found that degree-GV correlated nodes are more frequently observed in cancer cell lines than in normal cell lines (Figure 1D). First, we collected 21,064 nodes, which contain at least one sample-specific GV and show specific high/low degree in at least one of the 78 samples (73 normal and 5 cancer samples). Note that GV and high/low degree do not necessarily occur in the same cell line. For example, the degree of a node can be significantly high in GM12878, while the allele frequency of GV covered by this node is significantly high in K562. We found that the majority (62.59%) showed specificities in both cancer and normal cells, among which 18.36% were specific to cancer and 19.05% specific to normal cells (Figure 1F). Similarly, among the 629,547 cell-specific GVs, 58.86% showed specificities in both cancer and normal cells, 32.08% to only cancer cells and 9.06% to only normal cells (Figure 1E). However, the degree-GV-correlated nodes were dominated by cancer-specific nodes (87.18%), compared to 8.53% in both cancer and normal cells and 4.29% only in normal cells (Figure 1F). We observed the same trend for degree-GV-correlated hotspots, including 86.52% cancer-specific, 4.52% normal-specific, and 8.96% in both cancer and normal cells (Figure 1F). In summary, the majority of degree-GV-correlated nodes appear in cancer cells.

We further examined two groups of nodes in 8 distinct cell lines: one group had a significantly higher degree in one cell line than in other cell lines, which indicates cell-type-specific contact formation (one-cell-type-specific nodes), and the other had a significantly lower degree in one cell line than in the others, which indicates cell-type specific contact disruption (seven-cell-type-specific nodes). The percentages of HepG2-specific nodes and K562-specific nodes in the one-cell-type-specific group (cell-type-specific contact formation) are 28.5% and 7.4% in all nodes, 64.4% and 17.4% in degree-GV-correlated nodes and 64.3% and 18.1% in degree-GV-correlated hotspots, respectively (Figure 1G). Similarly, the percentages of HeLa-S3-specific nodes and K562-specific nodes in the seven-cell-type-specific group (cell-type-specific interaction disruption) are 41.8% and 13.2% in all nodes, 49.2% and 37.8% in degree-GV-correlated nodes and 61.3% and 27.1% in degree-GV-correlated hotspots, respectively (Figure 1H). Taken together, our analyses suggested that cancer-specific GVs are highly correlated with the node-degree change that alters the REIN.

CRISPR/Cas9 library screening identified hotspots essential for cell growth and survival

To further investigate the function of hotspots, 751 hotspots identified as enhancers were randomly selected for targeted deletion to analyse their impact on cell growth and survival (Supplementary Table S1). These hotspots do not overlap with coding regions of any protein-coding gene or noncoding RNA. In total, 14,399 paired gRNAs (pgRNAs) were designed to delete these loci (see Materials and Methods), including 473 positive control pgRNAs targeting 29 ribosomal genes, 100 negative control pgRNAs targeting the *AAVS1* locus and 100 non-targeting pgRNAs (Supplemen-

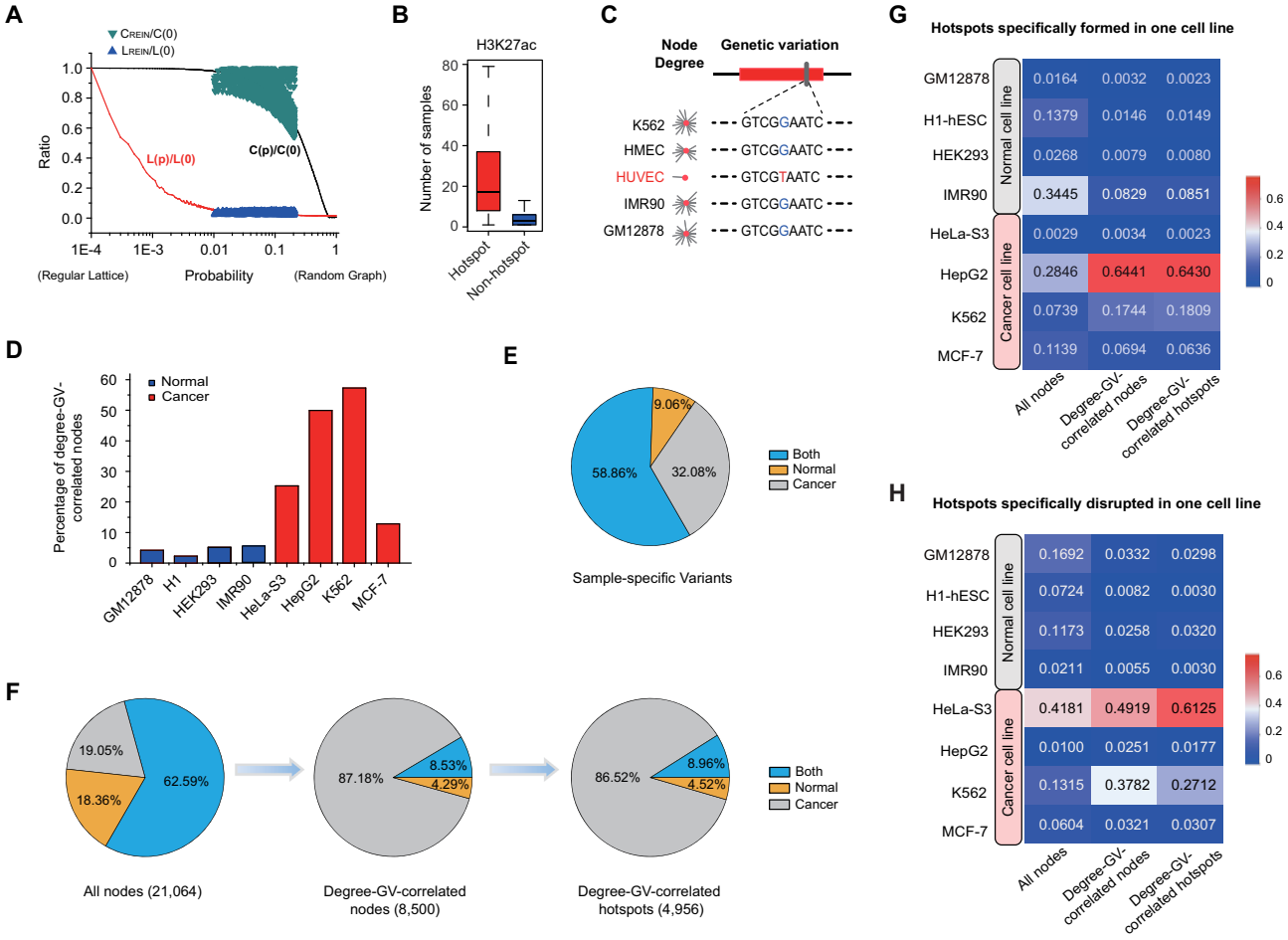

Figure 1. Small-world network analysis and mutation effects on 3D contact for hotspot enhancers and promoters. (**A**) The path length and cluster coefficient of REINs compared with equivalent regular lattice networks and equivalent random graph networks. (**B**) Comparison of H3K27ac peaks between hotspots and non-hotspots in 121 cell lines, primary cells and tissues characterized by the NIH Roadmap Epigenomics Project. (**C**) Definition of degree-GV-correlated nodes. In this example, the node has a low degree in the HUVEC cell line and a high degree in other cell lines, which is correlated with the GV profile with a G > T SNP in HUVEC that is not present in other cell lines. (**D**) The percentage of degree-GV-correlated nodes in normal cell lines and cancer cell lines. (**E**) The distribution of GV specificities in samples. Normal, cancer and both indicate GVs with specificities only in normal cells, only in cancer cells and in both cell types, respectively. (**F**) The distribution of normal or cancer cell line specificities in the nodes which contain at least one sample-specific GV and show specific high/low degree in at least one of the 78 samples (those nodes denoted as 'All nodes'), degree-GV-correlated nodes and degree-GV-correlated hotspots. Note that for the nodes in the first group (All nodes), GV and high/low degree do not necessarily occur in the same cell line. For example, the degree of a node can be significantly high in GM12878, while the allele frequency of GV covered by this node is significantly high in K562. (**G**) The distribution of one-cell-line hotspot formation in all nodes, degree-GV-correlated nodes and degree-GV-correlated hotspots. (**H**) The distribution of one-cell-line hotspot disruption in all nodes, degree-GV-correlated nodes and degree-GV-correlated hotspots.

