

## GENETICS

# Noncoding loci without epigenomic signals can be essential for maintaining global chromatin organization and cell viability

Bo Ding<sup>1†‡</sup>, Ying Liu<sup>2†</sup>, Zhiheng Liu<sup>2,3†</sup>, Lina Zheng<sup>4†</sup>, Ping Xu<sup>2</sup>, Zhao Chen<sup>1</sup>, Peiyao Wu<sup>1</sup>, Ying Zhao<sup>1</sup>, Qian Pan<sup>2</sup>, Yu Guo<sup>2</sup>, Wensheng Wei<sup>2\*</sup>, Wei Wang<sup>1,4,5\*</sup>

Most noncoding regions of the human genome do not harbor any annotated element and are even not marked with any epigenomic or protein binding signal. However, an overlooked aspect of their possible role in stabilizing 3D chromatin organization has not been extensively studied. To illuminate their structural importance, we started with the noncoding regions forming many 3D contacts (referred to as hubs) and performed a CRISPR library screening to identify dozens of hubs essential for cell viability. Hi-C and single-cell transcriptomic analyses showed that their deletion could significantly alter chromatin organization and affect the expressions of distal genes. This study revealed the 3D structural importance of noncoding loci that are not associated with any functional element, providing a previously unknown mechanistic understanding of disease-associated genetic variations (GVs). Furthermore, our analyses also suggest a possible approach to develop therapeutics targeting disease-specific noncoding regions that are critical for disease cell survival.

## INTRODUCTION

Noncoding sequences of the human genome, such as noncoding RNAs (ncRNAs), enhancers, and transposons, are known to be critical for many biological processes and are thus functionally important. Despite the great progress in uncovering new roles of these noncoding elements, most of the human genome remains unannotated. As the three-dimensional (3D) organization of the genome is essential for regulating transcription and other cellular functions (1–6), an overlooked aspect of noncoding sequences is their “structural importance” in forming and maintaining the proper 3D chromatin structure, particularly for those that are not marked by any epigenetic signal or annotated with any functional unit.

In protein function analysis, some residues could be important, if they are essential for maintaining the proper conformation (7), even though they may not be directly involved in the protein’s enzymatic activity or interaction with ligands. Similarly, noncoding genomic sequences could play critical roles in stabilizing the proper chromatin structure, although they do not harbor any enhancer or transcription factor (TF) binding site. Previous studies have shown that changing the noncoding sequences could alter chromatin organization; for instance, deletion of some boundary sequences of topologically associating domains (TADs) (1, 2) causes aberrant gene transcription, leading to disease (3). TAD boundaries can be considered a special

case, but the structural importance of noncoding sequences, particularly those not associated with TADs or any functional elements, has not been fully investigated.

Deleting a noncoding sequence and examining a phenotypic readout such as cell viability can directly assess its importance. High-throughput genetic screening by the CRISPR-Cas9 system has been effectively applied to analyzing long ncRNAs (lncRNAs) (8, 9), enhancers, and promoters (10–12). However, it is still prohibitive to delete each 5-kb segment in the genome for thorough screening, and random selection of deletion loci is inefficient. For example, less than 3% of lncRNAs were reported to be essential for cell growth and survival (8, 9), and this percentage is expected to be even lower for unannotated noncoding loci. A reasonable strategy is to start with the genomic loci involved in many chromatin contacts, hereinafter referred to as hubs, because disrupting these hubs would potentially lead to a relatively profound perturbation to the chromatin organization.

Here, we performed network analysis on Hi-C 3D contact data and identified a group of loci as hubs. Through a high-throughput CRISPR-Cas9 library screening by targeted deletion, we found that some hubs without any epigenetic marks were essential for cell growth and survival. We examined the impacts of hub deletion on the global chromatin structure and gene expression using Hi-C and single-cell RNA sequencing (scRNA-seq) technologies.

## RESULTS

We first downloaded the 5-kb resolution Hi-C data in seven human cell lines [GM12878, human mammary epithelial (HMEC), human umbilical vein endothelial (HUVEC), IMR90, normal human epidermal keratinocytes (NHEK), K562, and KBM7] (13) and identified significant intrachromosomal contact pairs ( $P$  value cutoff of  $e^{-20}$ ; see Materials and Methods). We next assembled all the contacts in a chromosome for a certain cell line into a network, which is hereinafter referred to as the fragment contact network (FCN). In the FCN, each node is a 5-kb fragment, and each edge represents a 3D contact. The

<sup>1</sup>Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093-0359, USA. <sup>2</sup>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. <sup>3</sup>Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China. <sup>4</sup>Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, CA 92093-0359, USA. <sup>5</sup>Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA 92093-0359, USA.

\*Corresponding author. Email: wswei@pku.edu.cn (W.We.); wei-wang@ucsd.edu (W.Wa.)

†These authors contributed equally to this work.

‡Present address: Thermo Fisher Scientific, 5781 Van Allen Way, Carlsbad, CA 92008, USA.

degree of a node reflects how many contacts it forms. We calculated the  $z$  score of each node's degree as  $z \text{ score} = \frac{d_i - \mu}{\sigma}$ , where  $d_i$  is the degree of the  $i$ th fragment and  $\mu$  and  $\sigma$  are the mean and SD of the degrees of all nodes in a chromosome of a cell line. The nodes with a  $z$  score  $\geq 2.0$  were considered "hubs," whereas the rest of the nodes were considered "nonhubs" [see Materials and Methods, the "FCN Network Analysis Results" section in the Supplementary Materials, table S2, and (14)]. The hubs count for less than 10% of the total nodes in a given FCN.

Note that these contacts indicate the spatial closeness of the contacting loci, and they are not necessarily mediated by proteins or ncRNAs to form specific chromatin loops. An analogy is the core residues of a protein, which are located in the interior and form many contacts with other residues but do not necessarily have specific residue-residue interactions mediated by such as hydrogen bonds and electrostatic interactions; however, deleting these residues can disrupt the packing of the interior residues and thus distort the proper conformation required for the protein's normal function. Similarly, perturbing a hub may have the same impacts on the 3D genome structure by disrupting chromatin organization.

To illustrate the importance of the hubs, we first investigated their contribution to stabilizing the FCN and their association with genetic variations (GVs; in this study, we focused on single nucleotide variations hereinafter) in cancer. Then, we identified hubs essential for cell viability using CRISPR screening. Lastly, we illustrated the impact of hub deletion on chromatin structure and gene expression using Hi-C and scRNA-seq.

### FCN networks are resistant to random attacks but vulnerable to targeted attacks

In this study, we focused on intrachromosomal contacts and constructed FCNs for each chromosome in each cell line, resulting in a total of 161 ( $= 23 \times 7$ ) FCNs for all chromosomes in the seven cell lines. We found that the degree distribution of FCN follows a power law (Fig. 1A), indicating that FCNs are scale-free networks. FCNs are resistant to random attacks (random removal of nodes in the network) but vulnerable to targeted attacks (targeted removal of specific nodes) against high-degree nodes, as scale-free networks (15). The 161 FCNs have similar network parameters, such as effective diameters, which is the path length such that 90% of node pairs are at a smaller or equal distance apart (see the "FCN Network Analysis Results" section in the Supplementary Materials). The most significant outlier was the FCN of chr9 in the leukemia cancer cell line K562, which had a significantly larger effective diameter than the rest (Fig. 1B and fig. S1A). We also calculated the diameter by considering the translocation between chr9 and chr22 (Philadelphia translocation), and it was still significantly different from other chromosomes. We found that computationally removing high-degree nodes from chr9 of GM12878 normal cells led to a similar degree distribution of chr9 in K562 cancer cells, which suggests that the targeted perturbation shifted the FCN of a normal cell toward that of a cancer cell (Fig. 1C). This analysis suggests that GV in K562 cells likely target the high-degree nodes of chr9 and thus alter the network properties. We also confirmed that the high degree nodes (hubs) are crucial for stabilizing the contacts between their connecting nodes in the network (hereinafter defined as "neighbors") (figs. S1, C to F).

We next investigated the genomic and epigenomic signals in the identified hub regions in six cell lines (no epigenomic data for

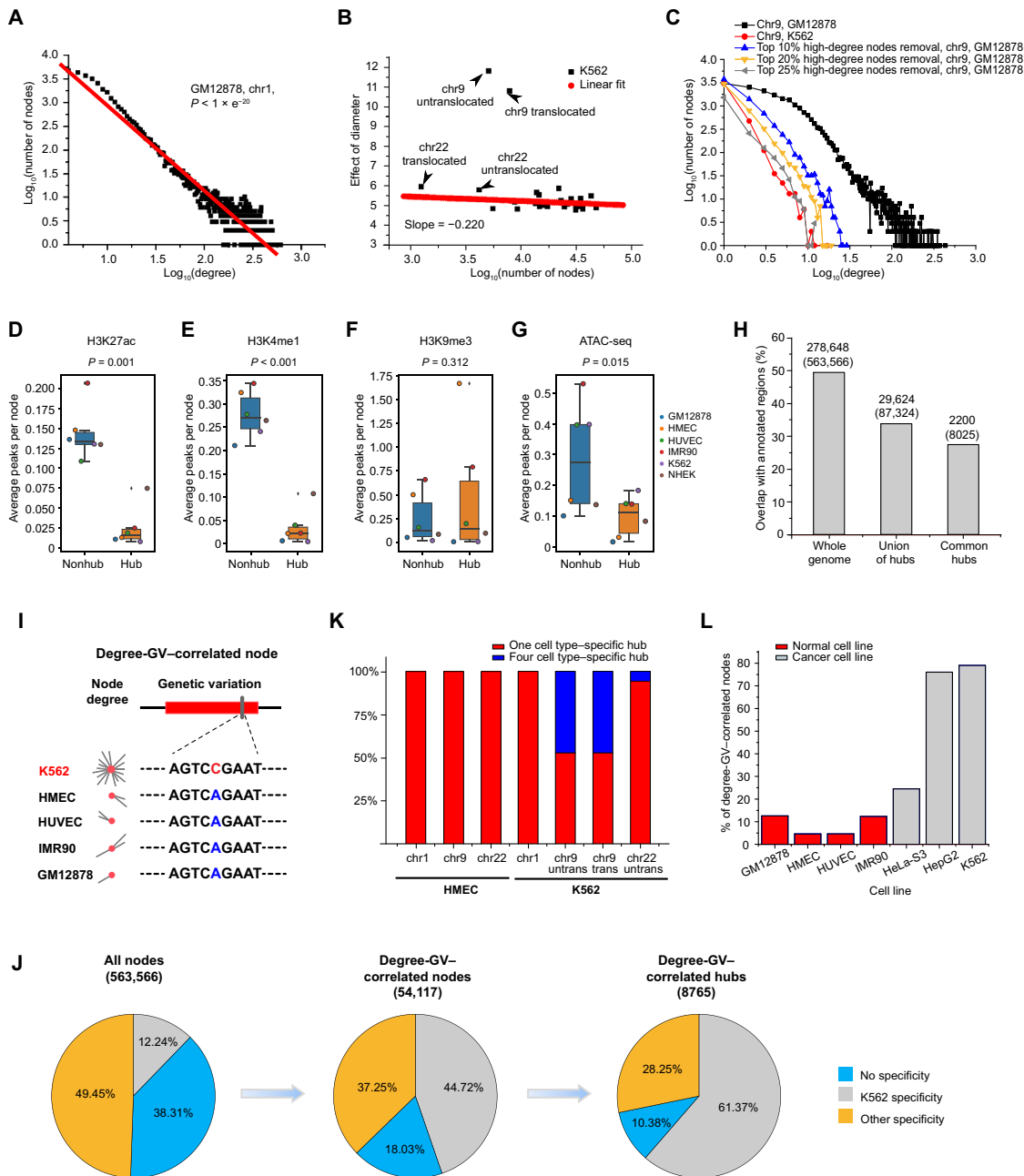
KBM7). Compared to the nonhub loci, hub loci had fewer peaks for five histone marks (H3K27ac, H3K27me3, H3K4me1, H3K4me3, and H3K36me3) and a comparable number of H3K9me3 peaks (Fig. 1, D to F; fig. S1, G to I; and the "FCN Network Analysis Results" section in the Supplementary Materials). We also observed less open chromatin (Fig. 1G) and fewer annotated regions (including coding genes, ncRNA, and other annotated regions downloaded from <https://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/refFlat.txt.gz>) in hubs than in nonhub regions (Fig. 1H) (see Materials and Methods and the "FCN Network Analysis Results" section in the Supplementary Materials). Furthermore, we compared the epigenetic marks (H3K27ac, H3K4me3, H3K4me1, H3K27me3, and H3K9me3) and assay for transposase accessible chromatin with high-throughput sequencing (ATAC-seq) peaks in the upstream and downstream of the hubs in multiple cell types. We considered different distances away from the hub regions ranging from 0 to 50 kb in linear distance (fig. S3). Comparing with the upstream and downstream regions, hubs also have lower H3K27ac, H3K27me3, H3K4me1, H3K4me3, H3K36me3, and open chromatin signals and comparable H3K9me3 peak numbers. We also identified A/B compartments for hub and nonhub loci and found that hubs are enriched in B compartments (fig. S1J), which is consistent with the histone modification analysis. These observations suggest that hubs are similar to the core residues in proteins, both densely packed in the interior of the 3D structure. Therefore, perturbation to hubs by GV and deletions could disrupt chromatin packing to affect the surrounding 3D organization of chromatin and propagate through the genome, leading to observable phenotypes such as disease formation and cell death.

