A genome-wide single-cell 3D genome atlas of lung cancer progression

Miao Liu†[1], Shengyan Jin†[1], Sherry S. Agabiti†[1,2], Tyler B. Jensen[1,3], Tianqi Yang[1], Jonathan S. D. Radda[1], Christian F. Ruiz[1,2], Gabriel Baldissera[1], Mandar Deepak Muzumdar*[1,2,3,4,5,6,7], Siyuan Wang*[1,3,6,7,8,9,10,11]

Affiliations:

[1]Department of Genetics, Yale School of Medicine, Yale University; New Haven, CT 06510, USA.

[2]Yale Cancer Biology Institute, Yale University; West Haven, CT 06516, USA.

[3]M.D.-Ph.D. Program, Yale University; New Haven, CT 06510, USA.

[4]Department of Internal Medicine, Section of Medical Oncology, Yale School of Medicine, Yale University; New Haven, CT 06510, USA.

[5]Yale Cancer Center, Smilow Cancer Hospital, New Haven, CT 06510, USA.

[6]Yale Combined Program in the Biological and Biomedical Sciences, Yale University; New Haven, CT 06510, USA.

[7]Molecular Cell Biology, Genetics and Development Program, Yale University; New Haven, CT 06510, USA.

[8]Department of Cell Biology, Yale School of Medicine, Yale University; New Haven, CT 06510, USA.

[9]Biochemistry, Quantitative Biology, Biophysics, and Structural Biology Program, Yale University; New Haven, CT 06510, USA.

[10]Yale Center for RNA Science and Medicine, Yale University School of Medicine; New Haven, CT 06510, USA.

[11]Yale Liver Center, Yale University School of Medicine; New Haven, CT 06510, USA.

*Corresponding authors. Email: mandar.muzumdar@yale.edu, siyuan.wang@yale.edu

†These authors contributed equally to this work.

Abstract:

Alterations in three-dimensional (3D) genome structures are associated with cancer[1–5]. However, how genome folding evolves and diversifies during subclonal cancer progression in the native tissue environment remains unknown. Here, we leveraged a genome-wide chromatin tracing technology to directly visualize 3D genome folding *in situ* in a faithful *Kras*-driven mouse model of lung adenocarcinoma (LUAD)[6], generating the first single-cell 3D genome atlas of any cancer. We discovered stereotypical 3D genome alterations during cancer development, including a striking structural bottleneck in preinvasive adenomas prior to progression to LUAD, indicating a stringent selection on the 3D genome early in cancer progression. We further showed that the 3D genome precisely encodes cancer states in single cells, despite considerable cell-to-cell heterogeneity. Finally, evolutionary changes in 3D genome compartmentalization – partially regulated by polycomb group protein Rnf2 through its ubiquitin ligase-independent activity – reveal novel genetic drivers and suppressors of LUAD progression. Our results demonstrate the importance of mapping the single-cell cancer 3D genome and the potential to identify new diagnostic and therapeutic biomarkers from 3D genomic architectures.

Main Text:

Cancer cells exhibit profound alterations in nuclear size, shape, and chromatin texture[7]. Microscopic examination of these structural features remains a gold standard for diagnosis and establishing pathologic cancer grade, which is frequently associated with prognosis[7]. It has become increasingly clear that the three-dimensional (3D) genome folding organization within the nucleus of cells plays a key role in both normal functions and disease[8–15]. Several recent studies have applied sequencing-based 3D genomics technologies – such as high-throughput chromosome conformation capture (Hi-C) – and reported systematically altered genome organization in cancer cells[1–5]. Yet, these studies relied on population averaging of cells and indirect inference of genome organization based on chromatin contacts. As a result, how the genome is folded in true 3D in individual cancer cells *in vivo* and how this folding organization evolves and is regulated during cancer progression from normal to preinvasive to heterogeneous invasive tumor cells within the native tissue environment remains elusive. To address these gaps in knowledge, we used an image-based 3D genomics method that we pioneered termed chromatin tracing[16,17] to develop the first *in situ* single-cell 3D genome cancer atlas.

Genome-wide chromatin tracing in a mouse model of lung adenocarcinoma

To directly visualize chromatin folding organization in single cells, we applied chromatin tracing to a genetically engineered mouse model (**Fig. 1a and Extended Data Fig. 1a, b**) that faithfully recapitulates the genetic and histologic progression of human lung adenocarcinoma (LUAD), a leading cause of cancer-related death worldwide[18]. We mapped genome-wide 3D chromatin organization by targeting a panel of 473 genomic loci spanning all 19 mouse autosomes at an average genomic resolution of 5 Mb (**Supplementary Tables 1 to 3**), including genomic regions harboring “classic” oncogenes, tumor suppressor genes, and super-enhancers. To unambiguously

distinguish each genomic locus in chromatin tracing, we adapted a previously published DNA multiplexed error-robust fluorescence *in situ* hybridization (DNA MERFISH) 100-choose-2 combinatorial barcoding design and assigned each target genomic locus a unique 100-bit binary barcode, with each barcode containing two "1" bits and 98 "0" bits[19]. Each "1" bit physically corresponded to a unique overhang readout sequence on the primary probes hybridized to the target genomic loci (**Fig. 1a**). The set of primary probes targeting each genomic locus contained two versions of readout sequences that corresponded to the two "1" bits in the barcode of the locus. We hybridized all primary probes to all target genomic loci and sequentially hybridized dye-labeled readout probes to image the 100 bits, with 50 hybridization rounds and two-color imaging (**Fig. 1a**). We then fitted the 3D positions of the imaged fluorescence *in situ* hybridization (FISH) foci, decoded the barcodes, and reconstructed chromatin traces in single cells (**Figs. 1b, c**). To validate the quality of the genome-wide chromatin tracing in the mouse lung, we measured the median spatial distance between each pair of target genomic loci of alveolar type 2 (AT2) cells – the putative cell-of-origin for LUAD[20] – in wild-type (WT) mice and constructed a median inter-loci distance matrix (**Fig. 1d**). The matrix showed lower *intra*-chromosomal distances and higher *inter*-chromosomal distances, consistent with chromosome territory organization[21]. Inter-loci distances measured from WT AT2 biological replicates were highly correlated with each other (Pearson correlation coefficient = 0.97; **Extended Data Figs. 1c, d**). Within each chromosome, inter-loci spatial distances scaled with their genomic distances, with longer chromosomes reaching larger spatial distances (**Extended Data Fig. 1e**). On average, inter-loci spatial distances followed a power-law function in comparison to genomic distances, with a scaling factor of about one-tenth (**Extended Data Fig. 2**), similar to scaling

factors calculated based on previous reports in other mouse tissue types (**Extended Data Fig. 3**)[22,23].

Genetically engineered mouse models allow for the study of autochthonous tumors that mimic the genetic and histologic features of human tumor progression in the native tissue environment. The proto-oncogene *KRAS* and the tumor suppressor gene *TP53* (mouse gene *Trp53*) are the most frequently mutated genes in human LUAD[24–26]. To trace the evolution of 3D genome organization during *Kras/Trp53*-driven LUAD progression in mice, we leveraged the **M**osaic **A**nalysis with **D**ouble **M**arkers (MADM) system[27] and established the *K-MADM-Trp53* lung cancer model that induces sparse and sequential mutagenesis of *Kras* and *Trp53* in the lung epithelium, faithfully mimicking the genetic progression of human LUAD development. To distinguish the genetic states of cells, the MADM system generates sibling green-fluorescent protein-expressing (GFP+) homozygous mutant and tdTomato-expressing (tdTomato+) homozygous WT cells from a rare Cre/loxP-mediated *inter*-chromosomal mitotic recombination (G2-X) event in a heterozygous parent, unambiguously coupling genotype to fluorescence labeling (**Extended Data Fig. 4)**[27,28]. MADM also generates yellow (GFP+/tdTomato+) heterozygous cells through non-mitotic inter-chromosomal recombination (G0/G1) or mitotic recombination followed by Z segregation (**Extended Data Fig. 4**). Therefore, the *K-MADM-Trp53* mouse model permits tracing of subclones of green (GFP+) *Trp53*$^{-/-}$, red (tdTomato+) *Trp53*$^{+/+}$, and yellow (GFP+/tdTomato+) *Trp53*$^{+/-}$ cells induced in preinvasive *Kras* mutant lung adenomas initiated by inhaled lentiviral Cre infection[6]. We observed that *Trp53*$^{-/-}$, but not *Trp53*$^{+/+}$ or *Trp53*$^{+/-}$ adenoma cells, progressed to advanced LUAD[6]. While a majority of preinvasive adenomas harbored *Trp53*$^{-/-}$ clones that were morphologically similar to their *Trp53*$^{+/+}$ or *Trp53*$^{+/-}$ counterparts, only a small minority of these *Trp53*$^{-/-}$ cells progressed to

advanced adenocarcinoma[6]. Thus, *Trp53* loss was permissive but insufficient to drive cancer progression, suggesting that additional cellular or molecular events must occur.

We hypothesized that changes in 3D genome organization might accompany and encode this histologic transition. To test this hypothesis, we applied the genome-wide chromatin tracing method to histologically defined normal AT2 cells, preinvasive adenoma (red *Trp53*$^{+/+}$, yellow *Trp53*$^{+/-}$, and green *Trp53*$^{-/-}$, hereafter referred to as AdenomaR, AdenomaY, and AdenomaG) and LUAD (green *Trp53*$^{-/-}$) cells within lung tissue sections. We restricted the 3D genome organization analyses to AT2/cancer cells by staining for the AT2 cell type marker surfactant protein C (SPC) (**Fig. 1e**). Furthermore, since advanced LUAD tumors tend to lose SPC expression and the MADM model labels both cancer cells and macrophages (presumably through lentiviral Cre infection), we excluded immune cells expressing the pan-immune marker CD45 (**Fig. 1e**). We combined multiplexed fluorescence imaging of GFP, tdTomato, SPC, and CD45 with genome-wide chromatin tracing on the same tissue sections to achieve single-cell cancer-state-specific 3D genome profiling (**Extended Data Fig. 1a**). To build a complete 3D genome atlas that captures the full spectrum of tumor development, we collected 17 replicates across multiple lung lobes (most containing several discrete tumors) from 13 mice harboring different stages of LUAD progression (normal AT2, preinvasive AdenomaR, AdenomaY, AdenomaG, and LUAD). A total of 1,540,582 chromatin traces in 84,074 cells were analyzed.

A structural bottleneck of the 3D genome in early lung cancer progression

Recent studies have shown that unlike stable protein structures, the precise 3D folding conformation of chromatin is highly variable even among cells of the same type[29–33]. We reasoned that the heterogeneous nature of cancer development[34] may similarly lead to

increasingly diverse chromatin conformations during subclonal progression. Alternatively, the selection pressure on cancer cells during progression may restrict the level of 3D genome conformation heterogeneity and favor cells with defined conformations required for cancer development. To test these opposing possibilities, we quantified the conformational heterogeneity of the 3D genome in each cell state by measuring the coefficient of variation (COV) of inter-loci distances along each autosome (**Figs. 2a, b and Extended Data Fig. 5a**). Strikingly, chromatin folding conformations became less heterogeneous during early progression from AT2 to preinvasive adenoma cells, which reversed upon progression to LUAD (**Fig. 2b and Extended Data Fig. 5a)**. Whereas more heterogeneous conformations were tolerated in both AT2 and LUAD cells, adenoma cells exhibited more constrained 3D genome architectures prior to progression to LUAD, arguing for early stringent selection on 3D genome organization during lung cancer development.

To further define the dynamics of 3D genome alterations during LUAD progression, we quantified the population-average of several structural features of chromosomes *in situ* for each cell state (AT2, Adenoma (R, Y, G), LUAD), including: (1) *intra*-chromosomal compaction (mean inter-loci distances on each autosome), (2) long-range intermixing (calculated using a demixing score metric benchmarked using published active and inactive X chromosome folding conformations in human cells[16]) (**Extended Data Fig. 5b**), (3) A (active chromatin) and B (inactive chromatin) compartment polarization (quantified using a polarization index metric[16]), (4) *inter*-chromosomal distance, (5) end-to-end chromosomal alignment (Rabl configuration score[35]), and (6) radial localization of each genomic locus in the nucleus (see **Methods**). When comparing all adenoma cell states to AT2 cells, we observed that chromosomes exhibited *increased* compaction (decreased decompaction score; **Figs. 2a, c and Extended Data Fig. 5c**)

but *reduced* long-range intermixing (increased demixing score; **Fig. 2d and Extended Data Fig. 5d**) and *increased* A-B compartment polarization (**Fig. 2e**). Chromosomes in adenoma states were globally less aligned (reduced Rabl score; **Fig. 2f**) and more spatially associated (reduced *inter*-chromosomal distance; **Extended Data Figs. 5e-g**), with a subset (*e.g.*, Chr 7, 13, and 19) showing consistent changes in radial distribution within the nucleus compared to AT2 cells (**Fig. 2g and Extended Data Fig. 5i**). Strikingly, all of these phenotypes largely reverted in LUAD cells (**Fig. 2 and Extended Data Fig. 5**). While *intra*-chromosomal distances in LUAD only partially reverted and were still lower than the level of AT2 cells (**Fig. 2c and Extended Data Fig. 5c**), *inter*-chromosomal distances in LUAD largely surpassed those of AT2 cells (**Extended Data Fig. 5h**), leading to an overall larger nuclear volume in LUAD than in AT2 and adenoma cells (**Extended Data Fig. 5j**). Together, these data argue that upon transformation of AT2 cells into preinvasive adenomas, they experience stereotypical, reversible alterations in their 3D genome conformations, which could represent a structural bottleneck for tumor development.