tary Table S1). Through lentivirus infection at a low MOI (Multiplicity of Infection), the pgRNA library was transduced into K562 cells stably expressing Cas9 protein. The pgRNA-infected samples were FACS-sorted 3 days post infection, serving as the control group, and then continuously cultured for 30 days to obtain the experimental group. The library cells from the control and experimental groups were sequenced to determine the abundance of each pgRNA (Figure 2A). The read distribution of pgRNAs showed a high correlation between the two biologically independent replicates for all groups (see Supplementary Figure S1A-C and Table S2), indicating high reproducibility.

Compared with the control group, pgRNAs targeting ribosomal genes and hotspots in Day 30 experimental cells were both decreased more than those targeting the *AAVS1* locus and non-targeting pgRNAs. For all the pgRNAs of each hotspot, we calculated their fold changes and *P* values by comparing them with the pgRNAs targeting *AAVS1* using the Mann-Whitney U test (53,54), which is focused on analysing screening data with the in-library controls and could more accurately reflect the fitness effect of each locus. By randomly sampling the pgRNAs targeting *AAVS1*, we generated a distribution of negative controls and further computed the hotspots' *P* values. The screen score of each hotspot was calculated by combining its mean fold change and corrected *P* values (see Materials and Methods and Supplementary Table S3), and 49 hotspots with screen scores $\leq$ -2.5 were considered to significantly affect cell fitness upon deletion (Figure 2B). To avoid cellular toxicity caused by potential off-target effects (55–58), we assessed the specificities of sgRNAs with 2 or 3 mismatches to off-target loci using the GuideScan specificity score and

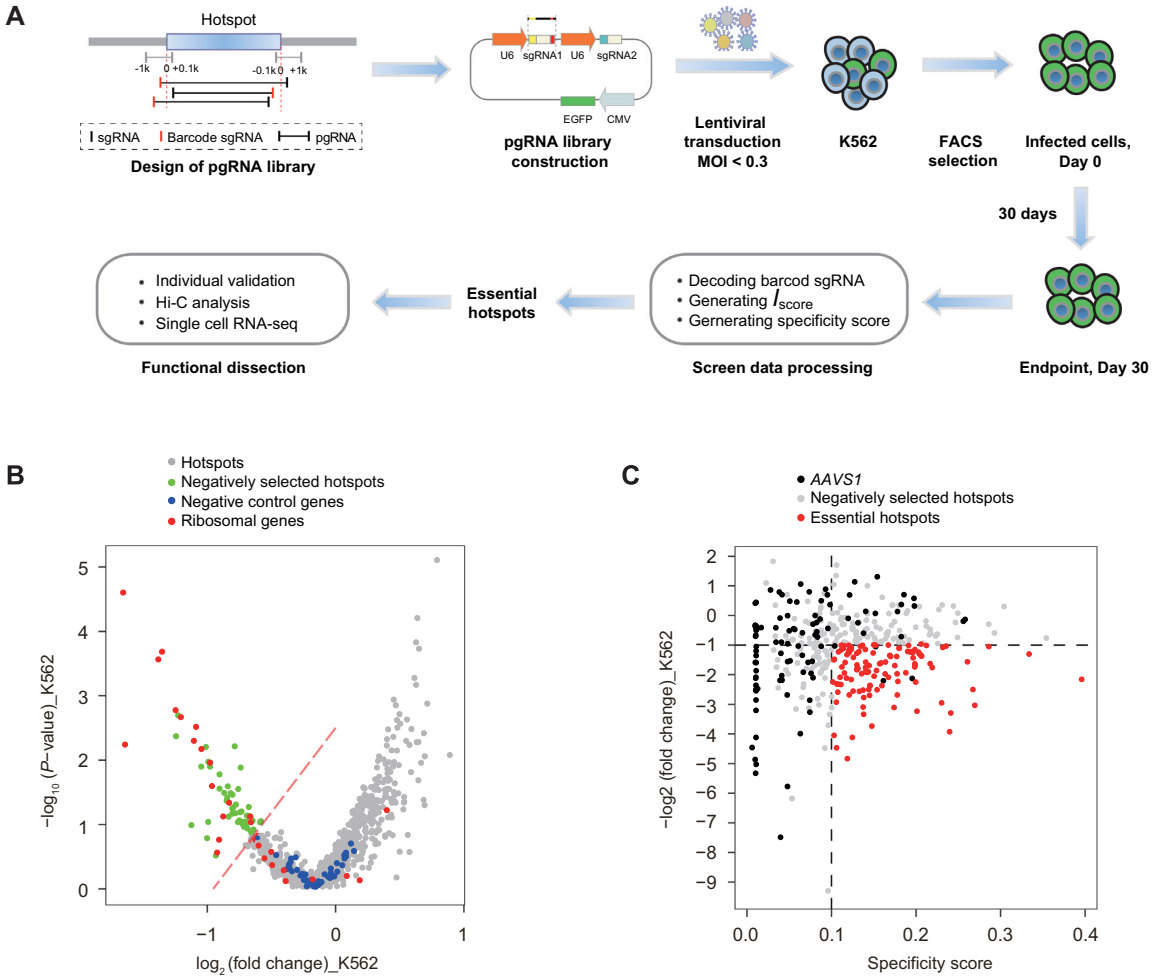

Figure 2. Identification of essential hotspots for cell growth and proliferation in the K562 cell line through pgRNA deletion-based CRISPR screening. (**A**) Schematic of the pgRNA library design, cloning and functional screening of selected hotspot loci. (**B**) Volcano plot of the fold change and *P-value* of hotspots in the K562 cell line. Negative control genes were generated by randomly sampling 20 *AAVS1*-targeting pgRNAs with replacement per gene, and ribosomal genes served as positive controls in the screening. The dotted red line represents a *screen score* = -2.5. (**C**) Selection of candidate essential hotspots by the fold change and specificity score of each pgRNA. These essential hits were selected under the threshold of a specificity score > 0.1 and $\log_2$(fold change) < -1.

calculated the specificity score for each pgRNA (see Materials and Methods) (28). Because *AAVS1*-targeting pgRNAs with specificity scores ≤ 0.1 could cause a dropout effect in K562 cells (Figure 2C and Supplementary Figure S1D), we only kept pgRNAs with specificity scores > 0.1 and $\log_2$(fold change) < -1 for subsequent analysis. Furthermore, hotspots with copy number amplification were also removed to avoid cell death caused by multiple cleavages (59). Using such stringent criteria, we identified 43 hotspots essential for the cell fitness of K562 cells (Figure 2C).

Based on the ranking of the screen score, 7 top-ranked hotspots in K562 cells were chosen for individual validation by cell proliferation assay. None of them overlapped with any promoter, protein-coding gene or noncoding RNA. Three or two pgRNAs with high targeting specificity were separately constructed for each hotspot, and the cell proliferation assay was performed as previously reported (54). We found that deletion of these hotspots led to significant cell death or cell growth inhibition (Figure 3A and Supplementary Figure S2A), which was consistent with the screening results, indicating that these hotspots played critical roles in cell fitness.

