Title: High-content image-based CRISPR screening reveals regulators of 3D genome architectures

Authors: Mengwei Hu#[1], Bing Yang#[1], Tyler B Jensen#[1,6], Jonathan S D Radda[1], Yubao Cheng[1], Shengyan Jin[1] and Siyuan Wang*[1, 2, 3, 4, 5, 6, 7, 8]
# These authors contributed equally
* Correspondence author

Affiliations: [1]Department of Genetics, Yale School of Medicine, Yale University, New Haven, CT 06510, USA; [2]Department of Cell Biology, Yale School of Medicine, Yale University, New Haven, CT 06510, USA; [3]Yale Combined Program in the Biological and Biomedical Sciences, Yale University, New Haven, CT 06510, USA; [4]Molecular Cell Biology, Genetics and Development Program, Yale University, New Haven, CT 06510, USA; [5]Biochemistry, Quantitative Biology, Biophysics, and Structural Biology Program, Yale University, New Haven, CT 06510, USA; [6]M.D.-Ph.D. Program, Yale University, New Haven, CT 06510, USA; [7]Yale Center for RNA Science and Medicine, Yale University School of Medicine, New Haven, CT 06510, USA; [8]Yale Liver Center, Yale University School of Medicine, New Haven, CT 06510, USA.

Abstract: Three-dimensional (3D) genome organization becomes altered during development, aging, and disease, but the factors regulating chromatin topology are incompletely understood and currently no technology can efficiently screen for new regulators of long-range chromatin organization. Here, we developed an image-based high-content CRISPR screening platform that combines a new FISH-based barcode readout method (BARC-FISH) with chromatin tracing. We performed a pooled loss-of-function genetic screen in human cells and visualized alterations to their genome organization from 13,000 imaging target-perturbation combinations, alongside perturbation-paired barcode readout in the same single cells. Using 1.4 million 3D positions along chromosome traces, we identified 26 regulators of chromatin architectures at different length scales. We found that the ATP-dependent helicase CHD7, the loss of which causes the congenital syndrome CHARGE, cooperates with CTCF to promote large-scale chromatin compaction. Altogether, our method enables scalable, high-throughput identification of chromatin topology regulators that will stimulate new insights into the 3D genome.

One-Sentence Summary: *In situ* CRISPR screen, cellular barcoding, and chromatin tracing identify regulators of higher-order chromatin organization.

Main Text:

Introduction

The spatial organization of chromatin is linked to many genomic functions and shows intriguing dynamics in a variety of biological processes and diseases (*1*–*23*). Chromatin organization occurs at many levels and length scales, from local accessibility to inter-domain contacts to global nuclear architecture. Studies on effectors such as histone modification enzymes or chromatin loop organizers *e.g.* CTCF have bolstered our understanding of chromatin architecture and provided new drug targets for a wide range of diseases, but our overarching understanding of how chromatin is regulated across length scales and in different cell types and conditions remains limited. It is a significant challenge to systematically identify new molecular regulators of complex three-dimensional (3D) genome architecture and form testable hypotheses about their mechanisms of action. Important 3D genome regulators have been primarily discovered by perturbing one candidate gene at a time (*24*–*26*) or by plate-based screening focusing on relatively low-content phenotypes such as the spatial distance between one pair of genomic loci (*27*, *28*). Recently, high-content CRISPR screens combining Perturb-seq and single-cell ATAC-seq technologies have enabled the high-throughput discovery of candidate regulators of local chromatin accessibility (*29*, *30*). However, we still lack scalable, broadly applicable methods to efficiently screen for regulators of higher-order 3D chromatin folding architectures, especially at the length scales of topologically-associating domains (TADs, also known as contact domains), chromatin compartments, and chromosome territories (*31*–*37*).

TADs largely confine the scope of promoter-enhancer interactions (*38*, *39*) and are structural units of chromatin with different DNA replication timing (*40*) and mutation susceptibility (*41*). TADs are further sorted into segregated A (active) and B (inactive) compartments in each chromosome territory (*32*). Whole chromosome territory compaction has been observed in X chromosome inactivation (*42*) and cellular senescence (*43*). Defining the regulatory landscape and architectural basis of chromatin folding at the length scales that are relevant to each genomic feature is critical to understanding their functions and dynamics in development, aging, and disease (*3*, *10*, *44*). Assessing multiple length scales in parallel is technically challenging and currently only possible with single-gene perturbations.

We endeavored to develop a technique that would allow for discovery of new factors influencing chromatin architecture as well as provide clues into which length scales and chromatin features they act on. To this end, here we present a new method that combines the power of CRISPR screening of candidate regulators with high-content readout of chromatin organization over multiple length scales. Our high-throughput, high-content pooled genetic screen is combined with super-resolved *in situ* tracing of complex chromatin folding conformations and imaging of nuclear architectures across multiple length scales in human cells. A key innovation of our technology is the decoding technique we devised, known as BARC-FISH, which enzymatically amplifies the barcode of each sgRNA *in situ* for robust decoding using FISH, which is compatible with the multi-scale 3D chromatin mapping required to assess high-content phenotypic readouts for each barcode. After validating our new method, we applied it to screen 137 candidate genes and identified 26 top hit novel regulators of 3D genome organization at different length scales. In particular, we identified CHD7 as a multi-scale chromatin compactor, in addition to its known activity in promoting local chromatin accessibility. Altogether, our work presented here provides a new platform for the discovery and characterization of new chromatin regulators across multiple length scales and paves the way to build a global map of how 3D genomic architecture is regulated.

Results

BARC-FISH and chromatin tracing enables image-based pooled CRISPR screen of chromatin and nuclear organization regulators

To systematically discover novel regulators of chromatin conformation, we developed an image-based method to screen chromatin structures in individual knockout cells. Briefly, we used CRISPR-Cas9 technology to generate a pooled library of knockout A549 human lung cancer cells that co-express Cas9 with a single guide RNA (sgRNA) and a unique RNA barcode (Fig. 1A and fig. S1, A and B). The barcode RNA comprised ten regions, each encoding one of three sequences (Fig. 1B), a design inspired by a recent report (*45*). Each region is analogous to a ternary digit (with values “0”, “1”, or “2”) in computation, and each ten-digit barcode was uniquely paired with an sgRNA in the same cell (Fig 1, A and B). We developed “Barcode Amplification by Rolling Circle and Fluorescence *In Situ* Hybridization” (BARC-FISH) to visually read out the barcode RNAs in single cells, and thereby identify the genetic perturbation in each cell (see below). To map the phenotypic effects of each knockout on 3D spatial organization of numerous genomic loci in single cells, we performed the highly multiplexed DNA FISH method known as chromatin tracing (*46*) on chromosome 22 (chr22) (Fig. 1A). Finally, we employed a high-throughput computational approach to identify conserved features and systematic changes of chromatin organization caused by the same genetic perturbation in multiple cells, for every gene knockout in the pooled screen.

For BARC-FISH, we adopted a high-efficiency rolling circle amplification (RCA) strategy from an *in situ* sequencing method (*47*). Specifically, a linear probe and a “padlock” probe are hybridized to each “digit” in the RNA barcode. Circularization and ligation of the padlock probe generate a template for RCA, primed by the linear probe, which in turn locally generates many copies of part of the digit sequence (Fig. 1B). For each of the ten digits, we hybridized dye-labeled secondary probes to the RCA product and observed a strong signal over background (Fig. 1, B and C). To detect the value of each digit in individual cells, we sequentially applied three-color secondary probes for each digit, and imaged the fluorescence signals from the labeled RCA products over ten rounds of three-color sequential FISH imaging (Fig. 1C). After each round of imaging, the fluorescence signals were removed from the cells before the next round of sequential FISH. Finally, the fluorescence signals from all rounds of imaging were computationally converted to barcode values, and subsequently to sgRNA identities for individual cells by referring to the barcode-sgRNA associations mapped by high-throughput sequencing (Materials and Methods). With high signal-to-background ratio and an error-correcting decoding algorithm (Materials and Methods), BARC-FISH achieved robust decoding while being compatible with chromatin tracing (Fig. 1, C and D, and fig S1, C and D), allowing us to match sgRNA identities underlying genetic perturbations to 3D genome phenotypes (Fig. 1E).

For our screen, we generated a plasmid library of 420 sgRNAs composed of 10 non-targeting control sgRNAs and 410 sgRNAs targeting 137 selected genes (a coverage of 2-3 sgRNAs per target gene) (table S1). The plasmid library was cloned using a high-throughput pooled cloning strategy (fig. S2A, Materials and Methods). This strategy randomly paired the 10-digit barcodes with the sgRNAs, and used a bottlenecking method to ensure unique mapping of each barcode to a single sgRNA. With this strategy, up to ~5000 sgRNAs can be distinguished with the current barcoding scheme (Materials and Methods). It is possible to extend the barcode sequence to allow for more perturbations in a single screen. The 137 selected genes included known chromatin conformation regulators such as NIPBL and CTCF as positive controls (*11*), and primarily genes encoding nuclear proteins that are differentially expressed by more than five-fold upon oncogene-induced senescence (*48*), given the extensive 3D genome reorganization during this process (*43*, *44*). We generated the cell knockout library by lentiviral transduction of the plasmid library and puromycin selection of the transduced cells, and detected 8 non-targeting sgRNA controls and 404 sgRNAs corresponding to all 137 targeted genes, suggesting an sgRNA dropout rate of 1.9% with our library construction procedure. To estimate the knockout efficiency in our cell background, we transduced A549-Cas9 cells with selected individual sgRNA constructs and performed next-generation sequencing on the target genomic DNA of the polyclonal cells after transduction and selection. The results showed knockout efficiencies of 43-70% for the selected sgRNAs (fig. S2B), close to the predicted knockout efficiencies in previously reported CRISPR editing prediction model (*49*, *50*) . This less-than-100% knockout efficiency is intrinsic to CRISPR and is expected to potentially lead to a mixture of wildtype and knockout phenotypes in subpopulations of cells carrying the same sgRNA, but we expect to still capture knockout phenotypes that are sufficiently strong in population analyses.

Identification of regulators of multi-scale chromatin folding

For chromatin tracing, we mapped the conformation of chr22 as a model (due to its short genomic length and the lack of known structural variations on chr22 in A549 cells) at the TAD-to-chromosome length scales by pinpointing the central 100-kb regions of all 27 TADs spanning chr22 (Fig. 1D). To exclude the potential influence of cell cycle on chromatin conformation, we incorporated Geminin staining and only included G1 phase cells in our analyses (fig. S2C) (*51*). Individual G1 phase A549-Cas9 cells contained 2-4 chr22 traces each. We analyzed 57,286 traces containing 1,407,797 3D positions from 17,304 cells.

To identify regulators of 3D genome organization, first we investigated the spatial distances between adjacent TADs on chr22. Consistent with recent imaging studies focusing on single pairs of adjacent TADs (*52*, *53*) , we found that loss of the known cohesin loader NIPBL led to a significant increase in adjacent TAD distance, whereas knocking out CTCF significantly decreased adjacent TAD distance as expected (*52*, *53*) (Fig. 2, A and B). These positive control results are consistent with the opposing roles of CTCF and NIPBL in loop extrusion (*24–26*, *52–55*), and cross-validated our screening method. Importantly, other knockout hits also altered the adjacent TAD distances, revealing new candidate regulators of chromatin organization and providing hints as to the length scale at which they operate. We observed that knocking out the tumor suppressor RB1, MRVI1 and PIP5K1B increased the adjacent TAD distance, while knocking out GLDC, the nuclear receptor NR4A1 and ZNF114 caused the opposite phenotype (Fig. 2, A to C).

We next explored the regulation of chromatin folding conformations at the length scale of A-B compartmentalization by measuring long-range chromatin contact frequencies for each perturbation within the same cells screened above. We defined two TADs spaced less than 500 nm apart as in contact with each other, and derived the long-range contact frequency between non-adjacent TADs along the chr22 genomic map. As A-B compartment organization stems from long-range chromatin contact (*32*), we further categorized the long-range contacts as contacts between compartment A regions (A-A), contacts between compartment B regions (B-B), and inter-compartmental contacts between A and B regions (A-B). We observed that loss of NR4A1, PDE1A or the homeobox transcription factor HOXB9 increased the long-range A-A contact frequency, while knocking out RB1, PCBP1 or LRRC10B showed the opposite phenotype (Fig. 2, D and E). Knocking out RFESD, HOXB9 or FAM69B increased the frequency of inter-compartmental A-B contacts, while knockout of C2CD2, the chromatin remodeler CHD7 or FAM13C decreased A-B contacts (Fig. 2, F and G). Finally, knocking out the AP-1 transcription factor subunit FOS, NR4A1, or the helicase DDX24 increased the long-range B-B contact frequency, while knocking out MYBPH decreased B-B contact frequency (Fig. 2, H and I).

Finally, we integrated all pairs of inter-TAD distances to measure and compare the overall compactness of the chr22 chromosome territory. The average fold changes of all 351 inter-TAD distances among the 27 TADs were calculated. We found that knocking out PCBP1, RB1, or CHD7 decompacted chr22 (Fig. 2, J to L), while GLDC, HOXB9 or C19orf57 knockout resulted in chr22 compaction (Fig. 2, J and K). Intriguingly, RB1 is known to promote the formation of senescence-associated heterochromatin foci (*56*), a highly compacted whole-chromosome conformation. Our results show that RB1 knockout decompacts chr22 across multiple scales, including increased adjacent TAD distance (Fig. 2, A to C), reduced long-range A-A contact frequency (Fig. 2, D and E), and overall decompacted chromosome territory (Fig. 2, J to L). The high-content readouts of our screen are therefore in agreement with this prior association but, importantly, provide new insights into the scales at which RB1 and the other hits impact chromatin organization.

Identification of regulators of chromosomal radial positioning and morphological properties of the nucleus

Chromatin association with the nuclear lamina is functionally linked to gene silencing (*57*, *58*). Gene-rich chromosomes (such as chr22) are known to be preferentially located towards the nuclear interior (*59*). Localization to the interior of the nucleus in interphase can also protect chromosomes from segregation errors during mitosis (*60*). To screen for factors that regulate chromatin-lamina association, we incorporated DAPI staining of whole nuclei together with our multi-scale chromatin tracing described above in the context of our CRISPR screen. We used the edges of the whole nuclear staining patterns to approximate lamina positions, and identified gene knockout hits that affected the distance between chr22 and the nuclear lamina (Fig. 3, A to C). We found that knocking out CUL1, PRSS22, or MAB21L2 brought chr22 closer to the nuclear lamina, as shown by shorter distances both between the centroid of the chromosome to the lamina (Fig. 3, A and B) and between each individual TAD to the lamina (Fig. 3C), while LRRC10B, DDX21, or HMGA2 knockout moved chr22 towards the nuclear interior (Fig. 3, A to C).

Even on its own, nuclear DAPI staining bears surprisingly rich information regarding the nuclear organization that can be used to distinguish cell states (*61*, *62*). Chromatin is distributed unevenly in the cell nucleus, and the “texture” of nuclear staining pattern is often used as a diagnostic marker of cancer (*63*). Moreover, cells with abnormal, non-spherically shaped nuclei are often seen in cancer and aging and may indicate genome instability (*64–66*). By analyzing changes in DAPI staining caused by each perturbation in our screen, we found that knockout of RB1 or MYBPH reduced the unevenness of nuclear DAPI staining, generating patterns with more homogeneous intensity within each nucleus (Fig. 3, D to G), whereas a more heterogenous intensity pattern with chromatin condensates is often associated with cancer (*63*) . In addition, knocking out TRIM36 and EEPD1 decreased the sphericity of the nuclei and led to multi-lobed nuclei shapes (Fig. 3, H to J).

In summary, our image-based pooled CRISPR screen using BARC-FISH and multi-scale chromatin profiling allowed us to systematically and simultaneously profile the effect of hundreds of candidate genes on many aspects of spatial genome organization, from short- and long-range chromatin folding and chromatin association with a nuclear landmark to whole-nucleus morphological characteristics. Overall, we identified 26 top candidate regulators of chromatin organization from our screen of 137 selected genes (table S2), and present a generalizable platform that can be used to screen thousands of genes of interest in different cell types and conditions in future work.

CHD7 is a long-range chromatin compactor

As we examined our hits for each chromatin organization phenotype, CHD7 drew our attention because it regulated chromatin conformation on different scales in opposing ways. CHD7, or chromodomain helicase DNA binding protein 7, is a chromatin remodeler known to promote local chromatin openness and is associated with CHARGE syndrome (*67–69*). The results of our screen showed that CHD7 knockout significantly reduced long-range A-B contact frequencies (Fig. 2, F and G) and resulted in decompaction of the whole chr22 territory (Fig. 2, J and K). To validate this large-scale chromatin-compaction function of CHD7 using an orthogonal method, we performed siRNA knockdown of CHD7 in the same A549-Cas9 cell line. Western blot analysis confirmed that the CHD7 protein level was reduced by >90% (fig. S3A). Chromatin tracing of chr22 in the CHD7 knockdown (siCHD7) and control (siCtrl) cells showed that while the compartment identities of TADs were largely identical between siCtrl and siCHD7 (fig. S3, B and C), and the A-B compartments were spatially positioned in a conventional, polarized manner (*46*) in both cases (fig. S3D), A-A, B-B, and A-B long-range contacts were less frequent in siCHD7 compared to siCtrl cells (fig. S3E). CHD7 knockdown also caused overall chromosome territory decompaction, represented by a global increase of inter-TAD distances (Fig. 4A and fig. S3F) and larger radii of gyration of the whole chromosome traces (fig. S3G). The spatial regulatory effect of CHD7 appeared stronger at long range (defined as spatial distances between genomic regions that are more than 3Mb apart) compared to short range (defined as spatial distances between genomic regions that are less than 3Mb apart) (Fig. 4B). Altogether, these results agree with the phenotypes we observed for CHD7 in our knockout screen, confirming that this factor facilitates chromatin compaction especially at long range.

To further validate the regulatory effect of CHD7 on chromatin organization, we overexpressed CHD7 in the A549-Cas9 cell line. In comparison to control GFP overexpression using the same vector, CHD7 overexpression significantly compacted chromatin and promoted chromatin contacts across larger length scales: While the A-B compartment identities of TADs and the polarized arrangement of A-B compartments largely remained unchanged (fig. S4, A to C), CHD7 overexpression led to significantly higher contact frequencies between A-A, B-B, and A-B compartment regions (fig. S4D), and significantly decreased global inter-TAD distances and radii of gyration (Fig. 4C and fig. S4, E and F), indicating overall chromatin compaction. Similar to the CHD7 knockdown scenario, the compaction effect upon CHD7 overexpression is more significant at long range (Fig. 4D). Together, these data further validated the findings from our image-based pooled screen, and show that CHD7 specifically promote long-range chromatin compaction and contact.

To investigate if the chromatin compaction function of CHD7 applies to a different cell background and different genomic context, we performed siRNA knockdown of CHD7 in human RPE-1 cells and conducted chromatin tracing on chromosome 21 (chr21). CHD7 knockdown caused an overall increase of inter-TAD distance (fig. S5, A and B) and an increase of the radius of gyration of chr21 territory (fig. S5C), which were mainly contributed by long-range chromatin decompaction (fig. S5D). CHD7 knockdown also led to decreased contact frequency between long-range A-A, A-B and B-B interactions (fig. S5E). Together, these results suggested that the chromatin compaction function of CHD7 exists in distinct cell lines and genomic contexts.

CHD7 compacts chromatin at long-range in a CTCF-dependent manner

CHD7 is known to promote local chromatin accessibility and regulate transcription (*69*) – while it is perhaps not surprising that CHD7 has an effect on genome organization in general, it was unintuitive and unexpected to discover that it promotes *compaction* of higher-order chromatin architectures. CHD7 belongs to the chromodomain helicase DNA-binding protein family, which are ATP-dependent chromatin remodelers (*70*). It contains two chromodomains, a helicase domain, SLIDE and SANT domains that are important for nucleosome sliding, and two BRK domains close to C terminus of largely unknown function (*68*). Consistent with previous reports in other cell contexts (*71*), we confirmed that CHD7 knockdown resulted in diverse gene expression changes (up- and down-regulation) in many cellular programs (fig. S6, A and B), and that CHD7 binds to a broad range of targets as shown by CUT&RUN (fig. S6, C to E). Intriguingly, a recent report showed that in the context of DNA damage, CHD7 binds to the damaged site and initially decompacts chromatin, and then recruits histone deacetylases (HDAC) to re-compact chromatin through histone deacetylation (*72*). To investigate the possibility that CHD7 could compact chromatin through regulating HDAC activity in the context of our undamaged cells, we treated wild-type cells with an HDAC inhibitor Trichostatin A (TSA). Immunofluorescence confirmed that TSA treatment increased global histone acetylation (fig. S7A), and chromatin tracing of chr22 showed that HDAC inhibition led to chromatin decompaction at short scales, including between adjacent TADs (fig. S7, B to D), but had the opposite effect at long range (Fig. 4, E and F, and fig. S7, E to G). Thus, the long-range chromatin compaction activity of CHD7 cannot be explained by its known HDAC recruitment capability.

It has been shown that CHD7 can directly bind CTCF (*73*). Indeed, our CUT&RUN profiling of CHD7 and CTCF showed that CHD7 binding peaks across the genome significantly overlap with CTCF binding peaks in the A549-Cas9 cell line (fig. S6F). A previous report showed that CTCF binding peaks along the genome were not affected by CHD7 knockout (*69*), suggesting that CHD7 does not modulate CTCF binding to the genome. To test if the chromatin compaction effect of CHD7 depends on CTCF, we performed chromatin tracing on cells with double knockdown of CHD7 and CTCF (doubleKD), and compared the traces to those from single knockdown of CTCF (siCTCF). Western blots confirmed the efficient depletion of CHD7 and CTCF (fig. S8A). Chromatin tracing showed that CTCF knockdown alone led to decreased inter-TAD distance at the short range, but increased inter-TAD distance at long range (Fig. 4, G and H). In the context of CTCF knockdown, siCHD7 still decompacted chromatin at short range distances, but on average had no effect for long range chromatin interaction (Fig. 4, I and J, and fig. S8, B to E). These results indicate that the long-range chromatin compaction function of CHD7 at least partially depends on CTCF.