Histologic transitions between AT2, adenoma, and LUAD cells are accompanied by differences in cell cycle progression and copy number variation (CNV), which could confound the 3D genomic structural findings. Our prior pulse EdU incorporation experiments in *K-MADM-Trp53* mice revealed that ~2% of preinvasive adenoma cells are actively proliferating compared to ~8% of LUAD cells[6]. Therefore, most of our profiled cancer cells were in the G0/G1 phase, which should have minimal confounding effects on chromatin folding and nuclear organization[36]. Furthermore, the monotonic increase of the proliferative proportion of cells in the AT2 to adenoma to LUAD populations cannot explain the reversible trend of 3D genome changes observed across these states. Whole-exome sequencing of AdenomaG and LUAD cells isolated from *K-MADM-Trp53* mice showed only a subset of LUAD tumors exhibiting evidence of gains

of Chromosome 6 and losses of Chromosomes 4, 9, and 11 (**Extended Data Figs. 6a, b**), as has been observed in other *Kras/Trp53* mutant murine LUAD models[37,38]. Given the lack of systematic gains or losses across most chromosomes and tumors, it is unlikely that CNV could account for the global chromatin structure differences observed between adenoma and LUAD cells.

In contrast to the cancer-state-specific differences in the 3D genome described above, we identified several cancer-state-independent 3D genome features. For example, we found that spatial proximity frequencies were the highest for *trans*-chromosomal A-A compartment interactions, followed by A-B and B-B interactions (**Extended Data Figs. 7a-i**). The same trend was also observed in long-range *cis*-chromosomal compartment interactions (**Extended Data Figs. 7j-l**). Furthermore, we identified a consistent negative correlation between radial scores and A-B compartment scores (**Extended Data Fig. 7m**). These observations align with previously reported findings in mouse fetal livers and human lung fibroblast cells[19,23], suggesting that these features likely represent general principles of 3D genome organization independent of cell state, cell type, and even mammalian species.

Organization of the 3D genome encodes specific cancer cell states at the single-cell level

Given the heterogeneous nature of the 3D genome, it remained unclear whether single-cell 3D genome conformations could faithfully encode and distinguish cell states, especially between AT2 and LUAD cells, as both states contain highly variant 3D genome structures with similar population-averaged features (**Fig. 2**). Nonetheless, subsampling analysis showed that as few as 250 cells per cell state were sufficient to recapitulate similar 3D genome organization changes for most structural features described above (**Extended Data Fig. 8**), indicating the potential of

cancer 3D genomic architectures to be developed into state-specific *in situ* single-cell biomarkers. To test this, we performed dimensionality reduction of single-cell 3D genome conformations to visualize clusters of different cell states in a low dimensional representation. We adopted a previously developed single-cell A/B compartment (scA/B) score metric for all target genomic loci (**Fig. 3a**)[35,39] and used these scores as high-dimensional input variables to generate a two-dimensional display with three independent dimensionality reduction methods[40–42]. On these plots, the single-cell 3D genome conformations of AT2 and LUAD cells were largely distinct, while those of preinvasive adenoma cells were clustered in a tight area at the transition between these populations (**Fig. 3a**), consistent with the structural bottleneck described above. We next trained a series of supervised machine learning models to classify the different cancer cell states based on the single-cell 3D genome data (see **Methods**). Among the models, a support vector machine model showed the best performance, with 91.6% overall accuracy in predicting cancer cell states (**Figs. 3b, c**). These results showed that 3D genome organization encodes and distinguishes specific histologic cell states during oncogenesis at the single-cell level.

Since the 3D genome specifies cancer states, we speculated that a subset of *Trp53*$^{-/-}$ AdenomaG cells may have acquired specific 3D genome features underlying their potential to progress into LUAD in the absence of clear additional mutational drivers[38]. Indeed, we observed that AdenomaG cells exhibited chromosome-scale structural features often intermediate to that of non-progressing adenoma red and yellow (AdenomaRY) cells and LUAD cells (**Figs. 2b-d**). To further test this hypothesis, we performed 3D genome dimensionality reduction analyses and found that preinvasive AdenomaRY cells and invasive LUAD cells formed distinct clusters, whereas AdenomaG cells were dispersed within and in between both clusters (**Fig. 3d**). With

either a linear division approach (**Figs. 3d, e**) or a k-nearest-neighbor method (**Fig. 3e**), we computationally segmented the low-dimensional space into two regions of adenoma-like and LUAD-like conformations and quantified the percentage of AdenomaG cells distributed in the two regions (see **Methods**). We found that a greater proportion of AdenomaG cells (vs. other adenoma cells) adopted LUAD-like 3D chromatin organization (*i.e.*, located in the LUAD-like region) (**Figs. 3d, e**). This result indicated that 3D chromatin organization changes associated with LUAD preceded histologic progression and may drive the adenoma-to-LUAD transition.

Alterations in 3D genome organization delineate genetic vulnerabilities in LUAD

To identify the genetic mechanisms linking 3D genome reorganization to cancer progression, we performed RNA-seq on distinct GFP+ tumors of *K-MADM-Trp53* mice (**Extended Data Fig. 6c and Supplementary Table 4**). Unsupervised hierarchical clustering revealed two distinct tumor groups (**Extended Data Fig. 6d**) that could be classified as AdenomaG and LUAD cells based on expected changes in gene expression with progression, including loss of lung epithelial markers (AT1/AT2), increased expression of gastric identity, epithelial-to-mesenchymal transition (EMT), and metastasis genes, and transcriptional alterations associated with enhanced Kras signaling (**Extended Data Figs. 6e, f**)[37]. We focused on gene expression changes within marker genomic loci with significantly altered scA/B scores between AdenomaG and LUAD cells (**Fig. 4a**). We found that in comparison to genes in non-marker loci with unchanged scA/B scores, genes in marker loci with increased scA/B scores showed significantly up-regulated expression levels, whereas those in marker loci with decreased scA/B scores had significantly reduced expression (**Fig. 4b**). We reasoned that genes with significantly elevated expression in marker genomic loci with increased scA/B scores from AdenomaG to LUAD cells could be candidate progression drivers (CPDs) of LUAD (**Fig. 4c and Supplementary Table 5**). In

contrast, genes with significantly decreased expression in marker loci with decreased scA/B scores could be candidate tumor suppressors (CTSs) of LUAD (**Fig. 4c and Supplementary Table 5**). In an analysis of LUAD tumors from the Cancer Genome Atlas (TCGA), individual patient tumor samples were scored for their correlation with the CPD and CTS gene signatures, respectively. For the CPD gene signature, high-scoring patients (top 40% most correlated tumor gene expression, n = 191) exhibited significantly worse survival compared to patients with lower scores (bottom 40%, n=191) whereas the inverse was true for the CTS gene signature (**Fig. 4d**). Among the 16 CTS genes, 7 of them have prior evidence as tumor suppressors (**Supplementary Table 5**).

To validate the function of the CPD genes in LUAD progression, we performed an arrayed lentiviral RNAi screen targeting 19 CPD genes with readily-available short hairpin RNAs (shRNAs, 3 per gene; **Supplementary Table 6**) in two independent primary LUAD mouse cell lines isolated from the *K-MADM-Trp53* mouse model (SA6082inf) and an analogous *Kras/Trp53* mutant (KP) mouse model[38]. We found that knockdown of 17 of 19 (89%) tested CPD genes reduced viability in both cell lines (**Fig. 4e**). In comparison, 16% of randomly selected genes would have been expected to affect cell proliferation based on an analysis of the Cancer Dependency Map[43–46] (see **Methods**). Importantly, clear pathogenic roles for several CPDs (*e.g.*, *Ephx2*, *Peg3*, *Olfm4*, *Oit1*, and *Slc6a19*) have not been previously reported in LUAD (**Supplementary Table 5**). Finally, we identified 16 candidate tumor initiation genes located in marker genomic loci with increased scA/B scores and significantly increased expression from AT2 to adenoma cells[37] (**Supplementary Table 5**). Strikingly, one of these genes, *Myc*, is a known downstream effector of *Kras* and is capable of initiating lung tumors in mice[38], arguing that 3D genome compartment alterations may enhance Myc activity to drive tumor initiation. Together, these results

demonstrated that single-cell *in situ* 3D genome profiling can identify cancer state-specific marker genomic loci, nominate novel epigenetically-regulated cancer regulators in tumor initiation and progression, and reveal genetic dependencies in LUAD.

Rnf2 partially regulates 3D genome changes during the adenoma-to-LUAD transition through a non-canonical function

We finally sought to elucidate the upstream regulatory mechanisms of 3D genome reorganization during cancer progression. We hypothesized that 3D genomic architectures in cancer cells were governed by cell-autonomous mechanisms and/or by neighboring cells, such as immune cells, a major component of the tumor microenvironment. To explore the latter possibility, we compared 3D genome organization of AT2/cancer cells during LUAD progression in relation to their spatial proximity to CD45+ immune cells. No or only modest differences were observed between AT2/cancer cells adjacent to or far from immune cells in 3D chromatin organization features, including chromatin conformation heterogeneity, chromatin decompaction, chromatin demixing, A-B compartment polarization, and Rabl configuration (**Extended Data Figs. 9a-e**). Furthermore, based on dimensionality reduction analyses, AT2/cancer cells adjacent to and far from immune cells contained similar spectra of single-cell 3D genome conformations (**Extended Data Figs. 9f**). These results indicated that the 3D genome organization changes during LUAD progression are likely independent of immune cell interactions.

To identify cell-autonomous regulators of cancer 3D genome reorganization, we leveraged a published Binding Analysis for Regulation of Transcription (BART) algorithm[47,48] to predict transcription factors and chromatin regulators that bind to upregulated genes located in marker loci with increased scA/B scores in the adenoma-to-LUAD transition. We identified a polycomb

group protein Rnf2 as a top candidate (**Fig. 5a**). Rnf2 is a major component of the polycomb repressive complex 1 (PRC1) and catalyzes monoubiquitination of lysine 119 of histone H2A (H2AK119ub1)[49]. Canonically, PRC1 works in concert with polycomb repressive complex 2 (PRC2), which modifies histones by depositing H3K27me3[50]. To study the 3D genome regulatory functions of Rnf2 in LUAD, we knocked down Rnf2 (shRnf2) in KP LUAD cells (**Fig. 5b**) and saw a significant negative effect on cell proliferation (**Fig. 5c, d**). We found that the scA/B score changes upon Rnf2 knockdown significantly correlated with scA/B score changes during LUAD progression (**Fig. 5e**), indicating that Rnf2 partially regulates 3D genome reorganization in LUAD cells. Since the predicted binding sites of Rnf2 were associated with increased scA/B scores (**Fig. 5a**), we speculated that Rnf2 may preferentially bind A compartment regions. Consistent with this hypothesis, CUT&RUN of Rnf2 revealed significantly higher Rnf2 peak density in compartment A TADs (**Fig. 5f).**

The canonical function of Rnf2 in the PRC1 complex is to mediate gene silencing through polycomb repression[49], but Rnf2 can also non-canonically associate with epigenetically active loci[51,52]. As both polycomb-repressed and active chromatin regions tend to localize in compartment A[53,54], it remained unclear whether the Rnf2-controlled regions in LUAD were mainly polycomb-repressed or active chromatin. To distinguish these possibilities, we aligned and clustered the Rnf2 CUT&RUN peaks in all target genomic regions with H3K4me3 and H3K27me3 CUT&RUN profiles. We found that the majority (51.6%) of Rnf2 peaks resided in active regions (H3K4me3+, H3K27me3-) (**Fig. 6a**). Rnf2 peaks were found less frequently (23.3% and 4.3%, respectively) in repressive (H3K4me3-, H3K27me3+) and bivalent (H3K4me3+, H3K27me3+) regions (**Fig. 6a**). CUT&RUN analysis of RNA polymerase II with phosphorylated S5 confirmed the active transcription states of the active regions (**Fig. 6a**). In

addition, scA/B score changes upon Rnf2 knockdown were better correlated with the scA/B score changes in LUAD progression, if only active Rnf2-bound regions were included in the correlation analysis (**Fig. 6b**). These results suggest that Rnf2 in active genomic loci regulates 3D genome reorganization associated with LUAD progression.

To further link the regulatory role of Rnf2 on the 3D genome to its molecular function, we performed H2AK119ub CUT&RUN, which showed that WT Rnf2 peaks in active regions do not colocalize with strong H2AK119ub peaks. In contrast, another core PRC1 component BMI1 colocalizes with Rnf2 in all regions (**Fig. 6a**). These data argued that the canonical ubiquitin ligase function of Rnf2 is dispensable for its role in shaping the 3D genome. To test this hypothesis, we performed rescue experiments upon Rnf2 knockdown with vectors expressing wild-type (WT) Rnf2 or a catalytic Rnf2 mutant (I53S) that lacks ubiquitin ligase activity[55] (**Fig. 6c**). We found that scA/B score changes upon rescue with WT Rnf2 significantly correlated with scA/B score changes during LUAD progression (**Fig. 6d**), validating the specific effect of Rnf2. Strikingly, mutant Rnf2 (I53S) also induced scA/B score changes that mirrored LUAD progression but with a stronger and more significant correlation than that of WT rescue. (**Fig. 6d**). These results demonstrated that Rnf2 partially regulates 3D genome changes during the adenoma-to-LUAD progression via a ubiquitin ligase-independent function.