In our design, we assured that the deletion regions were not associated with any coding regions of protein-coding genes, but there were a few essential hotspots located near the promoter regions or in the introns of coding genes. To rule out the possibility of affecting the expressions of certain genes essential for cell growth and survival (essential genes) after hotspot deletion, we further investigated two identified hotspots located near the gene promoter or in the intronic regions, whose deletion may affect the expression of the corresponding genes. For hotspot_19_32 located in the intron of an essential gene *GATAD2A*, we chose 2 highly specific pgRNAs to respectively delete this locus in K562 cells and observed significant cell growth inhibition (Figure 3B). Importantly, we found that overexpression of *GATAD2A* did not rescue the cell death caused by hotspot deletion (Figure 3B), indicating that the hotspot deletion itself has a profound impact on cell growth. By detecting the expression level of the *GATAD2A* gene under each condi-

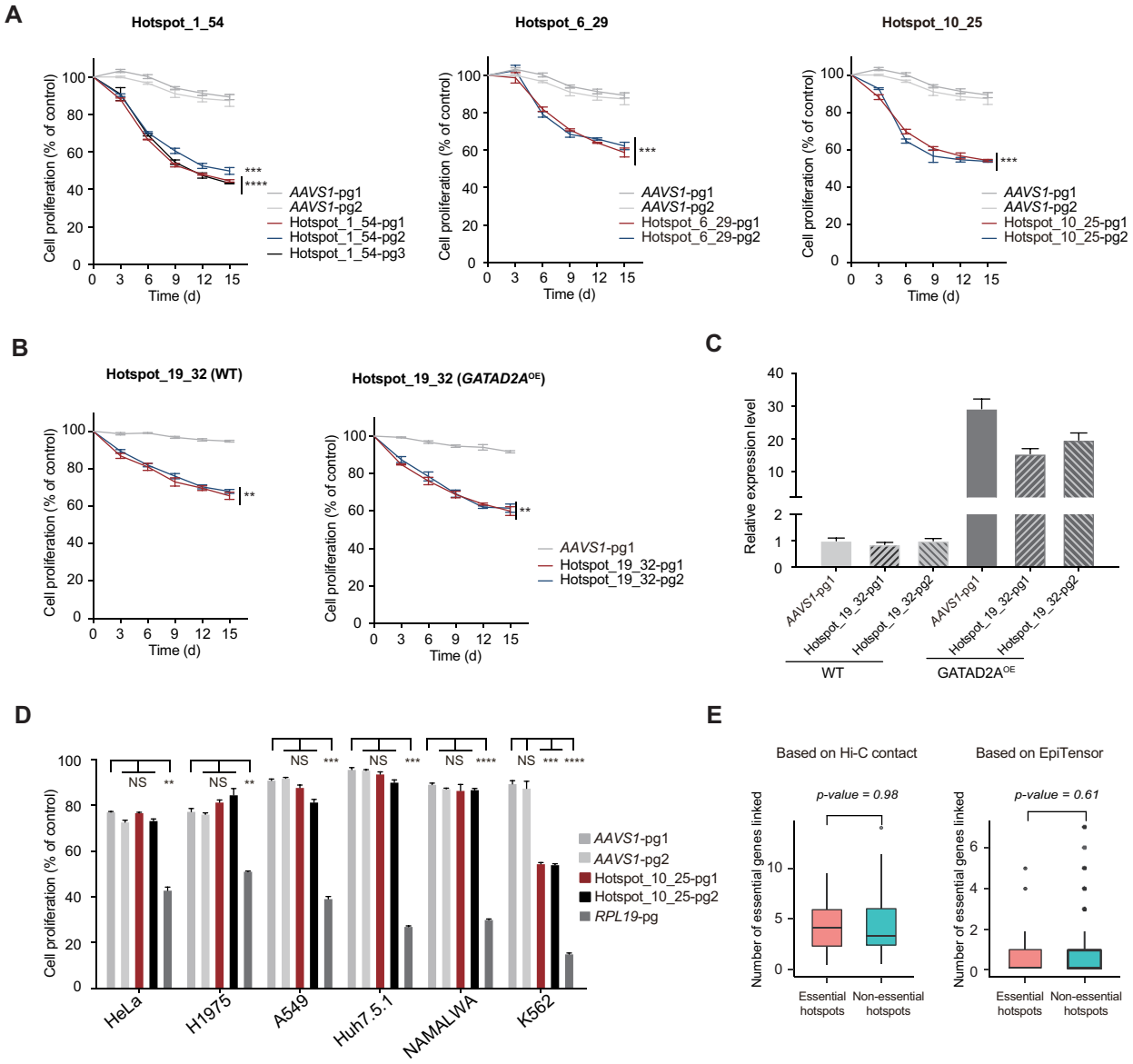

Figure 3. Validation of candidate essential hotspot loci in K562 cells and multiple cell lines. (**A**) Validation of the top-ranked essential hotspot in K562 cells by cell proliferation assay. *AAVS1*-pg1 and *AAVS1*-pg2 are pgRNAs targeting the *AAVS1* locus and serve as negative controls. The asterisk (*) represents *P-value* compared with pgRNAs targeting *AAVS1*-pg1 at Day 15, which were calculated by two-tailed Student's *t*-test and adjusted for multiple comparisons by Benjamini-Hochberg procedure. Data are presented as the mean ± s.d. (n = 3 biologically independent samples). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; NS, not significant. The pgRNAs for individual validation of each hotspot are listed in Supplementary Table S4. (**B**) Validation of essential hotspots overlapped with the intronic region of an essential gene in K562 cells by cell proliferation assay. Left: WT K562 cells infected with pgRNAs targeting hotspot_19_32. Right: *GATAD2A*-overexpressed K562 cells infected with pgRNAs targeting hotspot_19_32. (**C**) The expression levels of *GATAD2A* in WT and *GATAD2A*-overexpressed K562 cells infected with pgRNAs targeting *AAVS1* or hotspot_19_32. (**D**) Validation of hotspots_10_25 in multiple cancer cell lines, including A549, H1975, HeLa, Huh7.5.1 and NAMALWA cell lines. Asterisk (*) represents *P-value* compared with pgRNAs targeting *AAVS1*-pg1 at Day 15, which were calculated by two-tailed Student's *t*-test and adjusted by Bonferroni correction accounting for multiple testings. * $P < 0.05$; ** $P < 0.01$; *** $P < 0,001$; **** $P < 0.0001$; NS, not significant. (**E**) No significant difference between the numbers of essential genes contacting essential and non-essential hotspots from Hi-C or EpiTensor in K562 cells. The pgRNAs used above are listed in Supplementary Table S4.

tion by real-time qPCR, we confirmed that the gene was successfully overexpressed in K562 cells and that the cell death caused by the hotspot deletion was not rescued by *GATAD2A* overexpression (Figure 3C). A similar result was obtained for hotspot_1_36, which is located approximately 3 kb upstream of the transcriptional start site of an essential gene *SLC2A1*. We performed the cell proliferation assay using 2 pgRNAs in wild-type K562 cells and K562 cells stably overexpressing *SLC2A1*. A similar level of influence on cell fitness was observed in both conditions for each pgRNA deletion, and real-time qPCR further confirmed that the growth phenotype was not due to affecting the expression level of *SLC2A1* (Supplementary Figure S2B and C).