Next, we examined whether the hubs found in normal cells have significantly different 3D contacts in cancers and whether these changes are associated with GV. Then, we investigated whether and how deleting hubs can cause cell death.

### Cancer-related mutations alter 3D hub contacts

As K562 is a cancer cell line, we investigated whether K562-specific GV are related to changes in spatial contacts. After calculating the  $z$  score for each node's degree so that it is comparable across cell lines, we checked its specificity, i.e., whether the contact degree was specifically high in any particular cell line (see Materials and Methods). When considering all the nodes, we did not observe any specificity bias toward K562 cells: In the 563,566 total nodes of the whole genome, 38.3% showed no specificity, 12.2% showed specificity in K562 cells, and the largest of the other specificities was 13.4% (Fig. 1J and table S4).

Next, we calculated the Pearson correlation coefficient between a node's degree and GV occurrence in the node across cell lines. We found that K562-specific GV were associated with the degree changes in K562 cells: Among all 54,117 nodes with degree-GV Pearson correlation coefficients  $> 0.9$  (referred to as degree-GV-correlated nodes; Fig. 1I), 24,229 (44.72%) were K562 specific, i.e., the GV is only observed in K562, and the degree of the node shows significantly higher or lower degree in K562 than the other cell lines; as a comparison, the largest percentage for another cell type (HMEC) specificity was only 10.6% (5743 nodes) (Fig. 1J and table S4). This bias toward the only diploid cancer cell line K562 among the seven was even more obvious for hubs: For all the hubs identified in at least one of the seven cell lines, there were 8765 degree-GV-correlated hubs, among which 5379 (61.37%) were K562-specific compared to the largest percentage of 824 (9.4%) specific to another cell type



**Fig. 1. Characterization of the FCNs and hub nodes.** (A) Degree distribution of FCN. (B) The effective diameter of FCN remains largely unchanged with increasing network size. “Translocated” (untranslocated): network constructed by considering (not considering) the translocation between chr9 and chr22. (C) The degree distribution of chr9 in normal cell lines (GM12878 as an example) after removal of high-degree nodes is similar to that of K562 chr9. (D to G) Epigenomic signals in hubs and nonhubs: H3K27ac (D), H3K4me1 (E), H3K9me3 (F), and ATAC-seq (G) in diverse cell lines. (H) The percentages of the annotated regions (including coding genes, ncRNA, and other annotated regions at <https://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/refFlat.txt.gz>) in the whole genome, union of hubs (hubs appeared in at least one cell line), and common hubs (hubs appeared in all cell lines). Numbers above the bar plot are the number of nodes overlapping with gene regions (top) and the number of nodes in that category (bottom in parentheses). (I) Definition of degree-GV-correlated nodes. The example node has a high degree in K562 and low degrees in others, which is correlated with the GV profile with a SNP in K562 but none in others. (J) The distribution of cell line specificities in all nodes, degree-GV-correlated nodes, and degree-GV-correlated hubs. (K) The distribution of one cell type-specific hub and four cell type-specific hubs in chromosomes and cell lines. (L) The percentage of degree-GV-correlated nodes in normal cell lines and cancer cell lines.

(HMEC) (Fig. 1J and table S4). Together, these analyses suggest that K562-specific GVs tend to significantly change the contact degrees, particularly on hubs, which is consistent with the observation that the FCN is vulnerable to targeted GVs in hubs.

GVs can either disrupt hubs in normal cells or form new disease-specific hubs in cancer cells. We thus analyzed hub formation and disruption separately and found a strong correlation between GV and contact degree change in K562 cells for both scenarios (see the

“FCN Network Analysis Results” section in the Supplementary Materials). In particular, the percentages of hub disruption in chr9 of K562 cells (i.e., hubs found in the other four cell types but not in K562 cells) were 47.56 and 47.50% without and with consideration of translocation between chr9 and chr22, respectively (only the untranslocated part of chr9 was used for calculation). This was significantly higher than all other chromosomes in each cell line, whose range was between 0 and 18.6% (Fig. 1K). Our analyses clearly show that the GVs in K562 cells severely disrupted the hubs on chr9 shared by other cell lines.

To confirm the generality of this observation, we extended our analysis to four normal cell lines (GM12878, HMEC, HUVEC, and IMR90) and three cancer cell lines (HepG2, HeLa-S3, and K562) that had both 20-kb resolution Hi-C and GV data. We also found a strong correlation between the degree and GV in cancers (Fig. 1L), suggesting that cancer-specific GVs tended to significantly alter the 3D contacts of hubs.

### Targeted deletion of hubs can significantly affect cell viability

The above analyses indicated that hubs are not necessarily directly involved in functional activities, but they can be crucial for stabilizing the chromatin structure and are thus functionally important. To further test this hypothesis, we selected 960 hub regions (each 5 kb in length) to examine their impacts on cell growth and survival in a high-throughput deletion screen (table S8) with the highest partner linking tendency (PLT) (see details in the “FCN Network Analysis Results” section in the Supplementary Materials). These hubs are those likely to stabilize the contacts between neighbors, including 683 hubs present in all cell lines and 277 hubs specific to K562 cells. They are evenly distributed along the chromosomes (fig. S4E).

For screening, we constructed a paired-guide RNA (pgRNA) library (16) targeting the selected hubs mediated by the CRISPR-Cas9 system. Using lentiviral transduction at a low multiplicity of infection (MOI) of <0.3, we transfected the pgRNA library containing a total of 17,476 pgRNAs into K562 cells stably expressing the Cas9 protein. This library also included 473 pgRNAs targeting essential ribosomal genes as positive controls, 100 pgRNAs targeting the *AAVS1* locus, and 100 nontargeting pgRNAs as negative controls (table S9). The library cells were cultured for 30 continuous days after transduction. We sequenced cells at day 0 (controls) and day 30 to determine the abundance of barcode-gRNA regions, which represent the corresponding pgRNAs (Fig. 2A).

Distributions of pgRNA reads from the control/experimental group between two biological replicates were highly correlated (figs. S4, A and B), and the scatter plot of each hub's mean fold change between replicates also showed a high correlation (Pearson correlation coefficient = 0.75) (fig. S4C). In the day 30 cell population, compared with nontargeting pgRNAs or those targeting *AAVS1*, we identified hub regions with significant depletion in their targeting pgRNAs, consistent with positive controls that target essential ribosomal genes. The fold changes of all pgRNAs targeting each hub were calculated, and their *P* values were computed by comparison with the *AAVS1*-targeting pgRNAs using the Mann-Whitney *U* test (8, 9), which is focused on analyzing screening data with the in-library controls and could more accurately reflect the fitness effect of each locus. *AAVS1*-targeting pgRNAs were randomly sampled to generate a distribution of negative controls, which was used to compute the hubs' *P* values. Combining the mean fold change and corrected *P* values, an  $I_{\text{score}}$  was computed for each hub. Eventually,

the hubs whose  $I_{\text{score}}$  was less than or equal to  $-1$  were considered essential hits (see the Supplementary Materials and tables S10 and S11). Overall, 77 hubs were selected in K562 cells whose deletion led to cell death or growth inhibition (Fig. 2B).

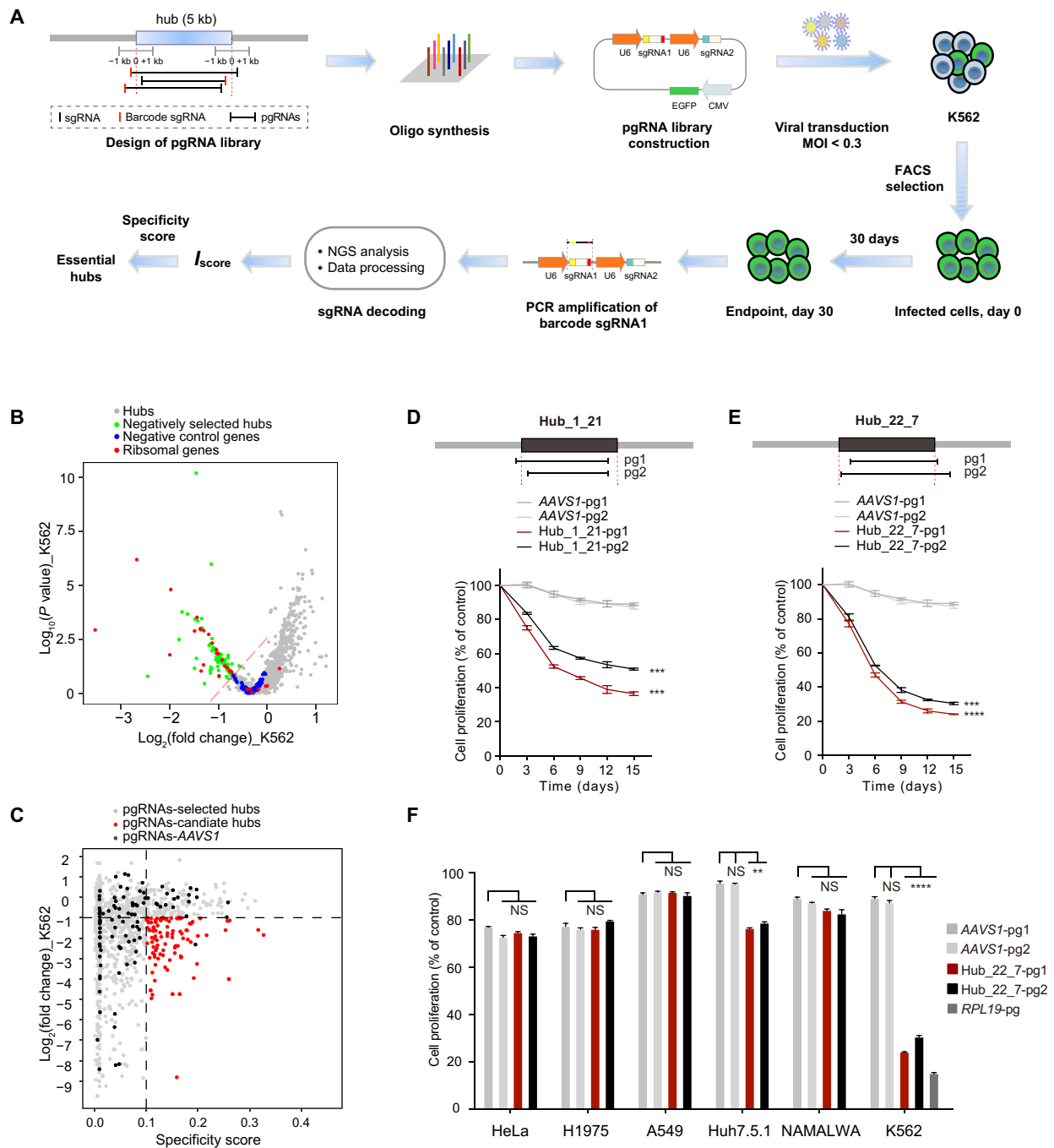
It has been reported that multiple cleavages in genomic loci generated by Cas9 activity could lead to cellular toxicity and thus affect growth screen measurements (17–20). To minimize the potential off-target effects, we calculated the GuideScan specificity score (21) for each single guide RNA (sgRNA) of every pgRNA, which focused on assessing the specificities of sgRNAs with two or three mismatches to off-target loci that are commonly used in library screens, and generated a specificity score for each pgRNA. We found that pgRNA targeting *AAVS1* with a specificity score  $\leq 0.1$  could lead to a significant dropout effect in K562 cells (fig. S4D). To further assure the target specificity, we selected only targeting pgRNAs with specificity scores  $> 0.1$  and  $\log_2$  (fold change) ( $\log_2\text{FC}$ )  $< -1$  for subsequent analysis (Fig. 2C). Furthermore, hub loci with copy number amplification were also filtered out to minimize the effect due to multiple cleavages by certain pgRNAs (22). Using these stringent criteria, we identified 35 essential hubs in K562 cells (Fig. 2C). We checked the location of essential hubs and found some of them located near the centromeres (fig. S4E), but they are not significantly closer to centromeres than the nonessential ones ( $P = 0.092$ , fig. S4F).