Discussion

In this study, we comprehensively profiled changes in genome-wide 3D chromatin organization in a physiologically relevant mouse model that faithfully recapitulates the genetic and histologic features of human LUAD progression, generating the first single-cell 3D genome atlas of cancer.

By combining unambiguous lineage markers with sequential mutagenesis of established cancer genes (*Kras* and *Trp53*) using MADM, we traced the evolution of 3D genome organization across the full spectrum of LUAD progression from normal (AT2) to precancerous stages (adenoma) and all the way through to advanced malignancy (LUAD). We identified a structural bottleneck of megabase-resolution chromatin folding in lung adenomas (**Extended Data Fig. 10**), which could reflect an early stringent selection of 3D chromatin organization in preinvasive tumors prior to progression to invasive LUAD. Analyses of human tumors subject to multiregional sequencing (MRS) revealed that different solid tumors undergo diverse evolutionary modes for subclonal selection[56–58]. While neutral evolution occurs in colorectal and esophageal adenocarcinomas, strong positive subclonal selection prevails in LUAD and brain tumors[58]. MRS data from nearly 250 LUAD tumors derived from prospectively enrolled patients in the longitudinal TRACERx study confirmed strong positive selection for truncal rather than subclonal mutations, with the exception of certain frequently mutated tumor-initiating genes, including *KRAS* and *TP53*[56,57], which we modeled here. This genomic bottleneck appears widespread across cancers, in which there is a transient restriction of genome diversity due to subclonal selection, followed by increased heterogeneity via mechanisms such as chromosomal instability[59,60]. Our results mirror this evolutionary genomic bottleneck and further argue that this stringent selection may also apply at the level of 3D genome structures and their resultant effects on the epigenome. The selection pressures that drive particular 3D genome conformations remain to be explored, but this state of limited heterogeneity may be susceptible to unique vulnerabilities, providing new avenues for cancer interception or therapy.

Beyond mapping 3D genomic architectures to cancer cell states during LUAD evolution, we established the utility of using this unique high-dimensional 3D genome data to identify both

candidate genetic dependencies and upstream 3D genome regulators in LUAD (**Extended Data Fig. 10**). We showed, for the first time, that single-cell 3D genome organization can distinguish and predict different cancer cell states with high prediction accuracy, underscoring its potential as a diagnostic biomarker. We further nominated candidate cancer progression drivers and suppressors based on reorganization of 3D genome compartments during cancer evolution and showed that their expression (or lack thereof) in tumors could predict survival in human LUAD patients. Perturbation of most of the nominated candidate progression drivers led to defects in cancer cell proliferation, demonstrating that the single-cell 3D genome organization contains rich information to uncover novel genetic vulnerabilities for drug target discovery.

By focusing on genes within loci that underwent genome compartment reorganization during LUAD progression, we identified an upstream chromatin regulatory factor Rnf2 that can partially reorganize the 3D genome during the adenoma-to-LUAD transition (**Extended Data Fig. 10**). As part of the canonical PRC1 (cPRC1) complex, Rnf2 deposits H2AK119ub1 to regulate PRC2 recruitment and transcriptional repression[49,50]. However, recent reports have also delineated a non-canonical gene activation function for Rnf2, though the precise mechanisms for this function remain controverial[51,52]. Here, we found that Rnf2 has a ubiquitin ligase-independent role in reorganizing the 3D genome in active chromatin regions. In comparison to WT Rnf2, re-expression of a catalytic mutant (I53S) of Rnf2 into LUAD cells subject to knockdown of endogenous Rnf2 more strongly recapitulated the 3D genome changes occurring in the adenoma-to-LUAD transition. One potential explanation is that Rnf2 I53S associates more with active chromatin regions than its WT counterpart, facilitating compartment transitions that drive LUAD progression. In support of this model, a recent study demonstrated that H2AK119ub1 and H3K27me3 are required to maintain Rnf2 at canonical PRC2-repressed sites,

as loss of Rnf2 catalytic activity displaces it from these sites[61]. An additional layer of regulation may occur downstream of KRAS signaling itself. Prior work in *BRAF*-mutant melanoma[52] suggested that phosphorylation of Rnf2 by MEK1/2 – a key mediator of the mitogen-activated protein kinase (MAPK) pathway downstream of oncogenic *KRAS* and *BRAF* – allows Rnf2 to recruit activating histone modifiers to poised promoters to induce gene expression. During murine LUAD progression, MAPK signaling downstream of oncogenic *Kras* increases[62,63], suggesting that a comparable mechanism of Rnf2 activation may be occurring. These data argue that MAPK hyperactivation – permitted by loss of *Trp53*[62,63] – may facilitate the capacity of Rnf2 to reorganize 3D genome compartmentalization to drive the adenoma-to-LUAD transition. Given increased MAPK signaling in human NSCLC samples[64], such mechanisms of Rnf2 regulation may also occur during human lung cancer progression.

Recent scRNA-seq[37] and scATAC-seq[65] studies in mouse models of *Kras*-driven LUAD characterized the heterogeneous transcriptomic and epigenomic landscapes that arise in cancer, leading to the identification of transitional and plastic cell states with molecular alterations that can drive tumor progression and metastasis. It remains unknown how 3D genomic landscapes map to these transcriptionally and epigenetically defined cell states. To address these questions, future studies can combine genome-wide chromatin tracing with image-based spatial transcriptomics[66–69], epigenomics[70], and proteomics[31,71] in the same single cancer cells within the native tumor microenvironment. In addition, finer-scale chromatin tracing with kilobase resolution may reveal gene-specific enhancer-promoter interactions further defining the molecular features and plasticity of cells during LUAD progression.

Finally, we demonstrated proof-of-principle of the utility of *in situ* chromatin tracing to garner insights into tumor biology, taking advantage of the capacity to visualize the entire spectrum of lung tumorigenesis in mouse models that include genetically-encoded lineage markers to faithfully trace subclonal evolution. Nonetheless, human tumors exhibit significantly more complex somatic genotypes compared with those in genetically engineered mouse models of a similar type[38]. Indeed, the human NSCLC genome is highly heterogeneous with different tumors displaying distinct combinations of point mutations, structural variations, and fusion genes[72]. Therefore, it remains to be explored whether the 3D genomic structural bottleneck during mouse LUAD progression is recapitulated in human lung tumors and how 3D genome evolution differs across driver oncogenes (e.g., *KRAS* vs. *EGFR*). To answer these questions, one can apply genome-wide chromatin tracing to clinical human lung tumor samples harboring a preinvasive component (carcinoma *in situ*) adjacent to invasive tumor regions and normal regions. As DNA FISH has been traditionally used to map structural variations in cancer, the FISH-based chromatin tracing technique has the potential to delineate multi-dimensional genomic alterations in single cancer cells, which may offer even more robust biomarkers for cancer diagnosis and therapeutics. Given the rich information one can derive from mapping genome-wide true-3D chromatin folding in subclonal cancer progression, we anticipate that comprehensive tissue-based chromatin tracing – when applied broadly in cancer and combined with multi-modal spatial profiling of the transcriptome, epigenome, proteome, and metabolome – will have transformative potential to reveal new aspects of cancer biology and novel biomarkers for cancer diagnosis, subtype stratification, prediction of treatment response, and therapeutic development.

Methods

Mice

Mouse strains: Animal studies were approved by the Institutional Animal Care and Use Committee of Yale University. Mice were housed in a specific-pathogen free facility with controlled temperature and day/night cycles and maintained in a mixed background. *MADM11-GT* (Stock #013749) and *MADM11-TG* (Stock #013751) were obtained from the Jackson Laboratory. *LSL-Kras*^{G12D} and *Trp53*^{KO} mice were a gift from Dr. Tyler Jacks. *LSL-Kras*^{G12D}*/Kras*^{WT}*; MADM11-TG,Trp53*^{KO}*/MADM11-TG,Trp53*^{WT} breeder mice were generated as previously described[6]. *LSL-Kras*^{G12D}*/Kras*^{WT}*; MADM11-TG,Trp53*^{KO}*/MADM11-TG,Trp53*^{WT} were crossed with *MADM11-GT* mice to generate *K-MADM-Trp53* experimental mice of both sexes for analysis. Genotyping primers and protocols have been previously reported[6].

Lentivirus production and infection*:* pPGK-Cre lentiviral backbone was amplified and transfected into 293T cells using VSV-G (Addgene 8454) and psPAX2 (Addgene 12260) packaging vectors and TransIT-LT1 kit (MirusBio). After 48 and 72 hours, virus was collected, filtered, and ultracentrifuged. Lentivirus was resuspended in OptiMEM (ThermoFisher) and administered intratracheally to *K-MADM-Trp53* mice at 6-10 weeks of age.

Probe design and synthesis

Template probe design: To select target genomic regions for mChr1-mChr19, first, we downloaded TAD coordinates in mouse embryonic stem cells from http://3dgenome.fsm.northwestern.edu/download.html. Second, we selected a list of target TADs containing one or more of the following features: 1) TADs containing MAPK pathway genes; 2) TADs containing “classic” oncogenes and tumor suppressor genes; 3) TADs containing super-

enhancers. Third, to select other target TADs, we selected 30 TADs that are equally spaced along each entire chromosome and removed those that are within an interval to the feature-containing target TADs. We then combined the remaining other target TADs and feature-containing target TADs as the target genomic regions. We designed about 400 oligos targeting a 60-kb feature-containing region within each feature-containing TAD, and about 400 oligos targeting the central 60-kb region of other target TADs. A total of 50 target TADs were designed for mChr6 and 18-25 target TADs were designed for each of the other autosomes. A total of 473 target TADs spanning along mChr1-mChr19 were selected. To distinguish the identities of each target TAD, we adapted a previously published Hamming weight 2 (HW2) binary barcode design strategy[19]. We rearranged the binary code assignment to 1) minimize the variance of the number of TADs imaged in each imaging round across different imaging rounds for each chromosome to make sure no more than 1 TAD was imaged per imaging round for each chromosome, except for mChr6; and 2) for mChr6, maximize the genomic distance between the TADs imaged in the same imaging round. The template oligos for the 473 target TADs were designed as follows: each template oligo consisted of 6 regions (from 5' to 3'): 1) a 20-nucleotide (20-nt) forward priming region; 2) a 20-nt adapter binding region; 3) a 40-nt genome targeting region; 4) a second 20-nt adapter binding region; 5) a third 20-nt adapter binding region; and 6) a 20-nt reverse priming region. The forward and reverse 20-nt priming sequences were generated from random sequences that were screened to lack homology to the mouse genome and optimized for PCR amplifications. The 20-nt adapter binding sequences bind to 60-nt adapter oligos consisting of one 20-nt adapter sequence and two 20-nt readout probe binding sequences. Both the adapter and readout probe binding sequences were generated from random sequences with minimum homology to the mouse genome and maximum performance in signal to noise ratios. The 40-nt genome targeting sequences were designed with OligoArray2.1 using the following parameters:

1) the melting temperatures of the target sequences were between 65 °C and 85 °C; 2) the melting temperatures of potential secondary structures were less than 76 °C; 3) the melting temperatures of potential cross-hybridizations were less than 72 °C; 4) the GC content of the sequences was between 20% and 90%; 5) no consecutive repeats of six or more A's, T's, C's and G's were identified; and 6) adjacent oligos were allowed to have 30-nt overlapping sequences. Genome targeting sequences were then screened against the mouse genome with BLAST+ to ensure single matches to the reference genome[73]. Genomic targeting sequences were further screened against a mouse repetitive database from Repbase (https://www.girinst.org/repbase/)[74]. Sequences with more than 16-nt homology to a list of repetitive sequences in the mouse repetitive database were removed. A monolayer of 400 oligos were then selected for each target TAD to construct the template probe library. The template probe library pool was purchased from Twist Biosciences. Template probe sequences are provided in **Supplementary Table 1**. Genome coordinates (mm9), codebook and target TAD features are provided in **Supplementary Table 2**.

Primary probe synthesis: Primary probes were synthesized from the template probe library using previously published protocols following a procedure of limited-cycle PCR, *in vitro* transcription, reverse transcription, and probe purification[16,67,75,76]. All PCR primers, reverse transcription primers, adapters, and readout probes were purchased from Integrated DNA Technologies (IDT). The sequences are provided in **Supplementary Table 3**.

Mouse *K-MADM-Trp53* lung tissue chromatin tracing experiments

Tissue preparation*: K-MADM-Trp53* mice were dissected at sign of respiratory distress via CO_2 asphyxiation. Lungs were perfused with ice cold 1×PBS, incubated in 4% (vol/vol)

paraformaldehyde overnight, and cryoprotected with 30% sucrose. Tissue was embedded in OCT (Tissue-Tek) and frozen on dry ice before storage at -80°C.

***Coverslip treatment**:* Prior to tissue sectioning, coverslips (Bioptechs, 40-mm-diameter, #1.5) were first silanized as previously described[77,78]. In brief, coverslips were immersed into a 1:1 mixture of 37% (vol/vol) hydrochloric acid (HCl) and methanol at room temperature for 30 min. Coverslips were then washed with deionized water three times, followed by a 70% ethanol wash. Coverslips were dried in a 70 °C oven for 1 h and immersed into chloroform containing 0.2% (vol/vol) allyltrichlorosilane (Sigma, 107778) and 0.1% (vol/vol) triethylamine (Millipore, TX1200) for 30 min at room temperature. Coverslips were then washed with chloroform and ethanol, and dried in a 70 °C oven for 1 h. Silanized coverslips can be stored in a desiccated chamber at room temperature for weeks. For tissue attachment, silanized coverslips were treated with 1% (vol/vol) polyethylenimine (Sigma, 408727) in water for 5 min and washed twice in water. Coverslips were then air-dried and ready for tissue attachment.