To further assess the essentialities of the identified K562-essential hotspots in other cancer cell lines, we chose hotspot_10_25 (chr10: 74,123,469–74,1248,68), which

showed a significant growth defect in K562 if deleted, for parallel validations in HeLa (cervical cancer cells), H1975 (non-small cell lung cancer cells), A549 (non-small cell lung cancer cells) and NAMALWA (Burkitt's lymphoma) cells. Surprisingly, compared with the negative control *AAVS1*-targeting pgRNAs, hotspot_10_25 showed no significant effect in any of the five tested cell lines, indicating that its role in K562 cells is cell-type specific (Figure 3D and Supplementary Figure S2D).

The essentiality of hotspots does not result from any association with essential genes

To understand how these identified essential hotspots exert their functional roles, we first examined whether direct interaction with essential genes determines the essentialities of these hotspots. We retrieved the essential genes whose knockdown would lead to cell death according to the CRISPRi-based screen (48) and identified all possible spatial contacts they formed that were detected by Hi-C (P-value ≤ 0.05) in wild-type K562 cells (60). There is no distinction between essential and non-essential hotspots regarding their association with essential genes (the *Wilcoxon* Rank-Sum test P-value $= 0.98$, indicating no significant difference) (Figure 3E). We also performed the same comparison using the spatial contacts predicted by EpiTensor and reached the same conclusion (the *Wilcoxon* Rank-Sum test P-value $= 0.61$) (Figure 3E). According to the above analysis, the essentiality of hotspots is not determined by their direct contact with essential genes.

Deleting essential hotspots can affect broad chromatin organization

We next investigated whether deleting hotspots affects chromatin organization. We selected hotspot_10_25 (chr10: 74,123,469–74,124,868) for further analysis, which showed unique essentiality in K562 cells (Figure 3A, D and Supplementary Figure S2D) yet does not interact with any essential protein-coding gene identified in the previous CRISPRi screening (48) in the Hi-C analysis. We first performed whole genome sequencing (WGS) to confirm that there was no off-target effect. The validated pgRNA hotspot_10_25-pg2 was chosen (Figure 3A), and the WGS library was generated 8 days after pgRNA infection in K562 cells. Compared to the hg19 human genome, we identified 4.1 million germline mutations in hotspot_10_25-deleted K562 cells, which showed 86.2% consistency with the published wild-type K562 WGS data. The high percentage of the germline mutation discovery rate indicated good quality of the library. We used Cas-OFFinder to identify 746 potential off-target loci with loose cut-off values (base mismatch ≤ 4, bulge ≤ 2) to avoid missing any possible off-target loci. We manually examined the putative off-target loci with the indels detected from the edited cells that differed from the wild-type cells (Supplementary Table S5). Except for the significant indels found in the two on-target loci with clear cleavages in the pileup reads in the genome browser view (Supplementary Figure S3), there was no cleavage in the pileup reads on any of the putative off-target loci (two examples of possible off-target loci are shown in Supplementary Figure S3). These analyses confirmed that the cell growth defects did not result from off-target effects.

We subsequently performed Hi-C analysis on the edited K562 cells and compared it with the wild-type cells (60) (see Materials and Methods). The 100 kb-resolution Hi-C contact maps of the wild-type and hotspot-deleted cells are overall similar (Supplementary Figure S4A), and no distinct flips between A and B compartments were observed on the entire chr10 at 50-kb resolution (Supplementary Figure S4B). We compared the chromosome-wide similarity and detected differential Hi-C contacts using HiCRep (33,34) and HiCcompare (35). The high Stratum-adjusted correlation coefficient (SCC > 0.7) and the small percentage of differential Hi-C contacts suggested an overall similarity between the wild-type and hotspot_10_25-deleted K562 cells (Supplementary Table S6). Topologically associated domains (TADs) largely remained similar, with a few TADs merge and split in the regions of chr10: 12230000–14540000, chr10: 15210000–15910000, chr10: 71220000–72220000, chr10: 89850000–91010000, chr10: 95290000–96350000 and chr10: 99450000–100140000 (Supplementary Figure S4C). Interestingly, using HiCCUPS (https://github.com/aidenlab/juicer/wiki/HiCCUPS), we found that hotspot deletion did affect chromatin loops (Supplementary Table S7). We next investigated whether deleting a hotspot could affect relatively broad genomic regions in spatial proximity. Using a sliding window with a bin step of 1 Mb and a flanking region of 2 Mb in the linear genome, we assembled all the Hi-C contacts (5-kb resolution with P-value ≤ 0.05) in each 5-Mb window into a sub-network. The modularity score and effective diameter were computed for each of these sub-networks in the wild-type and hotspot-deleted K562 cells. These two metrics of all the sliding windows of chr10 showed a high correlation between before and after hotspot deletion, with *Pearson* correlation coefficients of 0.84 and 0.91 for the effective diameter and modularity, respectively (Figure 4A and B). Notably, significant changes were observed on the 6–8 Mb regions surrounding the deleted hotspot for both effective diameter (chr10: 69–75 Mb for bin 71 and 72) and modularity (chr10: 68–76 Mb for bin 70, 71 and 73) (Figure 4A and B). Some other genomic regions interacting with the hotspot neighboring regions were also affected, such as chr10: 11–17 Mb (bin 13 and 14), showing a significant change in modularity (Figure 4B and Supplementary Figure S5). We further examined the Hi-C contact alteration within chr10: 11–17 Mb and chr10: 68–76 Mb (Figure 4C and E). In the region of chr10: 12–14 Mb with observed Hi-C contact changes (Figure 4C), we also found consistent TAD splits (Supplementary Figure S4C), disruption and formation of chromatin loops upon hotspot deletion (Supplementary Table S8). These chromatin changes led to alteration of promoter-enhancer interactions, such as the enhanced and weakened contacts between the *CELF2*, *RSU1*, *FAM149B1* and *CCAR1* promoters and their interacting enhancers upon hotspot deletion (Figure 4D and F). Notably, these affected promoters and enhancers are not only located close to but also can be as far as 62 Mb away from the deleted hotspot_10_25 located at chr10: 74,123,469–74,124,868. These observations showed that hotspot deletion resulted in broad alterations in chromatin structure beyond its linear neighbor genome.