Discussion

In this work, we report the development of a high-throughput, high-content, image-based genetic screening platform and demonstrate its ability to systematically identify regulators of 3D genome folding architectures from short-range to long-range and global nuclear organization. The platform integrates a pooled CRISPR knockout screen comprising hundreds of candidate genes with chromatin tracing and the development of a cellular barcoding and *in situ* decoding technique termed BARC-FISH. Our BARC-FISH technique combines *in situ* amplification and robust readout of a short RNA barcode associated with a CRISPR gRNA in each cell, thereby allowing us to pair the genetic perturbation through CRISPR (cause) with the 3D genome alteration (effect) observed in each cell. The BARC-FISH barcode design has the capacity to simultaneously assess libraries of up to ~5000 sgRNAs in pooled format with the current experimental design and can be further extended in the future with longer barcodes. Both BARC-FISH and chromatin tracing are based on highly multiplexed FISH, and thus are readily compatible with each other on the same sample. Importantly, the phenotypic readout of this screening platform is information-rich, allowing for categorization of a gene of interest based on its effects on many different length scales and aspects of chromatin organization. Overall, our current work included ~13,000 imaging target-perturbation combinations (420 gRNAs multiply 30 phenotypic imaging targets including 27 TADs, DAPI, total protein, and Geminin stains). We expect the entire screen methodology to be broadly adaptable to many other cell types and biological contexts.

In our initial screen of 137 candidate genes with multiple sgRNAs per gene, we identified 26 previously unknown or incompletely known regulators of 3D genome and nuclear organization features across multiple length scales, including the architecture of adjacent TADs, A-B compartment organization, entire chromosome conformation and localization, and whole nucleus morphology. We observed and validated a long-range chromatin compaction function of CHD7, which was striking because of its incongruence with the opposite chromatin accessibility-promoting effect of this protein at short scale. This observation suggests that CHD7-targeted genomic regions may be subject to a new type of “structural poising”, containing features of both open chromatin and compact chromatin at different length scales. We further showed that the long-range chromatin compaction function of CHD7 is independent of its regulation of HDAC activity, but is at least partially dependent on CTCF, indicating distinct mechanisms governing short- and long-range activities of CHD7. Our identification and characterization of CHD7 in this paper demonstrates how our screening methodology can generate new hypotheses with sufficient information to guide targeted follow-up experiments. We anticipate that the remaining hits from our screen, as well as hits from future broader screens encompassing a wider range of candidate genes, will provide a rich resource for discovery of new chromatin regulators at multiple scales and their mechanisms of action.

CHD7 belongs to a family of chromodomain helicases, which are DNA binding proteins with an ATPase domain, adjacent chromatin binding domains that contribute to their chromatin remodeling activity (*68*), and often C-terminal BRK domains that directly interact with CTCF (*73*). Our results suggest that CHD7 compacts chromatin at large length scales in a manner partially dependent on CTCF, and we hypothesize that this could be achieved by CHD7 tethering two DNA regions – one through its BRK domain binding to CTCF on DNA, and the other through direct binding to chromatin (Fig. 4K). While this model remains to be tested, it raises the intriguing possibility of single-sided loop extrusion by CHD7. *In vitro* reconstitution studies are needed to directly prove the existence of the single-sided loop extrusion activity, as was performed to prove the conventional loop extrusion mechanism (*74*), and while these are beyond the scope of this study, we believe that these new questions raised by our work highlight the broad utility of the resource presented here.

Interestingly, a congenital disorder known as CHARGE syndrome (named for its associated phenotypes of coloboma, heart defects, atresia, growth retardation, genital abnormalities, and ear abnormalities (*67*)) is highly attributed to mutations in CHD7 (*75*), and the new functional insights we gained through this work could advance understanding of how CHD7 contributes to these developmental abnormalities. A close paralog of CHD7 known as CHD8 is implicated in autism spectrum disorder (*76*), and while CHD8 was not represented in our screen, some of our insights and proposed mechanism for CHD7 function may translate to CHD8 as well.

The candidate 3D genome regulators identified in this work have diverse expression profiles in different tissues (*77*) (fig. S9), and 24 out of 26 of our reported top hit genes are associated with disease states (table S2). Our screen provides a potential mechanistic link between 3D genome regulation and the physiology/pathology of the diverse tissues and contexts. For example, NR4A1, an orphan nuclear receptor protein, has been found to be a signature of metastasis in certain primary solid tumors (*78*). Our screen found it to be a top candidate in regulating chromatin compartment interactions (Fig. 2, D, E, H and I). It is of great interest to investigate whether the chromatin conformation changes it mediates contribute to tumor metastasis. Our top candidates also included important genes in development, for example FOS, HOXB9 and MAB21L2. 3D genome reorganization has been shown to be crucial in regulating tissue-specific gene expression during development (*79*). We are intrigued by how these genes may regulate development in part through 3D genome regulation. ZNF114 was shown to associate with MDC1 in a two-hybrid screen, a critical regulator of DNA double-strand break (DSB) repair (*80*). Substantial data show that DSBs lead to a decompaction of heterochromatin and relocation to the periphery of heterochromatic domains (*81*). Given its association with MDC1 and observed compaction upon knockout, ZNF114 may regulate the topological reorganization of the genome after genotoxic stress and the generation of DSBs.

Besides the identification of novel 3D genome regulators, our work also raised new questions about previously known or speculated 3D genome regulators. For example, our model for CHD7-dependent long-range chromatin compaction (Fig. 4K) suggests a potential role of CTCF as an “anchor” in this mechanism. This potential mode of action of CTCF is distinct from and may happen in parallel with its canonical role as a “roadblock” in the cohesin-CTCF loop extrusion model. RB1 was known to recruit cohesin and condensin II complexes during DNA replication and mitosis for sister chromatid cohesion and chromosome condensation (*82*). Our results showed a prominent role of RB1 in promoting higher-order chromatin compaction during G1, likely through its regulation of heterochromatin formation and maintenance (*82*).

Overall, BARC-FISH combined with chromatin tracing enables mapping of the chromatin organization "regulatome" at scale, which will deepen our understanding of the regulatory landscape of the genome and of the relationships between genome architectures and diverse developmental and disease states.

References

1. P. Fraser, W. Bickmore, Nuclear organization of the genome and the potential for gene regulation. *Nature*. **447**, 413–417 (2007).
2. W. A. Bickmore, B. Van Steensel, Genome architecture: Domain organization of interphase chromosomes. *Cell*. **152**, 1270–1284 (2013).
3. P. H. L. Krijger, W. De Laat, Regulation of disease-associated gene expression in the 3D genome. *Nat. Rev. Mol. Cell Biol.* **17**, 771–782 (2016).
4. J. Dekker, A. S. Belmont, M. Guttman, V. O. Leshyk, J. T. Lis, S. Lomvardas, L. A. Mirny, C. C. O'Shea, P. J. Park, B. Ren, J. C. Ritland Politz, J. Shendure, S. Zhong, The 4D nucleome project. *Nature*. **549**, 219–226 (2017).
5. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, A Phase Separation Model for Transcriptional Control. *Cell*. **169**, 13–23 (2017).
6. M. Yu, B. Ren, The three-dimensional organization of mammalian genomes. *Annu. Rev.*

Cell Dev. Biol. **33**, 265–289 (2017).

7. M. Spielmann, D. G. Lupiáñez, S. Mundlos, Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467 (2018).
8. C. Y. Wang, T. Jégu, H. P. Chu, H. J. Oh, J. T. Lee, SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X. *Cell*. **174**, 406-421.e25 (2018).
9. R. Stadhouders, G. J. Filion, T. Graf, Transcription factors and 3D genome conformation in cell-fate decisions. *Nature*. **569**, 345–354 (2019).
10. H. Zheng, W. Xie, The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **20**, 535–550 (2019).
11. J. A. Beagan, J. E. Phillips-Cremins, On the existence and functionality of topologically associating domains. *Nat. Genet.* **52**, 8–16 (2020).
12. M. Lakadamyali, M. P. Cosma, Visualizing the genome in high resolution challenges our textbook understanding. *Nat. Methods*. **17**, 371–379 (2020).
13. C. A. De Graaf, B. Van Steensel, Chromatin organization: Form to function. *Curr. Opin. Genet. Dev.* **23**, 185–190 (2013).
14. T. Misteli, The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell*. **183**, 28–45 (2020).
15. R. K. A. Virk, W. Wu, L. M. Almassalha, G. M. Bauer, Y. Li, D. VanDerway, J. Frederick, D. Zhang, A. Eshein, H. K. Roy, I. Szleifer, V. Backman, Disordered chromatin packing regulates phenotypic plasticity. *Sci. Adv.* **6**, 1–18 (2020).
16. A. M. Oudelaar, D. R. Higgs, The relationship between genome structure and function.

Nat. Rev. Genet. **22**, 154–168 (2021).

17. M. Levine, C. Cattoglio, R. Tjian, Looping back to leap forward: Transcription enters a new era. *Cell*. **157**, 13–25 (2014).
18. G. Bonora, K. Plath, M. Denholtz, A mechanistic link between gene regulation and genome architecture in mammalian development. *Curr. Opin. Genet. Dev.* **27**, 92–101 (2014).
19. D. U. Gorkin, D. Leung, B. Ren, The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell*. **14**, 762–775 (2014).
20. A. Pombo, N. Dillon, Three-dimensional genome architecture: Players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **16**, 245–257 (2015).
21. B. Bonev, G. Cavalli, Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).
22. D. Hnisz, A. S. Weintrau, D. S. Day, A. L. Valton, R. O. Bak, C. H. Li, J. Goldmann, B. R. Lajoie, Z. P. Fan, A. A. Sigova, J. Reddy, D. Borges-Rivera, T. I. Lee, R. Jaenisch, M. H. Porteus, J. Dekker, R. A. Young, Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*. **351**, 1454–1458 (2016).
23. J. Dekker, L. Mirny, The 3D Genome as Moderator of Chromosomal Communication. *Cell*. **164**, 1110–1121 (2016).
24. E. P. Nora, A. Goloborodko, A. L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, B. G. Bruneau, Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*. **169**, 930-944.e22 (2017).

25. S. S. P. Rao, S. C. Huang, B. Glenn St Hilaire, J. M. Engreitz, E. M. Perez, K. R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, X. Huang, M. S. Shamim, J. Shin, D. Turner, Z. Ye, A. D. Omer, J. T. Robinson, T. Schlick, B. E. Bernstein, R. Casellas, E. S. Lander, E. L. Aiden, Cohesin Loss Eliminates All Loop Domains. *Cell*. **171**, 305-320.e24 (2017).

26. W. Schwarzer, N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, Y. Loe-Mie, N. A. Fonseca, W. Huber, C. H. Haering, L. Mirny, F. Spitz, Two independent modes of chromatin organization revealed by cohesin removal. *Nature*. **551**, 51–56 (2017).

27. S. Shachar, T. C. Voss, G. Pegoraro, N. Sciascia, T. Misteli, Identification of Gene Positioning Factors Using High-Throughput Imaging Mapping. *Cell*. **162**, 911–923 (2015).

28. D. S. Park, S. C. Nguyen, R. Isenhart, P. P. Shah, W. Kim, R. J. Barnett, A. Chandra, J. M. Luppino, J. Harke, M. Wai, R. Yang, Y. Lan, S. Yoon, R. Yunker, G. Vahedi, J. E. Phillips-Cremins, R. Jain, E. F. Joyce, High-throughput Oligopaint screen identifies druggable regulators of genome folding. *bioRxiv* (2022).

29. N. Liscovitch-Brauer, A. Montalbano, J. Deng, A. Méndez-Mancilla, H. H. Wessels, N. G. Moss, C. Y. Kung, A. Sookdeo, X. Guo, E. Geller, S. Jaini, P. Smibert, N. E. Sanjana, Profiling the genetic determinants of chromatin accessibility with scalable single-cell CRISPR screens. *Nat. Biotechnol.* **39**, 1270–1277 (2021).

30. S. E. Pierce, J. M. Granja, W. J. Greenleaf, High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. *Nat. Commun.* **12**, 1–8 (2021).

31. T. Cremer, C. Cremer, Chromosome territories, nuclear architecture and gene regulation

in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301 (2001).

32. 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).

33. 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).

34. C. Hou, L. Li, Z. S. Qin, V. G. Corces, Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. *Mol. Cell*. **48**, 471–484 (2012).

35. 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).

36. T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, G. Cavalli, Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*. **148**, 458–472 (2012).

37. 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).

38. Y. Shen, F. Yue, D. F. Mc Cleary, Z. Ye, L. Edsall, S. Kuan, U. Wagner, J. Dixon, L. Lee, B. Ren, V. V. Lobanenkov, A map of the cis-regulatory sequences in the mouse genome. *Nature*. **488**, 116–120 (2012).

39. T. Sexton, G. Cavalli, The role of chromosome domains in shaping the functional genome. *Cell*. **160**, 1049–1059 (2015).

40. B. D. Pope, T. Ryba, V. Dileep, F. Yue, W. Wu, O. Denas, D. L. Vera, Y. Wang, R. S. Hansen, T. K. Canfield, R. E. Thurman, Y. Cheng, G. Gülsoy, J. H. Dennis, M. P. Snyder, J. A. Stamatoyannopoulos, J. Taylor, R. C. Hardison, T. Kahveci, B. Ren, D. M. Gilbert, Topologically associating domains are stable units of replication-timing regulation. *Nature*. **515**, 402–405 (2014).

41. F. Senigl, Y. Maman, R. K. Dinesh, J. Alinikula, R. B. Seth, L. Pecnova, A. D. Omer, S. S. P. Rao, D. Weisz, J. M. Buerstedde, E. L. Aiden, R. Casellas, J. Hejnar, D. G. Schatz, Topologically Associated Domains Delineate Susceptibility to Somatic Hypermutation. *Cell Rep*. **29**, 3902-3915.e1-e8 (2019).

42. C. Naughton, D. Sproul, C. Hamilton, N. Gilbert, Analysis of Active and Inactive X Chromosome Architecture Reveals the Independent Organization of 30 nm and Large-Scale Chromatin Structures. *Mol. Cell*. **40**, 397–409 (2010).

43. T. Chandra, K. Kirschner, J. Y. Thuret, B. D. Pope, T. Ryba, S. Newman, K. Ahmed, S. A. Samarajiwa, R. Salama, T. Carroll, R. Stark, R. S. Janky, M. Narita, L. Xue, A. Chicas, S. Nũnez, R. Janknecht, Y. Hayashi-Takanaka, M. D. Wilson, A. Marshall, D. T. Odom, M. M. Babu, D. P. Bazett-Jones, S. Tavaré, P. A. W. Edwards, S. W. Lowe, H. Kimura, D. M. Gilbert, M. Narita, Independence of Repressive Histone Marks and Chromatin Compaction during Senescent Heterochromatic Layer Formation. *Mol. Cell*. **47**, 203–214

(2012).

44. S. W. Criscione, Y. V. Teo, N. Neretti, The Chromatin Landscape of Cellular Senescence. *Trends Genet.* **32**, 751–761 (2016).

45. C. Wang, T. Lu, G. Emanuel, H. P. Babcock, X. Zhuang, Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 10842–10851 (2019).

46. S. Wang, J. Su, B. J. Beliveau, B. Bintu, J. R. Moffitt, C. Wu, X. Zhuang, Spatial organization of chromatin domains and compartments in single chromosomes. *Science*. **353**, 598–602 (2016).

47. X. Wang, W. E. Allen, M. A. Wright, E. L. Sylwestrak, N. Samusik, S. Vesuna, K. Evans, C. Liu, C. Ramakrishnan, J. Liu, G. P. Nolan, F. A. Bava, K. Deisseroth, Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*. **361** (2018), doi:10.1126/science.aat5691.

48. A. Hernandez-Segura, T. V. de Jong, S. Melov, V. Guryev, J. Campisi, M. Demaria, Unmasking Transcriptional Heterogeneity in Senescent Cells. *Curr. Biol.* **27**, 2652-2660.e4 (2017).

49. M. W. Shen, M. Arbab, J. Y. Hsu, D. Worstell, S. J. Culbertson, O. Krabbe, C. A. Cassa, D. R. Liu, D. K. Gifford, R. I. Sherwood, Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature*. **563**, 646–651 (2018).

50. K. Labun, T. G. Montague, M. Krause, Y. N. Torres Cleuren, H. Tjeldnes, E. Valen, CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* **47**, W171–W174 (2019).

51. B. Bintu, L. J. Mateo, J. H. Su, N. A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A. N. Boettiger, X. Zhuang, Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*. **362** (2018), doi:10.1126/science.aau1783.

52. J. M. Luppino, D. S. Park, S. C. Nguyen, Y. Lan, Z. Xu, R. Yunker, E. F. Joyce, Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nat. Genet.* **52**, 840–848 (2020).

53. Q. Szabo, A. Donjon, I. Jerković, G. L. Papadopoulos, T. Cheutin, B. Bonev, E. P. Nora, B. G. Bruneau, F. Bantignies, G. Cavalli, Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat. Genet.* **52**, 1151–1157 (2020).

54. A. L. Sanborn, S. S. P. Rao, S. C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K. P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. K. Stamenova, E. S. Lander, E. L. Aiden, Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6456–E6465 (2015).

55. G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, L. A. Mirny, Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* **15**, 2038–2049 (2016).

56. M. Narita, S. Nũnez, E. Heard, M. Narita, A. W. Lin, S. A. Hearn, D. L. Spector, G. J. Hannon, S. W. Lowe, Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell*. **113**, 703–716 (2003).

57. C.-K. Chen, M. Blanco, C. Jackson, E. Aznauryan, N. Ollikainen, C. Surka, A. Chow, A. Cerase, P. McDonel, M. Guttman, Xist recruits the X chromosome to the nuclear lamina

to enable chromosome-wide silencing. *Science*. **354**, 468–472 (2016).

58. B. van Steensel, A. S. Belmont, Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell*. **169**, 780–791 (2017).

59. J. A. Croft, J. M. Bridger, S. Boyle, P. Perry, P. Teague, W. A. Bickmore, Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* **145**, 1119–1131 (1999).

60. S. J. Klaasen, M. A. Truong, R. H. van Jaarsveld, I. Koprivec, V. Štimac, S. G. De Vries, P. Risteski, S. Kodba, K. Vukušić, K. L. De Luca, J. M. Marques, E. M. Gerrits, B. Bakker, F. Foijer, J. Kind, I. M. Tolić, S. M. A. Lens, G. J. P. L. Kops, Nuclear chromosome locations dictate segregation error frequencies. *Nature*. **607**, 604–609 (2022).

61. Y. Takei, S. Zheng, J. Yun, S. Shah, N. Pierson, J. White, S. Schindler, C. H. Tischbirek, G. C. Yuan, L. Cai, Single-cell nuclear architecture across cell types in the mouse brain. *Science*. **374**, 586–594 (2021).

62. L. Funk, K. Su, J. Ly, D. Feldman, A. Singh, B. Moodie, P. C. Blainey, I. M. Cheeseman, The phenotypic landscape of essential human genes. *Cell*. **185**, 4634-4653.e22 (2022).

63. A. H. Fischer, K. A. Jacobson, J. Rose, R. Zeller, Hematoxylin and Eosin Staining of Tissue and Cell Sections. *Cold Spring Harb. Protoc.* **3**, 4986–4988 (2008).

64. A. Mattout, T. Dechat, S. A. Adam, R. D. Goldman, Y. Gruenbaum, Nuclear lamins, diseases and aging. *Curr. Opin. Cell Biol.* **18**, 335–341 (2006).

65. P. Scaffidi, T. Misteli, Lamin A-Dependent Nuclear Defects in Human Aging. *Science*. **312**, 1059–1064 (2006).

66. K. H. Chow, R. E. Factor, K. S. Ullman, The nuclear envelope environment and its cancer

connections. *Nat. Rev. Cancer*. **12**, 196–209 (2012).

67. L. E. L. M. Vissers, C. M. A. van Ravenswaaij, R. Admiraal, J. A. Hurst, B. B. A. de Vries, I. M. Janssen, W. A. van der Vliet, E. H. L. P. G. Huys, P. J. de Jong, B. C. J. Hamel, E. F. P. M. Schoenmakers, H. G. Brunner, J. A. Veltman, A. G. van Kessel, Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat. Genet.* **36**, 955–957 (2004).

68. K. Bouazoune, R. E. Kingston, Chromatin remodeling by the CHD7 protein is impaired by mutations that cause human developmental disorders. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19238–19243 (2012).

69. N. C. Reddy, S. P. Majidi, L. Kong, M. Nemera, C. J. Ferguson, M. Moore, T. M. Goncalves, H. K. Liu, J. A. J. Fitzpatrick, G. Zhao, T. Yamada, A. Bonni, H. W. Gabel, CHARGE syndrome protein CHD7 regulates epigenomic activation of enhancers in granule cell precursors and gyrification of the cerebellum. *Nat. Commun.* **12**, 1–17 (2021).

70. M. A. Basson, C. Van Ravenswaaij-arts, Functional Insights into Chromatin Remodelling from Studies on CHARGE Syndrome. *Trends Genet.* **31**, 600–611 (2015).

71. M. P. Schnetz, L. Handoko, B. Akhtar-Zaidi, C. F. Bartels, C. Filipe Pereira, A. G. Fisher, D. J. Adams, P. Flicek, G. E. Crawford, T. LaFramboise, P. Tesar, C. L. Wei, P. C. Scacheri, CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genet.* **6**, 1–15 (2010).

72. M. B. Rother, S. Pellegrino, R. Smith, M. Gatti, C. Meisenberg, W. W. Wiegant, M. S. Luijsterburg, R. Imhof, J. A. Downs, A. C. O. Vertegaal, S. Huet, M. Altmeyer, H. van Attikum, CHD7 and 53BP1 regulate distinct pathways for the re-ligation of DNA double-strand breaks. *Nat. Commun.* **11** (2020), doi:10.1038/s41467-020-19502-5.

73. M. D. Allen, T. L. Religa, S. M. V. Freund, M. Bycroft, Solution Structure of the BRK Domains from CHD7. *J. Mol. Biol.* **371**, 1135–1140 (2007).

74. Y. Kim, Z. Shi, H. Zhang, I. J. Finkelstein, H. Yu, Human cohesin compacts DNA by loop extrusion. *Science*. **366**, 1345–1349 (2019).

75. G. E. Zentner, W. S. Layman, D. M. Martin, P. C. Scacheri, Molecular and Phenotypic Aspects of CHD7 Mutation in CHARGE Syndrome. *Am. J. Med. Genet.*, 674–686 (2010).