***Tissue sectioning**:* Frozen mouse lung tissue blocks were sectioned at a thickness of 10 μm at -20 °C on a cryostat. Three consecutive tissue slices were sectioned at a time: one for hematoxylin and eosin (H&E) staining, one for whole-section fluorescence imaging, and one for chromatin tracing. The sections were air-dried at room temperature for 1 h and then directly used or stored at -20 °C for months.

***H&E staining and whole-section fluorescence imaging**:* H&E staining was performed by Yale Pathology Tissue Services (YPTS). For whole-section fluorescence imaging, the section was mounted in VECTASHIELD Vibrance Antifade Mounting Medium with DAPI (VectorLabs, H-1800-2). Stitched fluorescence images in 353-nm, 488-nm and 592-nm illumination channels for DAPI, GFP, and tdTomato, respectively, were collected with a Plan-Apochromat 10×/0.45 M27 objective on a Zeiss Axio Imager M2 microscope.

Co-immunofluorescence and fluorescent protein imaging of the chromatin tracing section: Frozen tissue sections were first balanced at room temperature for 10 min. Tissue sections were then hydrated with DPBS (Gibco, 14190-144) for 5 min twice. Tissue sections were permeabilized with 0.5% (vol/vol) Triton X-100 (Sigma-Aldrich, T8787) in DPBS for 30 min at room temperature and washed in DPBS for 2 min twice. Tissue sections were then blocked for 30 min at room temperature in blocking buffer containing 1% (wt/vol) BSA (Sigma-Aldrich, A9647-100G), 22.52 mg/mL glycine (AmericanBio, AB00730), 10% (vol/vol) donkey serum (Millipore Sigma, S30-100ml), 5% (vol/vol) goat serum (Invitrogen, 31873) and 0.1% (vol/vol) Tween-20 (Sigma-Aldrich, P7949) in DPBS. Tissue sections were incubated with rabbit anti-SPC antibody (Millipore, AB3786) at 1:50 dilution and rat anti-CD45 antibody (BioLegend, 103101) at 1:100 dilution in blocking buffer at 4 °C overnight. The samples were washed with DPBS for 5 min three times and then incubated with DyLight 800-labeled donkey anti-rabbit secondary antibody (Invitrogen, SA5-10044) at 1:1000 dilution and Alexa Fluor 647-labeled goat anti-rat secondary antibody (Invitrogen, A-21247) at 1:1000 dilution in blocking buffer for 1 h at room temperature. The samples were washed in DPBS for 5 min three times and then incubated with DAPI (Thermo Fisher, 62248) at 1:1000 dilution in 2× SSC (diluted from 20× SSC, Invitrogen, 15557-044) for 10 min. The samples were then mounted onto a Bioptech's FCS2 flow chamber and replenished with imaging buffer with an oxygen scavenging system (50 mM Tris-HCl pH 8.0 (AmericanBio, AB14043), 10% wt/vol glucose (Sigma-Aldrich, G5767), 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/mL glucose oxidase (Sigma-Aldrich, G2133), 40 μg/mL catalase (Sigma-Aldrich, C30)). The imaging buffer was freshly prepared for each experiment and was covered with a layer of mineral oil (Sigma, 330779) in the reservoir tube to prevent continuous oxidation. We then selected multiple fields of view (FOVs) at predefined tumor regions based on tumor grades of the adjacent whole-section fluorescence images. Tumor

grades were confirmed independently by two investigators (S.S.A. and M.D.M.). At each FOV, we sequentially took z-stack images of DAPI, GFP fluorescence, tdTomato fluorescence, anti-SPC immunofluorescence and anti-CD45 immunofluorescence with 405-nm, 488-nm, 560-nm, 647-nm and 750-nm laser illuminations. The z-stacks (7-10 μm total) had a step size of 200-nm and an exposure time of 0.4 s at each step.

Primary probe hybridization*:* After imaging, tissue sections on coverslips were de-assembled from the chamber, treated with 1 μg/mL proteinase K (Invitrogen, AM2546) in 2% (vol/vol) SDS (Sigma-Aldrich, 05030-1L-F) in 2× SSC at 37 °C for 10 min, and washed in DPBS for 2 min twice. Tissue sections were then treated in 0.1 M HCl for 5 min and washed in DPBS for 2 min twice. Tissue sections were digested with 0.1 mg/mL RNase A (ThermoFisher Scientific, EN0531) in DPBS for 45 min at 37 °C and washed with 2× SSC for 2 min twice. Tissue sections were treated with pre-hybridization buffer containing 50% (vol/vol) formamide (Sigma-Aldrich, F7503) and 0.1% (vol/vol) Tween-20 in 2× SSC. Synthesized primary probes were dissolved in 25 μL hybridization buffer containing 50% (vol/vol) formamide and 20% (vol/vol) dextran sulfate (Millipore, S4030) in 2× SSC. The total probe concentration was 30-40 μM. We then added the hybridization buffer containing probes to a 60 mm petri dish and flipped the coverslip onto it so that the sections were immersed into hybridization buffer. Heat denaturation was performed by incubating the petri dish in a 90 °C water bath for 3.5 min. We then incubated the petri dish at 47 °C in a humid chamber for 36-48 hours. We washed the samples with 50% (vol/vol) formamide in 2× SSC for 15 min twice at room temperature, and washed with 2× SSC for an additional 15 min at room temperature. We then incubated the sample with 0.1-μm yellow-green beads (Invitrogen, F8803) resuspended in 2× SSC as fiducial markers for drift correction. We washed the sample with 2× SSC briefly. We finally incubated the sample with

DAPI at 1:1000 dilution in 2× SSC for 10 min for image registration. The sample was washed with 2× SSC for 2 min twice.

***Readout probe hybridization and imaging**:* After the primary probe hybridization, the sample was repeatedly hybridized with adapters and readout oligonucleotide probes, imaged, and photobleached for a total of 50 rounds. Each adapter probe was 60-nt and consisted of a 20-nt primary probe binding region and two consecutive 20-nt readout probe binding regions. The readout probes were 30-nt oligos conjugated with Alexa Fluor 647 (or Cy5) or ATTO 565 (or Cy3) fluorophores with 20-nt complementary to the readout probe binding regions of adapters. The sequences of adapter and readout probes are provided in **Supplementary Table 3**. To perform automatic buffer exchange during multiple rounds of readout probe hybridization and imaging, we used a Bioptech's FCS2 flow chamber and a computer-controlled, custom-built fluidics system[16,67]. Prior to sequential readout probe hybridization and imaging, we used DAPI images for registration to images taken before primary probe hybridization. We assembled the sample onto the Bioptech's FCS2 flow chamber and further flowed 2 mL imaging buffer through the chamber. We selected the same FOVs as those selected during co-immunofluorescence and fluorescent protein imaging. At each FOV, we took z-stack images with 488-nm and 405-nm laser illuminations to image fiducial beads and DAPI respectively. The z-stacks (7-10 μm total) had a step size of 200-nm and an exposure time of 0.4 s at each step. For each round of readout probe hybridization, we flowed through the chamber 2 mL readout probe hybridization buffer (20% vol/vol ethylene carbonate (Sigma-Aldrich, E26258) in 2× SSC) containing two adapter probes each at 50 nM concentration and incubated for 15 min at room temperature. We flowed through the chamber 2 mL wash buffer (20% vol/vol ethylene carbonate in 2× SSC) for 2 min and further flowed through 2 mL readout probe hybridization buffer containing the two dye-labeled readout probes each at 75 nM concentration. We incubated the sample for 15 min at

room temperature and further flowed through 2 mL wash buffer for 2 min. We then flowed 2 mL imaging buffer through the chamber. At each FOV, we took z-stack images with 647-nm, 560-nm and 488-nm laser illuminations. Dye-labeled readout probes were imaged in the 647-nm and 560-nm channels and fiducial beads were imaged in the 488-nm channel. The z-stack (7-10 μm total) had a step size of 200-nm and an exposure time of 0.4 s at each step. After imaging, we switched buffer to readout probe hybridization buffer containing 1 μM dye-free readout probes (blocking oligos) and photobleached the sample by continuous simultaneous laser illuminations with 750-nm, 647-nm, and 560-nm lasers for 40 s. We then flowed 5 mL 2× SSC through the chamber to wash away the unbound dye-free readout probes before the next hybridization round. A total of 50 hybridization rounds was performed. The color shift between the 647-nm and 560-nm channels was canceled by taking z-stack calibration images of 100-nm Tetraspeck beads (Invitrogen, T7279) attached to a coverslip surface.

Mouse cell line chromatin tracing experiments

***Cell culture**:* The KP mouse primary adenocarcinoma cell line, 31671, was a gift from Dr. Nik Joshi and was derived from an autochthonous *LSL-Kras*^{G12D}*/Kras*^{WT}*,p53*$^{fl/fl}$ mouse administered with Adeno-Cre[79]. The *K-MADM-Trp53* mouse primary adenocarcinoma cell line, SA6082inf, was derived from collagenase-based dissociation of a large green tumor dissected from a *K-MADM-Trp53* mouse. *Kras* and *Trp53* genotype was confirmed by PCR. Both cell lines were cultured in DMEM (Corning, 10-013-CV) containing 10% (vol/vol) FBS (Gibco, 26140-079) and 1% (vol/vol) Penicillin-Streptomycin (ThermoFisher, 15140-122) at 37 °C with 5% CO_2. Cells were passaged whenever they reached confluency. Both cell lines were seeded onto UV-sterilized coverslips (Bioptechs, 40-mm-diameter, #1.5) and grown until 60%-70% confluency

prior to primary probe hybridization. All cell lines were determined to be mycoplasma free via PCR (ATCC, 30-1012K).

Stable Rnf2 knockdown and rescue cell line construction*:* Lentiviral supernatant for Sigma Mission^{TM} shRNAs targeting *Rnf2* and a non-targeting control (shNTC) were obtained from the Yale Cancer Center Functional Genomics Core. KP cells were seeded in 6-well plates (Falcon, 353046) with a density of 50,000 cells/well. Cells were infected with lentivirus at different titers 24 hours after seeding. Stable transfected cells were selected with 6 μg/mL puromycin (Gibco, A11138-03) for 48-72 hours after lentiviral transfection. Cell pellets were collected after 48 hours of selection for quantitative RT-PCR validation and after 72 hours of selection for western blot validation. For Rnf2 rescue experiments, Rnf2 WT and Rnf2 I53S cDNAs (both harboring a mutated Rnf2 shRNA target seed sequence) were synthesized from Genscript and inserted into the multiple cloning site of pLV-EF1a-IRES-Hygro (Addgene 85134). To construct rescue cell lines (shNTC+empty vector, shRnf2+empty vector, shRnf2+I53S Rnf2), stable Rnf2 knockdown KP cells were seeded in 6-well plates (Falcon, 353046) with a density of 50,000 cells/well. Cells were infected with lentivirus at different titers 24 hours after seeding. Stable transfected cells were selected with 700 μg/mL hygromycin (Gibco, 10687010) for 48-72 hours after lentiviral transfection. Rnf2 expression was confirmed by western blot.

Primary probe hybridization*:* Cells were washed with DPBS twice for 2 min each, fixed with 4% (vol/vol) paraformaldehyde (EMS, 15710) in DPBS for 10 min, and washed twice with DPBS for 2 min each. Cells were permeabilized with 0.5% (vol/vol) Triton-X (Sigma-Aldrich, T8787) in DPBS for 10 min and washed twice with DPBS for 2 min each. Next, the cells were treated with 0.1 M HCl for 5 min at room temperature, washed with DPBS twice for 2 min each, treated with 0.1 mg/mL RNase A in DPBS for 45 min at 37 °C, and washed twice with 2× SSC for 2 min each. The cells were subsequently incubated in pre-hybridization buffer containing

50% (vol/vol) formamide and 0.1% (vol/vol) Tween-20 in 2× SSC. Synthesized primary probes were dissolved in 25 μL hybridization buffer containing 50% (vol/vol) formamide and 20% (vol/vol) dextran sulfate in 2× SSC. The final probe concentration was 30-40 μM. We then added the hybridization buffer containing probes to a 60 mm petri dish and flipped the coverslip onto it so that the cells were immersed into hybridization buffer. Heat denaturation was performed by incubating the petri dish in a 90 °C water bath for 4 min. The petri dish was incubated at 47 °C in a humid chamber for two overnights (36-48 h). After hybridization, the cells were washed with 50% (vol/vol) formamide in 2× SSC for 15 min twice at room temperature and washed with 2× SSC for an additional 15 min. We then incubated each sample with 0.22-μm light yellow beads (Spherotech, FP-0245-2) resuspended in 2× SSC as fiducial markers for drift correction and washed the sample with 2× SSC briefly.

Readout probe hybridization and imaging*:* We followed the same "Readout probe hybridization and imaging" procedure, as performed in "Mouse *K-MADM-Trp53* lung tissue chromatin tracing experiments" described above, with the following exceptions: 1) DAPI images were acquired after sequential readout probe hybridization and imaging; 2) Both readout probe hybridization buffer and wash buffer were composed of 35% (vol/vol) formamide in 2× SSC instead of 20% (vol/vol) ethylene carbonate; 3) Fiducial beads were imaged in the 405-nm channel; and 4) We used tris(2-carbox-yethyl)phosphine (TCEP; Sigma-Aldrich, C4706) cleavage instead of photobleaching for fluorescence signal removal between the readout hybridization rounds. TCEP can reduce the disulfide bonds connecting fluorophores to readout probes and remove fluorophores from readout probes. The TCEP cleavage buffer was composed of 50 mM TCEP and 1 μM dye-free readout probes in 20% (vol/vol) ethylene carbonate to block unoccupied readout probe binding regions on adapters from interfering with the next round of hybridization and imaging.