We then chose seven candidate hubs for individual validation in K562 cells. For each hub, two or three pgRNAs with high specificity scores were selected (see Materials and Methods and table S12). All but two identified hubs were validated to severely affect cell growth and proliferation in K562 cells (Fig. 2, D and E, and fig. S5), indicating their functional roles in cell fitness. To further explore the cell type specificity of the essential hubs, we selected hub\_22\_7 (chr22: 17,325,000 to 17,330,000, hg19), which showed the most significant growth defect in K562 cells, and performed the same cell proliferation assay in five other cancer cell lines. Compared with negative controls targeting the *AAVS1* locus, targeted deletion of hub\_22\_7 did not lead to significant cell death or cell growth inhibition in the following four tested cell lines: HeLa (cervical cancer cells), H1975 (non-small cell lung cancer cells), A549 (non-small cell lung cancer cells), and NAMALWA (Burkitt's lymphoma) (Fig. 2F and fig. S6A). In the liver cancer cell line Huh7.5.1, deletion of hub\_22\_7 showed a weak effect on cell fitness compared with deletion of the essential gene *RPL19* serving as the positive control (Fig. 2F and fig. S6A). Overall, only the hub\_22\_7 locus exhibited a remarkable essential role in the K562 cell line. These results validate the essential hubs identified in the screen.

### Cell death caused by hub deletion does not result from disruption of functional elements or off-target effects

To illuminate the mechanism of cell death induced by hub deletion, we first examined the functional annotation and epigenetic modifications in these regions. None of the essential hubs overlap with gene coding regions, ncRNA regions, or TAD boundaries. A total of 77.1% (27 of 35 including 3 of 5 individually validated hubs) of the essential hubs did not overlap with any histone modification or TF chromatin immunoprecipitation sequencing (ChIP-seq) peak (Fig. 3A, an example of hub\_22\_7 in Fig. 3B and full genomics and epigenomics signals for hub\_22\_7 in fig. S7). We also checked the ChromHMM states (the 18-state data downloaded from the Roadmap Epigenomics project ([https://egg2.wustl.edu/roadmap/web\\_portal/chr\\_state\\_learning.html#exp\\_18state](https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#exp_18state))) in the K562 essential

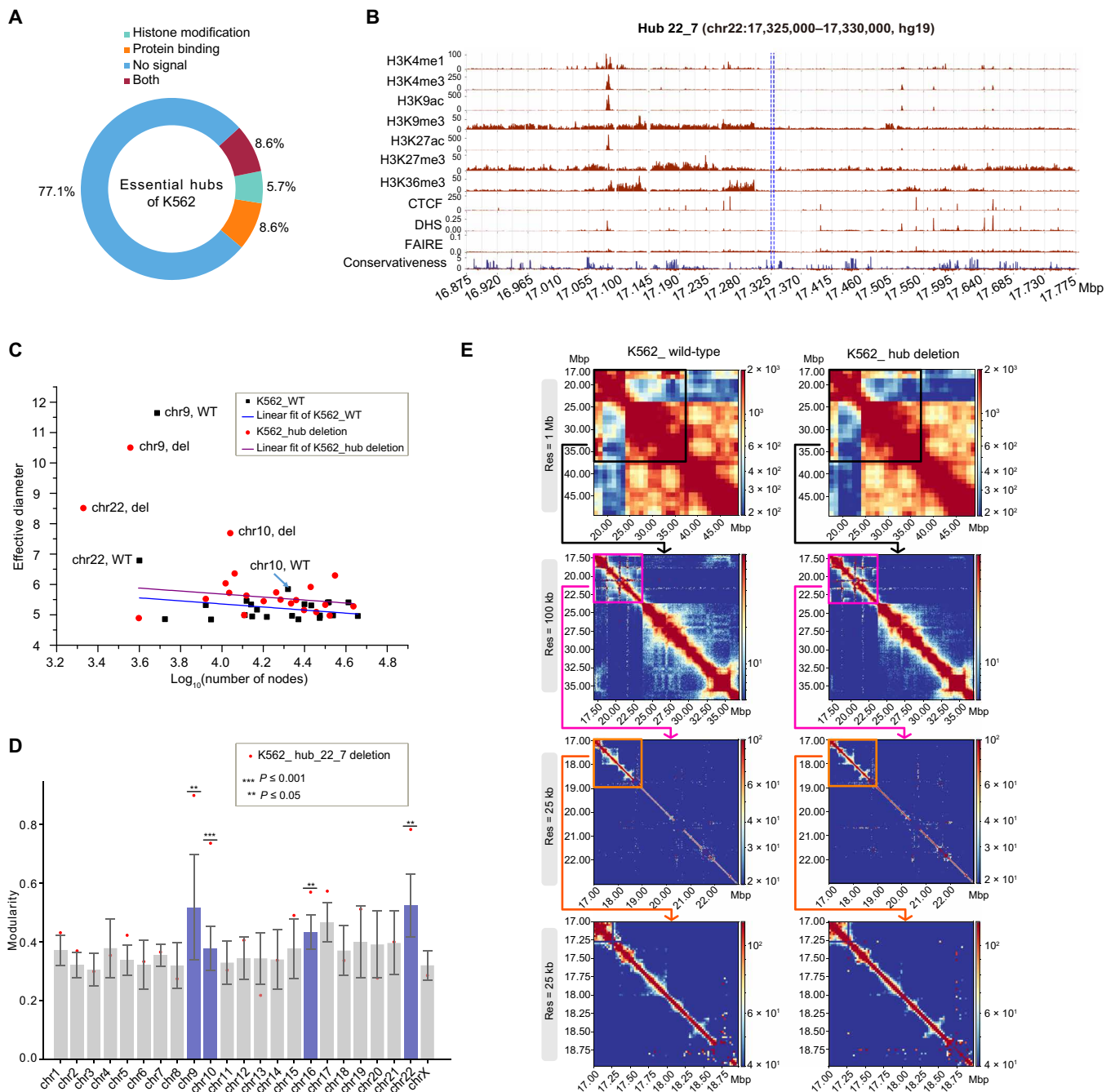




**Fig. 2. Identification of essential hubs for cell growth and proliferation in the K562 cell line through pgRNA-mediated fragment deletion.** (A) Schematic of the pgRNA library design, cloning, and functional screening of selected hub loci. CMV, cytomegalovirus. (B) Volcano plot of the fold change and P value of hubs in the K562 cell line. The dotted red line represents  $I_{score} = -1$ . (C) Selection of candidate essential hubs by pgRNA fold change and specificity score. Essential hits were selected with specificity score > 0.1,  $\log_2$  (fold change) ( $\log_2FC$ ) < -1. (D and E) Validation of top-ranked essential hubs in K562 by cell proliferation assay. AAVS1-pg1 and AAVS1-pg2 are pgRNAs targeting AAVS1 as negative controls. Asterisk (\*) represents P values compared with AAVS1-pg1 at day 15, calculated by two-tailed Student's t test, and adjusted by Benjamini-Hochberg procedure. (F) Validation of hub\_22\_7 in multiple cancer cell lines, including A549, H1975, HeLa, Huh7.5.1, and NAMALWA. Asterisk (\*) represents P values compared with AAVS1-pg1 at day 15, calculated by two-tailed Student's t test, and adjusted by Bonferroni correction accounting for multiple testings. Data are presented as the means  $\pm$  SD. (n = 3). \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , and \*\*\*\* $P < 0.0001$ . NS, not significant.

hubs and found that 82.286% of them are in the quiescent/low states (table S13). These observations indicated that the essentiality of these hubs did not result from the genes or regulatory elements they harbor.

We next evaluated the essential hub\_22\_7 to rule out the possibility that cell death was caused by off-target cleavage. Using the validated pgRNA hub\_22\_7-pg2 with high specificity (table S12), we first measured its deletion efficiency by real-time quantitative



**Fig. 3. Characterization of the deleted hub and the impact of its deletion on the global chromatin structure.** (A) Overlap of essential hubs in K562 cells with the peaks of 10 histone marks (H3K27ac, H3K4Me1, H3K4me3, H3K27me3, H3K9me3, H3K36me3, H3K4me2, H3K79me2, H3K9ac, and H3K20me1) and 151 TFs. (B) Histone marks, CTCF (CCCTC-binding factor) binding, open chromatin, DHS, DNase hypersensitivity; FAIRE, formaldehyde-assisted isolation of regulatory elements, and conservation score (100 vertebrates basewise conservation by PhyloP) on chr22: 17,325,000 to 17,330,000 (see fig. S6 for all signals). (C) Effective diameters versus log<sub>10</sub>(number of nodes) the wild-type (WT) and hub\_22\_7-deleted K562 cells. (D) Modularity scores in the seven wild-type cell lines for 23 chromosomes. Red dots, hub\_22\_7 deletion. (E) Hi-C contacts for wild-type and hub\_22\_7-deleted K562 cells at 1-Mb, 100-kb, and 25-kb resolutions.

polymerase chain reaction (PCR) (fig. S6B) at each time point after pgRNA transduction and then performed whole-genome sequencing (WGS) to evaluate its potential off-target effect on the day showing the highest deletion efficiency (see Materials and Methods). We identified >3.7 million single-nucleotide variants (SNVs) and >890,000 indels compared to the hg19 reference genome (table S14). The fact

that we could successfully identify 87.4% germline mutations found in the published wild-type K562 cells (ENCODE database with the accession codes ENCF0313MGL, ENCF004THU, ENCF0506TKC, and ENCF066GQD) suggests reliable library quality. We manually checked the indels on 455 potential off-target loci and 2 on-target loci identified by Cas-OFFinder (23) using loose criteria (bulge = 0,

mismatch  $\leq 4$ ; bulge  $\leq 2$ , mismatch  $\leq 2$ ) to avoid missing any possible off-target site (see Materials and Methods). Significant indels were found in only 2 on-target loci and not found in any of the 455 putative off-target loci, indicating no off-target cleavage. These analyses confirmed that cell death caused by hub deletion did not result from off-target effects.