76. B. J. O. Roak, L. Vives, S. Girirajan, E. Karakoc, N. Krumm, B. P. Coe, R. Levy, A. Ko, C. Lee, J. D. Smith, E. H. Turner, I. B. Stanaway, B. Vernot, M. Malig, C. Baker, B. Reilly, J. M. Akey, E. Borenstein, M. J. Rieder, D. A. Nickerson, R. Bernier, J. Shendure, E. E. Eichler, Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*. **485**, 246–250 (2012).

77. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. M. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P. H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. Von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. Von Heijne, J. Nielsen, F. Pontén, Tissue-based map of the human proteome. *Science*. **347** (2015), doi:10.1126/science.1260419.

78. S. Ramaswamy, K. N. Ross, E. S. Lander, T. R. Golub, A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**, 49–54 (2003).

79. 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).

80. J. Lee, S. Park, G. Hariharasudhan, M. Kim, S. M. Jung, S. Jeong, I. Chang, C. Kim, E. Kim, J. Yu, S. Bae, H. J. You, ID3 regulates the MDC1-mediated DNA damage response in order to maintain genome stability. *Nat. Commun.* **8**, 1–14 (2017).

81. A. Fortuny, A. Chansard, P. Caron, O. Chevallier, O. Leroy, S. E. Polo, Imaging the response to DNA damage in heterochromatin domains reveals core principles of heterochromatin maintenance. *Nat. Commun.* **12**, 1–16 (2021).

82. F. A. Dick, D. W. Goodrich, J. Sage, N. J. Dyson, Non-canonical functions of the RB protein in cancer. *Nat. Rev. Cancer*. **18**, 442–451 (2018).

83. J. R. Moffitt, J. Hao, G. Wang, K. H. Chen, H. P. Babcock, X. Zhuang, High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11046–11051 (2016).

84. C. Xia, H. P. Babcock, J. R. Moffitt, X. Zhuang, Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Sci. Rep.* **9**, 1–13 (2019).

85. Q. Xu, M. R. Schlabach, G. J. Hannon, S. J. Elledge, Design of 240,000 orthogonal 25mer DNA barcode probes. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 2289–2294 (2009).

86. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinformatics*. **10**, 1–9 (2009).

87. M. Liu, Y. Lu, B. Yang, Y. Chen, J. S. D. Radda, M. Hu, S. G. Katz, S. Wang, Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nat. Commun.* **11**, 1–14 (2020).

88. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*. **348**, 1360–1363 (2015).

89. M. Hu, B. Yang, Y. Cheng, J. S. D. Radda, Y. Chen, M. Liu, S. Wang, ProbeDealer is a convenient tool for designing probes for highly multiplexed fluorescence in situ hybridization. *Sci. Rep.* **10**, 1–7 (2020).

90. G. Emanuel, J. R. Moffitt, X. Zhuang, High-throughput, image-based screening of pooled genetic-variant libraries. *Nat. Methods*. **14**, 1159–1162 (2017).

91. N. E. Sanjana, O. Shalem, F. Zhang, Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods*. **11**, 783–784 (2014).

92. J. Joung, S. Konermann, J. S. Gootenberg, O. O. Abudayyeh, R. J. Platt, M. D. Brigham, N. E. Sanjana, F. Zhang, Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).

93. S. Wang, J. R. Moffitt, G. T. Dempsey, X. S. Xie, X. Zhuang, Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8452–8457 (2014).

94. L. K. Gopi, B. L. Kidder, Integrative pan cancer analysis reveals epigenomic variation in cancer type and cell specific chromatin domains. *Nat. Commun.* **12**, 1–20 (2021).

95. H. Kang, M. N. Shokhirev, Z. Xu, S. Chandran, J. R. Dixon, M. W. Hetzer, Dynamic regulation of histone modifications and long-range chromosomal interactions during postmitotic transcriptional reactivation. *Genes Dev.* **34**, 1–18 (2020).

96. 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).

97. Y. Liao, G. K. Smyth, W. Shi, The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41** (2013), doi:10.1093/nar/gkt214.

98. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

99. A. Zhu, J. G. Ibrahim, M. I. Love, Heavy-Tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics*. **35**, 2084–2092 (2019).

100. S. Durinck, Y. Moreau, A. Kasprzyk, S. Davis, B. De Moor, A. Brazma, W. Huber, BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics*. **21**, 3439–3440 (2005).

101. Z. Xie, A. Bailey, M. V. Kuleshov, D. J. B. Clarke, J. E. Evangelista, S. L. Jenkins, A. Lachmann, M. L. Wojciechowicz, E. Kropiwnicki, K. M. Jagodnik, M. Jeon, A. Ma'ayan, Gene Set Knowledge Discovery with Enrichr. *Curr. Protoc.* **1**, 1–51 (2021).

102. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**, 2114–2120 (2014).

103. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. **9**, 357–359 (2012).

104. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).

105. G. Yu, L. G. Wang, Q. Y. He, ChIP seeker: An R/Bioconductor package for ChIP peak

annotation, comparison and visualization. *Bioinformatics*. **31**, 2382–2383 (2015).

106. F. Ramírez, F. Dündar, S. Diehl, B. A. Grüning, T. Manke, DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, 187–191 (2014).

Acknowledgments: We thank Dr. Antonio Giraldez, Dr. Monkol Lek, and members of the Wang lab for helpful discussions. We thank members of the Yale Molecular Diagnostics Laboratory, Yale Center for Genome Analysis, the Keck DNA Sequencing Facility and the Yale Flow Cytometry Facility for their help.

Funding: S. W. was partly supported by NIH (UG3CA268202, U01DA052775, R01HG011245, R33CA251037, DP2GM137414) and Pershing Square Sohn Cancer Research Alliance. B.Y., M.H. and Y.C. were in part supported by the China Scholarship Council. T.B.J. was in part supported by 5T32GM007205. J.S.D.R. was supported by NIH Predoctoral Training Grant (2T32GM007499). This work was in part supported by NIH Director's New Innovator Award (DP2GM137414).

Author contributions: S.W. conceived the study. B.Y., M.H., T.B.J., J.S.D.R., and S.W. performed experiments. M.H., B.Y. and T.B.J. analyzed data with help from Y.C., S.J., and S.W.. M.H., B.Y., T.B.J., and S.W. wrote the manuscript with inputs from all authors.

Competing interests: S.W., B.Y. and M.H. are inventors on a patent applied for by Yale University related to this work. S.W. is a share-holder and consultant of Translura, Inc. The remaining authors declare no competing interests.

Data and materials availability: All sequencing data (genomic DNA sequencing of the CRISPR screen cell library, RNA-seq, and CUT&RUN data), analyzed imaging data, and original code generated for this study are available for download at https://campuspress.yale.edu/wanglab/BARCFISH. All sequencing data will also be available at the Gene Expression Omnibus (GEO) upon publication. All original code will also be available on GitHub upon publication. Open-source codes for imaging data collection are available at https://github.com/ZhuangLab/storm-control. Raw imaging data are available from the corresponding author upon request and are not deposited online due to prohibitively large size.

Supplementary Materials

Materials and Methods

Figures S1 to S9

Tables S1 to S13

Supplementary Files 1-7

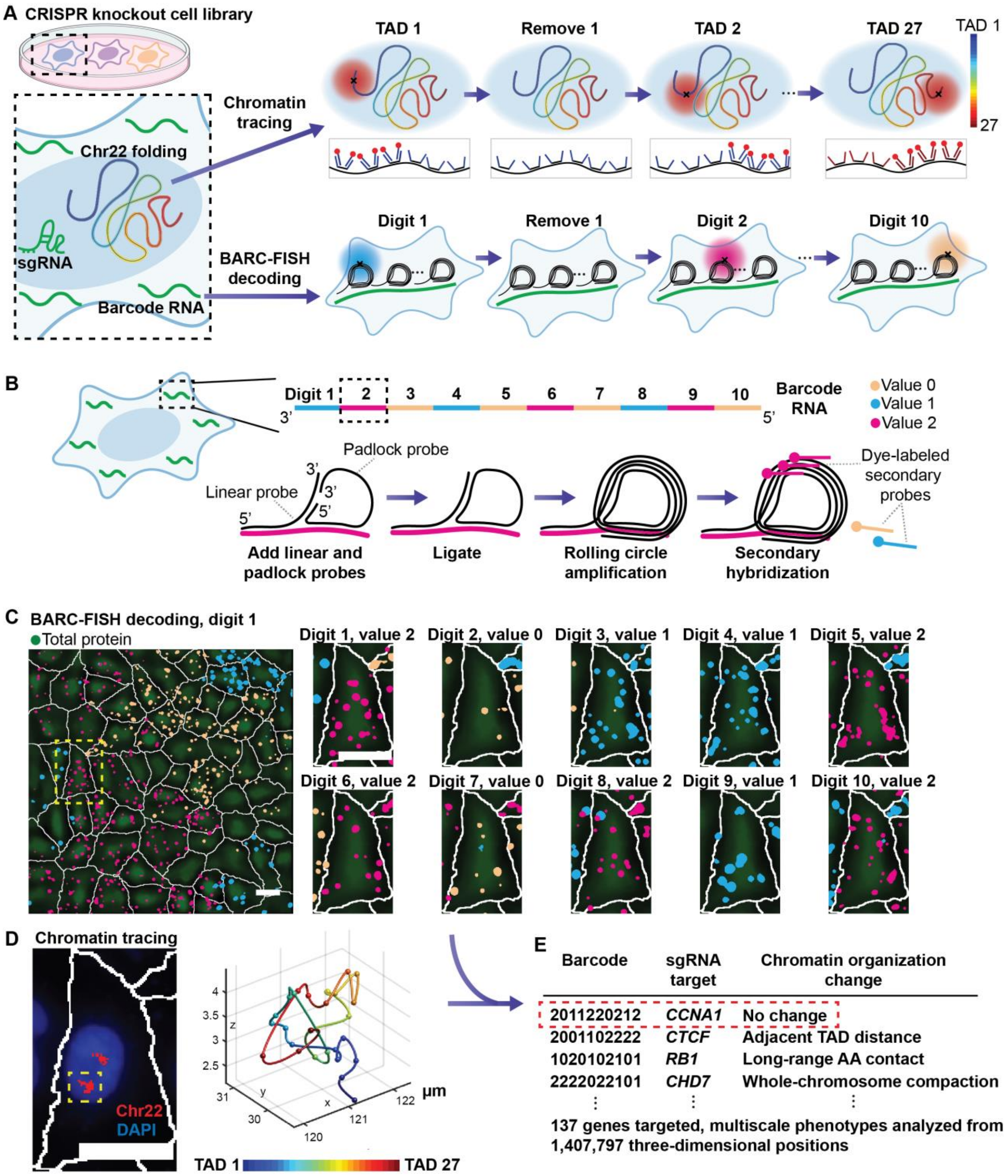

A CRISPR knockout cell library
Chromatin tracing
Chr22 folding
sgRNA
Barcode RNA
BARC-FISH decoding
TAD 1
Remove 1
TAD 2
TAD 27
TAD 1
27
Digit 1
Remove 1
Digit 2
Digit 10
B
Digit 1
2
3
4
5
6
7
8
9
10
3'
5'
Barcode RNA
Value 0
Value 1
Value 2
Padlock probe
Linear probe
3'
3'
5'
5'
Dye-labeled secondary probes
Add linear and padlock probes
Ligate
Rolling circle amplification
Secondary hybridization
C BARC-FISH decoding, digit 1
Total protein
Digit 1, value 2
Digit 2, value 0
Digit 3, value 1
Digit 4, value 1
Digit 5, value 2
Digit 6, value 2
Digit 7, value 0
Digit 8, value 2
Digit 9, value 1
Digit 10, value 2
D Chromatin tracing
Chr22
DAPI
4
3.5
3
2.5
z
31
30
y
x
120
121
122
µm
TAD 1
TAD 27
E
Barcode
sgRNA target
Chromatin organization change
2011220212
CCNA1
No change
2001102222
CTCF
Adjacent TAD distance
1020102101
RB1
Long-range AA contact
2222022101
CHD7
Whole-chromosome compaction
⋮
⋮
⋮
137 genes targeted, multiscale phenotypes analyzed from 1,407,797 three-dimensional positions

Figure 1. BARC-FISH and chromatin tracing enables image-based pooled CRISPR screen of chromatin and nuclear organization regulators. (**A**) Schematic of the screening approach. A lentivirus library encoding paired sgRNAs and barcode RNAs was transduced into human A549 cells expressing Cas9 protein. The organization of chr22 was determined by chromatin

tracing and the identity of the knockout gene was determined by BARC-FISH decoding of the barcode RNAs. For chromatin tracing, all 27 TADs spanning chr22 were sequentially visualized in a multiplexed DNA FISH procedure. For BARC-FISH decoding, 10 digits of the barcode were amplified and sequentially imaged. (**B**) A scheme of the BARC-FISH method. In each cell, the expressed barcode RNA was composed of 10 “digits”, and each digit had one of three different values (values 0, 1 and 2, represented by orange, cyan, and magenta respectively). Each digit was hybridized with a linear probe and padlock probe, enabling ligation of the padlock probe, which was then subjected to rolling circle amplification, generating an amplicon containing multiple copies of the digit sequences. Dye-labelled secondary probes were then introduced for imaging, reading out the value of the digit. (**C**) An example of BARC-FISH decoding. Left: A representative field of view from the screen with BARC-FISH signals shown in orange, cyan, and magenta, cell segmentation shown as white lines, and total protein stain in green. Right: The yellow-boxed cell in the left panel in 10 rounds of decoding. Scale bars: 20 µm. (**D**) Chromatin tracing of the yellow-boxed cell in C. Left: An image of the cell, with the traces of the two copies of chr22 shown in red and DAPI stain shown in blue. Right: 3D chromatin trace of the yellow-boxed chromosome in the left panel. The 3D positions of each TAD were shown as pseudo-colored spots, connected with a smooth curve. Scale bar: 20 µm. (**E**) A simplified table of screen results. The results of BARC-FISH decoding and chromatin tracing were integrated to match perturbations with chromatin organization phenotypes.

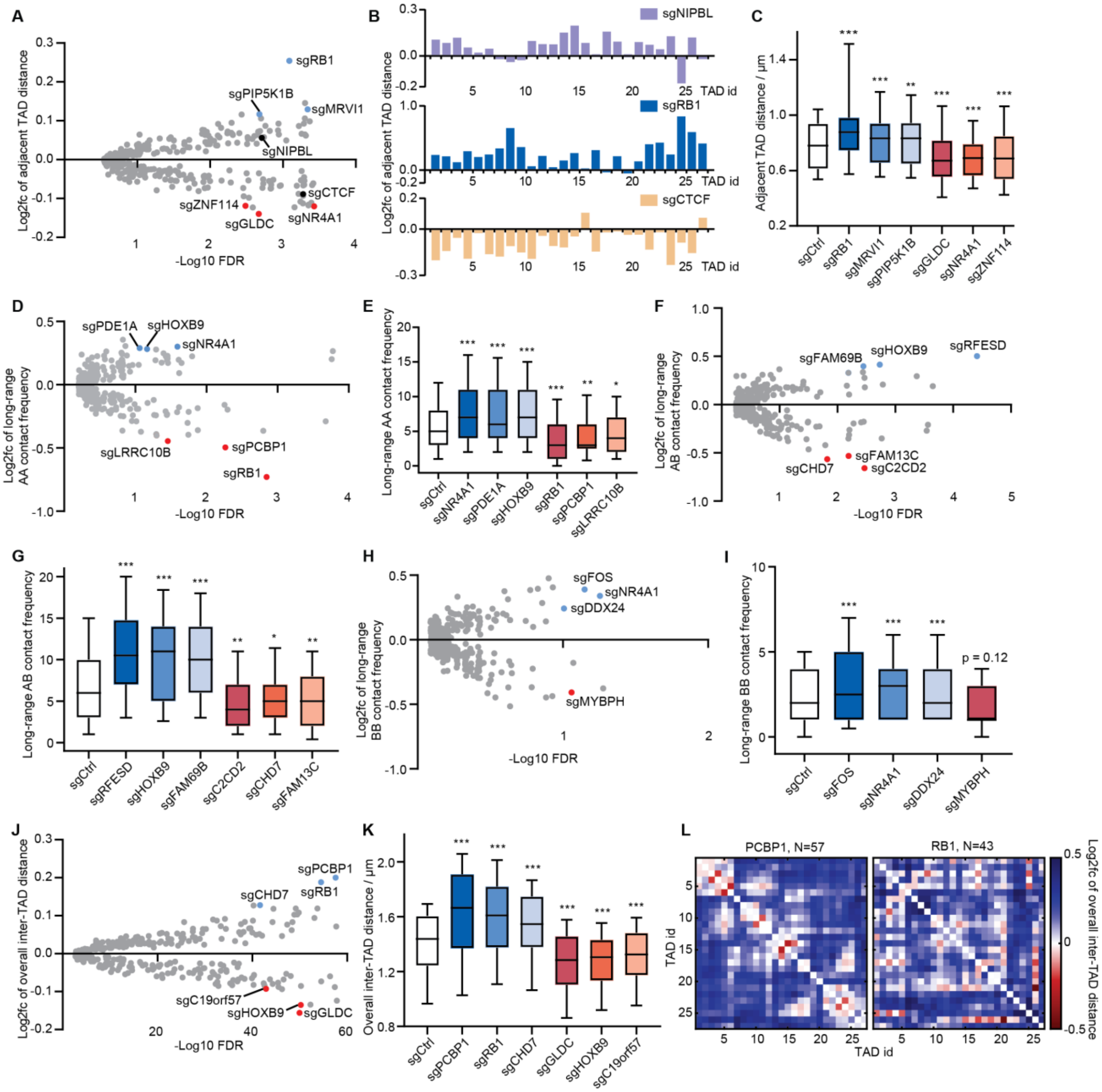


Figure 2. BARC-FISH screen identified regulators of multi-scale chromatin folding. (**A**) Log2 fold change (log2fc) of spatial distance between adjacent TADs versus -log10 false discovery rate (FDR) for each perturbation. Each dot represents a perturbation in the screen library. In all volcano plots, the top hits (nuclear proteins with the largest log2fc and FDRs<0.1) in both directions are indicated with blue (knockout leads to upregulation) and red dots (knockout leads to downregulation), respectively. The top candidate genes which when knocked

out led to increased adjacent TAD distances are: RB1, MRVI1 and PIP5K1B; the top candidate genes which when knocked out caused decreased adjacent TAD distances are: GLDC, NR4A1 and ZNF114. Positive controls (NIPBL and CTCF) are marked in black. (**B**) Log2 fold change of adjacent TAD distance across chr22 for selected hits. (**C**) Spatial distances between adjacent TADs for non-targeting control and selected hits. (**D**) Log2 fold change of long-range A-A contact frequency versus -log10 FDR for each perturbation. Top three hits in both directions including NR4A1, PDE1A, HOXB9, RB1, PCBP1 and LRRC10B are labeled. (**E**) Long-range A-A contact frequencies for non-targeting control and selected hits. (**F**) Log2 fold change of long-range A-B contact frequency versus -log10 FDR for each perturbation. Top three hits in both directions, including RFESD, HOXB9, FAM69B, C2CD2, CHD7 and FAM13C, are labeled. (**G**) Long-range A-B contact frequencies for non-targeting control and selected hits. (**H**) Log2 fold change of long-range B-B contact frequency versus -log10 FDR for each perturbation. Top hits in both directions, including FOS, NR4A1, DDX24 and MYBPH, are labeled. (**I**) Long-range B-B contact frequencies for non-targeting control and selected hits. (**J**) Log2 fold change of overall inter-TAD distances versus -log10 FDR for each perturbation. Top three hits in both directions, including PCBP1, RB1, CHD7, GLDC, HOXB9 and CUL1, are labeled. (**K**) Overall inter-TAD distances for non-targeting control and selected hits. (**L**) Log2 fold change of individual overall inter-TAD distances in chr22 for selected hits. P values in (C) and (K) were calculated by two-sided Wilcoxon signed rank test. P values in (E), (G) and (I) were calculated by two-sided Wilcoxon rank sum test. In all box plots throughout the manuscript, the boxes cover the 25th to 75th percentiles, the whiskers cover the 10th to 90th percentiles, and the line in the middle of the boxes represents the median value. For all relevant panels, significance is represented as *p<0.1. **p<0.05. ***p<0.01.