Microscope setup

For imaging, we used a custom-built microscope with a Nikon Ti2-U body, a Nikon CFI Plan Apo Lambda 60× Oil (NA1.40) objective lens, and an active auto-focusing system[80]. Different laser settings were applied for imaging tissues or cell lines. For *K-MADM-Trp53* lung chromatin tracing experiments, illumination lasers included: a 750-nm laser (2RU-VFL-P-500-750-B1R, MPB Communications) , a 647-nm laser (2RU-VFL-P-1000-647-B1R, MPB Communications), a 560-nm laser (2RU-VFL-P-1000-560-B1R, MPB Communications), a 488-nm laser (2RU-VFL-P-500-488-B1R, MPB Communications), and a 405-nm laser (OBIS 405 nm LX 50 mW, Coherent). The five laser lines were directed to the sample using a multi-band dichroic mirror (ZT405/488/561/647/752rpc-UF2, Chroma) on the excitation path. Laser intensities were controlled with an acousto-optic tunable filter (AOTF, 97-03309-01 Gooch & Housego), and laser on-off was controlled by mechanical shutters (LS3S2Z0, Vincent Associates). For cell line chromatin tracing experiments, we used a Lumencor CELESTA light engine for illumination, with the following laser wavelengths: 405-nm, 477-nm, 546-nm, 638-nm, and 749-nm. The lasers were directed to the sample using a corresponding penta-band dichroic mirror from Lumencor. Laser intensities and on-off were controlled by internal controls of the light engine. The 750/749-nm laser was used to excite and image DyLight 800-labeled donkey anti-rabbit secondary antibody. The 647/638-nm laser was used to excite and image Alexa Fluor 647 (or Cy5) on readout probes and on the anti-rat secondary antibody. The 560/546-nm laser was used to excite and image tdTomato fluorescence and ATTO 565 (or Cy3) on readout probes. The 488/477-nm laser was used to excite and image GFP fluorescence and the yellow-green fiducial beads for drift correction. The 405-nm laser was used to excite and image the DAPI stain and the

light-yellow fiducial beads. On the emission path, we had a multi-band emission filter (ZET405/488/561/647-656/752 m, Chroma for the tissue imaging setup or a corresponding penta-band emission filter for the cell line imaging setup) and a Hamamatsu Orca Flash 4.0 V3 camera. The pixel size of our system was 107.9 nm. To automatically scan and image multiple FOVs, we used a computer-controlled motorized x–y sample stage (SCAN IM 112×74, Marzhauser). For z-stepping and active auto-focusing, a piezo z positioner (Mad City Labs, Nano-F100S) was used.

Arrayed RNAi screen of cell proliferation

Lentiviral infection conditions were optimized in 96-well plates for initial cell seeding number, lentiviral dosage, antibiotic concentration, and assay time. KP and *K-MADM-Trp53* LUAD cells were seeded in complete cell culture media at a density of 250 cells per well in a 96-well assay plate (Corning 3903), incubated for 24 hours, and infected with lentiviral supernatant of Sigma Mission^{TM} shRNAs targeting CPDs obtained from the Yale Cancer Center Functional Genomics Core. A total of 200 uL media was added into each well, with 50 uL lentiviral supernatant and 150 uL complete cell culture media with 10 ug/mL polybrene (EMD Millipore, TR-1003-G). Each shRNA hairpin was tested in triplicate. All lentiviral infections were performed in duplicate to rule out the influence of lentiviral dosage on cell growth: one replicate with 6 ug/mL puromycin selection and the other replicate with no puromycin selection. After 24 hours of lentiviral incubation, corresponding wells were treated with or without puromycin selection for 48 hours. The cells were then incubated with complete cell culture media for 4-5 days. Total viable cell count was determined with CellTiter-Glo Luminescent Cell Viability Assay (Promega, G7572) using a Promega luminescence plate reader. For data analysis, we deducted luminescence readout values of blank wells from those of test wells. We then normalized

luminescence readout values of each target shRNA hairpin to those of shNTC. All shRNA hairpin sequences are provided in **Supplementary Table 6**.

Cellcyte cell proliferation assay

Cells were seeded into 12-well plates (Corning, 353043) with 20,000 cells per well and incubated in complete medium (DMEM containing 10% (vol/vol) FBS and 1% (vol/vol) Penicillin-Streptomycin) at 37 °C with 5% CO_2 for 3 hours for attachment. The plates were then transferred into the Cytena CellcyteX imaging system for phase contrast live cell imaging. Each condition had 3 replicate wells. Images were taken every 3 hours for 4-5 days until cells reached 100% confluency. To calculate the growth rate, we first plotted a delta confluency versus confluency curve at each time point and fitted a linear regression line in regions with 10% to 40% confluency. The slope of the fitted linear regression line is defined as the growth rate.

Western Blot

Cells were trypsinized, harvested, and lysed with RIPA buffer (ThermoFisher, 89900) with 1× protease inhibitors (ThermoFisher, 87786) at 4 °C for 30 min. Cell lysate supernatant was collected after centrifugation at 16,000 × g for 20 min at 4 °C. Supernatants were quantified using the BCA protein assay kit (Pierce, 23225). A total of 20 μg protein was denatured at 95 °C for 5 min and loaded on a 4-20% precast polyacrylamide gel (BioRad, 4568094) for gel electrophoresis. Proteins were transferred onto a PVDF membrane (ThermoFisher, IB24001) with an iBlot2 gel transfer device (Invitrogen, IB21001). The PVDF membrane was blocked with 5% (vol/vol) BSA in 1× TBST (AmericanBio, AB14330-01000), incubated with primary antibodies at 4 °C overnight, washed with 1× TBST for 5 min for three times, incubated with HRP-conjugated secondary antibodies, and washed with 1× TBST for 5 min for three times.

Membranes were treated with SuperSignal West Pico Plus chemiluminescent substrate (ThermoScientific, 34577) and imaged with a ChemiDoc imaging system (BioRad). Antibodies used: anti-Rnf2 rabbit antibody (Cell Signaling Technologies, 5694S) at 1:500 dilution, mouse anti-Hsp90 antibody at 1:10,000 dilution (BD Biosciences, 610418), HRP-conjugated goat anti-rabbit IgG secondary antibody at 1:3,000 dilution (Abcam, ab6721), HRP-conjugated goat anti-mouse IgG secondary antibody at 1:10,000 dilution (Biorad, STAR207P).

Bulk RNA-sequencing of lung tumors

Large green tumors from *K-MADM-Trp53* mice were microdissected under a Nikon SMZ1270 fluorescence dissection stereo microscope and flash frozen. Tissue was pulverized using a BioPulverizer that was sprayed down with RNAse Away (Molecular BioProducts) and cooled with liquid nitrogen. RNA and genomic DNA were extracted using the AllPrep DNA/RNA Mini Kit (Qiagen). Library preparation for RNA sequencing was performed by the Yale Center for Genome Analysis (YCGA), and libraries were sequenced on a NovaSeq S2 (Illumina) to obtain 100-bp paired-end reads. All reads that passed FASTQC quality metrics were mapped to the UCSC mm10 mouse genome and normalized gene count matrices were generated through STAR. Further analysis after trimming, alignment, and normalization were performed using DESeq2 on R[81]. Hierarchical clustering was done through the pheatmap package. Genes upregulated ($\log_2$ fold change >2, FDR<0.05) or downregulated ($\log_2$ fold change <-1, FDR<0.05) in LUAD compared to AdenomaG were compared to the MSigDB Hallmarks gene set collection (https://www.gsea-msigdb.org/gsea/msigdb) to determine enrichment (hypergeometric test). Normalized expression counts and differential expression analyses are included in **Supplementary Table 4**.

Copy number variation analysis

Tumor genomic DNA was obtained from flash frozen tumors using the AllPrep DNA/RNA Mini Kit, as described above. Paired normal DNA was obtained by extraction from FFPE slides of the same mouse lung by the YCGA. Whole exome sequencing (WES) was performed using the Mouse All Exon kit (Agilent) for target capture followed by next-generation sequencing by Psomagen. Mouse tumor samples were sequenced at 200× read coverage while healthy lung tissue was sequenced at 50×. Sample reads were mapped to the GRCm38 reference genome with BWA-mem[82], sorted based on coordinates with Picard SortSam tools (http://broadinstitute.github.io/picard/), and indexed with Samtools[83,84].

Image analysis

All image analysis was performed with MATLAB R2019b.

DAPI nucleus segmentation and registration*:* GFP and tdTomato fluorescence images and SPC and CD45 co-immunofluorescence images were aligned to genome-wide chromatin tracing images using intensity-based image registration of the DAPI images. To process the DAPI images, we first took the average z-projection of each DAPI image stack and normalized it to its maximum intensity. We then reduced the background by normalizing the average projection image to the background calculated by the adaptthresh function. We then adjusted the threshold of the image so that the maximum and minimum intensities corresponded to the 3rd and 1st quartiles of the pixel intensities. We then applied an "opening-by-reconstruction" technique with a disk-shaped morphological structural element with a radius of 25 pixels to reduce the noise. Next, to align the processed DAPI images, we applied an intensity-based image registration algorithm. We used the imregtform function to estimate the geometric transformation for image alignment, and used the imregconfig function to generate the optimizer and metric configurations

used by imregtform. We first optimized an initial transformation condition, and then used the optimized initial conditions to improve image alignment. For initial condition optimizations, we reduced the InitialRadius of the optimizer (generated by imregconfig) by a scale factor of 5, and set the MaximumIterations of the optimizer to 500. We then applied the optimizer and metric to the imregtform function with the "similarity" geometric transformation option to generate the initial geometric transformation object. We used the imregtform function to align the processed DAPI images with the "affine" geometric transformation option, the previously generated optimizer and metric, and the initial geometric transformation object. We finally generated a geometric transformation object to align DAPI images taken with fluorescent protein and co-immunofluorescence images to the DAPI images taken with the first hybridization round of genome-wide chromatin tracing.

GFP+ and tdTomato+ cell analysis*:* To identify cells with GFP or tdTomato fluorescence signals, we first generated maximum projections of GFP or tdTomato images along the z direction and aligned the images to the first-round readout hybridization images of genome-wide chromatin tracing. We then used an algorithm that can manually adjust the intensity threshold to determine GFP+ or tdTomato+ cells (Thresholding an image - File Exchange - MATLAB Central (mathworks.com)). We finally generated binary masks to distinguish GFP+ and tdTomato+ cells.

SPC+ and CD45+ cell extraction*:* To identify cells with SPC or CD45 immunofluorescence signals, we generated average projections of SPC or CD45 images along the z direction and normalized the images to the background calculated by the adaptthresh function. We then performed standard deviation filtering of the image with the stdfilt function, and filled holes with the imfill function, using a connectivity of 8 pixels. We then converted the SPC or CD45 images to binary images, used the regionprops function to identify SPC or CD45 patches, and excluded

patches smaller than 150 pixels. The remaining patches were used to generated binary masks to distinguish SPC+ or CD45+ cells. Finally, to match the signals to nuclei, we generated binary masks for each cell nucleus using the DAPI images as described in the "Nucleus segmentation" section below, and dilated the binary mask of each nucleus with a disk-shaped structural element of 10 pixels. We then multiplied each single-nucleus binary mask to the SPC or CD45 binary masks. Nuclei with more than 100 overlapping pixels were labeled as nuclei of SPC+ or CD45+ cells.

The genome-wide chromatin tracing image analysis pipeline consists of the following steps: color correction, drift correction, nucleus segmentation, foci fitting, decoding, and trace linking.

Color correction*:* The color shift between 647-nm and 560-nm laser channels was corrected by taking z-stack calibration images of Tetraspeck microspheres (0.1 μm, Invitrogen, T7279) attached to a coverslip surface. A polynomial spatial transformation structure in x and y was constructed with the cp2tform function and used for xy color shift correction. The color shift in z was corrected by calculating the mean z shift.

Drift correction*:* To correct for sample drifts between different rounds of hybridizations, we determined 3D positions (x, y, z) of fiducial beads with 3D Gaussian fitting for each hybridization round. We subtracted 3D positions (x, y, z) of the first hybridization round from each hybridization round to generate the drift correction profiles for all hybridization rounds.

Nucleus segmentation*:* Because the tissue sections largely consisted of a monolayer of cells, we segmented single cell nuclei in 2D based on DAPI staining patterns. We first applied drift corrections to the DAPI images and took their average projections along the z direction. We then normalized the DAPI average projection images to the background calculated by the adaptthresh function with a neighborhood size of 101 pixels. We further removed small "bright" objects using "opening-by-reconstruction" and small "dark" objects using "closing-by-reconstruction"

techniques, both with a disk-shaped structuring element of 15 pixels. These processed DAPI images were further analyzed to extract foreground and background markers for the watershed algorithm. To obtain foreground markers for each single nucleus, we calculated the regional maxima with the imregionalmax function. To acquire background markers, we binarized the processed DAPI images and calculated their complement. We then used the imimposemin function to modify the processed DAPI images so that the regional minima occurred at foreground and background marker pixels. Finally, we applied the watershed function to the modified DAPI images for nucleus segmentation.