### Deletion of essential hubs can alter the global chromatin structure

We next performed Hi-C analysis to examine the chromatin structure changes in hub\_22\_7-pg2-infected (hub\_22\_7-deleted) K562 cells. To characterize the global impact of hub deletion, we first constructed FCNs in the hub\_22\_7-deleted cells using the same criteria as in the wild-type cells and analyzed the changes in the network properties, including effective diameter and modularity, which is the difference between the fraction of edges observed within a group of nodes and the expected value in a random network.

By analyzing the effective diameters of the FCNs before and after the hub deletion, we found that chr9, chr10, and chr22 had significant changes ( $P < 0.05$ ; see Materials and Methods and Fig. 3C): chr22 and chr10 increased, while chr9 decreased upon hub deletion. The hub-deleted cells also showed significant changes in the modularity scores of chr9, chr10, chr16, and chr22 ( $P < 0.05$ ; see Materials and Methods and Fig. 3D). While the change in the hubs residing in chr22 and chr22-translocated chr9 in K562 cells was not unexpected, the unexpected impact on chr10 and chr16 illuminated the importance of the understudied interactions between chromosomes (fig. S8).

The increased diameter and modularity in chr22 suggest that hub deletion reduces long-range chromatin contacts and enhances modularization of the FCN, consistent with the overall Hi-C contact difference between the wild-type and hub-deleted cells (Fig. 3E). We did find newly formed and disrupted chromatin loops (examples in Fig. 3E) and merge or split of a small percent of TADs in the hub-deleted cells (examples of chr22: 24 to 26 Mb, 35 to 38 Mb, and 45 to 47 Mb in fig. S9). Together, deletion of a hub has a global impact on chromatin structure that can propagate to other chromosomes.

### Deletion of essential hubs can up-regulate apoptotic genes

Next, we set out to identify genes whose expression was significantly affected by hub deletion. Cells transduced with pgRNAs have various rates toward cell death, and the cell population is thus heterogeneous. Therefore, we used single-cell analysis to define the different cell states in the population. We performed Drop-seq analysis (24) on hub\_22\_7-pg2-infected K562 cells and collected scRNA-seq data for 393 cells passing the quality control criteria. The bulk RNA-seq data of the wild-type and *AAVS1*-deletion K562 cells were included as controls. All the RNA-seq data were normalized using counts per million (CPM), and the scaled  $z$  score for each gene in each individual cell or bulk sample was calculated by fitting a binomial distribution (see Materials and Methods). The scaled  $z$  score matrix of single-cell and bulk RNA-seq data was used for the following analysis.

We performed trajectory branching and pseudotime analysis using Monocle (25, 26). Given that cell viability was significantly affected upon hub\_22\_7 deletion, we analyzed 93 apoptosis genes documented in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database ([www.genome.jp/kegg/](http://www.genome.jp/kegg/)). The single cells together with bulk samples of the wild-type and *AAVS1*-deletion K562 cells were grouped into five cell states (Fig. 4A). Both *AAVS1*-deletion and wild-type samples were assigned to state 1, indicating that *AAVS1* deletion is a

valid control. Single cells in state 1 resemble the wild-type cells at the low value of pseudotime, which is understandable because hub deletions were not synchronized in all cells. States 4 and 5 have the highest pseudotime values and thus are the most distinct from the wild-type state. Overall, the apoptosis genes showed increasing expression levels from state 2 to state 5 (see examples in Fig. 4B). Because states 2, 4, and 5 are the leaf nodes in the trajectory tree that represent local minimum or maximum points, we clustered the apoptosis genes according to their expression profiles in these three states. Each of the gene clusters presented with unique patterns as they progressed toward cell apoptosis (Fig. 4C). This scRNA-seq analysis depicted the transcriptomic progression toward cell death upon hub deletion in K562 cells.

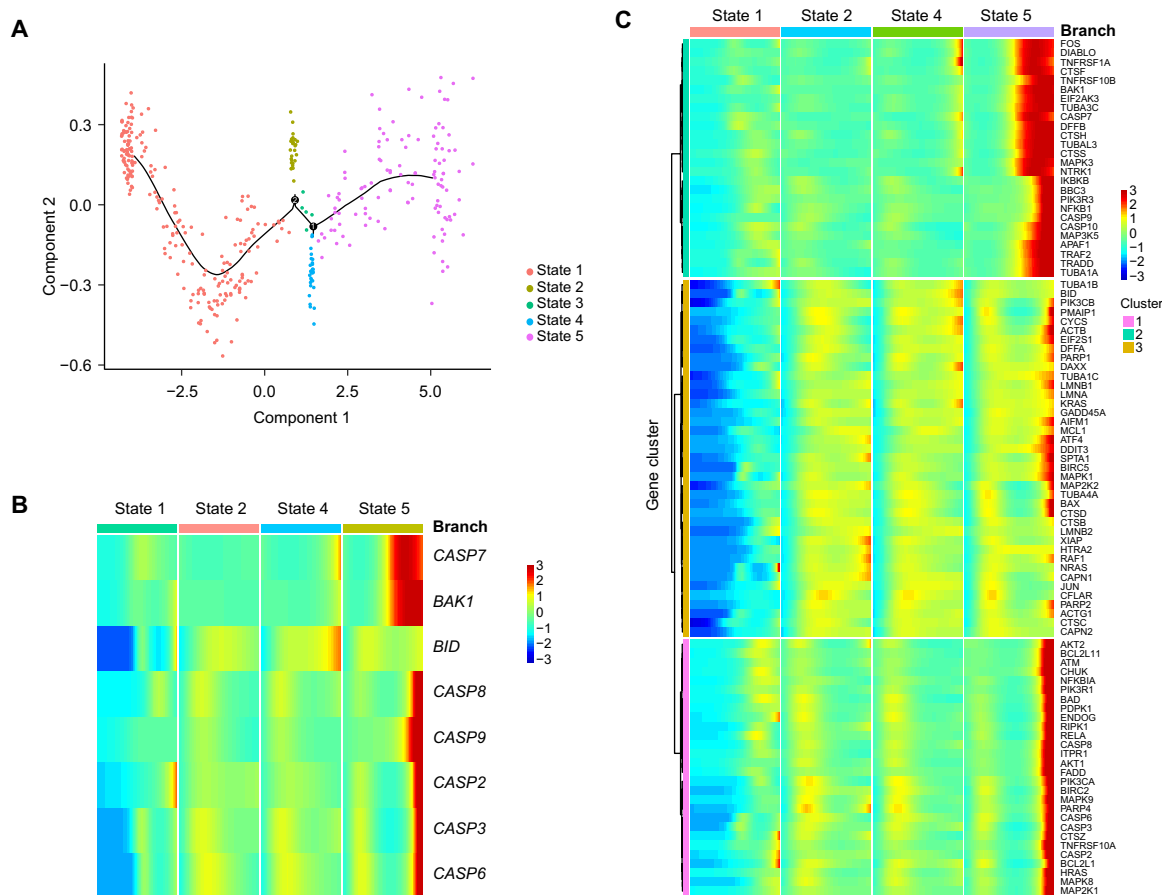
### Deletion of essential hubs can alter gene expression in distal regions

We noticed that multiple contacts between promoters and enhancers located at the opposite sides of the hub in the linear genome were disrupted upon hub deletion (Fig. 5A and fig. S10), indicating that deleting a hub could affect transcriptional regulation. To investigate whether important genes in chr22 (27, 28) were affected, we compared the expression profiles of state 2 and state 1 and found significantly down-regulated genes upon hub deletion, including multiple essential genes whose gene knockdown would significantly affect the K562 cell viability [identified from previous genome-wide CRISPRi screening (27)], such as *ATXN10*, *THOC5*, *CHEK2*, and *HSCB* (Fig. 5B). Notably, these genes are located distal (12 to 34 Mb away in the linear genome) from the deleted hub\_22\_7 loci. We confirmed the essentiality of *THOC5* in K562 cells through CRISPRi-based gene knockdown (Fig. 5B). Furthermore, the high-resolution Hi-C data indicated that its promoter's interaction with enhancers was disrupted upon hub deletion (Fig. 5C). They are located in compartment A (active) to compartment B (inactive) flip region (chr22: 29 to 32 Mb, Fig. 5C), consistent with *THOC5* repression. These observations suggest that the chromatin structure alteration induced by hub deletion could affect the expression of distal genes, including those essential for cell viability.

### The global impact of hub deletion suggests that hubs might be potential noncoding therapeutic targets

Given that deleting one essential hub can affect many genes, a new "one-drug-multiple-targets" therapeutic strategy may be developed to synergize different pathways. Namely, disease-specific noncoding regions, such as hubs that are essential in only cancer cells, could be potential therapeutic targets. In our screen, we identified a group of essential hubs specifically for K562 cells (Fig. 2, E and F, and fig. S6). The deletion of hub\_22\_7 resulted in an approximately 80% decrease in the cell proliferation rate of K562 cells but nearly no significant effects on the other analyzed cell lines (Fig. 2F and fig. S6A). As K562 cell is a leukemia cancer cell line, such K562-specific hubs could be potential therapeutic targets for chronic myelogenous leukemia. As shown above, deletion of this hub caused the down-regulation of many essential genes and the activation of apoptosis pathways. Therefore, this collective effect of killing cancer cells is more potent than targeting each individual pathway and would make it more difficult for cancer cells to develop drug resistance.

Furthermore, hub deletion also affected genes specifically expressed in K562 cells, although they are not essential for cell viability. For example, K562 cell-specific high expression of *TOP3B* (fig. S11),



**Fig. 4. Hub deletion induces global changes in gene expression.** (A) Pseudotime clusters of hub\_22\_7-deleted and wild-type K562 cells based on apoptosis gene expression. (B) Examples of typical apoptosis gene expressions (*CASP2*, *CASP3*, *CASP6*, *CASP7*, *CASP8*, *CASP9*, *BAK1*, and *BID*) under different cell states defined by pseudotime analysis. (C) Global analysis of 93 KEGG apoptosis gene expression levels in states 1, 2, 4, and 5. Genes were clustered into three groups.

which plays important roles in the maintenance of gene stabilities and chromosome bridging (28, 29), was down-regulated upon hub deletion due to the disruption of its promoter-enhancer interactions. By examining the ENCODE data in 23 cell lines/tissues, we found that the enhancers located at chr22: 17,125,000 to 17,130,000 were marked by H3K27ac in only K562 cells and another leukemia cell line, Dnd41 (fig. S11). The low expression of *TOP3B* in Dnd41 cells (fig. S11) (22) suggests that these enhancers may regulate only *TOP3B* in K562 cells. Therefore, deleting this hub can specifically down-regulate *TOP3B* in K562 cells.