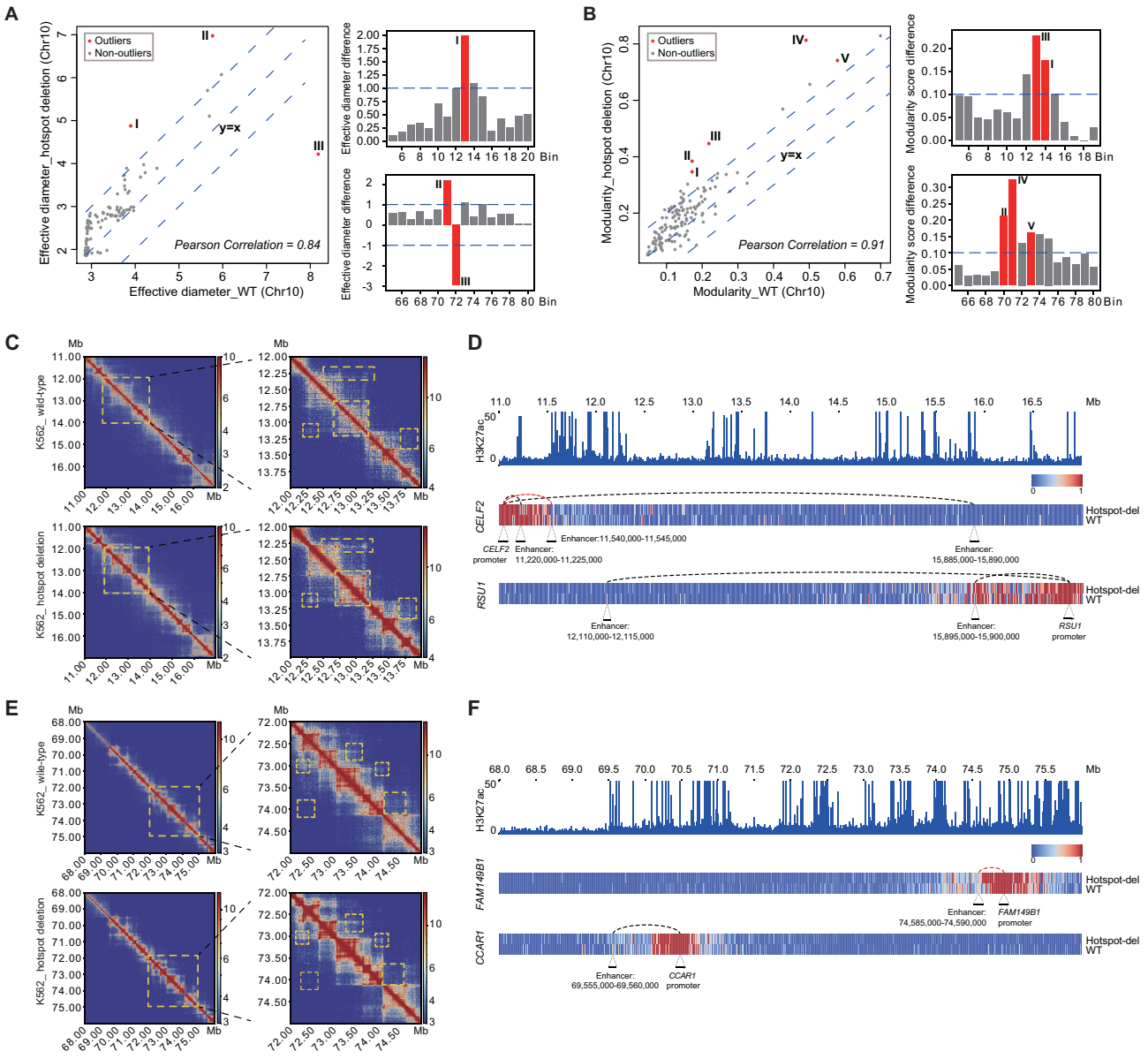

Figure 4. Deletion of an essential hotspot impacts broad chromatin structure. (**A-B**) The effective diameter (**A**) and modularity (**B**) before and after hotspot deletion in the sliding 5-Mb sub-networks on chr10 (left). The outliers are labeled, and their genomic locations are shown on the right. (**C**) Hi-C contact maps of chr10: 11–17 Mb at 5-kb resolution (left) and 12–14 Mb at 5-kb resolution (right), before and after hotspot deletion. (**D**) Two examples, *CELF2* and *RSU1*, for enhancer-promoter interactions altered after hotspot deletion within chr10: 11–17 Mb. (**E**) Hi-C contact maps of chr10: 68–76 Mb at 5-kb resolution (left) and chr10: 72–75 Mb at 5-kb resolution (right), before and after hotspot deletion. (**F**) Two examples, *FAM149B1* and *CCAR1*, for enhancer-promoter interactions altered after hotspot deletion within chr10: 68–76 Mb. In Figure 4D and F, Black dash line indicates decreased interactions in hotspot_10_25-deleted K562 cells, red dash line indicates enhanced interactions in hotspot_10_25-deleted K562 cells.

Essential hotspots tend to reside in dense chromatin structures

If essential hotspots are critical for maintaining the chromatin structure in the spatial neighbourhood, it is likely that the 3D contacts around them are dense. Therefore, we compared the sub-network effective diameters, modularity and chromatin loops in the 5-Mb regions centered at the essential and non-essential hotspots in wild-type K562. We found that essential hotspots were surrounded by higher effective diameters (*Wilcoxon* Rank-Sum test, *P*-value = 7.3E-6), higher modularities (*P*-value = 0.1) and higher loop densities (*P*-value = 4.7E-4) than non-essential hotspots (Supplementary Figure S6A-C). In fact, using these three metrics in wild-type K562 cells, a random forest classification model could distinguish essential and non-essential hotspots with an AUC of 0.73 in 10-fold cross validations. This result resonates with the above observations and suggests that hotspots are pivotal for stabilizing dense chromatin contacts in the spatial neighborhood.

Hotspot deletion synergistically affects gene expression

We next performed single-cell RNA-seq using Drop-seq (43) to analyse the changes in gene expression upon

hotspot_10_25 deletion. We transduced the individually validated pgRNA hotspot_10_25-pg2 (Figure 3A) targeting this essential hotspot into K562 cells, among which 482 single cells passed the quality control. We also included the bulk RNA-seq data of wild-type and *AAVS1*-deleted cells as controls. All the data were normalized together to make them comparable (see Materials and Methods). As deletion of this hotspot has an impact on cell viability or cell growth, we focused on genes related to apoptosis pathways to confirm their activation. We selected 99 apoptosis-related genes documented in the KEGG database and clustered the cells into five states by trajectory branching and pseudotime analysis using Monocle (Figure 5A) (45). The wild-type and *AAVS1*-deleted K562 cells (negative control) were located in state 1, suggesting that cells in this state resemble the wild-type cells. The apoptosis genes fell into three groups, with distinct expression patterns along the pseudotime but overall all increasing from cell state 1 to cell state 5, for example, *CASP2*, *CASP8*, *CASP9*, and *CASP10* in cluster 1, *CASP6* in cluster 2, and *CASP7* in cluster 3 (Figure 5B). Taken together, the single cell transcriptomic analysis showed that apoptosis pathways are activated upon hotspot deletion.

To investigate the impact of hotspot deletion on the spatial neighborhood, we analysed the genes whose promoters were predicted to interact with the essential hotspot_10_25 by EpiTensor. Among the 14 genes located within the same TAD of hotspot_10_25, 4 showed significantly downregulated (P-value < 0.05) in the progression from state 1 to 5, including *P4HA1* (downregulated from state 1 to 2, Figure 5C) and *DNAJB12*, *ASCC1* and *ECD* (downregulated from state 2 to 4, Figure 5D-E and Supplementary Figure S7A). By individually knocking down each gene by CRISPR interference (Figure 5F and Supplementary Figure S7B), only *ECD* knockdown showed a weak impact on cell growth, and all the other genes showed no detectable effects (Figure 5G and Supplementary Figure S7C). As the hotspot interacted with multiple genes, we investigated whether knocking down a pair of genes would have a synergistic effect on cell growth. Applying the CRISPRi strategy, we knocked down 6 pairs of genes (Supplementary Table S4) in K562 cells using paired gRNAs, respectively. We found that simultaneous knockdown of *P4HA1-ECD* and *ASCC1-ECD* showed a much more significant impact on cell growth (Figure 5G). These results indicated that disrupting hotspot_10_25 could affect the expression levels of multiple interacting genes, and their synergistic effect could lead to cell death. Note that we were limited to examining pairs of genes, but hotspot deletion can affect multiple genes together with more significant synergic effects.

DISCUSSION

In this study, we analysed the hotspot promoters/enhancers that were predicted by EpiTensor (25) to form a large number of 3D contacts with other promoters/enhancers. The unsupervised learning method EpiTensor predicts all possible 3D contacts of promoter-promoter, promoter-enhancer and enhancer-enhancer pairs. In a particular cell line, we focused on the predicted contacts between active promoters/enhancers denoted by histone marks. The hotspot promoters/enhancers are defined by their 3D contacts with many active promoters/enhancers, which makes them a class of high-degree nodes in the REIN. We showed that REIN is a small-world network that is vulnerable to targeted perturbation to high-degree nodes. Therefore, it is reasonable to infer that hotspots can be important for stabilizing REIN and the 3D contacts formed between active promoters/enhancers.