Foci fitting*:* To determine the intensity threshold for DNA foci identification, we adapted a previously developed adaptive thresholding procedure so that the fitted foci number matched the expected DNA loci count[23]. The expected DNA loci count per nucleus per bit was approximately 20 based on the probe design. We then fitted 3D positions (x, y, z) of all DNA foci using a 3D radial center algorithm[85] in each bit. We further applied color correction and drift correction to the fitted DNA foci in each bit, so that all fitted DNA foci were in the same 3D coordinate system as the first bit in the 560-nm laser channel. The signal intensity of each fitted DNA spot was normalized to the median signal intensities of all DNA foci in the corresponding image.

Decoding and trace linking*:* After we generated all fitted DNA foci in each bit in single nuclei, we adapted a previously reported expectation-maximization procedure for decoding[19]. First, we identified all valid spot pairs corresponding to a valid barcode whose two fitted DNA spots were within 500-nm spatial distance. For each spot pair, we calculated three quality metrics: 1) the distance between the 3D positions (x, y, z) of the spot pairs; 2) the difference between the signal intensities of the spot pairs; 3) the average signal intensities of the spot pairs. We then calculated the percentages of spot pairs with worse qualities than a given spot pair (e.g. larger distance, larger difference, smaller intensities), and calculated the product of the three percentages as the

quality score of the given spot pair. Next, for all spot pairs containing the same repetitive spot, we retained one spot pair with the highest quality score. Then, for each target genomic locus, we retained the top four spot pairs with the highest quality scores in each nucleus. After we obtained all processed spot pairs in all nuclei in each FOV, we linked the DNA loci positions into traces using a previously developed symmetric nearest neighbor approach[23]. We identified the 3D centroid positions of each linked chromosome territory and calculated a fourth quality metric: the distance between each spot pair to the nearest corresponding chromosome territory centroid. We then iteratively updated the quality scores, removed repetitive spot pairs, retained the top four spot pairs with the highest quality scores and performed trace linking. For each single nucleus, if more than 99% (iteration rounds no more than 10) or 97% (iteration rounds more than 10) of spot pairs in the current iteration were the same as the previous iteration, the nucleus would be labeled as "decoded". If more than 85% of nuclei in the FOV were labeled as "decoded", the iteration would be terminated and spot pairs in the current iteration were stored as the finalized spot pairs.

Data analysis and statistics

Heterogeneity*:* To compare chromatin conformational heterogeneity, we defined a heterogeneity score as the coefficient of variation (COV) of inter-loci distances between a pair of TADs on a chromosome of a cell state (variation among different copies of the chromosome). Wilcoxon signed-rank test was used to compare between cell states all heterogeneity scores of a chromosome.

Decompaction*:* To compare levels of chromatin compaction, we defined a decompaction score as the mean inter-loci distance between a pair of TADs on a chromosome of a cell state.

Wilcoxon signed-rank test was used to compare between cell states all decompaction scores of a chromosome.

***Demixing**:* To compare levels of chromatin intermixing or demixing, we defined a demixing score as the standard deviation of all normalized mean inter-loci distances on a chromosome of a cell state. The normalized mean inter-loci distance between a pair of TADs was calculated by normalizing the mean inter-loci distance between the pair of TADs to the average value of all mean inter-loci distances on the chromosome. Levene's test of normalized mean inter-loci distances of a chromosome was used for statistical comparisons between cell states for potential difference in chromatin intermixing/demixing.

***A and B compartment polarization analysis**:* To identify A and B compartments, we adapted a previously developed algorithm to determine A and B compartment scores[16,23]. In brief, we first generated a mean inter-loci spatial distance matrix for each chromosome, where each pixel corresponded to the mean spatial distance between a pair of TADs. We then normalized the mean spatial distances to the corresponding expected spatial distances calculated by power-law fitting of spatial versus genomic distances. Next, we calculated the Pearson correlation coefficient between each pair of columns of the normalized matrix. Finally, we applied a principal component analysis of the Pearson correlation matrix and used the coefficients of the first principal component as the compartment scores. We calculated the correlations between compartment scores and an averaged profile of H3K4me1, H3K4me3, DNase I hypersensitivity site, and gene densities[86,87], and flipped the signs of the compartment scores if necessary, so that the correlations were positive. Under this assignment, A compartment TADs had positive scores and B compartment TADs had negative scores. We then used a previously described polarization index metric to quantify the polarized arrangement of A and B compartments [23]. We downloaded the called H3K4me1 and H3K4me3 ChIP-seq peaks and DNA DNase I hypersensitivity sites

from the ENCODE portal with the following identifiers: ENCFF865RFC, ENCFF508WEP, ENCFF268DLZ. The gene density profile was downloaded from the UCSC table browser.

Rabl configuration score analysis*:* To measure the Rabl configuration score, in each single cell, we constructed a vector pointing from the first TAD to the last TAD for each chromatin trace, and normalized the vector length to the corresponding genomic distance. We then summed the normalized vectors of all chromatin traces in a cell and defined the length of the summed vector as the Rabl score. We used the Wilcoxon rank sum test to calculate p values when comparing Rabl scores across different cell states.

Radial score analysis*:* To calculate the radial score of each TAD in each single cell, we first measured the mean spatial distance between each TAD to the centroid of all target TADs, and normalized the distance to the average spatial distances from all TADs to the centroid. Wilcoxon signed-rank test was used to compare between cell states all radial scores of TADs on a chromosome.

Single-cell chromatin conformation cluster visualization*:* To visualize in low dimension potential clustering of single-cell 3D genome conformations, we constructed an input matrix where each row corresponded to a single cell and each column represented the single-cell A/B compartment (scA/B) score of each TAD[39]. The scA/B score of each TAD was calculated as the mean A/B score of all its spatially adjacent TADs within a 1200-nm radius neighborhood. Missing values in the matrix were replaced with 0's. The matrix was used as input in a principal component analysis, and the first 50 principal components were used for downstream visualizations. We scaled the 50 principle components with the PCA().fit-transform function in python, and visualized the single cell data with t-distributed stochastic neighbor embedding (t-SNE)[40].

***Cell state classification with supervised machine learning**:* To classify different cell states with supervised machine learning, we supplied the Classification Learner App of MATLAB with the same input scA/B score matrix as described in "Single-cell chromatin conformation cluster visualization". A Five-fold cross-validation method was used to evaluate the prediction accuracy of the machine learning model. All machine learning models in the App were trained and the model with the highest prediction accuracy (support vector machine) was retained to plot the confusion matrix and ROC curves.

***Marker TAD identification**:* To identify progression marker TADs, we performed Wilcoxon rank-sum tests of scA/B scores of each TAD comparing adenoma green (AdenomaG) cells and LUAD cells. TADs with p values smaller than 0.1 were defined as progression marker TADs. CPD genes were defined as genes located in progression marker TADs with increased scA/B scores and with significantly higher expression (mean expression count > 10, fold change > 3, FDR < 0.05) in LUAD compared to AdenomaG cells. CTS genes were defined as genes located in progression marker TADs with decreased scA/B scores and with significantly lower expression (mean expression count > 10, fold change < 0.5, FDR < 0.05) in LUAD compared to AdenomaG cells.

***Identification of AT2 cells spatially close to and far from immune cells**:* To distinguish AT2/cancer cells close to and far from immune cells, we extracted binary masks of nuclei of SPC+ cells (or GFP+CD45- cancer cells in LUAD tumors) and of CD45+ cells, and dilated SPC+/cancer cell binary mask by 50 pixels. The SPC+/cancer cells with overlapping pixel with the CD45+ mask were identified as AT2/cancer cells spatially close to immune cells, whereas SPC+/cancer cells with no overlapping pixel with the CD45+ mask were identified as AT2/cancer cells distant from immune cells.

***Probability for randomly selected genes to inhibit cell growth**:* To quantify the percentage of randomly selected genes that can affect cell growth, we downloaded the CRISPR_gene_dependency.csv table from DepMap Public 22Q2 release (https://depmap.org/portal/download/all/), which contained the probabilities that knocking out one gene has a cell growth inhibition or death effect. We identified all mouse LUAD cell lines from the table, and calculated the mean probability of all genes in mouse LUAD cells. 16% randomly selected genes are expected to affect cell growth.

***TCGA patient survival analysis**:* Clinical data for TCGA LUAD patient survival and RNA-seq data were obtained from the GDAC website from the Broad Institute (https://gdac.broadinstitute.org/). The identified CPD/CTS genes of LUAD were converted to human orthologs. The RNA-seq expression matrix and the gene list were applied as input to score the individual expression files using GSVA R package with ssGSEA scoring method[88]. TCGA LUAD patients were grouped based on high/low ssGSEA scores (most correlated/least correlated) with different thresholds (*e.g* top 40% vs bottom 40%). Kaplan-Meier survival analysis was carried out using survfit R function with log-rank significance test.

***Chromatin regulator prediction with the BART algorithm**:* To identify chromatin regulators that bind to genes in progression marker loci, we first identified genes with increased expression (FDR < 0.1, fold change > 1) in TADs with increased scA/B scores ($p < 0.1$) from adenoma green to LUAD cells. We then inputted the gene list into the BART algorithm[47,48] to generate a list of chromatin regulators predicted to bind to the input genes, using the default parameter settings.

CUT&RUN experiments and analysis

***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. Nuclei were prepared according to Epicypher Cut and Run Manual Appendix. Nuclei were incubated with activated beads. For *E. coli* spike in DNA, 0.1 ng was added to each sample. Antibodies were used as listed below, with 0.5 ug of antibody per reaction:

IgG Control antibody – CUTANA Kit Rabbit IgG CUT&RUN Negative Control Antibody

RNF2 antibody – CST RING1B (D22F2) XP® Rabbit mAb #5694

H3K4me3 antibody – Epicypher Rabbit Polyclonal H3K4me3 13-0041

H3K27me3 antibody – CST Tri-Methyl-Histone H3 (Lys27) (C36B11) Rabbit mAb #9733

H2AK119ub antibody – CST Ubiquityl-Histone H2A (Lys119) (D27C4) Rabbit mAb #8240

BMI-1 antibody – Active Motif BMI-1 antibody (mAb) 39993

RNA Pol II p-ser5 – Abcam Anti-RNA polymerase II phosphor-S5 EPR19015

***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. Reads were then aligned to mm9 and K-12 *E. coli* genome U00096.3 by Bowtie2 version 2.3.4.

Picard version 2.27.4 was used to down-sample the reads so that the read depths of *E. coli* sequence from different samples match each other. SAMTools version 1.11 was then used to convert to BAM format, index, isolate uniquely mapped paired reads, and remove duplicates. MACS2 version 2.2.7.1 was used to call sample narrow peaks, using IgG as an input. Read counts across genomic intervals and peak visualization were performed using deepTools version 3.3[89], and .bw files were visualized using Integrative Genomics Viewer.

Full Reference List

1. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).
2. Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in *IDH* mutant gliomas. *Nature* **529**, 110–114 (2016).
3. Flavahan, W. A. *et al.* Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature* **575**, 229–233 (2019).
4. Johnstone, S. E. *et al.* Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer. *Cell* **182**, 1474-1489.e23 (2020).
5. Xu, J. *et al.* Subtype-specific 3D genome alteration in acute myeloid leukaemia. *Nature* **611**, 387–398 (2022).
6. Muzumdar, M. D. *et al.* Clonal dynamics following *p53* loss of heterozygosity in *Kras*-driven cancers. *Nat. Commun.* **7**, 12685 (2016).
7. Zink, D., Fischer, A. H. & Nickerson, J. A. Nuclear structure in cancer cells. *Nat. Rev. Cancer* **4**, 677–687 (2004).
8. Krijger, P. H. L. & de Laat, W. Regulation of disease-associated gene expression in the 3D genome. *Nat. Rev. Mol. Cell Biol.* **17**, 771–782 (2016).
9. Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **20**, 535–550 (2019).
10. Yu, M. & Ren, B. The Three-Dimensional Organization of Mammalian Genomes. *Annu. Rev. Cell Dev. Biol.* **33**, 265–289 (2017).

11. Bickmore, W. A. & van Steensel, B. Genome Architecture: Domain Organization of Interphase Chromosomes. *Cell* **152**, 1270–1284 (2013).

12. van Steensel, B. & Furlong, E. E. M. The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* (2019) doi:10.1038/s41580-019-0114-6.

13. Bonora, G., Plath, K. & Denholtz, M. A mechanistic link between gene regulation and genome architecture in mammalian development. *Curr. Opin. Genet. Dev.* **27**, 92–101 (2014).

14. Mir, M., Bickmore, W., Furlong, E. E. M. & Narlikar, G. Chromatin topology, condensates and gene regulation: shifting paradigms or just a phase? *Development* **146**, dev182766 (2019).

15. Misteli, T. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* **183**, 28–45 (2020).

16. Wang, S. *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).

17. Hu, M. & Wang, S. Chromatin Tracing: Imaging 3D Genome and Nucleome. *Trends Cell Biol.* **31**, 5–8 (2021).

18. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).

19. Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182**, 1641-1659.e26 (2020).

20. Ferone, G., Lee, M. C., Sage, J. & Berns, A. Cells of origin of lung cancers: lessons from mouse studies. *Genes Dev.* **34**, 1017–1032 (2020).

21. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301 (2001).

22. Takei, Y. *et al.* Single-cell nuclear architecture across cell types in the mouse brain. *Science* **374**, 586–594 (2021).

23. Liu, M. *et al.* Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nat. Commun.* **11**, 2907 (2020).

24. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).