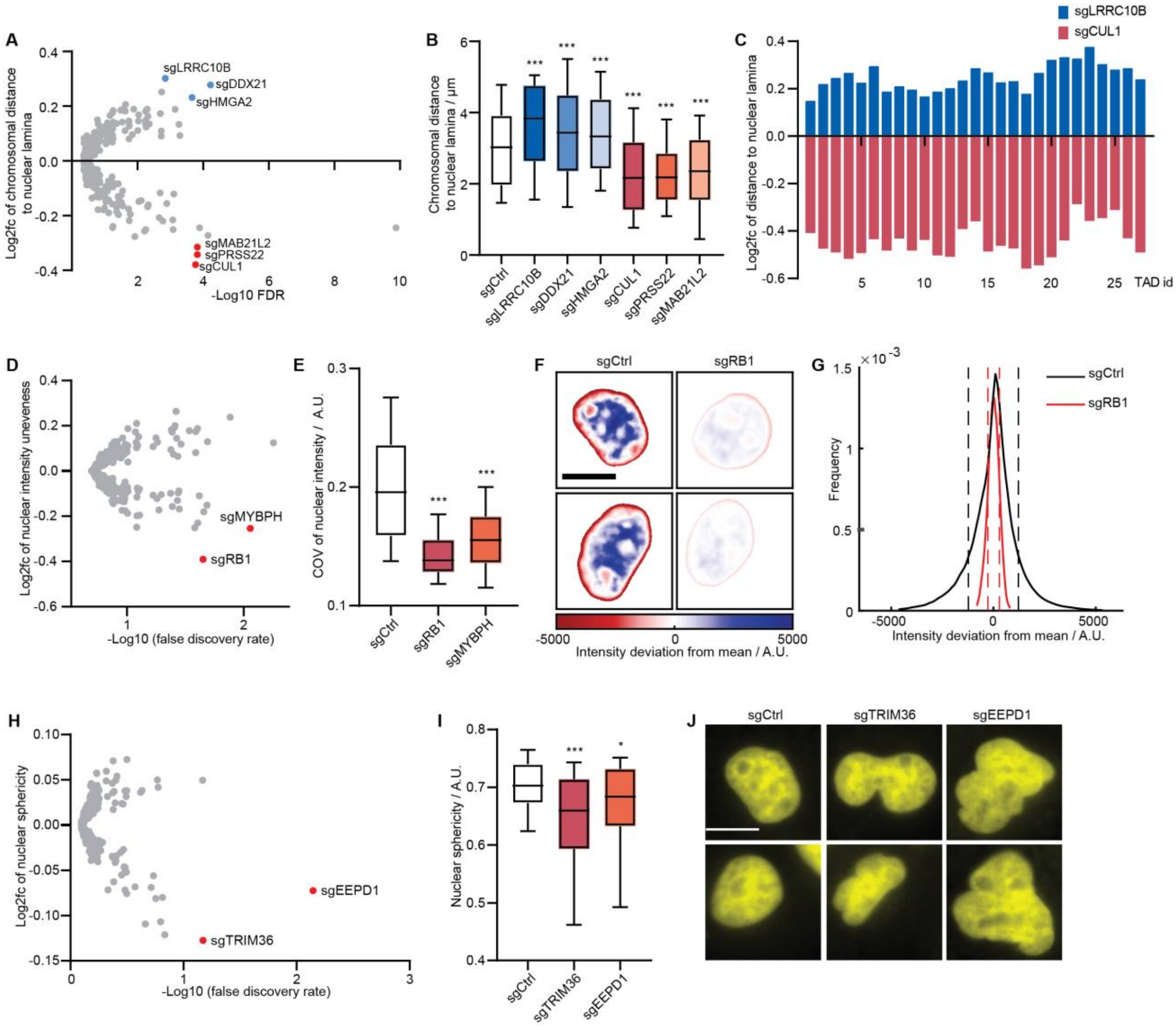


Figure 3. BARC-FISH screen identified hits that regulate chromosome association with nuclear lamina and the morphological properties of the nucleus. (**A**) Log2 fold change of chromosome distance to nuclear lamina versus -log10 FDR. Top hits in both directions, including LRRC10B, DDX21, HMGA2, CUL1, PRSS22 and MAB21L2, are labeled. (**B**) Chromosome distance to nuclear lamina of non-targeting control and selected hits. (**C**) Log2 fold change of distances between each TAD on chr22 to nuclear lamina of selected hits. (**D**) Log2 fold change of nuclear intensity unevenness (measured as coefficient of variation of nuclear voxel intensities) versus -log10 FDR. Top hits RB1 and MYBPH are labeled. (**E**) Nuclear

intensity unevenness of non-targeting control and selected hits. (**F**) Heatmap of nuclear intensity deviation from mean intensity of representative nuclei from non-targeting control (left column) and selected hit sgRB1 (right column). Scale bar: 10 μm. (**G**) Voxel intensity distribution of all nuclei from non-targeting control (black curve) and selected hit sgRB1 (red curve). Dashed lines indicate the standard deviations of the indicated distributions. (**H**) Log2 fold change of nuclear sphericity versus -log10 FDR. Top hits TRIM36 and EEPD1 are labeled. (**I**) Nuclear sphericity of non-targeting control and selected hits. (**J**) Representative nuclei images of non-targeting control (left column) and selected hits that regulate nuclear sphericity, TRIM36 (middle column) and EEPD1 (right column). Each column contains the DAPI staining of two representative cells from the indicated perturbation. Scale bar: 10 μm. P values in (B), (E) and (I) were calculated by two-sided Wilcoxon rank sum test.

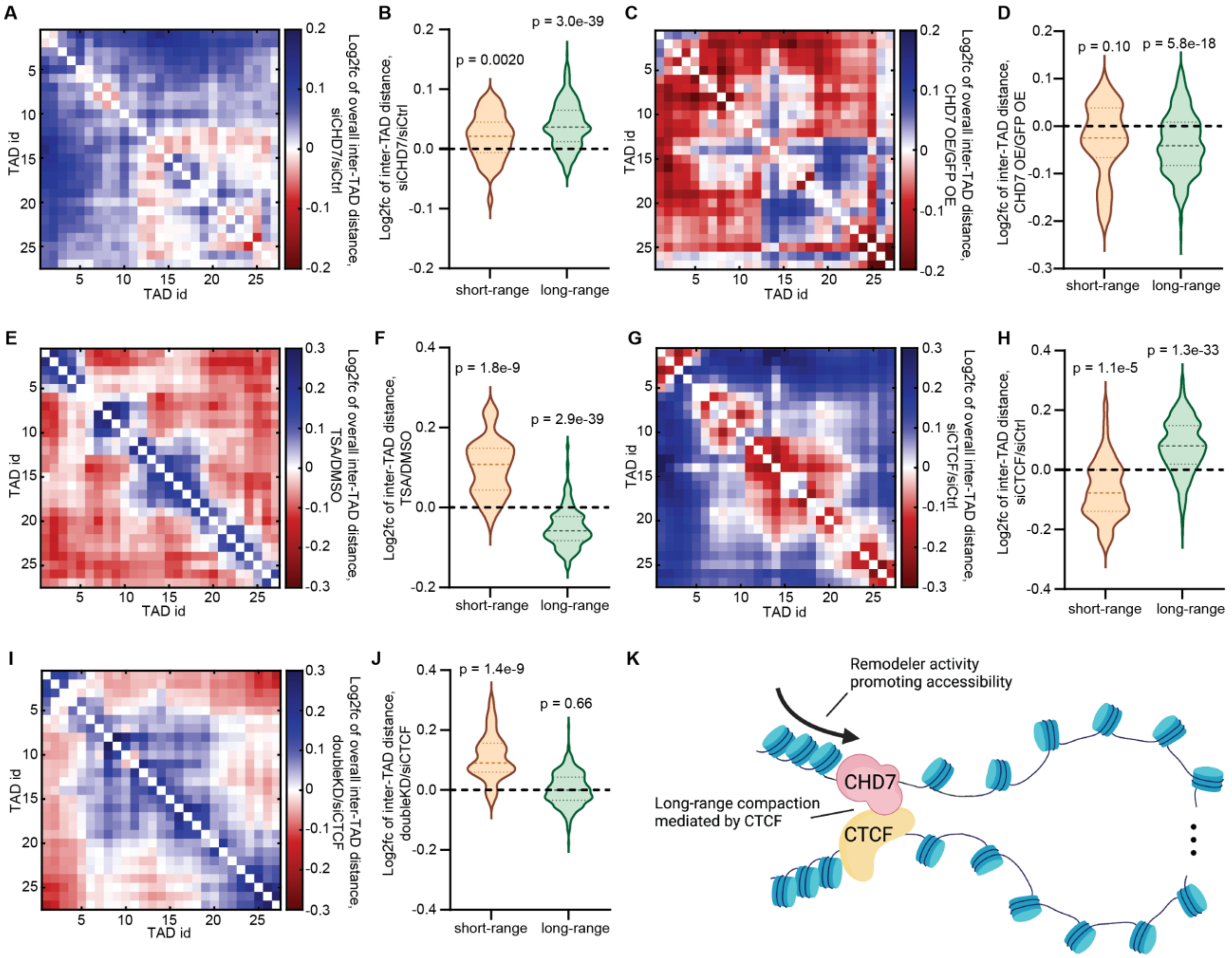


Figure 4. Long-range compaction by CHD7 does not operate through HDACs, but depends on CTCF. (**A**) Log2 fold change of inter-TAD distance of siCHD7 compared to siCtrl. Number of traces analyzed: 3,558 (siCtrl), 4,134 (siCHD7). (**B**) Log2 fold change of short-range (defined as spatial distances between genomic regions that are less than 3Mb apart) and long-range (defined as spatial distances between genomic regions that are more than 3Mb apart) inter-TAD distances between siCHD7 and siCtrl. (**C**) Log2 fold change of inter-TAD distance of CHD7 overexpression compared to GFP overexpression. Number of traces analyzed: 3,157 (GFP OE), 1,174 (CHD7 OE). (**D**) Log2 fold change of short-range and long-range inter-TAD distances between CHD7 and GFP overexpression. (**E**) Log2 fold change of inter-TAD distance of TSA-treated cells compared to DMSO-treated cells. Number of traces analyzed: 1,214 (DMSO), 2,223

(TSA). (**F**) Log2 fold change of short-range and long-range inter-TAD distances between cells with TSA and DMSO treatment. (**G**) Log2 fold change of inter-TAD distance of siCTCF compared to siCtrl. Number of traces analyzed: 5,768 (siCtrl), 4,226 (siCTCF). (**H**) Log2 fold change of short-range and long-range inter-TAD distances between siCTCF and siCtrl. (**I**) Log2 fold change of inter-TAD distance of double knockdown of CTCF and CHD7 (doubleKD) compared to siCTCF. Number of traces analyzed: 3,560 (doubleKD). (**J**) Log2 fold change of short-range and long-range inter-TAD distances between doubleKD and siCTCF. (**K**) Working model of CHD7 compacting chromatin over long range (cartoon created with Biorender.). All chromatin tracing experiments in this figure were done in the A549 cell background, targeting chr22. P values were calculated by two-sided Wilcoxon signed rank test. For all violin plots, the median log2 fold change values are indicated by colored dashed line, the quartiles are indicated by colored dotted line, and value 0 is marked by black dashed line.

Materials and Methods

Barcode and probe sequence design

Design of BARC-FISH barcodes, linear and padlock probes. We first designed 60 barcode-targeting regions of linear and padlock probes using 46 published 20-nt sequences that only contain A, T and C (*83*, *84*) and 14 newly designed sequences. The 14 extra sequences were generated using a previously introduced orthogonal 25-nt oligonucleotide dataset (*85*), by trimming off 5 consecutive nucleotides on the 5’ or 3’ end of the 25-nt sequences, and only sequences that met the following three criteria were retained: 1) the sequence only contains A, T and C; 2) the sequence does not contain 4 or more consecutive C’s; and 3) the percentage of C is between 40-45%. The newly designed 20-nt sequences were pooled with the 46 published sequences, and all these 20-nt sequences were BLASTed among themselves to ensure they were orthogonal to each other (*86*). Finally, we BLASTed all possible pairs of concatenated 20-nt sequences separated by a G spacer against human genome and transcriptome to ensure they were orthogonal to endogenous nucleotide sequences (*86*). We then grouped the 60 sequences into 30 pairs, and reverse-complemented the sequences to generate the DNA segments of barcode digits, each of which is 41-nt long (two 20-nt sequences with a 1-nt spacer). We adopted the backbone sequences of linear and padlock probes from a previous work (*47*) to generate the full-length linear and padlock probe sequences, with the following modification: The backbone of padlock probes of all value 0 sequences for all barcode digits carried a 20-nt secondary probe targeting region. For value 1 and 2 sequences of the barcode digits, the 20-nt region of the padlock probe that binds to the barcode RNA serves as the secondary probe targeting region. 27 secondary probe sequences were previously reported (*83, 87*). The other 3 secondary probe sequences were designed with a previously introduced procedure (*88*). The linear and padlock probes, and secondary probes labeled with ATTO565, Alexa Fluor 647 and Alexa Fluor 750 dyes were individually ordered from Integrated DNA Technologies, Inc (IDT). A list of DNA barcode segment sequences and their corresponding linear, padlock and secondary probe sequences are included in Table S3.

BARC-FISH helper probe design. To improve the hybridization efficiency of the linear and padlock probes, we designed helper probes targeting the flanking regions of the barcode RNA, which could help open potential secondary structures and linearize the RNA molecules. To design the helper probes, we adapted the algorithm of a published probe designing tool ProbeDealer (*89*). Briefly, three RNA regions were used as the input targeting regions for ProbeDealer: an 888-nt region upstream of the protospacer, a 108-nt region between the protospacer and barcode, and a 310-nt region downstream of a unique molecular identifier region (for the region layout, refer to later sections on plasmid library construction). The probes were generated with the following constraints: The probe length was 30-nt, with at least 1-nt spacing between adjacent probes; the allowed melting temperature range was 66-100°C; the allowed GC content range was 30-90%; the melting temperature of internal secondary structure of each probe was no greater than 76°C; the melting temperature of cross-hybridization regions among the probes was no greater than 72°C; and the probes containing consecutive repeat of more than five identical nucleotides were excluded. The probe sequences were further BLASTed against human genome and transcriptome to ensure the hybridization specificity (*86*). Probes with alignments to either human genome or transcriptome were excluded from the selection. In total, 36 helper probes were generated and ordered though IDT. The sequences of the helper probes are listed in Table S4.

Chromatin tracing primary probes. We used the primary probes of human chr21 and chr22 from a previous publication (*46*) with slight modifications on secondary probe binding regions. Specifically, the probes consisted of the following segments from the 5’ to 3’: 1) a 20-nt priming regions at 5’ end, 2) a secondary probe binding region of 17-19 nt, 3) a 30-nt primary binding region complementary to genomic DNA, 4) a same secondary probe binding region of 17-19 nt, and 5) a 20-nt priming region at 3’ end. The secondary probe binding regions were generated by trimming the 5’ or 3’ ends of previously published 30-nt secondary probe binding regions (*46*), so that the Tm of the trimmed secondary probe binding regions became 50-57°C.

Barcode plasmid library construction

The DNA sequences of the 10-digit barcodes each comprised ten 41-nt digit segments, with a nucleotide ‘C’ separating adjacent digits. Each digit contained one of three different 41-nt sequences, representing the three different values (0, 1, 2) of the ternary digit. The values from the 10 digits were randomly assembled. Therefore, theoretically 3^{10} or 59,049 possible unique combinatorial barcodes could be generated. To assemble the barcode library, we used a pooled barcode cloning strategy (fig. S2A) modified from approaches previously described (*45*, *90*): First, the 10 digits were divided into two fragments. Overlapping oligos encoding and bridging digit 1-5 were mixed to a final concentration of 100 nM each, and the mixture was subjected to phosphorylation and ligation. Oligos encoding digit 6-10 were treated in the same way in parallel. Then, the two 5-digit fragments were gel-purified and assembled into full-length barcodes by overlapping PCR. Next, the full-length barcodes were amplified by limited-cycle PCR. In this step, each barcode molecule was paired with a unique molecular identifier (UMI), a 20-bp random sequence, at the 3’ end for later next-generation sequencing (NGS), and overhangs at both ends for Gibson Assembly. The PCR products were gel-purified and then assembled into a plasmid backbone digested with MluI (New England BioLabs, R3198S) and SpeI (New England BioLabs, R3133S), through Gibson Assembly (New England Biolabs, E2621L). The backbone plasmid was modified from plasmid LentiGuide-Puro (Addgene, 52963), with an addition of a CMV promoter, an EGFP gene and two restriction sites, MluI and SpeI. Lastly, the Gibson reaction products were purified by isopropanol precipitation and electroporated into Endura electrocompetent cells (Lucigen, 60242-2). The transformed bacteria were spread onto LB agar plates containing 100 μg/mL ampicillin and incubated overnight at 37°C. On the next day, ~70 million colony forming units grown on the plates were collected and subjected to maxi-prep (Qiagen, 12362) to extract the barcode plasmid library. The barcode plasmid library was then used to generate CRISPR screen plasmid library. All oligos and primers were ordered from IDT. The oligo and primer sequences for constructing the barcode plasmid library are provided in Table S5. The modified backbone sequence with annotations is included in Supplementary File S1. An example barcode plasmid sequence with annotations is included in Supplementary File S2.

sgRNA plasmid library construction

To construct the sgRNA plasmid library, we selected protospacer sequences from a previous publication (*91*). The sgRNA fragments were amplified from a CustomArray oligo pool via limited-cycle PCR. The backbone was digested with FastDigest Esp3I (Thermo Fisher Scientific, FD0454) together with Alkaline Phosphatase (Thermo Fisher Scientific, EF0654) overnight at 37°C. The backbone plasmid was modified from plasmid LentiGuide-Puro with an addition of a CMV promoter and an EGFP gene. The sgRNA fragments and digested backbone were gel-purified and assembled through Gibson Assembly. The Gibson reaction products were purified by isopropanol precipitation and the electroporated into Endura electrocompetent cells. The cells were plated onto LB agar plates containing 100 µg/mL ampicillin and incubated at 37°C overnight. About 2.5 million colony forming units were collected and inoculated into 200 mL LB liquid medium containing 100 µg/mL ampicillin, and grown overnight to amplify the sgRNA plasmid library. The sgRNA plasmid library was maxi-prepped from the overnight culture and used for the construction of CRISPR screen plasmid library. All primers were ordered from IDT. The protospacer sequences, ordered CustomArray oligo sequences and primer sequences are provided in Table S6. The modified backbone sequence with annotations is included in Supplementary File S3. An example sgRNA plasmid sequence with annotations is included in Supplementary File S4.

CRISPR screen plasmid library construction

We adopted the cloning strategies from previous publications (*45*, *92*) to construct the CRISPR screen plasmid library (fig. S2A). In brief, sgRNA, barcode and backbone were assembled through Gibson Assembly. The barcode fragments were amplified by limited-cycle PCR from the pre-made barcode plasmid library. Similarly, the sgRNA fragments were PCR-amplified from the pre-made sgRNA plasmid library. The backbone plasmid was modified from plasmid CROPseq-Guide-Puro (Addgene, 86708), with an addition of a CMV promotor and removal of the gRNA scaffold. The modified plasmid was restriction digested with FastDigest Esp3I overnight at 37°C. The restriction reaction product was treated with Alkaline Phosphatase to remove the phosphate groups from the linearized backbone. The DNA fragments of sgRNA, barcode and backbone were gel purified and mixed for Gibson Assembly. The Gibson reaction products were purified by isopropanol precipitation and electroporated into Endura electrocompetent cells. Electroporated cells were grown overnight at 37°C on LB agar plates containing 100 µg/mL ampicillin. To restrict the number of barcodes paired with each sgRNA and allow unique projection from each barcode to a single sgRNA, we used a bottlenecking strategy: Only 4,000~5,000 bacterial colonies were collected from the plates and cultured in 200 mL LB liquid medium overnight at 37°C with shaking at 225 rpm. The CRISPR screen plasmid library was then extracted and purified by maxi-prep. All primer sequences are provided in Table S7. The modified backbone sequence with annotations is included in Supplementary File S5. An example CRISPR screen plasmid sequence with annotations is included in Supplementary File S6.

Plasmid construction for CHD7 and GFP overexpression

The GFP lentiviral overexpression vector, pLenti-GFP, was purchased from OriGene (OriGene, PS100071). To construct the CHD7 lentiviral overexpression construct, we cloned the human CHD7 open reading frame (ORF) into the same vector upstream of GFP, together with a P2A sequence placed between CHD7 and GFP for self-cleavage. Specifically, the CHD7 ORF was PCR amplified from a plasmid (Addgene, 89460) using Phusion High-Fidelity Master Mix (NEB, M0531L) following the manufacturer’s instructions. The P2A sequence was added to the 3’ end via reverse primer. Then, DpnI (NEB, R0176S) was directly added to the PCR mixture and incubated at 37°C for 1 hour to digest the plasmid template. The PCR products of correct size were confirmed by agarose gel electrophoresis and gel purified. To prepare the backbone, pLenti-GFP was digested by restriction enzymes AsiSI (NEB, R0630S) and MluI-HF (NEB, R3198S), and dephosphorylated by Shrimp Alkaline Phosphatase (NEB, M0371S) at 37°C for

3.5 hours. The resulting backbone was gel purified, mixed with the purified PCR products, and subjected to Gibson Assembly using NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621L) following the manufacturer's instructions. The reaction mixture was then column purified and transformed into chemically competent E. coli Stbl3 (ThermoFisher, C737303) following the manufacturer's instructions. The transformed cells were plated onto LB agar plates containing 34 µg/mL chloramphenicol and incubated overnight at 37℃. Several clones were picked and individually inoculated into LB liquid medium overnight at 37℃, ~225 rpm. The plasmids of individual clones were extracted using QIAprep Spin Miniprep Kit (Qiagen, 27106) following the manufacturer's instructions, and the correct sequence was confirmed by Sanger sequencing. The primer sequences for cloning and Sanger sequencing are provided in Table S8. The plasmid sequence with annotations of the reconstructed CHD7 lentiviral overexpression vector, pLenti-CHD7, is included in Supplementary File S7.

Lentivirus production and cell line construction

Lentivirus production. The HEK-293FT cells (Thermo Fisher Scientific, R70007) were cultured to be 70-90% confluent at the point of transfection to produce lentiviruses. The Cas9 plasmid lentiCas9-Blast (Addgene, 52962), the GFP overexpression plasmid pLenti-GFP, the CHD7 overexpression plasmid pLenti-CHD7, or the CRISPR screen plasmid library was mixed with helper plasmids psPAX2 (Addgene, 12260) and pVSV-G (Addgene, 138479) together with Lipofectamine 2000 (Thermo Fisher Scientific, 11668019) according to the manufacturer's instructions. The mixture was then added into the cell culture and incubated for two days. After the lentiviral transfection, the lentivirus supernatant was collected from the cell culture and filtered through a 0.45 µm strainer (Millipore, SLHAR33SS) to remove cell debris. The filtered lentivirus supernatant was concentrated using Amicon Ultra-15 Centrifugal Filter Unit (Millipore, UFC910024) following the manufacturer's instructions. The concentrated lentivirus supernatant was aliquoted and stored at -80℃.

Generation of clonal A549-Cas9 cells. A549 cells (ATCC, CCL-185) were infected with lentivirus generated from the lentiCas9-Blast plasmid. Cells were under Blasticidin selection at 10 µg/mL until all untransduced cells were killed. The Blasticidin-selected, polyclonal cells were sorted into single cells by flow cytometry and plated on 96-well plates. The single cells were clonally expanded and analyzed by immunofluorescence to verify Cas9 expression. The clonal selection was important to reduce the heterogeneity in the cell background in the screen.

Generation of BARC-FISH CRISPR screen cell library. Clonal A549-Cas9 cells were cultured to 60-80% confluency and transduced with the CRISPR screen lentivirus supernatant at an MOI < 0.3 to ensure that most cells received only one genetic perturbation. Two days after the lentiviral infection, Puromycin (Thermo Fisher Scientific, A1113803) was added into the media at 3 µg/mL to select the cells with resistance. Cells were under Puromycin selection for 10-12 days until the cell library was established. The cell library was then aliquoted into frozen stocks and stored in liquid nitrogen.