We found that the occurrence of genetic variations (GVs) is much more strongly correlated with the alteration of 3D contact degrees (degree-GV-correlated nodes) in the hotspots in cancer cells than in normal cells. Furthermore, we showed that cancer-specific hotspots (only formed or disrupted in one particular cancer cell) are enriched with degree-GV-correlated nodes. Taken together, these observations suggest that GVs occurring in hotspots can lead to chromatin structure changes and dysregulated cellular functions.

To confirm the functional importance of the hotspots, we performed CRISPR/Cas9 library screening on hotspot enhancers by paired-gRNA deletion in the K562 cancer cell line. By calculating the screen score for each hotspot and further filtering loci with potential off-target effects or copy number amplifications, we identified 43 hotspots essential for cell growth and survival. Nine randomly selected loci were individually validated by cell proliferation assay, including 7 top-ranked hotspots in K562 cells without any overlap with coding genes and 2 loci overlapped with promoter or intronic regions of certain genes. We further identified that hotspot_10_25 was essential for cell fitness specifically in K562 cells through multiple validations in four other cancer cell lines.

We thus selected a hotspot enhancer (hotspot_10_25) as a representative of cancer-specific hits for in-depth analysis, which ensured that no off-target cleavages occurred through WGS analysis. Note that this selected hotspot is not unique compared to the other essential hotspots, and the insights obtained here are expected to be generalizable. Hi-C and scRNA-seq analyses showed that deleting this 1.4-kb long hotspot could impact a broad chromatin structure of 8-Mb regions surrounding the hotspot and affect the expression of numerous distal genes not even directly associated with the hotspot. These observations indicate that the hotspot enhancer has a pivotal role in chromatin organization beyond forming chromatin loops.

Importantly, this hotspot does not directly interact with any essential gene, and thus, the cell death resulting from its deletion is not due to directly disrupting the expression of essential genes. Single cell RNA-seq revealed that hotspot_10_25 deletion could affect the expression levels of multiple interacting genes located within the same TAD of the hotspot. By knocking down individuals and pairs of these genes, we found that although none of these dysregulated genes has a significant impact on cell fitness individually, altered expression of gene pairs showed significant synergistic effects leading to cell death.

We have revealed the understudied 'structural importance' of noncoding regulatory elements, especially enhancers. We are aware that establishing the causal relationship between broad chromatin organization changes and cell proliferation or gene expression remains technically challenging. However, to our knowledge, this is the

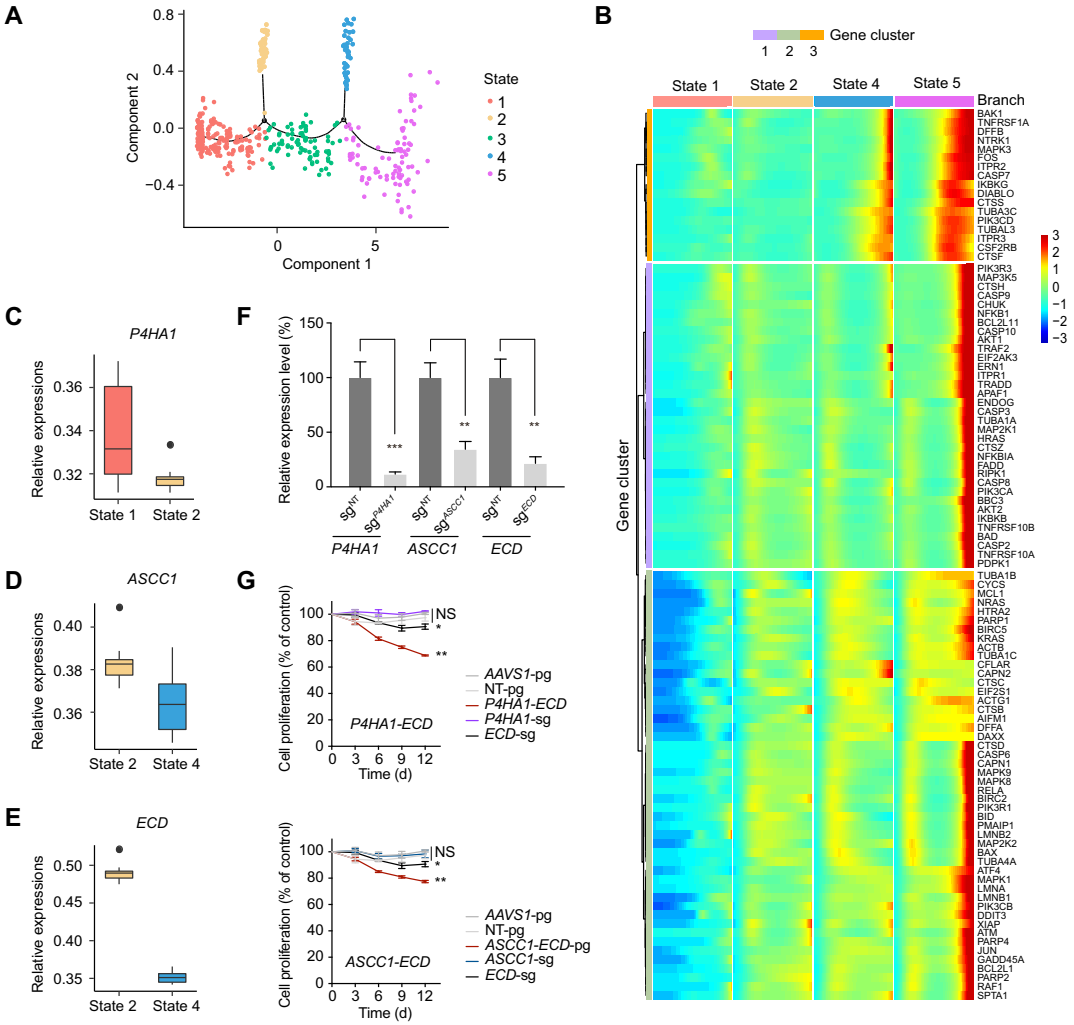

Figure 5. Synergistic change in gene expression after hotspot deletion. (**A**) Pseudotime clusters of hotspot_10_25-deleted and wild-type K562 cells based on apoptosis gene expression. (**B**) Global analysis of the expression levels of 99 KEGG apoptosis genes in state 1, 2, 4 and 5. Genes were clustered into 3 groups. (**C-E**) The relative expression levels of three representative downregulated genes *P4HA1, ASCC1, ECD* in different states as determined by single cell RNA-seq. (**F**) The knockdown efficiency of the indicated sgRNAs targeting each downregulated gene in K562 cells stably expressing KRAB-dCas9. The expression level of each gene was detected by real-time qPCR. $sgRNA^{NT}$ represents the non-targeting sgRNA serving as the negative control. (**G**) Validation of the synergistic effects of two sets of gene pairs on K562 cell fitness by cell proliferation assay. Asterisk (*) represents *P-value* compared with pgRNAs targeting *AAVS1*-pg at Day 12, which were calculated by two-tailed Student's *t*-test and adjusted for multiple comparisons by Benjamini–Hochberg procedure. * $P < 0.05$; ** $P < 0.01$; NS, not significant. The sgRNAs, pgRNAs and primers used are listed in Supplementary Table S4.

first report about the observations that enhancers could maintain a broad chromatin organization, which goes far beyond the direct interaction between promoters and enhancers. A logical inference towards the causal relationship based on our observations is that the impact of hotspot deletion is propagated through the 3D contact network and could impact distal genes that are important for cell fitness.