25. Slebos, R. J. C. *et al.* p53 alterations in atypical alveolar hyperplasia of the human lung. *Hum. Pathol.* **29**, 801–808 (1998).

26. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science* **339**, 1546–1558 (2013).

27. Zong, H., Espinosa, J. S., Su, H. H., Muzumdar, M. D. & Luo, L. Mosaic Analysis with Double Markers in Mice. *Cell* **121**, 479–492 (2005).

28. Muzumdar, M. D., Luo, L. & Zong, H. Modeling sporadic loss of heterozygosity in mice by using mosaic analysis with double markers (MADM). *Proc. Natl. Acad. Sci.* **104**, 4495–4500 (2007).

29. Ou, H. D. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, eaag0025 (2017).

30. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).

31. Takei, Y. *et al.* Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).

32. Cheng, Y., Liu, M., Hu, M. & Wang, S. TAD-like single-cell domain structures exist on both active and inactive X chromosomes and persist under epigenetic perturbations. *Genome Biol.* **22**, 309 (2021).

33. Payne, A. C. *et al.* In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* **371**, eaay3446 (2021).

34. Tammela, T. & Sage, J. Investigating Tumor Heterogeneity in Mouse Models. *Annu. Rev. Cancer Biol.* **4**, 99–119 (2020).

35. Tan, L., Xing, D., Chang, C.-H., Li, H. & Xie, X. S. Three-dimensional genome structures of single diploid human cells. *Science* **361**, 924–928 (2018).

36. Gibcus, J. H. *et al.* A pathway for mitotic chromosome formation. *Science* **359**, eaao6135 (2018).

37. Marjanovic, N. D. *et al.* Emergence of a High-Plasticity Cell State during Lung Cancer Evolution. *Cancer Cell* **38**, 229-246.e13 (2020).

38. McFadden, D. G. *et al.* Mutational landscape of *EGFR*- , *MYC*- , and *Kras*- driven genetically engineered mouse models of lung adenocarcinoma. *Proc. Natl. Acad. Sci.* **113**, E6409–E6417 (2016).

39. Tan, L. *et al.* Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell* **184**, 741-758.e17 (2021).

40. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

41. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).

42. Wang, Y., Huang, H., Rudin, C. & Shaposhnik, Y. Understanding how dimension reduction tools work: an empirical approach to deciphering t-SNE, UMAP, TriMAP, and PaCMAP for data visualization. *J. Mach. Learn. Res.* **22**, 9129–9201 (2021).

43. Dempster, J. M. *et al. Chronos: a CRISPR cell population dynamics model*. http://biorxiv.org/lookup/doi/10.1101/2021.02.25.432728 (2021) doi:10.1101/2021.02.25.432728.

44. Pacini, C. *et al.* Integrated cross-study datasets of genetic dependencies in cancer. *Nat. Commun.* **12**, 1661 (2021).

45. Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).

46. Dempster, J. M. *et al. Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines*. http://biorxiv.org/lookup/doi/10.1101/720243 (2019) doi:10.1101/720243.

47. Ma, W. *et al.* BARTweb: a web server for transcriptional regulator association analysis. *NAR Genomics Bioinforma.* **3**, lqab022 (2021).

48. Wang, Z. *et al.* BART: a transcription factor prediction tool with query gene sets or epigenomic profiles. *Bioinformatics* **34**, 2867–2869 (2018).

49. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004).

50. Blackledge, N. P., Rose, N. R. & Klose, R. J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat. Rev. Mol. Cell Biol.* **16**, 643–649 (2015).

51. Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by Polycomb proteins. *Sci. Adv.* **1**, e1500737 (2015).

52. Rai, K. *et al.* Dual Roles of RNF2 in Melanoma Progression. *Cancer Discov.* **5**, 1314–1327 (2015).

53. Vieux-Rochas, M., Fabre, P. J., Leleu, M., Duboule, D. & Noordermeer, D. Clustering of mammalian *Hox* genes with other H3K27me3 targets within an active nuclear domain. *Proc. Natl. Acad. Sci.* **112**, 4672–4677 (2015).

54. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).

55. Ben-Saadon, R., Zaaroor, D., Ziv, T. & Ciechanover, A. The Polycomb Protein Ring1B Generates Self Atypical Mixed Ubiquitin Chains Required for Its In Vitro Histone H2A Ligase Activity. *Mol. Cell* **24**, 701–711 (2006).

56. Frankell, A. M. *et al.* The evolution of lung cancer and impact of subclonal selection in TRACERx. *Nature* **616**, 525–533 (2023).

57. Martínez-Ruiz, C. *et al.* Genomic–transcriptomic evolution in lung cancer and metastasis. *Nature* **616**, 543–552 (2023).

58. Sun, R. *et al.* Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nat. Genet.* **49**, 1015–1024 (2017).

59. Swanton, C. Intratumor Heterogeneity: Evolution through Space and Time. *Cancer Res.* **72**, 4875–4882 (2012).

60. Gerlinger, M. & Swanton, C. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *Br. J. Cancer* **103**, 1139–1143 (2010).

61. Tamburri, S. *et al.* Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression. *Mol. Cell* **77**, 840-856.e5 (2020).

62. Junttila, M. R. *et al.* Selective activation of p53-mediated tumour suppression in high-grade tumours. *Nature* **468**, 567–571 (2010).

63. Feldser, D. M. *et al.* Stage-specific sensitivity to p53 restoration during lung cancer progression. *Nature* **468**, 572–575 (2010).

64. Vicent, S. *et al.* ERK1/2 is activated in non-small-cell lung cancer and associated with advanced tumours. *Br. J. Cancer* **90**, 1047–1052 (2004).

65. LaFave, L. M. *et al.* Epigenomic State Transitions Characterize Tumor Progression in Mouse Lung Adenocarcinoma. *Cancer Cell* **38**, 212-228.e13 (2020).

66. Moffitt, J. R. *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).

67. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).

68. Fang, R. *et al.* Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. *Science* **377**, 56–62 (2022).

69. Allen, W. E., Blosser, T. R., Sullivan, Z. A., Dulac, C. & Zhuang, X. Molecular and spatial signatures of mouse brain aging at single-cell resolution. *Cell* **186**, 194-208.e18 (2023).

70. Lu, T., Ang, C. E. & Zhuang, X. Spatially resolved epigenomic profiling of single cells in complex tissues. *Cell* **186**, 2275–2279 (2023).

71. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife* **7**, e31657 (2018).

72. Govindan, R. *et al.* Genomic Landscape of Non-Small Cell Lung Cancer in Smokers and Never-Smokers. *Cell* **150**, 1121–1134 (2012).

73. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

74. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).

75. Moffitt, J. R. *et al.* High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci.* **113**, 11046–11051 (2016).

76. Liu, M. *et al.* Chromatin tracing and multiplexed imaging of nucleome architectures (MINA) and RNAs in single mammalian cells and tissue. *Nat. Protoc.* **16**, 2667–2697 (2021).

77. Moffitt, J. R. *et al.* High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. *Proc. Natl. Acad. Sci.* **113**, 14456–14461 (2016).

78. Lu, Y. *et al.* Spatial transcriptome profiling by MERFISH reveals fetal liver hematopoietic stem cell niche architecture. *Cell Discov.* **7**, 47 (2021).

79. Damo, M. *et al.* Inducible de novo expression of neoantigens in tumor cells and mice. *Nat. Biotechnol.* **39**, 64–73 (2021).

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

81. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

82. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

83. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

84. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

85. Parthasarathy, R. Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nat. Methods* **9**, 724–726 (2012).

86. Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).

87. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

88. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).

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

Acknowledgements: We thank Drs. Andrew Xiao, Megan King, Bogdan Bintu, Pu Zheng, Robert Homer, Katerina Politi, Roy Herbst, and members of the Wang and Muzumdar labs for helpful discussions, the Yale Cancer Center (YCC) Functional Genomics Core for assistance with RNAi screening assays, Yale Pathology Tissue Services (YPTS) for histology support, the Yale Center for Genome Analysis (YCGA) for library preparation and sequencing of isolated mouse tumors, Dr. Arjun Bhutkar and Sarah Blatt for sharing the computational pipeline for TCGA survival analyses, the ENCODE Consortium and the labs of Drs. Bing Ren and John Stamatoyannopoulos for generating the H3K4me1 and H3K4me3 ChIP-seq and DNase-seq datasets in adult mouse lung, Dr. Nik Joshi for providing the KP cell line, Drs. Robert Weinberg and Didier Trono for constructs, and Drs. Tyler Jacks and Liqun Luo for mouse strains.

Funding: S.S.A. and C.F.R. were supported by postdoctoral fellowships through the Yale Cancer Biology Training Program (T32CA193200). T.B.J. was in part supported by 5T32GM007205. J.S.D.R. was supported by a NIH Predoctoral Training Grant (2T32GM007499). M.D.M. acknowledges support from a NIH Director's New Innovator Award (DP2CA248136), AACR-Genentech NextGen Grant for Transformative Cancer Research (19-

20-18-MUZU), Yale SPORE in Lung Cancer Career Enhancement Program award (P50CA196530), and in part, the Yale Comprehensive Cancer Center Support Grant (P30CA016359). S.W. was partly supported by NIH (UG3CA268202, U01DA052775, R01HG011245, DP2GM137414) and Pershing Square Sohn Cancer Research Alliance. This work was largely supported by an NCI IMAT program award (R33CA251037) to M.D.M. and S.W.

Author contributions: M.D.M. and S.W. conceived of and supervised the study. M.L., M.D.M., and S.W. designed the study with help from S.J. and S.S.A. M.L., S.J., S.S.A., T.B.J., and J.S.D.R. performed experiments. M.L., S.J., S.S.A., T.B.J., T.Y. C.F.R., and G.B. analyzed data with help from M.D.M. and S.W. M.L., S.J., S.S.A., T.B.J., M.D.M. and S.W. wrote the manuscript with input from all authors.

Competing interests: S.W., M.L., M.D.M., S.J., and S.S.A. are inventors on a patent applied for by Yale University related to this work. M.D.M. received research funding from a Genentech supported AACR grant. S.W. is a share-holder and consultant of Translura, Inc. The remaining authors declare no competing interests.

Materials and Correspondence: Correspondence and requests for materials should be addressed to Siyuan Wang or Mandar Deepak Muzumdar.

Data and materials availability: All generated cell lines, constructs, raw and analyzed imaging data, and code for data analysis are available upon request. All sequencing data will be available at the Gene Expression Omnibus (GEO) upon publication. Open-source code for imaging data collection are available at https://github.com/ZhuangLab/storm-control.

Figures

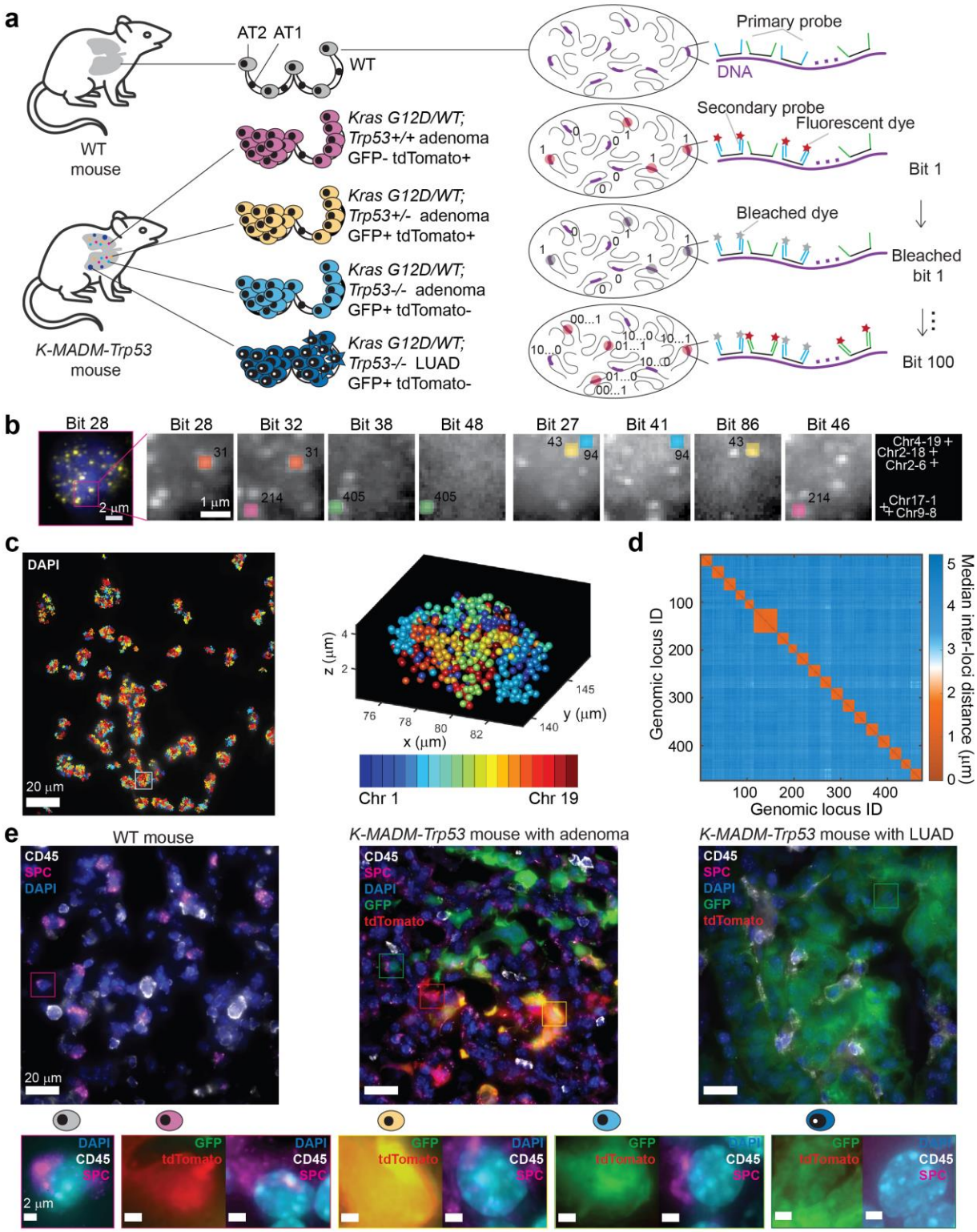

a
AT2 AT1
WT
WT mouse
K-MADM-Trp53 mouse
Kras G12D/WT; Trp53+/+ adenoma GFP- tdTomato+
Kras G12D/WT; Trp53+/- adenoma GFP+ tdTomato+
Kras G12D/WT; Trp53-/- adenoma GFP+ tdTomato-
Kras G12D/WT; Trp53-/- LUAD GFP+ tdTomato-
Primary probe
DNA
Secondary probe
Fluorescent dye
Bit 1
Bleached dye
Bleached bit 1
Bit 100
b
Bit 28
Bit 32
Bit 38
Bit 48
Bit 27
Bit 41
Bit 86
Bit 46
2 μm
1 μm
Chr4-19 + Chr2-18 + Chr2-6 + Chr17-1 + Chr9-8
c
DAPI
20 μm
z (μm)
x (μm)
y (μm)
Chr 1
Chr 19
d
Genomic locus ID
Median inter-loci distance (μm)
e
WT mouse
K-MADM-Trp53 mouse with adenoma
K-MADM-Trp53 mouse with LUAD
CD45
SPC
DAPI
GFP
tdTomato
20 μm
2 μm