Generation of CHD7 and GFP OE cell line. Clonal A549-Cas9 cells were infected with the concentrated lentivirus produced from the overexpression plasmid pLenti-GFP or pLenti-CHD7. Puromycin selection started two days after lentiviral infection at a concentration of 3 µg/mL in growth media, and lasted for 5-6 days until the stable cell lines were established.

Cell culture for cloning and imaging

A549 cells and the clonal A549-Cas9 cells (introduced in the section above, derived from the A549 cells) were maintained in F-12K media (Corning, 10-025-CV) supplemented with 10% FBS (VWR, 97068-091) and 1% Pen-Strep (Thermo Fisher Scientific, 15140122) at 37°C in 5% CO_2. hTERT RPE-1 cells (ATCC, CRL-4000) were maintained in DMEM:F-12 media (ATCC, 30-2006) supplemented with 10% FBS (VWR, 97068-091) and 1% Pen-Strep (Thermo Fisher Scientific, 15140122) at 37°C in 5% CO_2. To prepare imaging samples, coverslips (Bioptechs, 40-1313-03193) were disinfected by 15-min UV treatment on each side prior to use. To culture cells for conducting the screen, A549-Cas9 cells transduced with BARC-FISH screen library at the same passage were seeded onto the disinfected coverslips at 10% density in F-12K media supplemented with 15% FBS and 1% Pen-Strep and cultured for 6 days. The higher percentage of FBS in the media allowed the cells to produce more barcode RNA copies. The media was also supplemented with 3 µg/mL Puromycin and 10 µg/mL Blasticidin (Thermo Fisher Scientific, R21001) to ensure sgRNA and Cas9 expression. Media were refreshed every 3-4 days. For lentivirus production, HEK-293FT cells were cultured in DMEM medium supplemented with GlutaMAX-I (Thermo Fisher Scientific, 10569-010), 1× MEM Non-Essential Amino Acids (Thermo Fisher Scientific, 11140-050), 10% FBS and 1% Pen-Strep at 37°C in 5% CO_2, and were used within 10 passages. All cells were routinely tested of mycoplasma contamination via Yale Molecular Diagnostics Laboratory.

Detection of sgRNA-barcode associations in the cell library

Next-generation sequencing (NGS) library preparation for mapping sgRNA-barcode associations. To determine the sgRNA-barcode correspondence in the screen library, we PCR-amplified the cell genomic DNA for NGS. Because the total length of the sgRNA-barcode-UMI cassette is longer than the maximum sequencing length of NGS (300 bp paired-end sequencing on Illumina MiSeq), we generated two sequencing libraries: an sgRNA-barcode-UMI library and a barcode-UMI library. The sgRNA-UMI correspondence can be determined by the sgRNA-barcode-UMI library and the barcode-UMI correspondence can be determined by the barcode-UMI library. Then the sgRNA-barcode correspondence can be determined through their common UMI associated with both the sgRNA and the barcode on the same molecule. Briefly, the genomic DNA was extracted from the CRISPR screen cell library using Zymo Quick-DNA Miniprep Kit (Zymo Research, D4068) following the manufacturer’s protocol. To achieve a sequencing coverage of >500, more than 2.5 million cells were harvested to cover the ~5,000 barcode varieties. We amplified all harvested genomic DNA and limited the number of amplification cycles to 20-23 to reduce amplification bias. For PCR amplification, each 50 µL reaction mixture was composed of 1 µg of genomic DNA, 500 nM of each forward and reverse primers, 5% DMSO and 1× NEBNext High-Fidelity PCR master mix (New England BioLabs, M0541L) following the manufacturer’s instructions. The PCR primers contained flow-cell binding sequence, sequencing index, sequencing primer binding region, and a 9-13 bp random sequence that improved the library diversity. After the reactions were completed, the PCR products were purified and concentrated using Zymo DNA Clean & Concentrator (Zymo Research, D4030) following the manufacturer’s instructions. Lastly, the concentrated products were run on a 2% agarose gel to validate the correct DNA amplicon size. The DNA library amplicons with the correct size were then gel-purified and eluted with elution buffer (10 mM Tris-HCl, pH 8.5) or MilliQ water. The DNA libraries were sequenced on Illumina MiSeq system in 2×300 bp format. The PCR primers were all purchased from IDT and the sequences are provided in Table S9.

sgRNA-barcode NGS analysis. As described above, we generated two sequencing libraries, a sgRNA-barcode-UMI library and a barcode-UMI library, and determined the sgRNA-barcode correspondence by the common UMI sequence. Briefly, the sequencing reads were filtered by read length and quality score. From the sgRNA-barcode-UMI library, protospacer and UMI sequences were extracted from the reads, and the protospacer-UMI look-up table was generated accordingly. The protospacer sequences were aligned with the sequences in the pre-designed sgRNA oligo library such that the reads with improper protospacers were removed. From the barcode-UMI library, the UMI sequences were extracted and compared to the UMI found in the sgRNA-barcode-UMI library. Only reads with common UMIs were retained for further analysis. Because the read length from a single end did not cover the full length of the barcode, the sequencing read from each end only contained a partial barcode sequence. The digit values from the partial barcode sequences were extracted by BLAST (*86*) against the pre-designed digit sequences. The two decoded partial codes assigned to the two ends of the same sequence were then assembled into one full-length complete code by the overlapping region of the two partial codes. The partial codes that failed to overlap were excluded. The UMI-barcode lookup table was then established. The two look-up tables were merged to determine the sgRNA-barcode correspondence (codebook) through the shared UMIs. This codebook allowed analyses of the percentage of good codes (codes uniquely associated with one sgRNA) versus bad codes (codes associated with multiple sgRNAs) (fig. S1D). The final codebook containing both the good codes that each uniquely associate with one sgRNA and the bad codes that each have multiple sgRNA projections is listed in Table S10.

Validation of CRISPR knockout efficiency

To validate the CRISPR knockout efficiency, we constructed individual CRISPR knockout cell lines for selected sgRNAs and utilized NGS of genomic DNA to analyze the frameshift indel mutations. The cloning procedure of individual sgRNA plasmids was similar to previously described in “sgRNA plasmid library construction”, except that we directly ordered double-stranded sgRNA fragments from IDT for Gibson Assembly. Cell line generation procedure was identical to “Generation of BARC-FISH CRISPR screen cell library”. We then performed genomic DNA extraction and PCR amplification following a procedure similar to previously described in “Next-generation sequencing (NGS) library preparation for mapping sgRNA-barcode associations”. The amplified region was 250-450 bp centered around the sgRNA targeting site. The PCR amplicons were sequenced on Illumina MiSeq system in 2×250 bp format, achieving ~50,000 reads per sgRNA. The indel rate analysis was performed and delivered by Genewiz.

Probe synthesis

The template oligo pool for synthesizing chromatin tracing primary probes was ordered from CustomArray, GenScript. We adopted a previously described probe synthesis workflow (*87*, *88*) of limited-cycle PCR, *in vitro* transcription, reverse transcription, alkaline hydrolysis and column purification. The PCR primers and reverse transcription primers were ordered from IDT. The sequences of the template oligo library and primers are included in Table S11.

Imaging sample preparation

Geminin antibody staining. All steps were performed at room temperature unless otherwise described. Cells grown on coverslips were briefly rinsed with DPBS after media removal, and then fixed in 4% paraformaldehyde (PFA) (Electron Microscopy Sciences, 15710) diluted in DPBS for 10 minutes followed by three DPBS washes. Cells were then permeabilized with 0.5% v/v Triton X-100 in DPBS for 10 minutes followed by three DPBS washes. After permeabilization, cells were blocked in blocking buffer (1% w/v bovine serum albumin (Sigma-Aldrich, A9647), 0.1% v/v Tween-20 in DPBS) supplemented with 0.1% v/v murine RNase inhibitor (MRI) (New England Biolabs, M0314L) for 30 minutes, followed by primary antibody incubation for 1 hour with 1:100 diluted anti-Geminin antibody (Abcam, ab195047) in blocking buffer supplemented with 1% v/v MRI. Unbound primary antibody was washed three times with 0.1% v/v Tween-20 in DPBS (DPBSTw), 5 minutes each, followed by a 1-hour incubation with 1:1000 diluted Alexa Fluor 488-labeled secondary antibody (Invitrogen, A11034) in blocking buffer supplemented with 1% v/v MRI. Starting from secondary antibody incubation, samples were protected from light during all steps. Excessive secondary antibody was washed off by washing three times, 5 minutes each with DPBSTw, and samples were post-fixed in 4% PFA in DPBS for 10 minutes followed by 3 DPBS washes. Post-fixed samples were next used for the BARC-FISH procedure in the screen experiments and for chromatin tracing primary probe hybridization in the validation experiments.

BARC-FISH. Helper probes and BARC-FISH linear and 5’-phosphorylated padlock probes were ordered from IDT. A total of 36 helper probes were pooled in an equimolar manner to a concentration of 5.6 µM for each probe. A total of 60 probes for BARC-FISH (30 linear probes and 30 padlock probes) were pooled in an equimolar manner to a concentration of 1.67 µM for each probe. Immediately prior to use, the pooled BARC-FISH probes were heat-denatured at 90°C for 2-5 minutes, and then cooled down to room temperature. Samples with Geminin staining were pre-hybridized in 2× saline-sodium citrate buffer (2×SSC) containing 20% v/v formamide and 0.1% v/v Tween-20 for 10 minutes, and then 50 µL of primary hybridization buffer was added to each sample, which contained 20% v/v formamide, 0.1 mg/mL salmon sperm DNA (Invitrogen, 15632-011), 800 nM each of the 60 BARC-FISH probes, 100 nM each of the 36 helper probes and 1% v/v MRI in 2×SSC. Samples were then incubated at 37°C for 16-20 hours. To remove excessive primary probes, samples were washed twice in 2×SSC containing 40% v/v formamide and 0.1% v/v Tween-20, 15 minutes each, followed by a third wash for 20 minutes at 37°C in buffer containing 4×SSC, 1×DPBS, 0.1% v/v Tween-20 and 0.1% v/v MRI. Samples were then briefly rinsed twice with DPBSTw prior to the T4 ligation step. The T4 ligation mixture contained 0.2 mg/mL BSA (New England Biolabs, B9000S), 1 mM extra-supplemented dithiothreitol (DTT) (Thermo Scientific, R0861), 1 mM extra-supplemented ATP (Thermo Scientific, R0441), 0.5 U/µL T4 DNA ligase (Thermo Scientific, EL0014) and 1% v/v MRI in 1× T4 ligase buffer (Thermo Scientific, EL0014). To perform the T4 ligation step, samples were incubated with 50 µL of T4 ligation mixture for 2 hours, followed by two washes in DPBSTw. Samples were then incubated in 30°C water bath for 5 hours with 50 µL of RCA mixture that contained 250 µM dNTP (New England Biolabs, N0447L), 1 mM extra-supplemented DTT, 0.2 mg/mL BSA, 1 U/µL phi29 enzyme (Thermo Scientific, EP0092) and 1% v/v MRI in 1× phi29 buffer (Thermo Scientific, EP0092). After the RCA step, samples were washed in DPBSTw, post-fixed with 4% PFA in DPBS for 30 minutes, and washed with DPBS. Post-BARC-FISH samples were used for the chromatin tracing primary probe hybridization in the screen experiments.

Chromatin tracing primary probe hybridization. Post-BARC-FISH samples (for screen experiments) or post-Geminin staining samples (for validation experiments) were briefly rinsed in DPBS, and then incubated with 0.1M HCl diluted in water for 5 minutes. After two DPBS washes, samples were treated with 0.1 mg/mL RNase-A (AB12023-00100) diluted in DPBS for 45 minutes at 37°C, followed by two DPBS washes and one 2×SSC wash. Subsequently, samples were pre-hybridized in 2×SSC containing 50% v/v formamide and 0.1% v/v Tween-20 for 30 minutes and carefully dried by dipping on tissue paper to remove excessive pre-hybridization buffer. 25 µL of hybridization buffer containing 50% v/v formamide, 20% v/v dextran sulfate (Millipore, S4030) and 4 µM (total concentration) chromatin tracing primary probes in 2×SSC was applied to a glass slide, and the sample coverslip was carefully flipped and placed on top so that the glass slide and coverslip "sandwiched" the hybridization buffer, with the cells submerged into the hybridization buffer. The samples were then heat-denatured on an 86°C heat block (with a surface temperature of ~80°C) for 3 minutes with the glass slide touching the heat block, and incubated in a humid chamber for 16-20 hours at 37°C. To remove excessive primary probes, samples were washed twice in a 60°C water bath with 0.1% v/v Tween-20 in 2×SSC, 15 minutes each, followed by a third 15-minute wash with 0.1% v/v Tween-20 in 2×SSC at room temperature. Samples were then briefly rinsed with 2×SSC, and 1:250,000 diluted yellow-green fiducial beads (Invitrogen, F8803) in 2×SSC were applied to the samples and incubated for 10 minutes. Excessive beads were removed by two 2×SSC washes. Post-hybridization samples were used for automated sequential imaging.

Automated sequential imaging

Imaging system setup. Four home-built fluorescence microscopes were used for image acquisition (*87*). Each microscope had a Nikon Ti2-U body, a Nikon CFI Plan Apo Lambda 60× oil objective lens (NA 1.40) and an automated focus-lock system (*87*, *93*). One setup has identical lasers, light paths and filters as introduced previously (*87*), and the image size was 1536×1536 pixels with a pixel size of 108 nm. For the other three setups, a Lumencor CELESTA light engine was used for illumination, with the following laser wavelengths: 405 nm, 477 nm, 546 nm, 638 nm and 749 nm. The 405 nm laser was used to excite and image the nuclear stain with diamidino-2- phenylindole (DAPI). The 477 nm laser was used to excite and image the yellow-green fiducial beads and the Geminin stain. The 546 nm laser was used to excite and image ATTO565 dye on secondary probes for BARC-FISH and chromatin tracing, and CF568 dye for total protein stain. The 638 nm laser was used to excite and image Alexa Fluor 647 dye on secondary probes for BARC-FISH and chromatin tracing. The 749 nm laser was used to excite and image Alexa Fluor 750 dye on secondary probes for BARC-FISH. A pentaband dichroic mirror supplied by Lumencor for the light engine was installed on the excitation path to direct the lasers to the sample, together with a ND0.6 neutral density filter to reduce the laser intensity. A pentaband emission filter supplied by Lumencor for the light engine and a Hamamatsu Orca Flash 4.0 V3 camera were each installed on the emission path. The image size was 2048×2048 pixels with a pixel size of 108 nm. A motorized x-y stage (SCAN IM 112×74, Marzhauser) was used to automatically image different fields of view (FOVs). Samples cultured on 40 mm #1.5 coverslips (Bioptechs, 40-1313-03193) were assembled in a Bioptech's FCS2 flow chamber, mounted onto the microscope stage and connected with a previously described automated fluidic system (*87*, *88*) for liquid handling during sequential hybridization and imaging.

Chromatin tracing sequential imaging. We conducted 17 and 14 rounds of three-color imaging for chromatin tracing of chr21 and chr22, respectively. During each round of hybridization, the sample was first incubated with 20% v/v ethylene carbonate (EC) (Sigma-Aldrich, E26258) in 2×SSC containing two secondary probes (one labeled by ATTO565 and one labeled by Alexa Fluor 647) for 25 minutes. The final concentration of each secondary probe was 6 nM except the following exceptions for chr22 tracing to tune signal intensity differences: TAD4 (0.5 nM), TAD14 (12 nM), TAD18 (12 nM) and TAD26 (12 nM). In 6 initial screen replicates out of the 17 total replicates, we used a TAD4 probe concentration of 6 nM, which led to a bleedthrough of TAD4 signal (labeled with Alexa Fluor 647 dye) into the TAD18 fluorescent channel (546 nm laser channel). This bleedthrough issue was computationally addressed as described later. The dye-labeled secondary probe sequences were selected from a previous study (*46*). The secondary probes were ordered from IDT and their sequences are listed in Table S12. 2 mL of wash buffer containing 20% v/v EC in 2×SSC was flowed through the chamber to remove unbound secondary probes, followed by 2 mL anti-photobleaching oxygen-scavenging imaging buffer that contained 50 mM Tris-HCl pH 8.0, 5% w/v glucose, 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/mL glucose oxidase (Sigma-Aldrich, G2133) and 40 µg/mL catalase (Sigma-Aldrich, C30) in 2×SSC. During the overnight imaging process, the imaging buffer in the input tube was covered by a layer of mineral oil (Sigma-Aldrich, 330779-1L) to prevent oxidization. At each FOV, three z image stacks were taken sequentially with 638-nm, 546-nm and 477-nm laser illumination, with a step size of 200 nm, 0.4 s exposure time at each step, and a total z range of 7 µm. After all the FOVs were imaged for the current imaging round, the secondary probes were stripped off by slowly flowing 4 mL of 65% v/v formamide in 2×SSC for 10 minutes and incubating for additional 5 minutes. Extra formamide was then removed by flowing 2 mL of wash buffer, and the next round of hybridization and imaging followed. To correct for color shift between 546-nm and 638-nm lasers, a #1.5 coverslip was coated with 100-nm Tetraspeck beads (Invitrogen, T7279) diluted at 1:200 in the imaging buffer, and z-stack calibration images were taken using the two lasers (*46*, *87*).

BARC-FISH sequential imaging. For screen samples, we conducted 10 rounds of four-color imaging to detect the barcode after chromatin tracing. During each round, the sample was first incubated for 25 minutes with 20% v/v EC in 2×SSC containing three secondary probes that correspond to the three values in each barcode digit, and the three probes were labeled with ATTO565, Alexa Fluor 647 and Alexa Fluor 750, respectively. The concentrations of the Alexa Fluor 750- and Alexa Fluor 647-labeled probes were 3 nM each, and the concentrations of ATTO565-labeled probes were 6 nM each. Excessive probes were washed away by flowing 2 mL of wash buffer, followed by application of 2 mL imaging buffer. At each FOV, four z image stacks were taken sequentially with 749-nm, 638-nm, 546-nm and 477-nm lasers. Each z-stack had a step size of 1.5 µm, exposure time at each step of 0.4 s, and a total range of 9 µm. After each imaging round, 11 mL of 90% v/v formamide in DPBS was flowed through the sample chamber for 28 minutes and then let stand for 100s to remove the bound fluorescent probes. Excessive formamide was removed with 2 mL wash buffer.

DAPI and total protein staining. For imaging screen samples, after BARC-FISH imaging was completed, the nuclei were stained by applying 3 mL of 1:1000 diluted DAPI (Thermo Scientific, 62248) in 2×SSC in 4 minutes, and incubated for 3.5 minutes. The sample was then washed by applying 2 mL of wash buffer. Total protein stain was then conducted by diluting CF568-labeled succinimidyl ester (Biotium, 92131) at 1:100,000 in water containing 0.1 M $NaHCO_3$ and 25 mM Na_2CO_3, flowing 3 mL of the buffer through the sample chamber over 4 minutes, and incubating for 3.5 minutes. After further application of 2 mL wash buffer and 2 mL imaging buffer, three z image stacks were taken at each FOV sequentially with 546-nm, 477-nm and 405-nm lasers. Specifically, the z-stack of 546-nm and 477-nm had the same step size, exposure time and total range as that used in BARC-FISH imaging, and the z-stack of 405-nm lasers had the same parameters as those used in chromatin tracing imaging. For imaging siRNA validation samples, DAPI staining was applied and imaged after chromatin tracing in a similar manner in the 477-nm and 405-nm channels, with parameters identical to those mentioned above.

siRNA knockdown

siRNA in A549 cells. The clonal A549-Cas9 cells used to construct the screen cell library were seeded at 6% density one day before siRNA transfection. The siCHD7 single knockdown experiments were conducted using Dharmacon ON-TARGETplus siRNA pools and DharmaFECT transfection reagent (Dharmacon, T-2001-03) following the manufacturer's protocol. Briefly, for each 10-cm dish, 3 µL of 100 µM siCtrl (Dharmacon, D-001810-10) or siCHD7 (Dharmacon, L-025947-01) was diluted to 1.2 mL Opti-MEM reduced-serum medium (Thermo Fisher Scientific, 31985-070). In a separate tube, 24 µL of transfection reagent was diluted to 1.2 mL Opti-MEM reduced-serum medium. The diluted siRNA and the diluted transfection reagent were incubated for 5 minutes at room temperature. Then the diluted siRNA was added into the diluted transfection reagent and the mixture was incubated for 20 minutes at room temperature. Lastly, the antibiotics-free growth medium was added to a final volume of 12 mL. Cells were cultured with the 12 mL transfection medium at 37°C for 96 hours and used for follow-up analysis, including Western blot, chromatin tracing, RNA-seq, and CUT&RUN.
For siCHD7+siCTCF (Dharmacon, L-020165-00) double knockdown experiments, we used Lipofectamine RNAiMAX transfection reagent (ThermoFisher, 13778150) following the manufacturer's protocol optimized for A549 cells. To control for equal amount of siRNA across all conditions, we set up the following conditions: 1) 175nM siCtrl; 2) 25nM siCHD7 + 150nM siCTCF; 3) 150nM siCTCF + 25nM siCtrl. siRNAs were gently mixed with Lipofectamine RNAiMAX in Opti-MEM medium and incubated for 10-20 minutes at room temperature. The complexes were then mixed with growth media and plated onto the cells. We incubated the cells at 37°C in a CO_2 incubator for 48 hours and then repeated the siRNA transfection to ensure sufficient CTCF knockdown. The cells were incubated for another 48 hours until ready for downstream assays.

siRNA in hTERT-RPE1 cells. The hTERT-RPE1 cells were seeded at 6% density one day before siRNA transfection. The same siRNA pools for siCtrl and siCHD7 in A549 RNAi experiments and RNAiMAX transfection reagents (Invitrogen, 13778-075) were used. Transfection was done following RNAiMAX manufecturer's protocol with 50nM final concentration of siCtrl or siCHD7 siRNA. After siRNA transfection, cells were cultured with the transfection mix for 4 days before downstream assays.

HDAC inhibition drug treatment

A549 cells were seeded at 40% density one day prior to the drug treatment. Trichostatin A (TSA) (Sigma-Aldrich, T8552) was dissolved in DMSO at 1mM to make drug stocks. To inhibit HDAC, cells were cultured in media supplemented with 150 nM TSA or equivalent volume of DMSO (as a solvent-only control) for 2 hours before PFA fixation.