DATA AVAILABILITY

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The sequencing data in this paper are accessible through NCBI Sequence Read Archive (SRA) under BioProject ID PRJNA749478.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We acknowledge the staff of the BIOPIC High-throughput Sequencing Center (Peking University) for their assistance in next-generation sequencing analysis, the National Center for Protein Sciences (Beijing) at Peking University for their assistance with fluorescence-activated cell sorting and analysis, and Dr. Hongxia Lv and Ms. Liying Du for their technical help. We acknowledge Dr. Ying Yu (Peking University) for her assistance in preparing the NGS library. We acknowledge the staff of the UC San Diego IGM Genomics Center for sequencing services and the UC San Diego Human Embryonic Stem Cell Core Facility for cell sorting ser-

vices. We acknowledge Ms. Jia Xu (UC San Diego) for her assistance in preparing a single cell RNA-seq library.
Author contributions: W.Wei and W.Wang conceived and supervised the project. W.Wei, W.Wang, Y.L. and B.D. designed the experiments. B.D. and L.Z. constructed network analysis and identified and characterized hotspot regions. Y.G. designed the pgRNA library for hotspot screening. Y.L. and P.X. performed the pgRNA library construction and screening. Y.L. performed the experiments, including individual validation of candidate hotspots in multiple cell lines, whole-genome sequencing (WGS), bulk RNA-seq and examination of the synergistic effects with the help of P.X. and Q.P. Z.L. performed the bioinformatics analysis of the screening data and designed the pgRNAs used for individual validation. P.W. and Z.C. performed the Hi-C experiments on hotspot-deleted K562 cells. P.W. and Y.Z. performed single cell RNA-seq on hotspot-deleted K562 cells. L.Z. and B.D. performed the bioinformatics analyses of the WGS, Hi-C and single-cell RNA-seq data. Y.L., B.D., L.Z., W.Wang and W.Wei wrote the manuscript with contributions from all other authors.

FUNDING

This project was supported by funds from CIRM (RB5-07012) and the NIH (R01HG009626) (to Wei Wang); the National Science Foundation of China (NSFC31930016), Beijing Municipal Science & Technology Commission (Z181100001318009), the Beijing Advanced Innovation Center for Genomics at Peking University and the Peking-Tsinghua Center for Life Sciences (to Wensheng Wei); China Postdoctoral Science Foundation (2020M670031, to Ying Liu).
Conflict of interest statement. None declared.

REFERENCES

1. Andersson,R., Gebhard,C., Miguel-Escalada,I., Hoof,I., Bornholdt,J., Boyd,M., Chen,Y., Zhao,X., Schmidl,C., Suzuki,T. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455–461.
2. Furlong,E.E.M. and Levine,M. (2018) Developmental enhancers and chromosome topology. *Science*, **361**, 1341–1345.
3. Tippens,N.D., Vihervaara,A. and Lis,J.T. (2018) Enhancer transcription: what, where, when, and why?*Genes Dev.*, **32**, 1–3.
4. Plank,J.L. and Dean,A. (2014) Enhancer function: mechanistic and genome-wide insights come together. *Mol. Cell*, **55**, 5–14.
5. van Arensbergen,J., van Steensel,B. and Bussemaker,H.J. (2014) In search of the determinants of enhancer-promoter interaction specificity. *Trends Cell Biol.*, **24**, 695–702.
6. Ernst,J., Kheradpour,P., Mikkelsen,T.S., Shoresh,N., Ward,L.D., Epstein,C.B., Zhang,X., Wang,L., Issner,R., Coyne,M. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**, 43–49.
7. Chepelev,I., Wei,G., Wangsa,D., Tang,Q. and Zhao,K. (2012) Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res.*, **22**, 490–503.
8. Fitz,J., Neumann,T., Steininger,M., Wiedemann,E.-M., Garcia,A.C., Athanasiadis,A., Schoeberl,U.E. and Pavri,R. (2020) Spt5-mediated enhancer transcription directly couples enhancer activation with physical promoter interaction. *Nat. Genet.*, **52**, 505–515.
9. Pombo,A. and Dillon,N. (2015) Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.*, **16**, 245–257.
10. Long,H.K., Prescott,S.L. and Wysocka,J. (2016) Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell*, **167**, 1170–1187.
11. Schoenfelder,S. and Fraser,P. (2019) Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.*, **20**, 437–455.
12. Ghavi-Helm,Y., Klein,F.A., Pakozdi,T., Ciglar,L., Noordermeer,D., Huber,W. and Furlong,E.E.M. (2014) Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*, **512**, 96–100.
13. Sanyal,A., Lajoie,B.R., Jain,G. and Dekker,J. (2012) The long-range interaction landscape of gene promoters. *Nature*, **489**, 109–113.
14. Rao,S.S.P., Huntley,M.H., Durand,N.C., Stamenova,E.K., Bochkov,I.D., Robinson,J.T., Sanborn,A.L., Machol,I., Omer,A.D., Lander,E.S. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.
15. Hansen,A.S., Pustova,I., Cattoglio,C., Tjian,R. and Darzacq,X. (2017) CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife*, **6**, e25776.
16. Shen,Y., Yue,F., McCleary,D.F., Ye,Z., Edsall,L., Kuan,S., Wagner,U., Dixon,J., Lee,L., Lobanenkov,V.V. *et al.* (2012) A map of the cis-regulatory sequences in the mouse genome. *Nature*, **488**, 116–120.
17. Siersbæk,R., Madsen,J.G.S., Javierre,B.M., Nielsen,R., Bagge,E.K., Cairns,J., Wingett,S.W., Traynor,S., Spivakov,M., Fraser,P. *et al.* (2017) Dynamic rewiring of promoter-anchored chromatin loops during adipocyte differentiation. *Mol. Cell*, **66**, 420–435.
18. Morcillo,P., Rosen,C., Baylies,M.K. and Dorsett,D. (1997) Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in drosophila. *Genes Dev.*, **11**, 2729–2740.
19. Deng,W., Lee,J., Wang,H., Miller,J., Reik,A., Gregory,P.D., Dean,A. and Blobel,G.A. (2012) Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*, **149**, 1233–1244.
20. Kong,S., Bohl,D., Li,C. and Tuan,D. (1997) Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. *Mol. Cell. Biol.*, **17**, 3955–3965.
21. Rickman,C. and Bickmore,W.A. (2013) Transcription. Flashing a light on the spatial organization of transcription. *Science*, **341**, 621–622.
22. Teves,S.S., An,L., Hansen,A.S., Xie,L., Darzacq,X. and Tjian,R. (2016) A dynamic mode of mitotic bookmarking by transcription factors. *Elife*, **5**, e22280.
23. Liu,Z., Legant,W.R., Chen,B.-C., Li,L., Grimm,J.B., Lavis,L.D., Betzig,E. and Tjian,R. (2014) 3D imaging of Sox2 enhancer clusters in embryonic stem cells. *Elife*, **3**, e04236.
24. Heinz,S., Texari,L., Hayes,M.G.B., Urbanowski,M., Chang,M.W., Givarkes,N., Rialdi,A., White,K.M., Albrecht,R.A., Pache,L. *et al.* (2018) Transcription elongation can affect genome 3D structure. *Cell*, **174**, 1522–1536.
25. Zhu,Y., Chen,Z., Zhang,K., Wang,M., Medovoy,D., Whitaker,J.W., Ding,B., Li,N., Zheng,L. and Wang,W. (2016) Constructing 3D interaction maps from 1D epigenomes. *Nat. Commun.*, **7**, 10812.
26. Roadmap Epigenomics Consortium, Kundaje,A., Meuleman,W., Ernst,J., Bilenky,M., Yen,A., Heravi-Moussavi,A., Kheradpour,P., Zhang,Z., Wang,J. *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.
27. Zhu,S., Li,W., Liu,J., Chen,C.-H., Liao,Q., Xu,P., Xu,H., Xiao,T., Cao,Z., Peng,J. *et al.* (2016) Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nat. Biotechnol.*, **34**, 1279–1286.
28. Perez,A.R., Pritykin,Y., Vidigal,J.A., Chhangawala,S., Zamparo,L., Leslie,C.S. and Ventura,A. (2017) GuideScan software for improved single and paired CRISPR guide RNA design. *Nat. Biotechnol.*, **35**, 347–349.
29. Durand,N.C., Shamim,M.S., Machol,I., Rao,S.S.P., Huntley,M.H., Lander,E.S. and Aiden,E.L. (2016) Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.*, **3**, 95–98.
30. Ramírez,F., Bhardwaj,V., Arrigoni,L., Lam,K.C., Grüning,B.A., Villaveces,J., Habermann,B., Akhtar,A. and Manke,T. (2018) High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.*, **9**, 189.