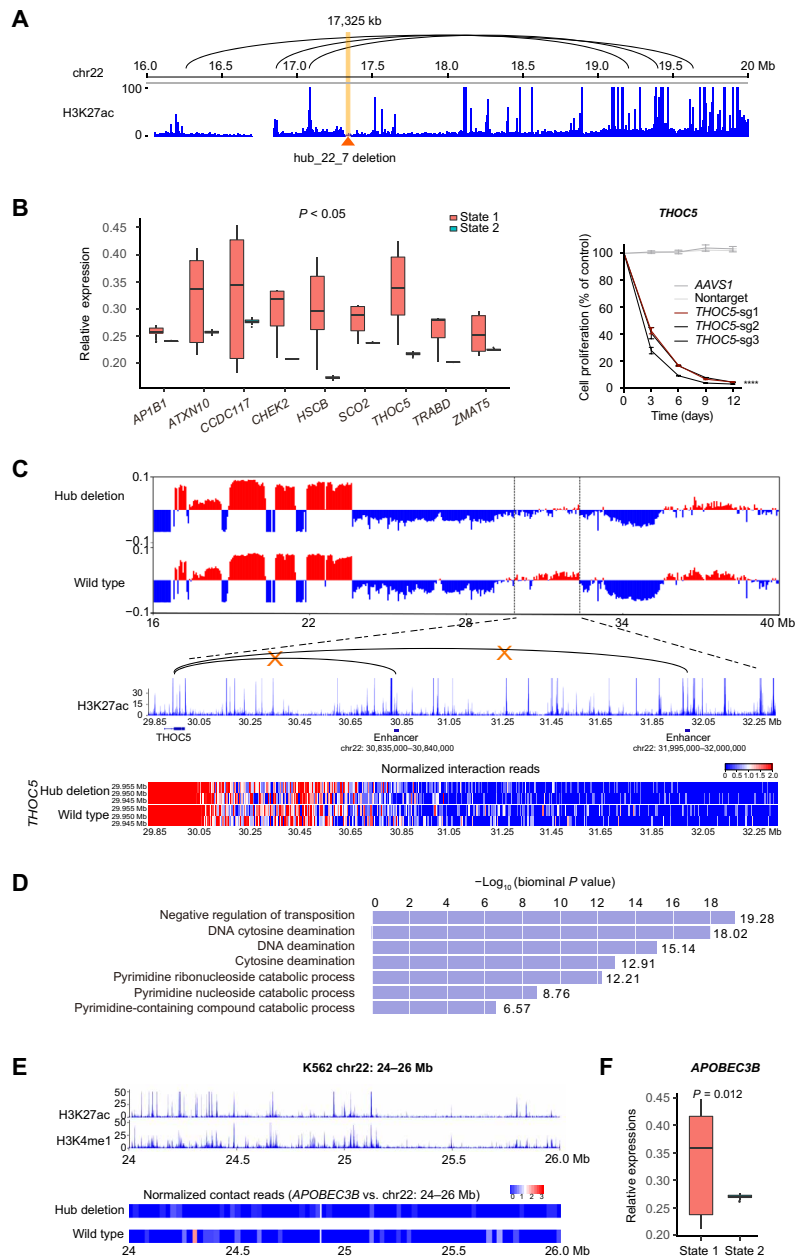
We also used Genomic Regions Enrichment of Annotations Tool (GREAT) (30) to search for pathways enriched (binomial false discovery rate  $Q \leq 1 \times e^{-5}$ ) in the loci whose Hi-C contacts ( $P \leq 10^{-20}$ ) were significantly reduced upon hub\_22\_7 deletion in chr22 (Fig. 5D). Notably, the *APOBEC3* family genes stood out, and in particular, *APOBEC3B* was significantly down-regulated from state 1 to state 2 (Fig. 5F). This is likely due to the reduced interaction between the *APOBEC3B* promoter and its enhancers upon hub\_22\_7 deletion (Fig. 5E). *APOBEC3* enzymes were reported as therapeutic targets for cancer treatment (31, 32), and their aberrant expression (e.g., higher expression of *APOBEC3B*) could cause cancerous mutagenesis leading to drug resistance or metastasis (33–35). Although *APOBEC3B* is not essential for K562 cell viability, its down-regulation

could effectively reduce the mutation rate, which is crucial for developing a potent therapy. Together, deleting one hub may synergize with multiple pathways to kill cancer cells and simultaneously reduce the cancer's mutation capability. This example suggests that the identification and deletion of cancer-specific hubs could open a new avenue for developing potent therapeutics.

## DISCUSSION

Noncoding genomic regions without any epigenetic mark, open chromatin, or TF binding have been overlooked in functional analysis. By analyzing the 3D contact networks derived from Hi-C data, we found that such noncoding regions without any mark can be in contact with many other loci and thus become hubs in the 3D contact network. Our simulated deletion of hubs in normal GM12878 cells shifted the 3D contact network toward the K562 cancer cell line. Our analysis also showed a strong correlation between 3D contact change and GV occurrence in the hubs of cancer cell lines, suggesting that cancer-specific GVs tend to significantly alter the 3D contacts of hubs. These results indicate that hubs likely play critical roles in normal cells, and noncoding disease-associated GVs can occur in hub regions to form or disrupt hubs in normal cells, which may cause aberrant cellular functions leading to diseases. Therefore,





**Fig. 5. The concurrent alterations of 3D chromatin structure and gene expressions after hub deletion.** (A) Disruption of enhancer-promoter interactions upon hub<sub>22\_7</sub> deletion. (B) Essential genes of K562 cells located on chr22 with significantly down-regulated expression ( $P < 0.05$ ) in state 2 compared to state 1. (C) A/B compartment change (50-kb resolution) upon hub deletion. Multiple enhancer-promoter contacts with *THOC5* were disrupted in the compartment changing region (chr22: 29,850,000 to 32,350,000). (D) Gene Ontology biological process pathways associated with loci whose 3D contacts were disrupted by hub deletion. (E) The 3D contacts between the *APOBEC3B* promoter and enhancers located on chr22: 24,000,000 to 26,000,000 were significantly decreased in hub-deleted cells. The enhancers were identified using the overlapping peaks of H3K27ac and H3K4me1 in the wild-type cells. (F) The relative expression level of *APOBEC3B* from state 1 to state 2.

our analysis provides a new perspective to understand the mechanisms of noncoding GV that do not overlap with any epigenetic mark, TF binding, or open chromatin but are tightly associated with diseases.

To further examine the importance of the hub regions, we deleted 960 hubs in K562 cells using a pgRNA CRISPR-Cas9 library. Through computational analysis combined with the in-library *AAVS1* controls and stringent filtering to avoid the potential issues of off-target effects and copy number amplifications, we found that 35 hubs could affect cell growth or viability after targeted deletion. The percentage

of hubs essential for cell fitness is comparable to those of essential lncRNAs (<3%) (8, 9) and protein-coding genes (< 3%) (36), which further supports the importance of hubs. Five of seven loci were individually validated with multiple pgRNAs, and hub<sub>22\_7</sub> was further validated to be specifically essential for cell fitness in the K562 cell line. Using WGS analysis, we also confirmed that the targeting pgRNA of hub<sub>22\_7</sub> has no off-target effect across the genome.

To understand the impact of hub deletion, we focused on 5 validated hub hub<sub>22\_7</sub> that has no epigenetic mark, TF binding,

or open chromatin signal in K562 cells. This hub was randomly selected from the K562 essential hubs and could serve as a representative group of cell type-specific essential hubs. Hi-C analysis showed that deleting the 5-kb hub significantly altered the 3D contact networks, as quantified by the significant change in FCN properties, including diameter and modularity. The hub deletion effects were far beyond the contacting loci of the hub and indicate that the impact of hub deletion is global.

We speculate that this global impact may start from the disruption of chromatin packing around the deleted hub and propagate to affect distal chromatin looping and promoter-enhancer interactions. An analogy is mutation of a residue in the interior of a protein's structure that can significantly change the protein conformation, leading to protein dysfunction. Therefore, although hubs do not host or interact with any gene, the propagated effect can alter the transcription of distal genes, as shown by the scRNA-seq data, which are essential for cell viability by themselves or in combination with other affected nonessential genes. We recognize that it is difficult to prove the causal relationship between global chromatin organization change and cell proliferation or gene expression, which still remains technically challenging and worthy of future investigation. Nevertheless, this is the first study to observe that noncoding loci without any epigenetic signals are not junk DNA, which could contribute to maintaining the global chromatin structure.

Furthermore, we showed that hubs can be cancer specific, which indicates a possibility of developing treatments to target a specific cancer. We are aware that the present studies are in cell lines and that further analysis in tumor tissues is necessary to confirm the translational value. However, it is worth noting that, because the global impact of hub deletion can affect many genes located distal from each other in the genome, the identified cancer-specific hubs could be potential new therapeutic targets. Targeting these noncoding loci could leverage the synergistic effects of multiple mechanisms to develop potent therapeutics, and treatment resistance is harder to develop because it requires mutations to interfere with the large number of genes affected by hub inhibition. There is a long way to go to translate this discovery, and there are possible roadblocks such as targeting multiple genes/pathways that may lead to lack of specificity for developing new therapeutics. As there are much more noncoding loci than the genes, overcoming the potential pitfalls requires additional effort to better understand the mechanisms of these "dark matter" in the genome for treating disease. Our findings here suggest an exciting direction for further exploration given the fast advancement of genome editing and delivery technologies.

Together, we report here the first study to reveal that noncoding loci without any epigenetic mark, TF binding, or open chromatin signal can be essential for cell viability. The importance of these loci for global chromatin organization and their impact on distal gene expression upon deletion make them a potential new class of therapeutic targets that have not yet been found.

## MATERIALS AND METHODS

### Network construction and hub identification

#### Evaluating the significance of Hi-C interaction pairs

We collected the raw reads, scale factors for vanilla coverage (VC) normalization, and the expected normalized reads for interaction pairs from the Hi-C experiments provided by Rao *et al.* (13)(GSE63525). The raw read,  $R_{ij}$ , between fragments  $F_i$  and  $F_j$  was first divided by

both sequence distances between  $F_i$  and  $F_j$  and obtained the expected normalized reads for the scale factors  $S_{F_i}$  and  $S_{F_j}$  for VC normalization,  $R_{ij}^{\text{norm}} = \frac{R_{ij}}{S_{F_i} S_{F_j}}$ . Then, we calculated the distance  $R_{ij}^{\text{exp}}$ . Last, the significance of the interaction between  $F_i$  and  $F_j$  was evaluated using the  $P$  value of the normalized read  $R_{ij}^{\text{norm}}$  calculated on the basis of a Poisson distribution (4) with an expectation equal to  $R_{ij}^{\text{exp}}$ .

#### Consideration of translocation in K562 cells

When processing the Hi-C reads in K562 cells, we took the reciprocal translocation between chr9 (9q34) and chr22 (22q11) into account. We mapped the reads to chr9 and chr22 in the reference genome and then translocated them. Next, the reads were normalized to the scale factors for each fragment provided by the study from Rao *et al.* (13). The  $P$  value was then calculated in the same way as described above. The translocated chr9 is ~10 Mbp longer than the reference chr9. As the expected reads and the genomic distance follow a power law (4), we fitted a linear model between the logarithm of expected reads and the logarithm of genomic distance and estimated the expected reads for longer genomic distances.

#### Hub identification

We identified hubs in each FCN using a  $z$  score of its degree,  $z \text{ score} = \frac{d_i - \mu}{\sigma}$ , where  $d_i$  is the degree of the  $i$ th fragment and  $\mu$  and  $\sigma$  are the average and SD of the degrees of all nodes in a chromosome of a cell line. We used a  $z$  score cutoff of 2.0 to select hubs that accounted for less than 10% of the total nodes (see the "FCN Network Analysis Results" section in the Supplementary Materials and table S2).

#### Epigenetic signal/gene enrichment in hubs and nonhubs

The peaks of six histone modifications (H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K27me3, and H3K9me3) and ATAC-seq peaks were counted in the hub/nonhub regions in the six cell lines (GM12878, HMEC, HUVEC, IMR90, NHEK, and K562). They were downloaded from [www.encodeproject.org/](http://www.encodeproject.org/). The KBM7 cell line was not included in the analysis because of the lack of enough histone modification ChIP-seq data. Distributions of the overlapping histone modification and ATAC-seq peaks were compared between hubs and nonhubs, and  $P$  values were calculated using matched-pairs  $t$  test.

To check the gene enrichment in the hub region and the entire genome, the annotated genes in hg19 genome downloaded from the UCSC genome browser overlapped with the whole genome (all 563,566 5-kb fragments covered in the Hi-C data in the entire genome), union of the hubs (union of all 87,324 hubs in the seven cell lines), and common hubs (8025 common hubs were found in the seven cell lines).