Western blot

Cells were trypsinized by TrypLE Express (Thermo Fisher Scientific, 12605-010) and pelleted through centrifugation at 500g for 5 minutes at 4°C. To enrich the nuclear proteins, cell nuclei were isolated and lysed using NE-PER Nuclear and Cytoplasmic Extraction Kit (Thermo Fisher Scientific, 78833) following the manufacturer's instructions. The reagents were supplemented with 1× protease inhibitors (Thermo Fisher Scientific, 87786) to prevent the protein degradation during extraction. The protein concentrations of the nuclear extract were quantified using Pierce BCA protein assay kit (ThermoFisher, 23225) following the manufacturer's instructions. The nuclear extract was denatured at 95-100°C for 5 minutes in 1× NuPAGE LDS sample buffer (ThermoFisher, NP0007) containing 10% v/v β-Mercaptoethanol (Millipore Sigma, M6250). For detection of CHD7 protein, the denatured nuclear extract was subjected to electrophoresis on a precast 3-8% Tris-Acetate polyacrylamide gel (Thermo Fisher Scientific, EA0378BOX) in 1× Tris-Acetate SDS running buffer (Thermo Fisher Scientific, LA0041) at 100V for 2.5 hours. For detection of CTCF and Actin, the protein electrophoresis was performed using 4-15% precast polyacrylamide gels (Biorad, 4568084) in 1× Tris/Glycine/SDS buffer (diluted from Biorad, 1610732) at 100V for 1.5 hours. The proteins were transferred from the gel onto a nitrocellulose membrane (ThermoFisher, IB301032) using iBlot2 dry blotting system (Thermo Fisher Scientific, IB21001S) following the manufacturer's instructions. The membrane was blocked in 5% BSA in TBST for 1 hour at room temperature with gentle shaking. Primary antibodies against human CHD7 (Thermo Fisher Scientific, PA5-72964), CTCF (Millipore Sigma, 07-729) and Actin (Abcam, ab179467) were incubated with the membrane at recommended concentrations at 4℃ overnight with gentle shaking. The following day, the membrane was washed three times in TBST for 5 minutes each, followed by incubation with horseradish peroxidase-conjugated secondary antibodies (Abcam, ab6721) diluted at 1:3000 ratio in 5% non-fat milk in TBST for 1 hour at room temperature. The blots were then washed three times in TBST for 5 minutes each. CTCF and Actin blots were treated with SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific, 34580) and CHD7 blot was treated with SuperSignal West Femto Maximum Sensitivity Substrate (ThermoFisher, 34094) for 2-5 minutes at room temperature. The blots were imaged using a CCD camera-based imager, ProteinSimple FluorChem E system or Bio-Rad ChemiDoc MP Imaging System.

Immunofluorescence of Cas9 and H3K27ac

Cells were fixed in 4% PFA in DPBS for 10 minutes and washed twice with DPBS. Cells were then permeabilized in 0.5% v/v Triton X-100 in DPBS for 10 minutes and washed twice with DPBS. Cells were blocked in blocking buffer (1% w/v BSA and 0.1% v/v Tween-20 in DPBS) for 30 minutes at room temperature with gentle agitation. Then the cells were incubated for 1 hour at room temperature with 1:500 diluted anti-Cas9 primary antibody (Sigma, SAB4200701-25UL, for detection of Cas9 expression) or 1:100 diluted anti-H3K27ac primary antibody (Active Motif, 91193, for detection of H3K27ac level in TSA- and DMSO-treated samples) in blocking buffer, followed by three DPBS washes for 5 minutes each. The cells were incubated with 1:1000 diluted Alexa Fluor 647-labeled secondary antibody (Thermo Fisher Scientific, A21237, for anti-Cas9 primary antibody) or Alexa Fluor 488-labeled secondary antibody (Invitrogen, A21202, for anti-H3K27ac primary antibody) in blocking buffer for 1 hour at room temperature, followed by three DPBS washes for 5 minutes each. Starting from incubation with secondary antibody, the samples were covered from the light with aluminum foil.

Data analysis

Determination of chromatin traces in 3D. Image and phenotype analyses were performed using MATLAB version R2020a. We adopted a previously reported pipeline to determine the 3D chromatin traces (*87*). Briefly, to correct the drift between sequential chromatin tracing secondary hybridization rounds, we determined the 3D positions of fiducial beads during each round of hybridization by fitting their z-stack images to Gaussian functions in 3D. We then fitted the center position of each DNA FISH foci in 3D using the same algorithm, and the sample drift (represented by bead movements) was subtracted from the 3D position of each DNA locus. The drift-corrected DNA loci positions were then linked into traces based on their expected spatial proximity in each chromosome territory. To remove false traces generated from non-specific target objects or genomic DNA released from dead cells, linked traces were excluded if 1) they located outside of the nucleus, represented by 2D-projected DAPI staining patterns, or 2) their radii of gyration was smaller than 0.5 μm. In a small number of datasets (see *Chromatin tracing sequential imaging* section for explanation), we noticed that due to extra-strong labeling of TAD4, signal from TAD4 (638-nm channel) could be excited with 546-nm laser and thus be misidentified as TAD18. To correct this in the affected datasets, after the traces were linked, we calculated the 3D distance between TAD4 and TAD18 on each trace in those datasets. If the distance was smaller than 100 nm, the 3D position recognized as TAD18 were regarded as bleedthrough from TAD4, and the 3D positions of TAD18 were assigned to NaN.

BARC-FISH pattern extraction. The decoding process consisted of the following steps: 1) drift correction, 2) identifying all possible BARC-FISH patterns for each value of each barcode digit, 3) filtering mis-identified BARC-FISH patterns based on intensity thresholding, and 4) assigning the digit identity to each cell. To correct the drift between sequential BARC-FISH imaging rounds, the center positions of the fiducial beads were determined by fitting to Gaussian functions in 2D, and the movement was subtracted from BARC-FISH images. To extract the BARC-FISH patterns for each value in each digit, BARC-FISH images were first projected along z dimension with average intensity projection, and filtered by a 2D median filter and a Gaussian low-pass filter. The image background was derived by image opening with a disk-shaped structural element with a 20-pixel radius, and the background was subtracted. Potential BARC-FISH areas were identified by binarizing the background-subtracted image and selecting objects that were larger than 50 pixels. Due to the strong labeling intensity of BARC-FISH foci, bleedthrough was often seen in channels excited with lasers of shorter wavelength. To correct this, the bleedthrough of Alexa Fluor 647 into 546-nm channel was identified by the shared areas between 546-nm and 638-nm channel in the same BARC-FISH round, and the areas were subtracted from the 546-nm BARC-FISH pattern. Similarly, the bleedthrough of Alexa Fluor 750 into 638-nm channel was subtracted from the 638-nm BARC-FISH pattern. Regions that were positive for all three channels (546-nm, 638-nm and 749-nm) during one BARC-FISH round were subtracted from all three channels because such regions could be due to broad-spectrum autofluorescent objects or non-specific binding of all probes to the same sticky objects. These steps generated high-fidelity binary masks for BARC-FISH signals. To further increase the reliability of our identified BARC-FISH foci masks, we applied the binary masks to the median- and Gaussian-filtered BARC-FISH images collected in 15-20% of the FOVs collected in each screen dataset. For each value of each digit, we calculated the normalized intensity ([total intensity] / [pattern size]) of each BARC-FISH focus identified through the binary mask, and sorted the normalized intensity. A normalized intensity cutoff was then derived for each value of each digit, and applied to all the FOVs in the dataset to only include the BARC-FISH patterns of which the normalized intensity was above this threshold.

Cell segmentation and decoding. Drift between the DAPI/total protein stain imaging round and previous rounds were corrected using the same algorithm as in BARC-FISH pattern extraction. Z-stacks of DAPI and total protein stain was projected along the z dimension with max and average intensity projections, respectively. A median filter was applied to the z-projected DAPI images. To remove the bleedthrough of the fiducial beads pattern into total protein stain, total protein images was eroded by a disk-shaped structural element with a radius of 5 pixels. We then performed watershed-based cell segmentation of the total protein stain profile using DAPI patterns as foreground markers. To determine the barcode identity of each cell, for each digit, we calculated the total pixel counts from BARC-FISH pattern of each of the three values, value 0 (546-nm channel), 1 (638-nm channel) or 2 (749-nm channel), and assigned the value with the largest pixel count to the digit of the cell. Cells with BARC-FISH signal in none of the 3 values were assigned a fourth value of 10000. We then calculated the Hamming distance between the current decoded barcode to those in the codebook. A barcode was decoded/corrected if: 1) the current decoded barcode has a unique closest match in the codebook, 2) this unique closest match is a good code (a code that uniquely associates with only one sgRNA), and 3) the Hamming distance between the decoded barcode and the closest match was not larger than 1 (maximum 1 mismatching digit allowed). The cell carrying the decoded/corrected barcode was assigned the corresponding sgRNA identity.

Cell cycle stage classification. We only included G1-phase cells in phenotypic analyses of both the screen and the validation. G1 cells were indicated by an absence or low intensity of Geminin stain in the nucleus. The foreground DAPI marker generated for the watershed-based cell segmentation was used as a binary mask of nuclear regions. Two methods were used to determine the cell cycle status: 1) We calculated the normalized Geminin intensity for each cell by dividing the total nuclear Geminin fluorescence intensity by the area of the nucleus, and empirically determined the threshold for each imaging experiment between Geminin-positive and Geminin-negative cells based on the distribution of the normalized intensity. Cells of which the normalized nuclear Geminin intensity was below the threshold were included in phenotypic analyses. This method was used for 13 out of 17 screen experiments and all validation experiments. For 4 early replicates, due to area-to-area variation of Geminin staining, a uniform normalized nuclear Geminin intensity could not be applied to all FOVs for G1 cell identification. For these 4 replicates, we determined the cutoff for G1 versus non-G1 cells by calculating the ratio between the normalized Geminin intensity in the nucleus and the whole cell area (determined by watershed boundaries) for each cell, and empirically determined the threshold for each dataset based on the distribution of this ratio.

Analysis of adjacent TAD distance and contact. Table S13 includes the p-value, false discovery rate (FDR) and log2 fold change (log2fc) value for each analyzed sgRNA and for each phenotype analyzed in the screen. We grouped traces from the same sgRNA (determined by the cell barcode detected via BARC-FISH) across different datasets. To increase the statistical power in identifying novel hits, for phenotype analyses regarding chromatin folding and nuclear lamina association, traces from all 412 observed sgRNAs in the screen were pooled as control (termed as “whole population”), and 341 sgRNAs with at least 40 chromatin traces were compared to the whole population, which were used to generate volcano plots (Fig. 2, A, D, F, H and J, and Fig. 3A) and Fig. 2B. In other panels, only non-targeting control sgRNAs with no fewer than 40 traces were pooled and plotted as sgCtrl, and statistics were performed between selected hits and the included non-targeting control sgRNAs (Fig. 2, C, E, G, I, K and L, and Fig. 3, B and C).

Adjacent TADs were defined as TADs that were next to each other along the genomic map. To compare adjacent TAD distance, we calculated the distance between each pair of adjacent TADs on each trace, and generated a mean adjacent TAD distance profile for each sgRNA by calculating the mean distance for each pair of adjacent TADs. We then conducted two-sided Wilcoxon signed rank test in MATLAB to compare the mean adjacent TAD distance between each sgRNA and the control group, calculated FDR from the resulting p values, and selected top hits that 1) have FDRs<0.1, 2) have the largest absolute values of log2fc, and 3) have protein localization in the cell nucleus based on The Human Protein Atlas (proteinatlas.org) to prioritize potential direct regulators of 3D genome. To calculate the log2 fold change in Fig. 2A, we divided the mean adjacent TAD distance between the same pair of TADs from the sgRNA by that from the control group, calculated the log2 fold change value for each TAD pair, and averaged the log2 fold change values to generate the log2 fold change for each sgRNA. Log2 fold change was calculated in a similar manner in Fig. 2J by averaging the log2 fold change value of all pairwise comparisons. To analyze adjacent TAD contact frequency (fig. S7, B and C), we defined a contact event as when the spatial distance between two TADs is less than 500 nm, and counted the frequency of contacts between each pair of adjacent TADs.

Determination of compartment identity of TADs. We assigned A/B compartment identity to the TADs using a previously introduced algorithm (*87*). Briefly, we fitted a power-law function to the data points of the mean inter-TAD spatial distances versus their genomic distances, which yielded the expected inter-TAD spatial distance for each pair of TADs according to their genomic distance. We then normalized the observed mean inter-TAD spatial distance by the expected spatial distance, yielding a normalized inter-TAD distance matrix. We then calculated the Pearson correlation coefficient between each pair of rows/columns in this matrix, generating a Pearson correlation matrix. We next applied principal component analysis to the Pearson correlation matrix, and took the coefficients of the first principal component as compartment scores. The compartment score profiles were cross-validated with published ChIP-seq profiles (*94, 95*), so that the compartment A regions are enriched in active histone modifications, while compartment B regions are enriched in repressive histone modifications. If the trend was opposite, the signs of the compartment scores were flipped. In the CRISPR screen, traces from the whole population were pooled to generate the compartment identity, which was applied to different perturbations when evaluating long-range AA, AB and BB compartment contacts (Fig. 2, D to I).

Analysis of long-range contact in different compartments. Long-range contacts were defined as the spatial contact events between TAD pairs that are not adjacent TADs on the genomic map. We used the same distance threshold for calling contacts as that used in the adjacent TAD contact, and calculated the numbers of AA, AB and BB long-range TAD contact events per trace. In Fig. 2D to I, AB compartment profile generated from all traces were used to define A and B compartments. In figs. S3E, S4D and S5E, AB compartment profile of the according control group (siCtrl of A549-Cas9 cells, GFP-overexpressed A549-Cas9 cells and siCtrl of hTERT-RPE1, respectively) were used to define A and B compartments.

Analysis of overall chromosome compaction. We calculated the inter-TAD distances between all pairs of TADs on chr22 for each sgRNA, and generated mean inter-TAD distance matrix for each sgRNA by calculating the mean distance between each pair of TADs. We then performed two-sided Wilcoxon signed rank test between the mean inter-TAD distances of each sgRNA and control, and calculated FDR and log2 fold change values.

Determination of chromatin association with nuclear lamina. The 2D-projected nuclear lamina profile was derived from 2D-projected DAPI images by adopting a previously reported algorithm (*87*). Briefly, we extracted the pixels on the edges of average z-projected nuclei and measured the x-y 2D distance between the positions on each trace (trace centroid or each TAD position) and the nearest nuclear edge pixels. The trace centroid distances to nuclear lamina of each sgRNA and control were compared.

Analysis of nuclear morphological properties. To increase the statistical power of the screen, similar to analyzing chromatin phenotypes, 3D nuclear DAPI patterns from all 412 sgRNAs were pooled as controls in the volcano plots (Fig. 3, D and H), and only sgRNAs with more than 10 nuclei were included in the screen analyses; 3D nuclear DAPI patterns from non-targeting control sgRNAs with more than 10 nuclei were pooled as controls in the box plots (Fig. 3, E to G, I and J). We compared the 3D coefficient of variation of nuclear stain intensity (nuclear intensity unevenness) and nuclear sphericity between each sgRNA and control. To abstract the nucleus shape in 3D, we first calculated a normalization factor of each FOV by selecting the max intensity of the median-filtered z-stack images of the current FOV. We normalized the max z-projected DAPI image by this normalization factor, and calculated a background by applying adaptive thresholding on the normalized image using a sensitivity of 0.4. We then processed each layer of the z-stack, normalized the median-filtered current slice of z-stack image by the normalization factor, and binarized it using the threshold generated from the previous adaptive thresholding step. Only binary objects at each slice that were larger than 5,000 pixels were retained. Partial nuclei at the edges of each FOV were excluded. We then matched the cell identity to each segmented 3D nucleus, and analyzed the coefficient of variation of voxel intensity within each nucleus of each sgRNA. The nuclear volume and surface area were calculated by the *regionprops3* function in MATLAB. The nuclear sphericity was determined by the following equation:

$$Sphericity = \frac{\pi^{\frac{1}{3}}\,(6 * Volume)^{\frac{2}{3}}}{Surface\ area}$$

A sphericity of 1 indicates the shape is a perfect sphere. A sphericity smaller than 1 indicates reduced roundness. Sphericity values exceeding 1 due to inaccuracy of surface area calculation by the *regionprops3* function were excluded from the analysis. For volcano plots, p values were calculated by unpaired t test followed by FDR calculation (Fig. 3, D and H). For box plots, p values were calculated by two-sided Wilcoxon rank sum test (Fig. 3, E and I).

Calculation of radius of gyration. We only retained traces with more than 80% of TADs detected for analyzing this phenotype. The radius of gyration was defined as the root mean square of the distance between each TAD to the centroid of the trace, as shown in the following equation:

$$Radius\ of\ gyration = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i^2}$$

where n is the number of detected TADs in the given trace, and d_i represents the spatial distance between TAD i and the centroid of the trace.

Quantification of polarized organization of A-B compartments. Only traces with more than 80% of TADs detected were included in this analysis. We used a previously developed algorithm (*87*) to quantify the polarized organization of A-B compartments, yielding a polarization index. Briefly, we constructed 3D convex hulls for all A compartment TADs and B compartment TADs on each individual trace, with the *convhull* MATLAB function. We then calculated the volume of these two hulls (referred to as V_A and V_B, respectively) and the volume of the overlapped spaces between A and B hulls (referred to as V_S). Polarization index was defined as:

$$Polarization\ index = \sqrt{\left(1-\frac{V_S}{V_A}\right)\left(1-\frac{V_S}{V_B}\right)}$$

A polarization index of 1 indicates that the A and B compartments are completely spatially separated from each other in a side-by-side manner. A polarization index of 0 indicates that A and B compartments completely overlap, or one is surrounded by the other. The A-B compartment profile of the according treatment was used to calculate polarization indices for each condition in figs. S3D and S4C.

RNA-seq

Library preparation. Total RNA from siCtrl and siCHD7 A549-Cas9 cells was extracted using Qiagen RNeasy Mini Kit (Qiagen, 74104) with on-column DNase treatment (Qiagen, 79254) according to the manufacturer's specifications. RNA integrity was determined using Agilent Fragment Analyzer RNA Kit (DNF-471) according to the manufacturer's specifications. To prepare the sequencing library, for each sample, 1 µg of RNA was prepared using Kapa Biosystems KAPA mRNA HyperPrep kit with 10 PCR cycles according to the manufacturer's specifications. The final library quality control was run on the Agilent TapeStation D1000 assay, and the concentration was determined with the KAPA Library Quantification Kit. Sequencing was performed with 25 million reads per sample, 100 bp paired-end reads, on an Illumina NovaSeq 6000. 4 replicates of siCtrl and 4 replicates of siCHD7 cells were processed in this study.

RNA-seq data analysis. RNA-seq reads from siCHD7 and siCtrl-treated A549-Cas9 cells were aligned to Ensembl genome GRCh38 by STAR version 2.7.9 (*96*). Counts of each unique gene were performed by FeatureCounts contained in the Subread package (2.0.0) (*97*). Differentially-expressed genes (DEGs) were detected using the DESeq2 R package from Bioconductor (*98*). Cutoffs were placed at > 20 counts per gene, Log2FC > 0.5, and P Value > 10e-10. The expression levels were normalized and adjusted using Apeglm (*99*). Bioconductor EnhancedVolcano (http://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html) was used to generate the volcano plot and previously listed cutoffs were added to the plot. Ensembl IDs were converted to gene names using Bioconductor package biomaRt (version 2.28.0) (*100*). Gene ontology (GO) enrichment analysis was performed using Enrichr GO:BP 2021 (*101*). Log10 P values were generated and listed on the graph with the top 10 up and down GO terms after analysis of all significantly upregulated genes and downregulated genes according to cutoffs.

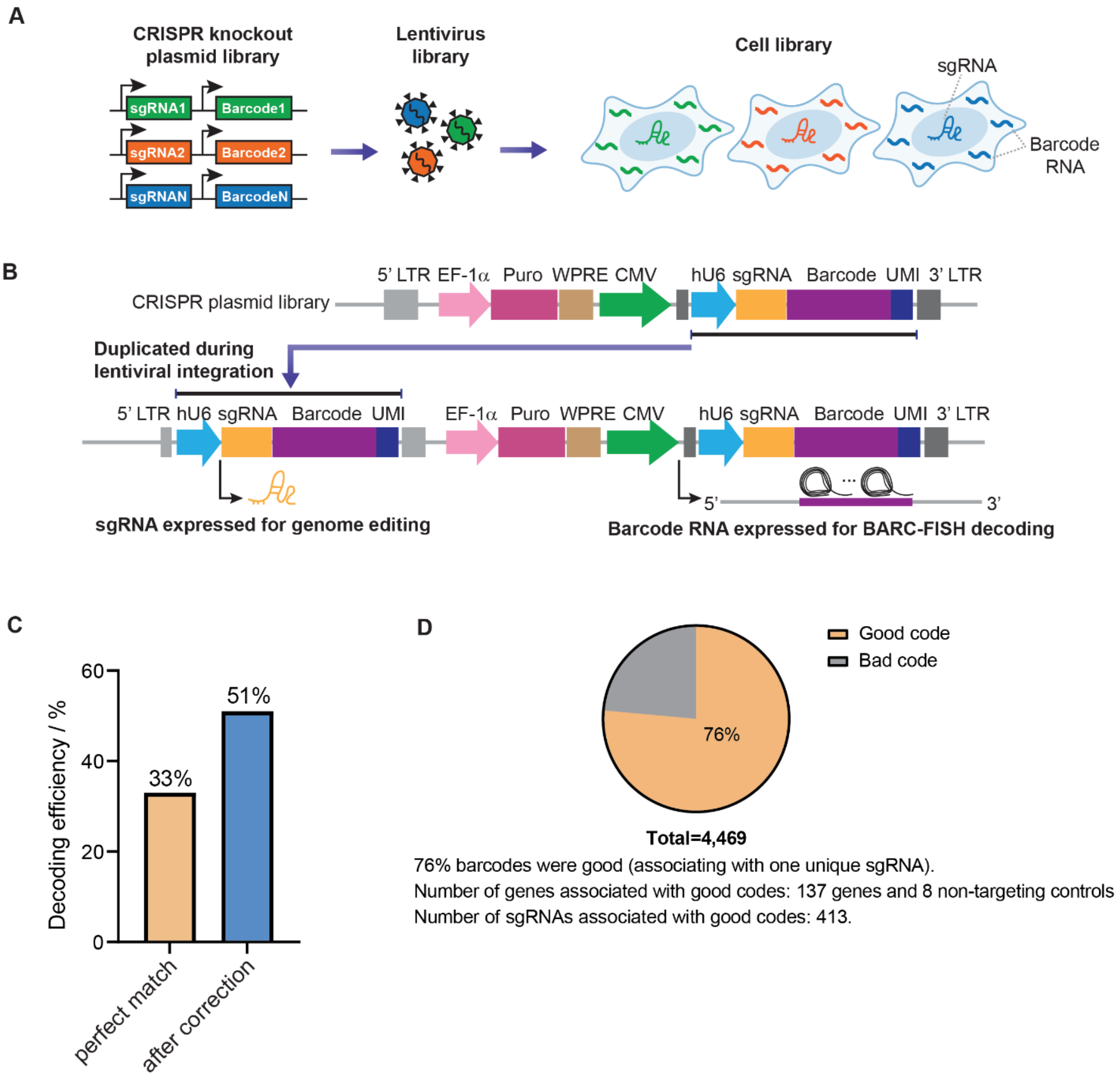
CUT&RUN

CUT&RUN protocol. CUT&RUN was performed using the Epicypher CUT&RUN Kit (14-1048) according to the manufacturer's specifications with the following conditions and modifications: For binding to the Concanavalin beads, 500,000 cells per sample were prepared, counted twice by hemacytometer and averaged, and incubated with activated beads. Instead of spike-in nucleosomes, an H2A antibody was used to validate successful CUT&RUN and used later to align reads as input control. Antibodies were used as listed below, with 0.5 ug of antibody per reaction.