31. Wolff,J., Bhardwaj,V., Nothjunge,S., Richard,G., Renschler,G., Gilsbach,R., Manke,T., Backofen,R., Ramírez,F. and Grüning,B.A. (2018) Galaxy hicexplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.*, **46**, W11–W16.
32. Akdemir,K.C. and Chin,L. (2015) HiCPlotter integrates genomic data with interaction matrices. *Genome Biol.*, **16**, 198.
33. Lin,D., Sanders,J. and Noble,W.S. (2021) HiCRep.py: fast comparison of Hi-C contact matrices in python. *Bioinformatics*, **37**, 2996–2997.
34. Yang,T., Zhang,F., Yardımcı,G.G., Song,F., Hardison,R.C., Noble,W.S., Yue,F. and Li,Q. (2017) HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.*, **27**, 1939–1949.
35. Stansfield,J.C., Cresswell,K.G., Vladimirov,V.I. and Dozmorov,M.G. (2018) HiCcompare: an R-package for joint normalization and comparison of HI-C datasets. *BMC Bioinf.*, **19**, 279.
36. Kim,D., Bae,S., Park,J., Kim,E., Kim,S., Yu,H.R., Hwang,J., Kim,J.-I. and Kim,J.-S. (2015) Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods*, **12**, 237–243.
37. Smith,C., Gore,A., Yan,W., Abalde-Atristain,L., Li,Z., He,C., Wang,Y., Brodsky,R.A., Zhang,K., Cheng,L. *et al.* (2014) Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell*, **15**, 12–13.
38. Ding,B., Liu,Y., Liu,Z., Zheng,L., Xu,P., Chen,Z., Wu,P., Zhao,Y., Pan,Q., Guo,Y. *et al.* (2021) Noncoding loci without epigenomic signals can be essential for maintaining global chromatin organization and cell viability. *Sci. Adv.*, **7**, eabi6020.
39. Pertea,M., Kim,D., Pertea,G.M., Leek,J.T. and Salzberg,S.L. (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, stringtie and ballgown. *Nat. Protoc.*, **11**, 1650–1667.
40. Kim,D., Paggi,J.M., Park,C., Bennett,C. and Salzberg,S.L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.*, **37**, 907–915.
41. Kim,D., Langmead,B. and Salzberg,S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.
42. Pertea,M., Pertea,G.M., Antonescu,C.M., Chang,T.-C., Mendell,J.T. and Salzberg,S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.*, **33**, 290–295.
43. Macosko,E.Z., Basu,A., Satija,R., Nemesh,J., Shekhar,K., Goldman,M., Tirosh,I., Bialas,A.R., Kamitaki,N., Martersteck,E.M. *et al.* (2015) Highly parallel Genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, **161**, 1202–1214.
44. Kanehisa,M., Furumichi,M., Tanabe,M., Sato,Y. and Morishima,K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.
45. Trapnell,C., Cacchiarelli,D., Grimsby,J., Pokharel,P., Li,S., Morse,M., Lennon,N.J., Livak,K.J., Mikkelsen,T.S. and Rinn,J.L. (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.*, **32**, 381–386.
46. Qiu,X., Hill,A., Packer,J., Lin,D., Ma,Y.-A. and Trapnell,C. (2017) Single-cell mRNA quantification and differential analysis with census. *Nat. Methods*, **14**, 309–315.
47. Haynes,W. (2013) Wilcoxon rank sum test. *Encyclopedia of Systems Biology*, 2354–2355.
48. Horlbeck,M.A., Gilbert,L.A., Villalta,J.E., Adamson,B., Pak,R.A., Chen,Y., Fields,A.P., Park,C.Y., Corn,J.E., Kampmann,M. *et al.* (2016) Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife*, **5**, e19760.
49. Benson,A.R., Gleich,D.F. and Leskovec,J. (2016) Higher-order organization of complex networks. *Science*, **353**, 163–166.
50. Ball,W.W.R. and Coxeter,H.S.M. (1987) In: *Mathematical Recreations and Essays Courier Corporation*.
51. Cohen,R. and Havlin,S. (2010) In: *Complex Networks: Structure, Robustness and Function Cambridge University Press*.
52. Grabow,C., Grosskinsky,S., Kurths,J. and Timme,M. (2015) Collective relaxation dynamics of small-world networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **91**, 052815.
53. Liu,S.J., Horlbeck,M.A., Cho,S.W., Birk,H.S., Malatesta,M., He,D., Attenello,F.J., Villalta,J.E., Cho,M.Y., Chen,Y. *et al.* (2017) CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science*, **355**, eaah7111.
54. Liu,Y., Cao,Z., Wang,Y., Guo,Y., Xu,P., Yuan,P., Liu,Z., He,Y. and Wei,W. (2018) Genome-wide screening for functional long noncoding RNAs in human cells by cas9 targeting of splice sites. *Nat. Biotechnol.*, **36**, 1203–1210.
55. Munoz,D.M., Cassiani,P.J., Li,L., Billy,E., Korn,J.M., Jones,M.D., Golji,J., Ruddy,D.A., Yu,K., McAllister,G. *et al.* (2016) CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov.*, **6**, 900–913.
56. Aguirre,A.J., Meyers,R.M., Weir,B.A., Vazquez,F., Zhang,C.-Z., Ben-David,U., Cook,A., Ha,G., Harrington,W.F., Doshi,M.B. *et al.* (2016) Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.*, **6**, 914–929.
57. Morgens,D.W., Wainberg,M., Boyle,E.A., Ursu,O., Araya,C.L., Tsui,C.K., Haney,M.S., Hess,G.T., Han,K., Jeng,E.E. *et al.* (2017) Genome-scale measurement of off-target activity using cas9 toxicity in high-throughput screens. *Nat. Commun.*, **8**, 15178.
58. Tycko,J., Wainberg,M., Marinov,G.K., Ursu,O., Hess,G.T., Ego,B.K., Aradhana, Li,A., Truong,A., Trevino,A.E. *et al.* (2019) Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. *Nat. Commun.*, **10**, 4063.
59. The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
60. Rao,S.S.P., Huntley,M.H., Durand,N.C., Stamenova,E.K., Bochkov,I.D., Robinson,J.T., Sanborn,A.L., Machol,I., Omer,A.D., Lander,E.S. *et al.* (2015) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **162**, 687–688.