#### Cell line specificity of the node degree distribution

For 5-kb resolution Hi-C data in the five cell lines, we used a correlation-based method to evaluate cell type specificity. (i) The degree of each node was represented as a vector containing the degree  $z$  score values calculated for the five cell lines that had both GV and Hi-C data (GM12878, HMEC, HUVEC, IMR90, and K562). (ii) For cell type specificities, there are  $2^5 = 32$  possible vectors, including 2 with no cell line specificity (0,0,0,0,0), (1,1,1,1,1); 5 specific to one cell line (1,0,0,0,0), (0,1,0,0,0)...(0,0,0,0,1); 10 specific to two cell lines (1,1,0,0,0), (1,0,1,0,0)...(0,0,0,1,1); 10 specific to three cell lines (1,1,1,0,0), (1,0,1,1,0)...(0,0,1,1,1); and 5 specific to four cell lines (1,1,1,1,0), (1,0,1,1,1)...(0,1,1,1,1). (iii) For each node, we calculated the Pearson correlation between the degree vector and these cell line specificity vectors. If the best correlation coefficient was larger than a threshold of 0.9 ( $P < 0.006$ ), then we assigned the node with the corresponding cell line specificity.

For 20-kb resolution Hi-C data in 12 normal cell lines and 2 cancer cell lines, we used a distribution-based method to evaluate the cell type specificities. (i) The degree of each node was represented as a vector containing the degree  $z$  score values calculated in all cell lines that had both GV and Hi-C data. (ii) For each node, we assumed that the normalized degrees obey a Gaussian distribution across normal cell lines, and we calculated the mean and SD. (iii) On the basis of the mean and SD for each node, we calculated the  $z$  score for each cell line, i.e., the cell line specificity  $z$  score. A node was considered cell line-specific if the absolute value of the cell line specificity  $z$  score was greater than 1. The “Network construction and hub identification” section was presented in an earlier and limited preprint version of this study deposited in BioRxiv (14)

## Hub screening and validation

### Cell culture

K562, H1975, and NAMALWA cells were cultured in RPMI 1640 medium (Gibco). 293T, HeLa, A549, and Huh7.5.1 cells were cultured in Dulbecco’s modified Eagle’s medium (Gibco). All cells were supplemented with 10% fetal bovine serum (Biological Industries) with 1% penicillin/streptomycin and cultured in 5% CO<sub>2</sub> at 37°C.

### Design and construction of the CRISPR-Cas9 pgRNA library

To validate the importance of the hub regions, we sorted the hub regions with PLT and selected the top 700 all-cell line hubs and top 300 K562-specific hubs. Among them, 960 hubs were suitable for designing pgRNAs for CRISPR-Cas9 screening. For each hub, up to 20 pgRNAs were designed to target 1-kb upstream and 1-kb downstream regions flanking the two boundaries of the 5-kb segment. To ensure the cleavage accuracy and efficacy, we required sgRNAs in each pair to contain at least two mismatches to any other loci in the human genome, and their GC contents are between 0.2 and 0.8. For all the possible pgRNAs obtained from the selected sgRNAs, we removed those that may delete any promoter or exon of protein-coding genes, and we ensured that the cut site of each sgRNA is at least 30 base pairs (bp) away from the exon-intron boundary of the coding genes. We also designed 473 pgRNAs deleting the promoter region and first exon of 29 ribosomal genes as positive controls, and 100 pgRNAs targeting the *AAVS1* locus as well as 100 nontargeting pgRNAs as negative controls, which were obtained from our previous library (16). As a result, the hub deletion library contained 17,476 pairs of gRNAs targeting 960 hub loci. The 128-nt oligonucleotides containing pgRNA coding sequences were designed, synthesized (Agilent Technologies Inc.), and cloned into the lentiviral expression vector following the two-step cloning method as previously described (16), with a minimum representation of 150 transformed colonies per pgRNA in each cloning step.

### CRISPR-Cas9 pgRNA library screening

K562 cells stably expressing Cas9 were infected with pgRNA library lentiviruses at an MOI of <0.3 (1000× to 1500× coverage of the library), and two replicates were arranged. Seventy-two hours after infection, enhanced green fluorescent protein-positive (EGFP<sup>+</sup>) cells were selected by fluorescence-activated cell sorting (FACS; day 0). For each replicate, the harvested cells were divided into a day 0 control group and an experimental group, which was further maintained at a minimum coverage of 1500× for 30 days. Then, cells from each group with 1500× library coverage were, respectively subjected to genomic DNA extraction, PCR amplification of sgRNA-coding sequences, and high-throughput sequencing analysis (Illumina HiSeq 2500 and HiSeq X Ten platform) as previously described (16).

## Identification of functional hubs

Sequencing reads were mapped to the pgRNA library and further normalized to reads per million for each barcoded gRNA. After calculating the quantile of pgRNA counts from two replicates, we removed noisy pgRNAs if a pgRNA’s quantile difference of two replicates was in either 3% tail of the distribution. Then, log<sub>2</sub>FC between the experimental and control groups was calculated for each pgRNA, and 100 negative control genes were generated by randomly sampling 20 *AAVS1*-targeting pgRNAs with replacement. Two scores for each set of hubs were calculated: (i) the mean log<sub>2</sub>FC of all pgRNAs in the set, denoted by FC<sub>hub</sub>; and (ii)  $-\log_{10}P_{\text{value}}$  of the one-sided Mann-Whitney  $U$  test of all pgRNAs in the set compared with pgRNAs targeting the *AAVS1* locus, denoted by  $P_{\text{hub}}$ . The background distribution of these two scores was represented by the mean ( $\mu_{\text{FC}}$  and  $\mu_{\text{P}}$ ) and SD ( $\sigma_{\text{FC}}$  and  $\sigma_{\text{P}}$ ) of all negative control genes. Then, the essentiality of hubs was evaluated by the following function

$$I_{\text{score}} = \text{sign}\left(\frac{\text{FC}_{\text{hub}} - \mu_{\text{FC}}}{\sigma_{\text{FC}}}\right) \times \left| \frac{\text{FC}_{\text{hub}} - \mu_{\text{FC}}}{\sigma_{\text{FC}}} + \frac{P_{\text{hub}} - \mu_{\text{P}}}{\sigma_{\text{P}}} \right|$$

and hubs with the lowest  $I_{\text{score}}$  ( $\leq -1$ ) were identified as essential hubs.

To further avoid the potential issue of cell toxicity generated from multiple cleavages by some pgRNAs, we retrieved the GuideScan specificity score to evaluate each sgRNA (21). By calculating the harmonic mean of the two sgRNAs for each pgRNA, a specificity score was generated for each pgRNA. We kept only the identified essential hubs if their targeting pgRNAs had specificity scores > 0.1 and log<sub>2</sub>FC < -1. Furthermore, to avoid the copy number effect on dropout screening, the copy number of each hub locus in the K562 cell line was analyzed on the basis of ENCODE consortium copy number data ([www.encodeproject.org/files/ENCFF486MJU/](http://www.encodeproject.org/files/ENCFF486MJU/)). After further filtering hub loci with copy number amplification, the remaining hits were regarded as essential hubs.

### Distance between hubs and centromeres

We calculated the distances between hubs and centromeres using their nearest boundaries and compared the distance distributions for essential and nonessential hubs. Chi-squared goodness of fit test was used to calculate the  $P$  value.

### Individual validation of essential hubs by cell proliferation assay

For each candidate hub locus, two pgRNAs were used for the individual validations, and they were either newly designed or selected from the library showing consistent depletion in replicates. To ensure high targeting specificity of all the selected pgRNAs, we required that their specificity scores are all greater than 0.15, and the score of at least one pgRNA for each hub is greater than 0.2. For the newly designed pgRNA, we further required that they do not include  $\geq 4$ -bp homopolymer stretches and that their GC contents are between 0.4 and 0.7. We also changed the deletion regions, which included each sgRNA targeting -1 to +0.5 kb, flanking the two boundaries of the 5-kb hub loci (- and + refer to the outer and inner hub directions, respectively). Other rules were the same as those used for the pgRNA design in the library screening.

All the pgRNAs targeting each hub to be validated were individually cloned into a lentiviral expression vector containing an EGFP selection marker. After virus packaging, the pgRNA lentiviruses were respectively transduced into K562 cells at an MOI of <1. The percentages of EGFP-expressing cells indicating the fraction of pgRNA-containing cells were quantified every 3 days by FACS. Cell

proliferation of each sample was measured by normalizing the percentage of EGFP<sup>+</sup> cells at each time point to that at 3 days after infection (labeled day 0), which was the same as previously described (9, 16). The experiments lasted for 15 days after the first FACS analysis, and at least 100,000 cells were analyzed.

### WGS to evaluate off-target effects

K562 cells were lentivirally transduced with the pgRNA hub\_22\_7-pg2. The EGFP<sup>+</sup> cells were collected by FACS sorting at day 8 after pgRNA infection at an MOI of <1, and the sorted cells were subjected to genomic DNA extraction. The WGS library was prepared following the manufacturer's instructions and sequenced using the Illumina HiSeq 4000 platform. Using the WGS data, we evaluated the deletion efficiency at the targeted locus and off-target effects.

We downloaded the K562 (wild-type) WGS data from ENCODE with accession codes ENCF0313MGL, ENCF004THU, ENCF0506TKC, and ENCF066GQD and then evaluated the potential off-target effects following the published procedures (37). We first generated putative off-target sites for hub\_22\_7 in the hg19 genome using Cas-OFFinder (23). We called the base mismatch type with at most four mismatches without considering any bulge (mismatch  $\leq 4$ , bulge = 0). We also called bulge mismatch type with at most two mismatches with at maximum two bulges (mismatch  $\leq 2$ , bulge  $\leq 2$ ). In total, we examined 455 potential off-target loci. To detect the candidate mutations and indels in the hub-deleted cells, variant calling was performed as described in genome analysis toolkit (GATK) Best Practices (<https://gatk.broadinstitute.org/hc/en-us>). Briefly, reads were aligned to the human reference genome (hg19) using BWA-0.7.17. Duplicated reads were then removed using GATK4 MarkDuplicatesSpark (<https://gatk.broadinstitute.org/hc/en-us/articles/360037224932-MarkDuplicatesSpark>). The reads were then processed via base quality score recalibration using GATK4. Germline mutations (compared to the hg19 reference genome) were called in both wild-type and hub-deleted cells by GTAK HaplotypeCaller (version 4.1.4.1) with the default parameters. SNVs and indels called by GATK4 Mutect2 (version 4.1.4.1) with the default parameters were used to assess off-target deletions.

We further confirmed no off-target effects using a different analysis software, BCFTOOLS suite (version 1.9, [www.htslib.org/doc/bcftools.html](http://www.htslib.org/doc/bcftools.html)), to reexamine the single-nucleotide polymorphisms (SNPs) and indel sites from the WGS data. The mapped BAM file of K562 cells was piped into bcftools mpileup and bcftools call with default parameters. The called raw variant call format (VCF) file was filtered by a bcftools filter with “%QUAL < 30 || DP < 30” marked as low-quality variants. Homozygous variants were also removed from the raw VCF file with the parameter “GT = 1/1.” Gold standard indels VCF of Mills and 1000G were downloaded from GATK Resource Bundle (<https://gatk.broadinstitute.org/hc/en-us/articles/360035890811-Resource-bundle>). The gold standard indels were also removed from the VCF file using bcftools isec with parameter “-n -1 -c all.” There were no putative off-target sites found in the 13,809 indels obtained using bedtools intersect (<https://bedtools.readthedocs.io>).

### Hi-C library preparation and data analysis

#### Hi-C library preparation

The pgRNA Hub\_22\_7-pg2 was delivered into K562 cells through lentiviral infection at an MOI of <1. EGFP<sup>+</sup> cells were collected by FACS sorting at day 9 after infection, and the sorted cells were allowed to recover under normal cell culture conditions for 2 hours

before proceeding to conduct the Hi-C library. One million cells were used for each Hi-C library preparation using an Arima-HiC kit (Arima Genomics, San Diego) following the manufacturer's instructions. Hi-C libraries were sequenced using the Illumina NovaSeq platform.