H2A antibody – CST Histone H2A (D6O3A) Rabbit mAb #12349
IgG Control antibody – CUTANA Kit Rabbit IgG CUT&RUN Negative Control Antibody
H3K4me3 antibody – CST Tri-Methyl-Histone H3 (Lys4) (C42D8) Rabbit mAb #9751
CTCF antibody – CTCF CUTANA™ CUT&RUN Antibody 13-2014
CHD7 antibody – Invitrogen CHD7 Polyclonal Antibody PA5-72964
RAD21 antibody – Active Motif Rad21 #91245

Library preparation. Libraries were prepared using the Epicypher CUT&RUN Library Prep Kit (14-1001) according to the manufacturer's specification with the following modification: SPRI-select beads were used after library preparation to perform right and left-handed selections at size 200 bp – 700 bp to enrich for DNA fragments from CUT&RUN and remove any adapter dimers or high molecular weight DNA. Library quality was analyzed using Agilent Tapestation D1000 High Sensitivity Tapes #5067-5584. Sequencing was performed with 10 million reads ordered per sample, 150 bp paired-end reads, on an Illumina NovaSeq 6000.

CUT&RUN data analysis. CUT&RUN reads were analyzed for quality using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed using Trimmomatic (*102*). Reads were then aligned to hg18 by Bowtie2 version 2.3.4 (*103*). SAMTools version 1.11 was then used to convert to BAM format, index, isolate uniquely mapped paired reads, and remove duplicates (*104*). Picard version 2.27.4 was used to downsample all samples to the same read depth (6.5 million reads) for direct comparison. MACS2 version 2.2.7.1 was used to call sample narrow peaks, using H2A as an input. Peaks were called using IgG control against H2A input, and these were subtracted from the final peaks counted. The Bioconductor package ChIPseeker was then used to annotate peaks using RefSeq files for hg18 and was used for the calculation of significant overlap between features (*105*). Read counts across genomic intervals and peak visualization were performed using deepTools version 3.3 (*106*), and .bw files were visualized using Integrative Genomics Viewer.


A
CRISPR knockout plasmid library
sgRNA1
Barcode1
sgRNA2
Barcode2
sgRNAN
BarcodeN
Lentivirus library
Cell library
sgRNA
Barcode RNA
B
CRISPR plasmid library
5' LTR EF-1α Puro WPRE CMV
hU6 sgRNA Barcode UMI 3' LTR
Duplicated during lentiviral integration
5' LTR hU6 sgRNA Barcode UMI
EF-1α Puro WPRE CMV
hU6 sgRNA Barcode UMI 3' LTR
5'
3'
sgRNA expressed for genome editing
Barcode RNA expressed for BARC-FISH decoding
C
Decoding efficiency / %
0
20
40
60
33%
51%
perfect match
after correction
D
Good code
Bad code
76%
Total=4,469

76% barcodes were good (associating with one unique sgRNA).
Number of genes associated with good codes: 137 genes and 8 non-targeting controls
Number of sgRNAs associated with good codes: 413.

Fig. S1. CRISPR screen library allows for sgRNA expression for genome editing and the barcode RNA expression for BARC-FISH decoding. (**A**) Schematic of the construction of CRISPR screen library. A CRISPR knockout plasmid library containing sgRNA-barcode associations was constructed to generate a lentivirus library, which was transduced into human A549-Cas9 cells to produce a cell library. (**B**) Design of the CRISPR screen plasmid and the lentiviral integration strategy. The sgRNA-barcode cassette was composed of human U6 promoter (hU6, blue), sgRNA (yellow), barcode (purple) and UMI (dark blue) sequences and placed within 3' long terminal repeat (3' LTR, dark gray), downstream of a strong RNA Pol II promoter (CMV, green). This cassette was duplicated and inserted within 5' LTR (light gray) during lentiviral integration. Therefore, the cassette within 5' LTR was able to express the sgRNA for genome editing, while the other copy was driven by CMV promoter to express a high level of barcode RNA for BARC-FISH decoding. Other elements on the plasmid including EF-

1α promoter (pink), Puromycin resistance gene (magenta) and WPRE element (brown) were shown. (**C**) BARC-FISH decoding efficiency in the CRISPR screen cell library. After BARC-FISH decoding procedure, the decoded barcodes were compared and matched to the barcodes determined by NGS. 33% of the imaged cells contained barcodes with perfect matches. After the error correction, 51% of the cells contained matched barcodes. (**D**) Analysis of barcode quality determined by NGS. The sgRNA-barcode associations in the CRISPR screen cell library were determined by NGS (see Materials and Methods, "Detection of sgRNA-barcode associations in the cell library" section). In total, 4,469 barcodes were detected, among which 76% were good codes associating with one unique sgRNA. 413 sgRNAs targeting 137 genes and 8 non-targeting controls were found to be associated with these good codes. 412 of the 413 sgRNAs were observed in the image-based screen.

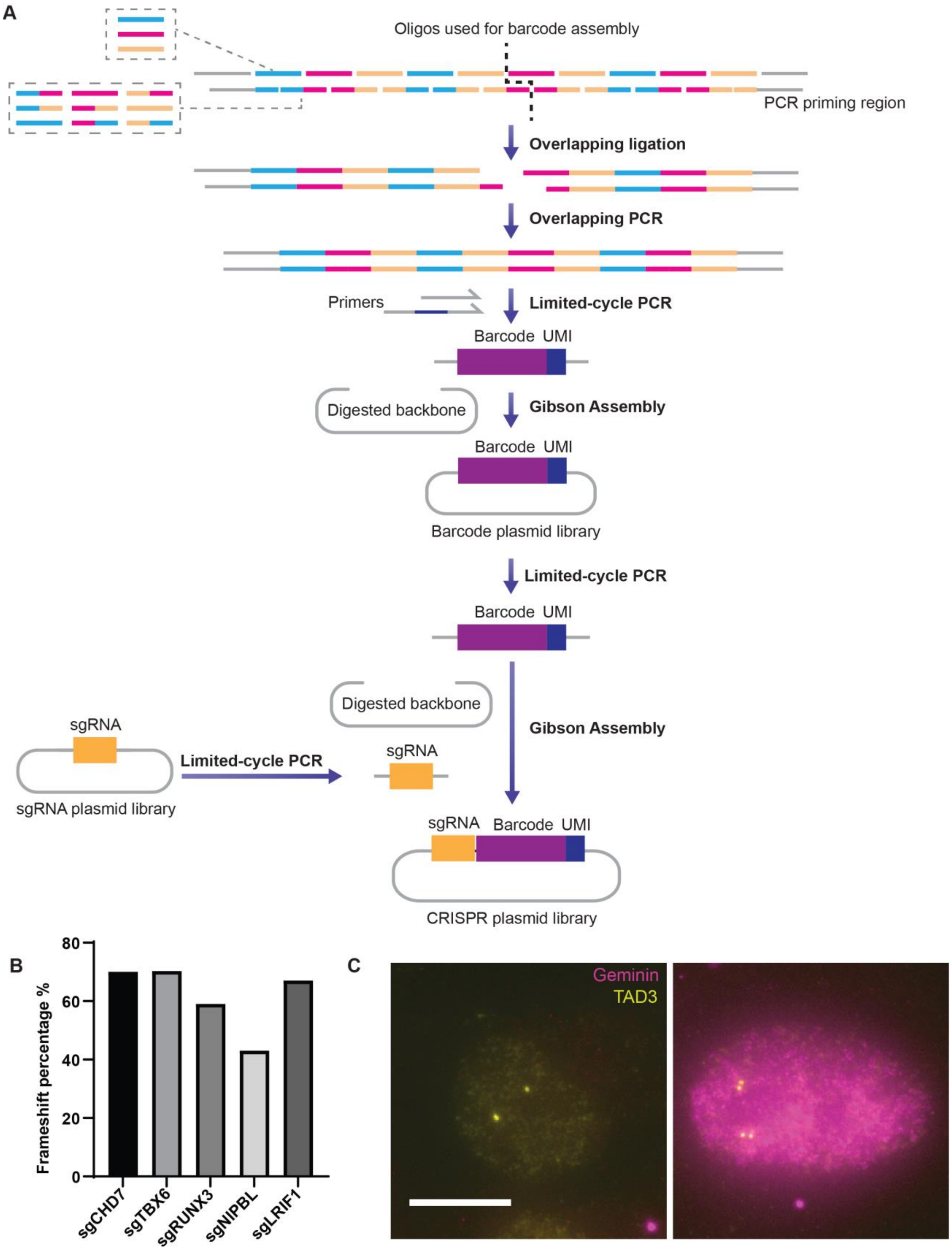

A
Oligos used for barcode assembly
PCR priming region
Overlapping ligation
Overlapping PCR
Primers
Limited-cycle PCR
Barcode
UMI
Digested backbone
Gibson Assembly
Barcode
UMI
Barcode plasmid library
Limited-cycle PCR
Barcode
UMI
Digested backbone
Gibson Assembly
sgRNA
Limited-cycle PCR
sgRNA
sgRNA plasmid library
sgRNA
Barcode
UMI
CRISPR plasmid library
B
Frameshift percentage %
80
60
40
20
0
sgCHD7
sgTBX6
sgRUNX3
sgNIPBL
sgLRIF1
C
Geminin
TAD3

Fig. S2. Cloning strategy of plasmid libraries, CRISPR knockout efficiency, and the Geminin-based cell cycle identification. (**A**) The barcode plasmid library was assembled from individual oligos through overlapping ligation, overlapping PCR, limited-cycle PCR and Gibson

Assembly (see Materials and Methods, “Barcode plasmid library construction” section). Each of the forward-strand oligos contained three alternative sequences (in the smaller gray dashed box), represented by three different colors cyan, magenta, and yellow. The overlapping oligos in the reverse strand contained 9 alternative sequences (in the larger gray dashed box). Oligos at the two ends carried PCR priming regions (straight gray lines). The barcode was divided into two halves which were subjected to overlapping ligation to form two double-stranded fragments. The two fragments were assembled by overlapping PCR to form a full-length barcode. The barcode was then amplified and added with UMI (unique molecular identifier) by limited-cycle PCR primers. The barcode-UMI fragments were inserted into a digested plasmid backbone through Gibson Assembly to construct the final barcode plasmid library. To clone the CRISPR screen plasmid library, sgRNA fragments and barcode-UMI fragments were amplified from the premade sgRNA plasmid library and barcode plasmid library respectively, through limited-cycle PCR. The sgRNA and barcode-UMI were then Gibson Assembled into a digested lentiviral plasmid backbone to generate the final CRISPR screen plasmid library (see Materials and Methods, “CRISPR screen plasmid library construction” section). The UMI was necessary for sequencing-based mapping of barcode-sgRNA associations (see Materials and Methods, “Next-generation sequencing (NGS) library preparation for mapping sgRNA-barcode associations”). (**B**) Percentage of frameshift mutations of sgCHD7, sgTBX6, sgRUNX3, sgNIPBL and sgLRIF1. (**C**) Two representative cells from the screen datasets were shown to demonstrate the Geminin staining strategy for G1 phase cell detection. Geminin antibody stain (magenta) is absent in a G1 phase cell (left), which showed two DNA FISH foci of TAD3 (yellow) of chr22. The S/G2 phase cell (right) is positive for Geminin stain and have four DNA FISH foci of TAD3 in two pairs, indicating replicated TAD3 DNA. Because Geminin and the yellow-green fiducial beads were imaged using the same laser channel, bead patterns were seen in both images (small, round magenta spots outside of the nuclei). Scale bar: 10 μm.

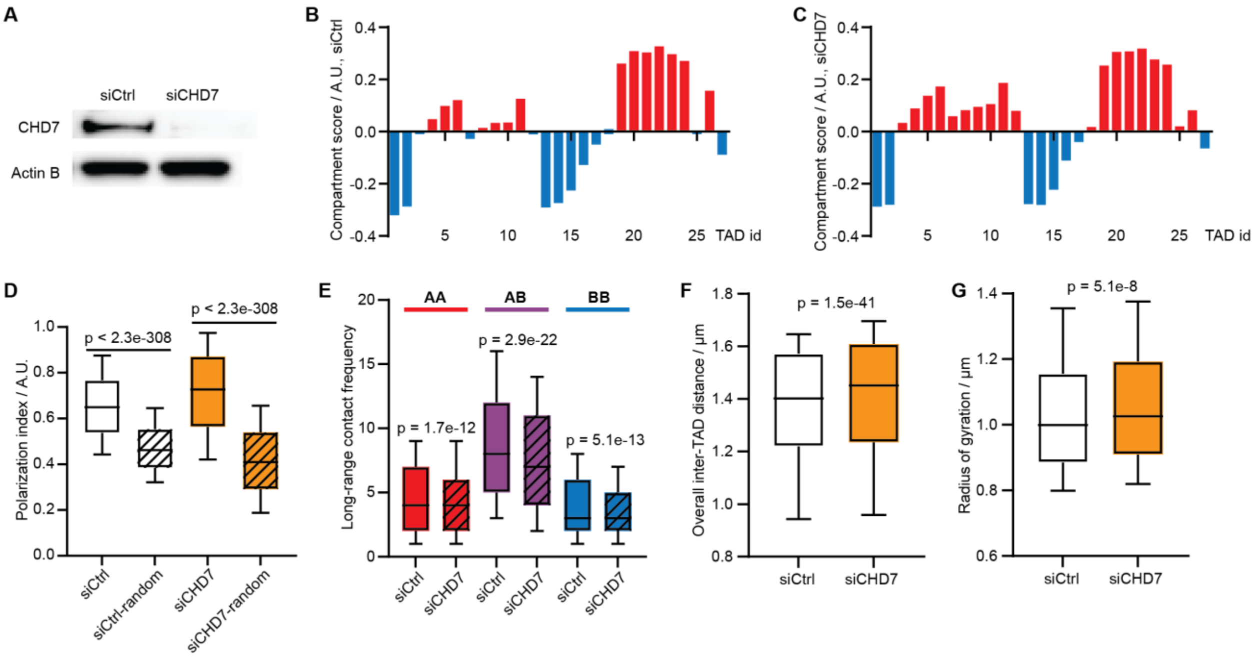


Fig. S3. Validation of CHD7 perturbation phenotypes using RNA interference. (**A**) Western blot of siCtrl- and siCHD7-treated A549-Cas9 nuclear extracts. Top: anti-CHD7 antibody; bottom: anti-Actin B antibody. (**B**) A-B compartment profile of chr22 in siCtrl cells. (**C**) A-B compartment profile of chr22 in siCHD7 cells. (**D**) Polarization indices of chr22 A-B compartments of siCtrl (white) and siCHD7 (orange). Shadowed boxes show the polarization indices from randomized controls, where the compartment identities of TADs are scrambled. (**E**) Long-range contact frequency of siCtrl and siCHD7 (shadowed) among A compartments (red), between A and B compartments (purple), and among B compartments (blue). (**F**) Overall inter-TAD distance of siCtrl and siCHD7. (**G**) Radii of gyration of siCtrl and siCHD7. P values in (D), (E) and (G) were calculated by two-sided Wilcoxon rank sum test. P value in (F) was calculated by two-sided Wilcoxon signed rank test.

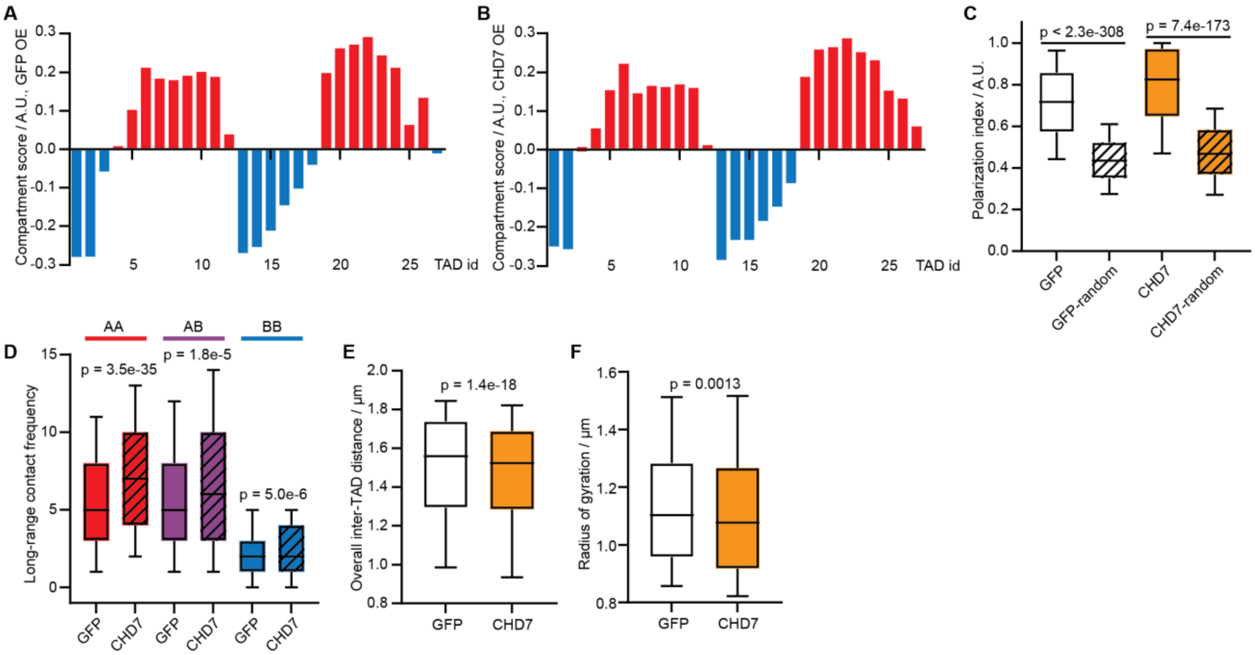


Fig. S4. Validation of CHD7 perturbation phenotypes using overexpression. (**A**) A-B compartment profile of chr22 in A549-Cas9 cells with GFP overexpression. (**B**) A-B compartment profile of chr22 in A549-Cas9 cells with CHD7 overexpression. (**C**) Polarization indices of cells with GFP (white) and CHD7 (orange) overexpression and the according randomized controls (shadowed). (**D**) Long-range contact frequency of cells with GFP of CHD7 (shadowed) overexpression in A compartments (red), across A and B compartments (purple) and in B compartments (blue). (**E**) Overall inter-TAD distance of chr22 in cells with GFP and CHD7 overexpression. (**F**) Radii of gyration of chr22 in cells with GFP and CHD7 overexpression. P values in (C), (D) and (F) were calculated by two-sided Wilcoxon rank sum test. P value in (E) were calculated by two-sided Wilcoxon signed rank test.