#### Hi-C data processing

The Hi-C raw FASTQ data were processed by the Juicer pipeline (38) with the default parameters. Hi-C reads were aligned to hg19 (GRCh37), and the reads with mapping quality score (MAPQ) < 30 were further trimmed (table S15). The output bam files were transformed into 5-kb, 10-kb, 25-kb, 50-kb, 100-kb, and 1-Mb resolution contact matrix. The contact matrix was then normalized by the VC method (13). The significance level of a given interaction pair was calculated from Poisson distribution fitting between the measured interaction reads and the expected reads by VC normalization. Juicebox (<https://aidenlab.org/juicebox/>) and HiCExplorer (39, 40) were used to visualize the processed Hi-C data.

#### Loop calling

In both wild-type K562 and hub\_22\_7-deleted K562 cells, the VC normalized Hi-C contact reads were processed by HiCCUPS with default parameters at 25-kb resolution for calling loops. (<https://github.com/aidenlab/juicer/wiki/HiCCUPS>).

#### TAD calling

We used insulation score (41) to identify the TADs for K562 wild-type and hub\_22\_7 deletion cells in 10-kb resolution data. The HiCExplorer software was used to plot the TADs. (39)

#### A/B compartment analysis

The A/B compartment analysis was conducted using 50-kb bins. The eigenvectors for each chromosome in both K562 wild-type and hub-deleted cells were extracted from the VC normalized Hi-C counts processed by the Juicer pipeline with the default parameters (38). The polymerase II (Pol II) ChIP-seq data in K562 cells were downloaded from ENCODE (42). The correlation between the first eigenvector of each chromosome and the Pol II peaks density was calculated, on the basis of which we determined the A and B compartments (43). We repeated this analysis in GM12878, HUVEC, IMR90, and NHEK. For HMEC, there were no Pol II ChIP-seq data available, and thus, we used TSS density for hg19 genome to assign A/B compartments.

#### Effective diameter comparison

The effective diameter was computed by SNAP software. We calculated the effective diameter deviation for each chromosome both before and after hub deletion and found that the deviation followed a Gaussian distribution by the Shapiro-Wilk normality test ( $P = 0.27$  so that the null hypothesis of being normal distribution was accepted). Then, we calculated the  $P$  value for the deviation of each chromosome on the basis of a Gaussian distribution and identified the significantly changed chromosome with  $P < 0.05$ .

#### Modularity comparison

The modularity was computed by SNAP software. We collected the modularity scores of each chromosome in the seven wild-type cell lines (GM12878, K562, HUVECs, IMR90, NHEK, KBM7, and HMEC) and found that the modularity score for each chromosome followed a Gaussian distribution (all  $P$  values  $\geq 0.01$  to accept the null hypothesis of being a Gaussian distribution in the Shapiro-Wilk normality test). Then, for each chromosome in hub-deleted K562 cells, we calculated the  $P$  value of its modularity score on the basis of chromosome-specific modularity distribution and identified significantly changed chromosomes with  $P < 0.05$ .



**Bulk RNA-seq and data analysis****Bulk RNA-seq library preparation**

The pgRNA AAVS1-pg1 targeting the AAVS1 locus was delivered into K562 cells at an MOI of <1. Then,  $2 \times 10^6$  EGFP<sup>+</sup> K562 cells were sorted by FACS 8 days after transfection. Total RNA was extracted using the RNeasy Mini Kit (QIAGEN, 79254) with three replicates. The RNA-seq libraries were further prepared following the NEBNext PolyA mRNA Magnetic Isolation Module [New England Biolabs (NEB), E7490S], NEBNext RNA First Strand Synthesis Module (NEB, E7525S), NEBNext mRNA Second Strand Synthesis Module (NEB, E6111S), and NEBNext Ultra DNA Library Prep Kit for Illumina (NEB, E7370L). All samples were subjected to next-generation sequencing (NGS) analysis using the Illumina HiSeq 4000 platform.

**Bulk RNA-seq data processing**

In the bulk RNA-seq library, the sequencing reads with Phred scores of  $\geq 30$  were aligned to the human reference genome (GRCh37/hg19) using HISAT2 (2.0.4) (44, 45) and assembled and quantified by StringTie (1.3.5) (44, 46). The gene read counts for each sample were further normalized by CPM.

**scRNA-seq and data analysis****Single-cell library preparation**

K562 cells infected with Hub\_22\_7-pg2 were FACS-sorted 8 days after lentivirus transduction for single-cell library preparation. The single-cell library was prepared with the established protocol described previously (24). Briefly, polyadenylated RNA was reverse transcribed through tailed oligo(dT) priming directly in whole-cell lysate (single droplet) using Moloney murine leukemia virus reverse transcriptase (MMLV RT) and temperature switch oligos. The resulting full-length complementary DNA (cDNA) contained the complete 5' end of the mRNA, as well as an anchor sequence that served as a universal priming site for second-strand synthesis. The cDNA was preamplified using 15 cycles with Kapa HiFi HotStart ReadyMix. We used the Nextera DNA Sample Preparation Kit to generate single-cell libraries. The amplified cDNA was tagged at 55°C for 5 min in a 20- $\mu$ l reaction with 0.25  $\mu$ l of transposase and 5  $\mu$ l of Nextera reaction buffer. Five microliters of neutralization buffer was added to the tagmentation reaction mix to strip the transposase off the DNA, and the tagmented DNA was amplified by 12 cycles of standard Nextera PCR. Then the DNA was purified with 20  $\mu$ l of Ampure beads (sample to beads ratio of 1:0.6). The prepared libraries were sequenced on an Illumina HiSeq 4000 instrument.

**scRNA-seq processing**

The FASTQ files were first mapped to the human reference genome (GRCh37/hg19) using Picard (2.17.0) (47) and STAR (2.5.3a) (48). We used the Drop-seq processing pipeline developed by the McCarroll laboratory (24) to remove low-quality reads (lower than Q10) and PCR duplicates (identified by cell barcodes and molecular barcodes). The cells were descendingly ordered by read count. Reads from all the cells were pooled together to form a cumulative distribution. Cells with the most reads before the inflection point “knee” of the cumulative distribution were kept for the following analysis (table S16).

We calculated a *P* value for each gene to assess whether the change was significant. Each cell was first normalized by CPM. We calculated  $E_i$ , which is the sum of CPMs for a given gene across all the cells, and  $E_{\text{total}}$ , which is the sum of  $E_i$  for all the genes. We then computed  $P_i = E_i/E_{\text{total}}$ . In a given cell *j*, the normalized gene expression

of all genes was assumed to follow a binomial distribution  $G_{ij} \sim B(N_j, P_i)$  independently and identically, where  $G_{ij}$  is the expected reads of gene *i* in cell *j* and  $N_j$  is the total reads for cell *j*. We calculated a *P* value to evaluate how significantly each gene expression in each cell deviated from the expected value on the basis of the binomial distribution, which indicates its differential expression across cells. We also calculated the *P* value for genes in the negative control ( $\Delta$ AAVS1) and wild-type bulk RNA-seq data the same way.

**Single-cell trajectory branching and pseudotime analysis**

Because hub deletion affected cell proliferation, we focused on analyzing the apoptosis genes annotated in the KEGG database (www.genome.jp/kegg/). Considering the noise in the scRNA-seq data, we selected apoptosis genes that showed differential expression in at least 10 to 15% of cells ( $P < 0.05$ ). As a result, 93 apoptosis genes were identified in K562 cells with the essential hub chr22: 17,325,000 to 17,330,000 deleted. All the single-cell and bulk data were clustered with trajectory branching and pseudotime analysis using the Monocle R package (25, 26). Monocle (25, 26) assigned each cell a pseudotime value and a “state” on the basis of the segment of the trajectory according to the PQ tree algorithm. Cells with the same state were clustered together (26), and then relative gene expression in each cluster was computed.

**DEGs identified from pseudotime analysis**

To identify differentially expressed genes (DEGs) between state 1 and state 2 defined in the pseudotime analysis, a Wilcoxon rank sum test was applied to identify DEGs in state 2 compared to those in state 1 using a *P* value cutoff of 0.05. The chromosome distributions for these DEGs are listed in table S16.

**Investigation of the essentialities of DEGs from scRNA-seq data**

Among the DEGs in chr22 upon hub\_22\_7 (chr22: 17,325,000 to 17,330,000) deletion, which were significantly decreased from state 1 to state 2, a top-ranked DEG *THOC5* was selected to analyze its importance on cell growth and proliferation in K562 cells. Three sgRNAs were designed to knock down its expression through the CRISPRi strategy, which were selected from the hCRISPRi-v2 library (27). These sgRNAs were also individually cloned into the lentiviral expression vector with an EGFP marker and then respectively transduced into K562 cells stably expressing dCas9-KRAB (Krüppel-associated box) protein at an MOI of <1. The cell proliferation assay was performed as previously described (9, 16). The first time point of FACS analysis was 6 days after lentiviral infection, and the experiment lasted for 12 days.

**SUPPLEMENTARY MATERIALS**

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abi6020>

[View/request a protocol for this paper from Bio-protocol.](#)

**REFERENCES AND NOTES**

1. J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
2. E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, E. Heard, Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
3. D. G. Lupiáñez, K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J. M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S. A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel, S. Mundlos, Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).