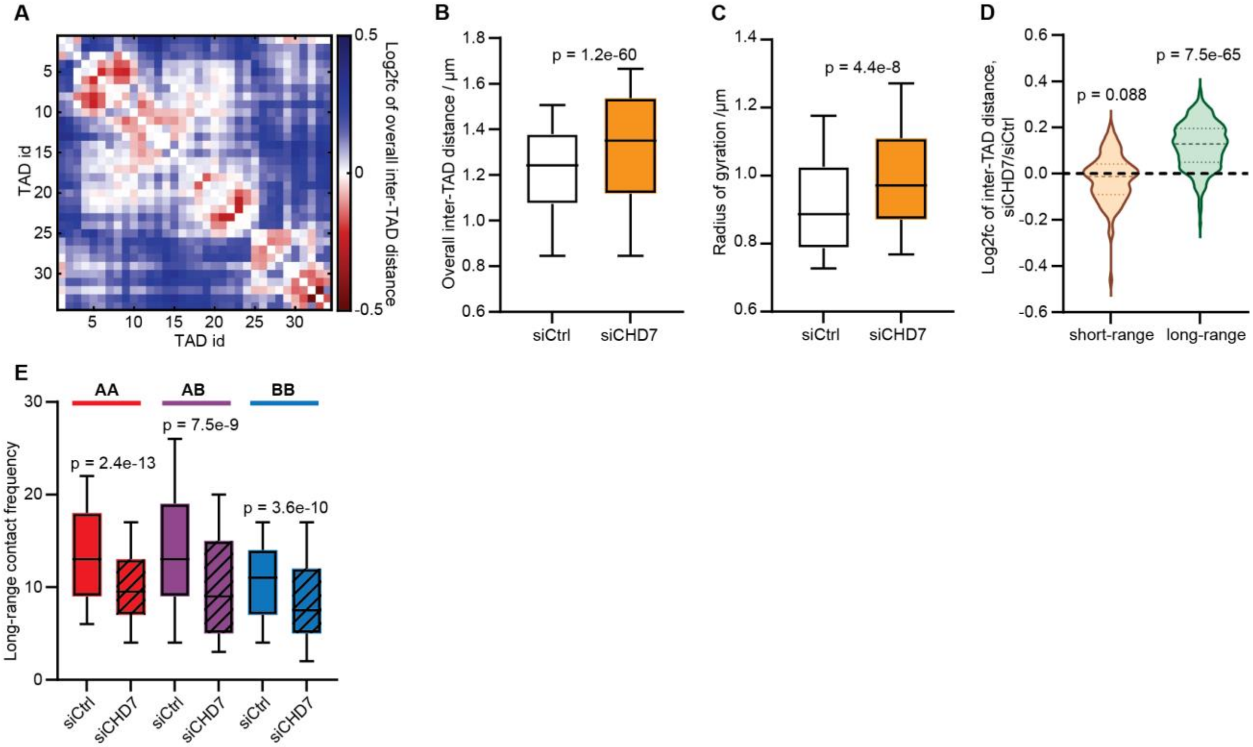


Fig. S5. Validation of CHD7 perturbation phenotypes in a different cell background and genomic context. (**A**) Log2 fold change matrix of overall inter-TAD distance of chr21 between siCHD7 and siCtrl in hTERT-RPE1 cells. (**B**) Overall inter-TAD distance of chr21 in siCtrl and siCHD7 cells. (**C**) Radii of gyration of chr21 in siCtrl and siCHD7 cells. (**D**) Log2 fold change of short-range and long-range inter-TAD distances between siCHD7 and siCtrl. (**E**) Long-range contact frequency in A compartments (red), across A and B compartments (purple) and in B compartments (blue) of chr21 in siCtrl and siCHD7 cells. P values in (B) and (D) were calculated by two-sided Wilcoxon signed rank test. P values in (C) and (E) were calculated by two-sided Wilcoxon rank sum test. Number of traces analyzed: 904 (siCtrl), 210 (siCHD7).

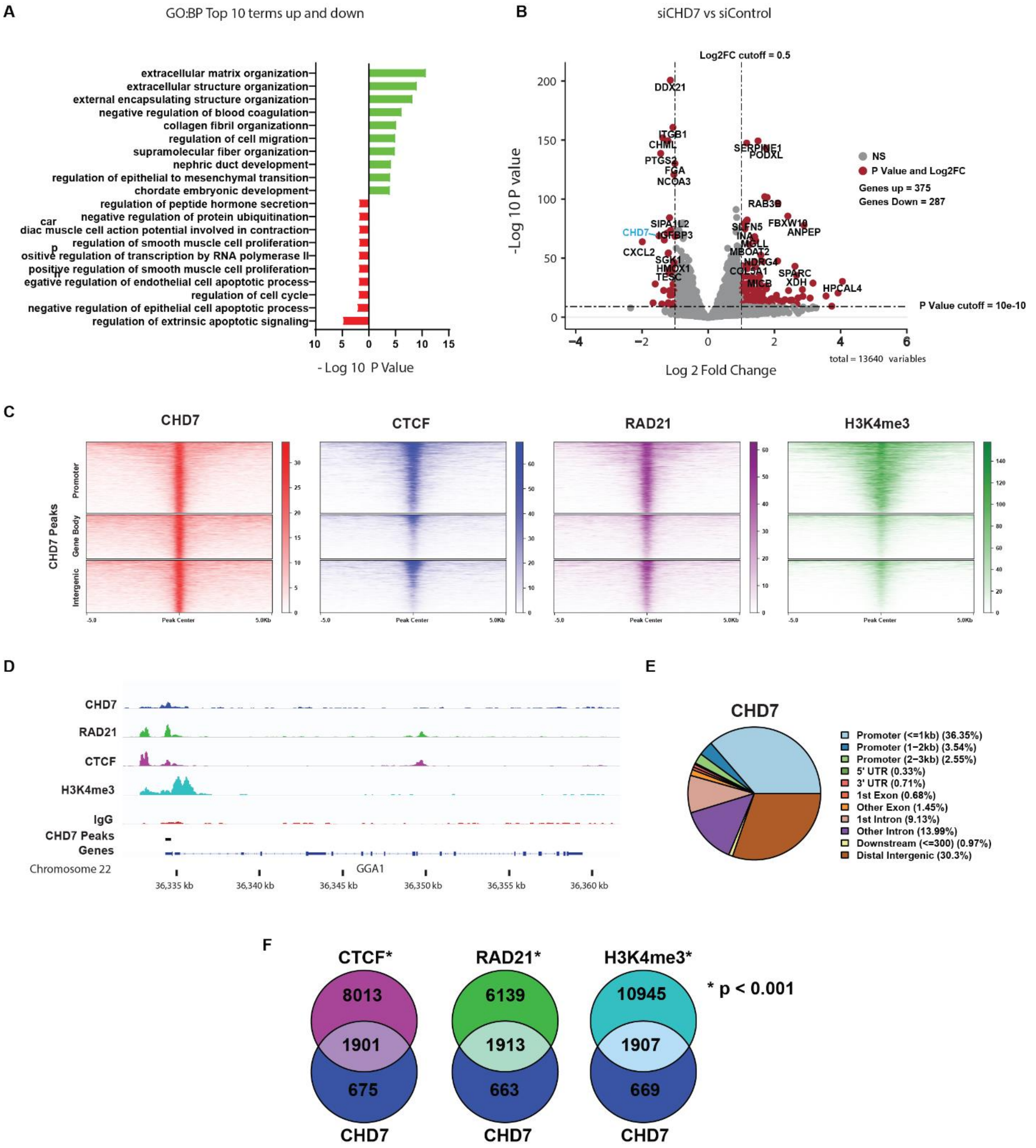


Fig. S6. CHD7 has diverse effects on transcription and binds diverse genomic regions. (**A**) Top 10 gene ontology terms up and down in siCHD7 cells versus siControl cells. Gene ontology was performed using Enrichr. (**B**) Volcano plot of RNA-seq of siCHD7 and siControl. Top differentially expressed genes are displayed on the graph as labels. CHD7 is highlighted and is a top differentially downregulated gene in the siCHD7 cells, validating the knockdown. (**C**) Heat map of other proteins/epigenetic mark localized to CHD7 peaks by CUT&RUN. (**D**) Example tracks of CUT&RUN peak profiles of CHD7 and other proteins/epigenetic mark over a gene. (**E**) Peak annotation for all CHD7 peaks. (**F**) Overlap of CUT&RUN peaks of CTCF, RAD21, and H3K4me3 with CHD7 peaks on promoter regions.

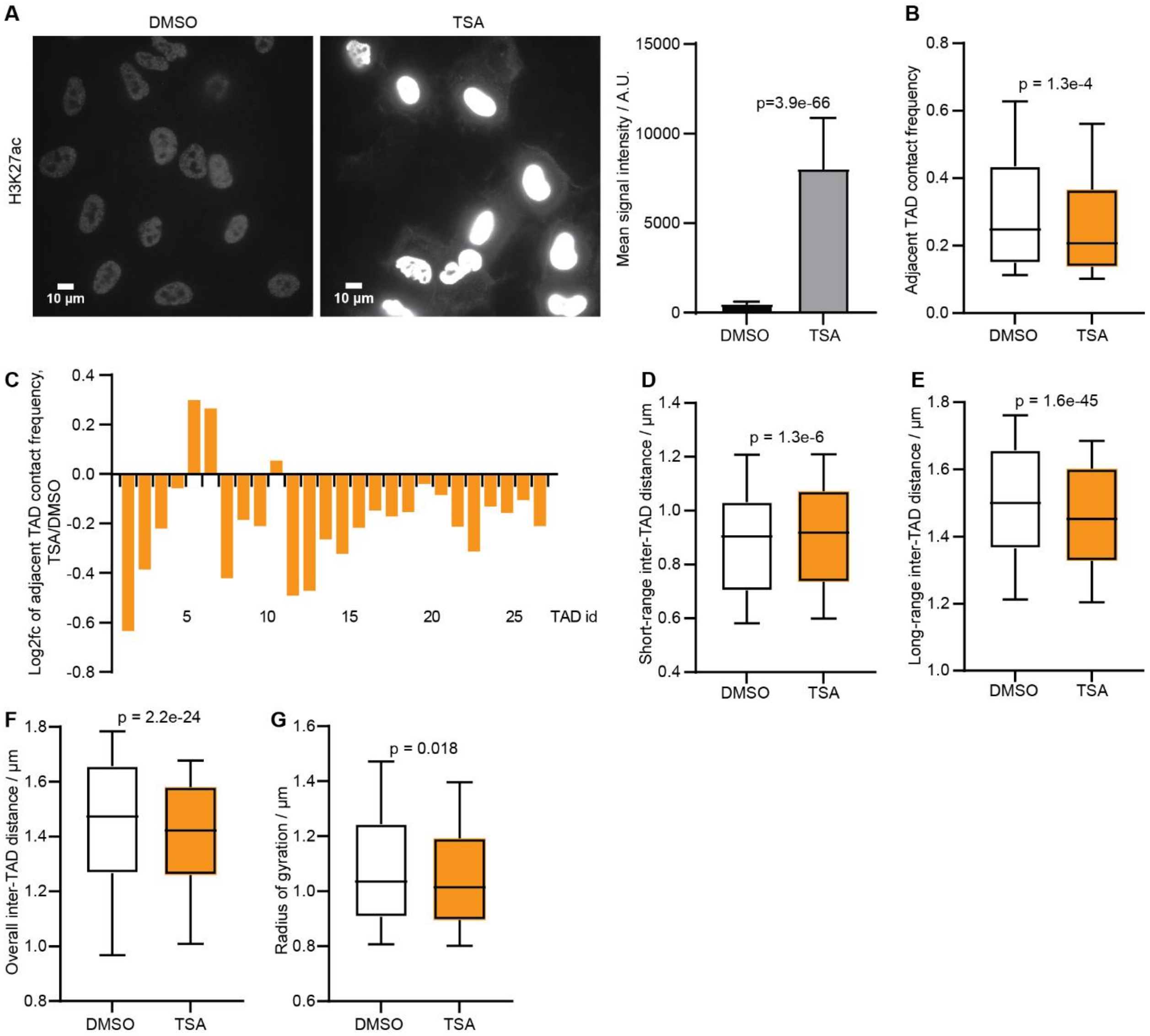


Fig. S7. HDAC inhibition causes short-range chromatin decompaction and long-range chromatin compaction. (**A**) H3K27ac IF measurements of DMSO-treated control (left) and TSA-treated (middle) A549 cells, and the quantification of mean signal intensity (right). (**B**) Adjacent TAD contact frequency of DMSO- and TSA-treated cells. (**C**) Log2 fold change of contact frequency between each pair of adjacent TADs along chr22 of TSA-treated cells compared to DMSO-treated cells. (**D**) Short-range inter-TAD distance along chr22 of DMSO- and TSA-treated cells. (**E**) Long-range inter-TAD distance along chr22 of DMSO- and TSA-treated cells. (**F**) Overall inter-TAD distance of chr22 of DMSO- and TSA-treated cells. (**G**) Radii of gyration of chr22 in DMSO- and TSA-treated cells. P value in (A) was calculated by unpaired t test. P values in (B) and (D) to (F) were calculated by two-sided Wilcoxon signed rank test. P value in (G) was calculated by two-sided Wilcoxon rank sum test.

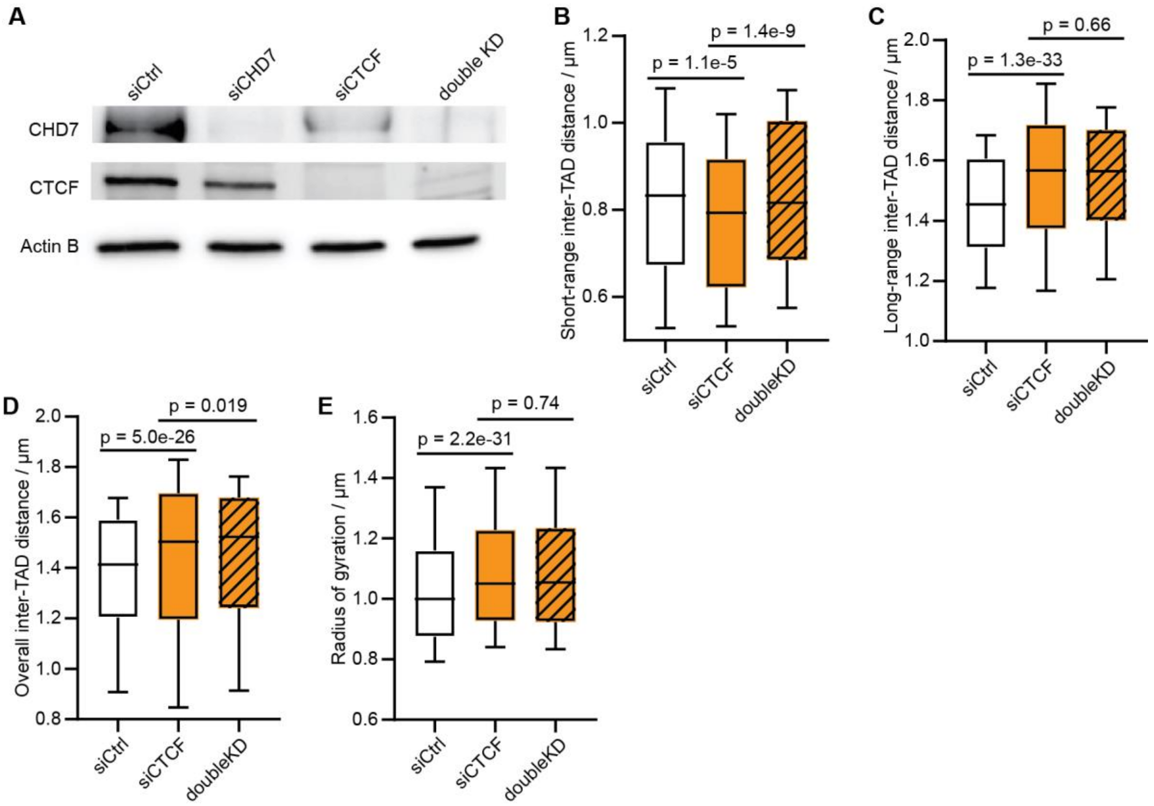


Fig. S8. CTCF knockdown revealed mechanism of CHD7-mediated large-scale chromatin compaction. (**A**) Western blot of siCtrl, siCHD7, siCTCF and doubleKD (siCHD7+siCTCF) in A549-Cas9 cells. Top: anti-CHD7 antibody; middle: anti-CTCF antibody; bottom: anti-Actin B antibody. (**B**) Short-range inter-TAD distance of chr22 in siCtrl, siCTCF and doubleKD cells. (**C**) Long-range inter-TAD distance of chr22 in siCtrl, siCTCF and doubleKD cells. (**D**) Overall inter-TAD distance of chr22 in siCtrl, siCTCF and doubleKD cells. (**E**) Radii of gyration of chr22 in siCtrl, siCTCF and doubleKD cells. P values in (B) to (D) were calculated by two-sided Wilcoxon signed rank test. P values in (E) was calculated by two-sided Wilcoxon rank sum test.

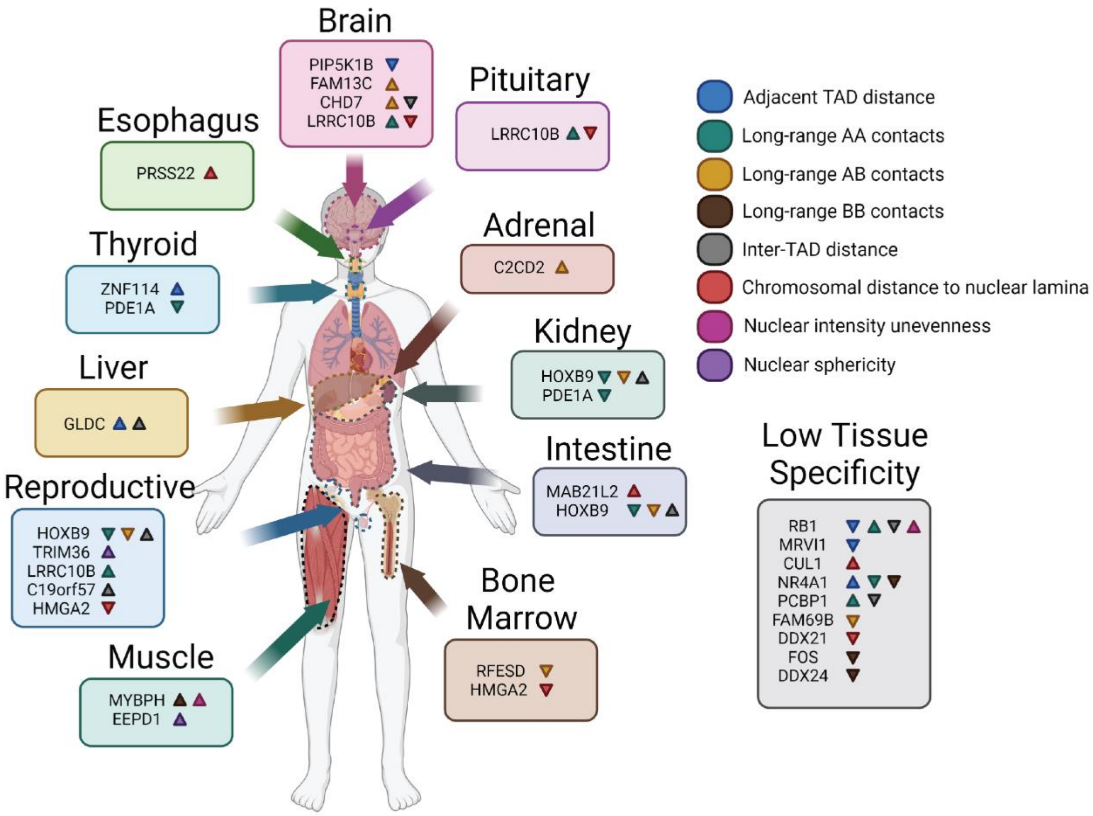

Brain
PIP5K1B ▼
FAM13C ▲
CHD7 ▲▼
LRRC10B ▲▼
Pituitary
LRRC10B ▲▼
Esophagus
PRSS22 ▲
Thyroid
ZNF114 ▲
PDE1A ▼
Adrenal
C2CD2 ▲
Kidney
HOXB9 ▼▼▲
PDE1A ▼
Liver
GLDC ▲▲
Intestine
MAB21L2 ▲
HOXB9 ▼▼▲
Reproductive
HOXB9 ▼▼▲
TRIM36 ▲
LRRC10B ▲
C19orf57 ▲
HMGA2 ▼
Bone Marrow
RFESD ▼
HMGA2 ▼
Muscle
MYBPH ▲▲
EEPD1 ▲
● Adjacent TAD distance
● Long-range AA contacts
● Long-range AB contacts
● Long-range BB contacts
● Inter-TAD distance
● Chromosomal distance to nuclear lamina
● Nuclear intensity unevenness
● Nuclear sphericity
Low Tissue Specificity
RB1 ▼▲▼▲
MRVI1 ▼
CUL1 ▲
NR4A1 ▲▼▼
PCBP1 ▲▼
FAM69B ▼
DDX21 ▼
FOS ▼
DDX24 ▼

Fig. S9. Tissue specific expression of top hits based on bulk RNA sequencing results from the Human Protein Atlas. Arrow head colors indicate different 3D genome architecture metrics. Arrow head directions (up and down) indicate the directions of the regulatory effects of the genes on the metrics (opposite to the directions of the knockouts). This figure was created with Biorender.

Table S1 (Separate file).
Target gene summary. Three categories of target genes are included in the screen, as described in the table.

Table S2 (Separate file).
Disease relevance of top hits identified in the screen.

Table S3 (Separate file).
Digits and values for DNA barcode sequences in BARC-FISH screen, the corresponding linear and padlock probe sequences, and dye-labeled secondary probes.

Table S4 (Separate file).

Oligonucleotide sequences of helper probes.

Table S5 (Separate file).

Oligonucleotide sequences and PCR primers for barcode library assembly.

Table S6 (Separate file).

Protospacer sequences for gene targets in the screen, and the oligonucleotide sequences and PCR primers for sgRNA library assembly.

Table S7 (Separate file).

PCR primers for sgRNA-barcode CRISPR screen library cloning.

Table S8 (Separate file).
Primers for CHD7 overexpression plasmid cloning.

Table S9 (Separate file).

Oligonucleotide sequences for sgRNA-barcode sequencing library preparation.

Table S10 (Separate file).

Codebook determined by sequencing. The “Goodcodes” worksheet contains good barcodes that each uniquely associate with an sgRNA, in which the first column represents the 10-digit barcode, and the second column represents the name of the sgRNAs listed in Table S6. The “Badcodes” worksheet contains barcodes that associate with multiple sgRNAs.

Table S11 (Separate file).

Template oligonucleotide sequences, PCR primers and reverse transcription primer used for chromatin tracing primary probe library synthesis.

Table S12 (Separate file).

Secondary probe sequences for chromatin tracing.

Table S13 (Separate file).

BARC-FISH screen results summary. 8 worksheets regarding 8 phenotype analyses were included, and each worksheet consists of the following columns:

sgRNA id: sgRNA ranking alphabetically, representing their orders in Table S6;

sgRNA name: name of the sgRNAs as represented in Table S6;

copy number/cell number: number of traces (for chromatin-related phenotypes) or cells (for nuclear morphological properties) analyzed;

p, fdr, log2fc: p value, false discovery rate, and log2 fold change of the analyzed phenotypes.

Supplementary File S1 (Separate file).
Backbone sequence used for barcode plasmid library construction.

Supplementary File S2 (Separate file).
Example barcode plasmid sequence.

Supplementary File S3 (Separate file).
Backbone sequence used for sgRNA plasmid library construction.

Supplementary File S4 (Separate file).
Example sgRNA plasmid sequence.

Supplementary File S5 (Separate file).
Backbone sequence used for CRISPR screen plasmid library construction.

Supplementary File S6 (Separate file).
Example CRISPR screen plasmid sequence.

Supplementary File S7 (Separate file).
Sequence of plasmid for CHD7 overexpression.