4. E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
5. J. Dekker, M. A. Marti-Renom, L. A. Mirny, Exploring the three-dimensional organization of genomes: Interpreting chromatin interaction data. *Nat. Rev. Genet.* **14**, 390–403 (2013).
6. T. Nagano, Y. Lubling, T. J. Stevens, S. Schoenfelder, E. Yaffe, W. Dean, E. D. Laue, A. Tanay, P. Fraser, Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
7. G. D. Rose, P. J. Fleming, J. R. Banavar, A. Maritan, A backbone-based theory of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 16623–16633 (2006).
8. S. J. Liu, M. A. Horlbeck, S. W. Cho, H. S. Birk, M. Malatesta, D. He, F. J. Attenello, J. E. Villalta, M. Y. Cho, Y. Chen, M. A. Mandegar, M. P. Olvera, L. A. Gilbert, B. R. Conklin, H. Y. Chang, J. S. Weissman, D. A. Lim, CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* **355**, aah7111 (2017).
9. Y. Liu, Z. Cao, Y. Wang, Y. Guo, P. Xu, P. Yuan, Z. Liu, Y. He, W. Wei, Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. *Nat. Biotechnol.* **36**, 1203–1210 (2018).
10. C. P. Fulco, M. Munschauer, R. Anyoha, G. Munson, S. R. Grossman, E. M. Perez, M. Kane, B. Cleary, E. S. Lander, J. M. Engreitz, Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
11. D. R. Simeonov, B. G. Gowen, M. Boontanrart, T. L. Roth, J. D. Gagnon, M. R. Mumbach, A. T. Satpathy, Y. Lee, N. L. Bray, A. Y. Chan, D. S. Lituiev, M. L. Nguyen, R. E. Gate, M. Subramaniam, Z. Li, J. M. Woo, T. Mitros, G. J. Ray, G. L. Curie, N. Naddaf, J. S. Chu, H. Ma, E. Boyer, F. Van Gool, H. Huang, R. Liu, V. R. Tobin, K. Schumann, M. J. Daly, K. K. Farh, K. M. Ansel, C. J. Ye, W. J. Greenleaf, M. S. Anderson, J. A. Bluestone, H. Y. Chang, J. E. Corn, A. Marson, Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549**, 111–115 (2017).
12. Y. Diao, R. Fang, B. Li, Z. Meng, J. Yu, Y. Qiu, K. C. Lin, H. Huang, T. Liu, R. J. Marina, I. Jung, Y. Shen, K.-L. Guan, B. Ren, A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat. Methods* **14**, 629–635 (2017).
13. S. S. P. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, E. L. Aiden, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
14. B. Ding, L. Zheng, D. Medovoy, W. Wang, Targeted mutations on 3D hub loci alter spatial interaction environment. *bioRxiv* 030999 (2015); <https://doi.org/10.1101/030999>.
15. R. Albert, H. Jeong, A. L. Barabasi, Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
16. S. Zhu, W. Li, J. Liu, C.-H. Chen, Q. Liao, P. Xu, H. Xu, T. Xiao, Z. Cao, J. Peng, P. Yuan, M. Brown, X. S. Liu, W. Wei, Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nat. Biotechnol.* **34**, 1279–1286 (2016).
17. D. M. Munoz, P. J. Cassiani, L. Li, E. Billy, J. M. Korn, M. D. Jones, J. Golji, D. A. Ruddy, K. Yu, G. McAllister, A. DeWeck, D. Abramowski, J. Wan, M. D. Shirley, S. Y. Neshat, D. Rakiec, R. de Beaumont, O. Weber, A. Kauffmann, E. R. McDonald, N. Keen, F. Hofmann, W. R. Sellers, T. Schmelzle, F. Stegmaier, M. R. Schlabach, CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov.* **6**, 900–913 (2016).
18. A. J. Aguirre, R. M. Meyers, B. A. Weir, F. Vazquez, C.-Z. Zhang, U. Ben-David, A. Cook, G. Ha, W. F. Harrington, M. B. Doshi, M. Kost-Alimova, S. Gill, H. Xu, L. D. Ali, G. Jiang, S. Pantel, Y. Lee, A. Goodale, A. D. Cherniack, C. Oh, G. Kryukov, G. S. Cowley, L. A. Garraway, K. Stegmaier, C. W. Roberts, T. R. Golub, M. Meyerson, D. E. Root, A. Tsherniak, W. C. Hahn, Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* **6**, 914–929 (2016).
19. D. W. Morgens, M. Wainberg, E. A. Boyle, O. Ursu, C. L. Araya, C. K. Tsui, M. S. Haney, G. T. Hess, K. Han, E. E. Jeng, A. Li, M. P. Snyder, W. J. Greenleaf, A. Kundaje, M. C. Bassik, Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun.* **8**, 15178 (2017).
20. J. Tycko, M. Wainberg, G. K. Marinov, O. Ursu, G. T. Hess, B. K. Ego, Aradhana, A. Li, A. Truong, A. E. Trevino, K. Spees, D. Yao, I. M. Kaplow, P. G. Greenside, D. W. Morgens, D. H. Phanstiel, M. P. Snyder, L. Bintu, W. J. Greenleaf, A. Kundaje, M. C. Bassik, Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. *Nat. Commun.* **10**, 4063 (2019).
21. A. R. Perez, Y. Pritykin, J. A. Vidigal, S. Chhangawala, L. Zamparo, C. S. Leslie, A. Ventura, GuideScan software for improved single and paired CRISPR guide RNA design. *Nat. Biotechnol.* **35**, 347–349 (2017).
22. The ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
23. S. Bae, J. Park, J.-S. Kim, Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
24. E. Z. Macosko, A. Basu, R. Satija, J. Nemeshegyi, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
25. C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, J. L. Rinn, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
26. X. Qiu, A. Hill, J. Packer, D. Lin, Y.-A. Ma, C. Trapnell, Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
27. M. A. Horlbeck, L. A. Gilbert, J. E. Villalta, B. Adamson, R. A. Pak, Y. Chen, A. P. Fields, C. Y. Park, J. E. Corn, M. Kampmann, J. S. Weissman, Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* **5**, e19760 (2016).
28. T. Wang, K. Birsoy, N. W. Hughes, K. M. Krupczak, Y. Post, J. J. Wei, E. S. Lander, D. M. Sabatini, Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
29. T. Zhang, M. Wallis, V. Petrovic, J. Challis, P. Kalitsis, D. F. Hudson, Loss of TOP3B leads to increased R-loop formation and genome instability. *Open Biol.* **9**, 190222 (2019).
30. C. Y. McLean, D. Bristol, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, G. Bejerano, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
31. S. Venkatesan, R. Rosenthal, N. Kanu, N. McGranahan, J. Bartek, S. A. Quezada, J. Hare, R. S. Harris, C. Swanton, Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. *Ann. Oncol.* **29**, 563–572 (2018).
32. M. E. Olson, R. S. Harris, D. A. Harki, APOBEC enzymes as targets for virus and cancer therapy. *Cell Chem. Biol.* **25**, 36–49 (2018).
33. C. Swanton, N. McGranahan, G. J. Starrett, R. S. Harris, APOBEC enzymes: Mutagenic fuel for cancer evolution and heterogeneity. *Cancer Discov.* **5**, 704–712 (2015).
34. N. Roper, S. Gao, T. K. Maity, A. R. Banday, X. Zhang, A. Venugopalan, C. M. Cultraro, R. Patidar, S. Sindiri, A.-L. Brown, A. Goncarenko, A. R. Panchenko, R. Biswas, A. Thomas, A. Rajan, C. A. Carter, D. E. Kleiner, S. M. Hewitt, J. Khan, L. Prokunina-Olsson, U. Guha, APOBEC mutagenesis and copy-number alterations are drivers of proteogenomic tumor evolution and heterogeneity in metastatic thoracic tumors. *Cell Rep.* **26**, 2651–2666.e6 (2019).
35. Y. Zhang, R. Delahanty, X. Guo, W. Zheng, J. Long, Integrative genomic analysis reveals functional diversification of APOBEC gene family in breast cancer. *Hum. Genomics* **9**, 34 (2015).
36. F. M. Behan, F. Iorio, G. Picco, E. Gonçalves, C. M. Beaver, G. Migliardi, R. Santos, Y. Rao, F. Sassi, M. Pinnelli, R. Ansari, S. Harper, D. A. Jackson, R. McRae, R. Pooley, P. Wilkinson, D. van der Meer, D. Dow, C. Buser-Doeppner, A. Bertotti, L. Trusolino, E. A. Stronach, J. Saez-Rodriguez, K. Yusa, M. J. Garnett, Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**, 511–516 (2019).
37. C. Smith, A. Gore, W. Yan, L. Abalde-Atristain, Z. Li, C. He, Y. Wang, R. A. Brodsky, K. Zhang, L. Cheng, Z. Ye, Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* **15**, 12–13 (2014).
38. N. C. Durand, M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, E. L. Aiden, Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
39. F. Ramirez, V. Bhardwaj, L. Arrigoni, K. C. Lam, B. A. Grünig, J. Villaveces, B. Habermann, A. Akhtar, T. Manke, High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
40. J. Wolff, V. Bhardwaj, S. Nothjunge, G. Richard, G. Renschler, R. Gilsbach, T. Manke, R. Backofen, F. Ramirez, B. A. Grünig, Galaxy HiCExplorer: A web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **46**, W11–W16 (2018).
41. E. Crane, Q. Bian, R. P. McCord, B. R. Lajoie, B. S. Wheeler, E. J. Ralston, S. Uzawa, J. Dekker, B. J. Meyer, Condensin-driven remodeling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
42. K. Y. Yip, C. Cheng, N. Bhardwaj, J. B. Brown, J. Leng, A. Kundaje, J. Rozowsky, E. Birney, P. Bickel, M. Snyder, M. Gerstein, Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol.* **13**, R48 (2012).
43. R. Kalhor, H. Tjong, N. Jayathilaka, F. Alber, L. Chen, Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* **30**, 90–98 (2011).
44. M. Perlea, D. Kim, G. M. Perlea, J. T. Leek, S. L. Salzberg, Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).

45. D. Kim, B. Langmead, S. L. Salzberg, HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
46. M. Pertea, G. M. Davis, C. M. Antonescu, T.-C. Chang, J. T. Mendell, S. L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
47. Picard Tools - By Broad Institute; <https://broadinstitute.github.io/picard/>.
48. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
49. A.-L. Barabási, Z. N. Oltvai, Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
50. R. Albert, A.-L. Barabási, Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
51. C. B. Luzzio, B. B. Luzzio, Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood* **45**, 321–334 (1975).
52. X. Zhu, M. Gerstein, M. Snyder, Getting connected: Analysis and principles of biological networks. *Genes Dev.* **21**, 1010–1024 (2007).
53. P. M. Kim, L. J. Lu, Y. Xia, M. B. Gerstein, Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* **314**, 1938–1941 (2006).
54. M. Vidal, M. E. Cusick, A.-L. Barabási, Interactome networks and human disease. *Cell* **144**, 986–998 (2011).
55. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A.-L. Barabási, The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
56. J. Förster, I. Famili, P. Fu, B. Ø. Palsson, J. Nielsen, Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244–253 (2003).
57. G. Balázsi, A.-L. Barabási, Z. N. Oltvai, Topological units of environmental signal processing in the transcriptional regulatory network of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7841–7846 (2005).
58. N. M. Luscombe, M. Madan Babu, H. Yu, M. Snyder, S. A. Teichmann, M. Gerstein, Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**, 308–312 (2004).
59. X. Wang, X. Wei, B. Thijssen, J. Das, S. M. Lipkin, H. Yu, Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat. Biotechnol.* **30**, 159–164 (2012).
60. S. B. Petersen, M. T. Neves-Petersen, S. B. Henriksen, R. J. Mortensen, H. M. Geertz-Hansen, Scale-free behaviour of amino acid pair interactions in folded proteins. *PLOS ONE* **7**, e41322 (2012).
61. R. Cohen, S. Havlin, Scale-free networks are ultrasmall. *Phys. Rev. Lett.* **90**, 058701 (2003).

**Acknowledgments:** We acknowledge the staff of the BIOPIC High-throughput Sequencing Center (Peking University) for assistance in NGS analysis; the National Center for Protein Sciences (Beijing) at Peking University for assistance with FACS and analysis; and H. Lv and L. Du for technical help. We acknowledge Y. Yu (Peking University) for 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 services. We acknowledge J. Xu (UC San Diego) for assistance in preparing a single-cell RNA-seq library. **Funding:** This project was supported by funds from CIRM (RB5-07012) and the NIH (R01HG009626) (to W.Wa.); the National Science Foundation of China (NSFC31930016), Beijing Municipal Science and Technology Commission (Z181100001318009), the Beijing Advanced Innovation Center for Genomics at Peking University and the Peking-Tsinghua Center for Life Sciences (to W.We.); and China Postdoctoral Science Foundation (2020 M670031 to Y.L.). **Author contributions:** W.Wa. and W.We. conceived and supervised the project. W.Wa., W.We., B.D., and Y.L. designed the experiments. B.D. and L.Z. constructed the network analysis and identified and characterized hub regions. Y.G. designed the pgRNA library for hub screening. Y.L. and P.X. performed the pgRNA library construction and screening. Y.L. performed the experiments, including individual validation of candidate hubs, DEGs, WGS, and bulk RNA-seq, 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 hub-deleted cells. P.W. and Y.Z. performed scRNA-seq on hub-deleted cells. L.Z. and B.D. performed the bioinformatics analysis of the WGS, Hi-C, and single-cell RNA-seq data. B.D., Y.L., L.Z., W.We., and W.Wa. wrote the manuscript with contributions from all other authors.

**Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The data here are accessible through GEO accession code GSE176503 and NCBI Sequence Read Archive (SRA) under BioProject ID PRJNA732577.

Submitted 19 March 2021  
Accepted 15 September 2021  
Published 3 November 2021  
10.1126/sciadv.abi6020

## Noncoding loci without epigenomic signals can be essential for maintaining global chromatin organization and cell viability

Bo DingYing LiuZhiheng LiuLina ZhengPing XuZhao ChenPeiyao WuYing ZhaoQian PanYu GuoWensheng WeiWei Wang

*Sci. Adv.*, 7 (45), eabi6020.

### View the article online

<https://www.science.org/doi/10.1126/sciadv.abi6020>

### Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of think article is subject to the [Terms of service](#)

---

*Science Advances* (ISSN ) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.  
Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).