Fig. 1. A genome-scale chromatin tracing strategy to visualize cancer 3D genomes. a, Schematic illustration of the experimental procedure. **b,** (Left panel) Raw FISH foci in bit 28 in a WT AT2 cell. (Right panels) Zoom-in images of raw FISH foci showing the decoding procedure and reconstructed genomic loci for 5 loci (shown as dots with 5 pseudo-colors each appearing twice). **c,** (Left panel) Reconstructed chromatin traces superimposed with DAPI staining. The traces are 2D projections of x, y coordinates. The DAPI image is a maximum-intensity z-projection from a 10-μm z-stack. (Right panel) The 3D positions of all decoded genomic loci in a single-cell nucleus. Different pseudo-colors represent different autosomes. **d,** Matrix of median inter-loci distances between all genomic loci in AT2 cells. n = 5,146 AT2 cells. **e,** (Upper panels) Immunofluorescence staining of cell-type markers, DAPI staining, and fluorescent protein imaging of lung tissue from a WT mouse, a *K-MADM-Trp53* mouse with adenomas, and a *K-MADM-Trp53* mouse with LUAD. (Lower panels) Representative cells of each state. The images are maximum-intensity z-projections from 10-μm z-stacks.

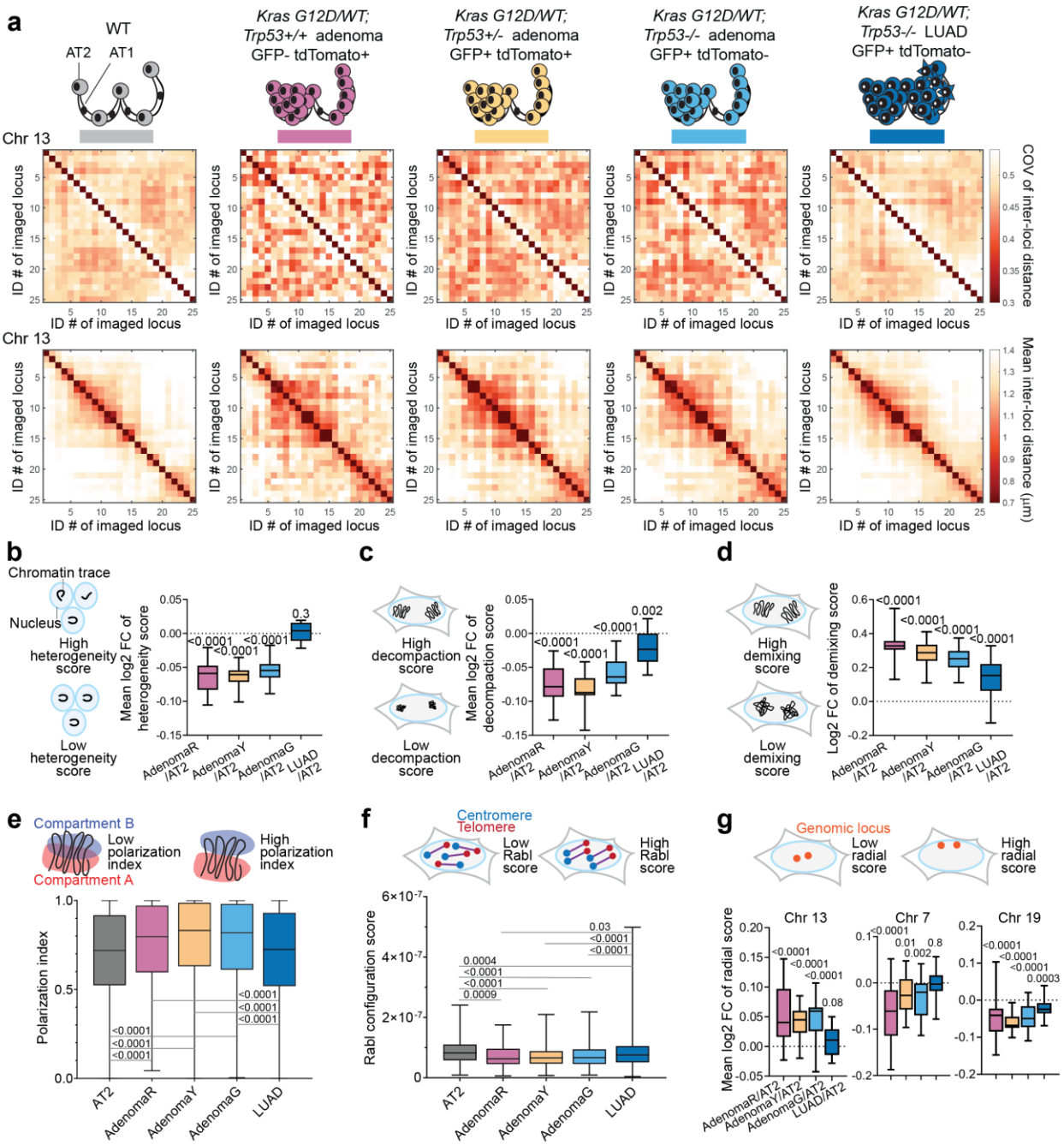


Fig. 2. Systematic changes of 3D genome conformations during lung cancer progression. a, Coefficient of variation (COV) of inter-loci distances (upper panels) and mean inter-loci distances (lower panels) of mouse Chr13 in WT AT2 (n = 5,146), *Trp53*$^{+/+}$ adenoma (AdenomaR, n = 1,063), *Trp53*$^{+/-}$ adenoma (AdenomaY, n = 1,631), *Trp53*$^{-/-}$ adenoma (AdenomaG, n = 1,995), and *Trp53*$^{-/-}$ LUAD (n = 18,470) cells. **b-d,** Distributions of the (mean)

$\log_2$ fold change of heterogeneity (COV of inter-loci distance), decompaction (mean inter-loci distance), and demixing scores of each autosome (n = 19), comparing each cancer state to AT2 state. **e,** Distribution of polarization indices of A and B compartments in the five cell states. **f,** Distribution of Rabl configuration scores of single cells in the five cell states. **g,** Distribution of the mean $\log_2$ fold change of radial scores of Chr13, Chr7 and Chr19, comparing each cancer state to AT2 state. Cell numbers in **b-g** are the same as in **a**. p values of two-sided Wilcoxon signed-rank test (**b-d, g**) or two-sided Wilcoxon rank-sum test (**e**, **f**) are displayed. In **b-g**, the horizontal lines of each box from top to bottom represent the 75th percentile, median, and 25th percentile. Whiskers extend to the maximum and minimum values.

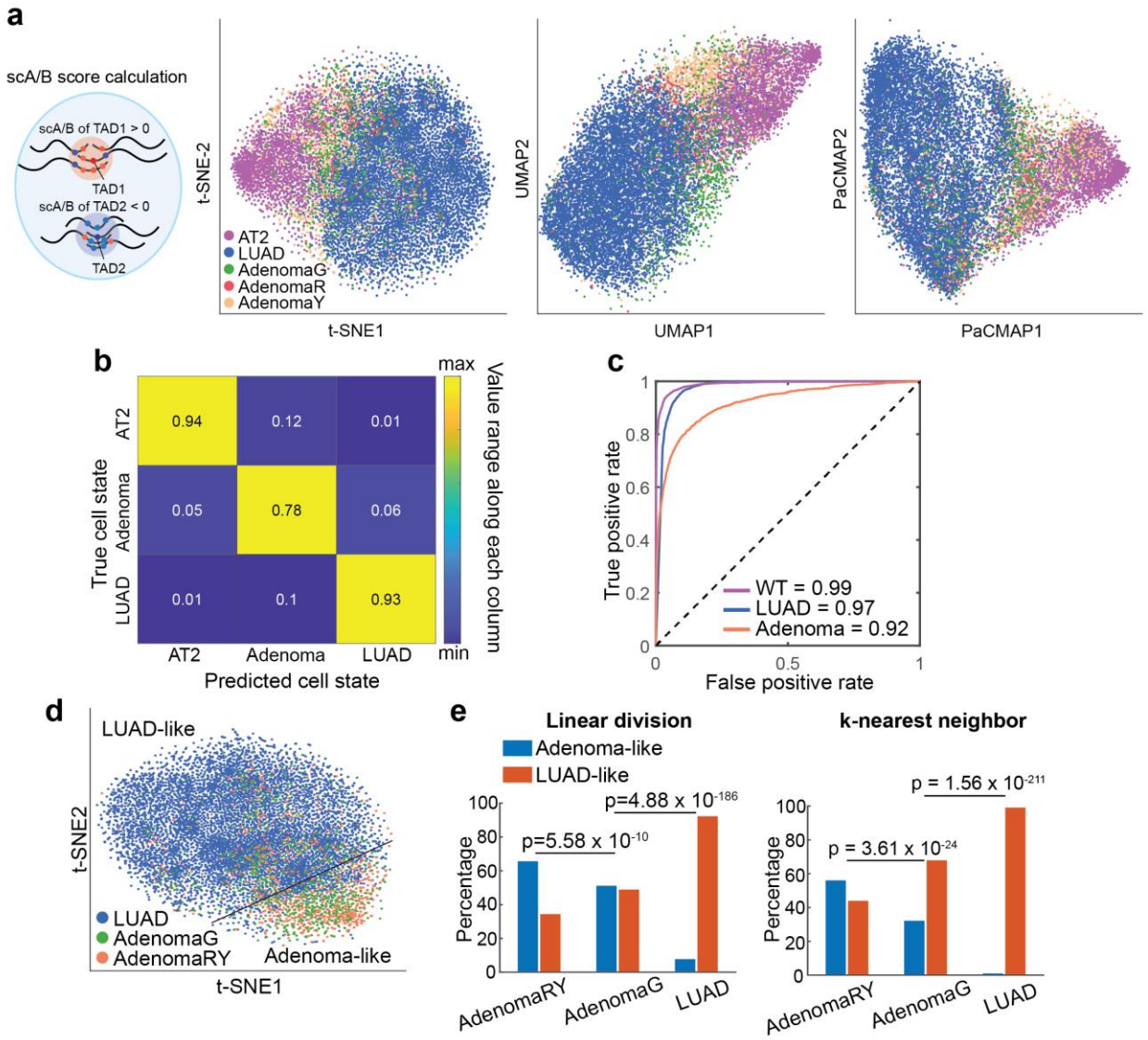


Fig. 3. The single-cell 3D genome distinguishes and encodes cancer progression states. a, (Left panel) Cartoon illustration of scA/B score calculation. (Right panels) t-SNE, UMAP and PaCMAP plots of single-cell 3D genome conformations. n = 3,430, 170, 741, 938, 9,677 for WT AT2, AdenomaR, AdenomaG, AdenomaY, LUAD cells. Cell numbers of each cell state are the same in **b-d**. **b,** Confusion matrix of supervised machine learning. The number in each matrix element represents the precision in each predicted state. **c,** Receiver operating characteristic (ROC) curves of the machine learning model. The area under curve (AUC) values are shown. **d,** t-SNE plot of single-cell 3D genome conformations of adenoma and LUAD cells. The black line

best separates adenoma red and yellow cells (AdenomaRY) from LUAD cells. **e,** Percentages of cells with adenoma-like or LUAD-like 3D genome conformations in **d**, in each of the AdenomaRY, AdenomaG, and LUAD states. The adenoma-like or LUAD-like conformation state for each cell is assigned based on a linear division approach (left, cells below or above the black line in **d**) or the majority state of its five nearest neighbors (k-nearest neighbor, right) in **d**. p values from two-sided Fisher's exact test are shown.

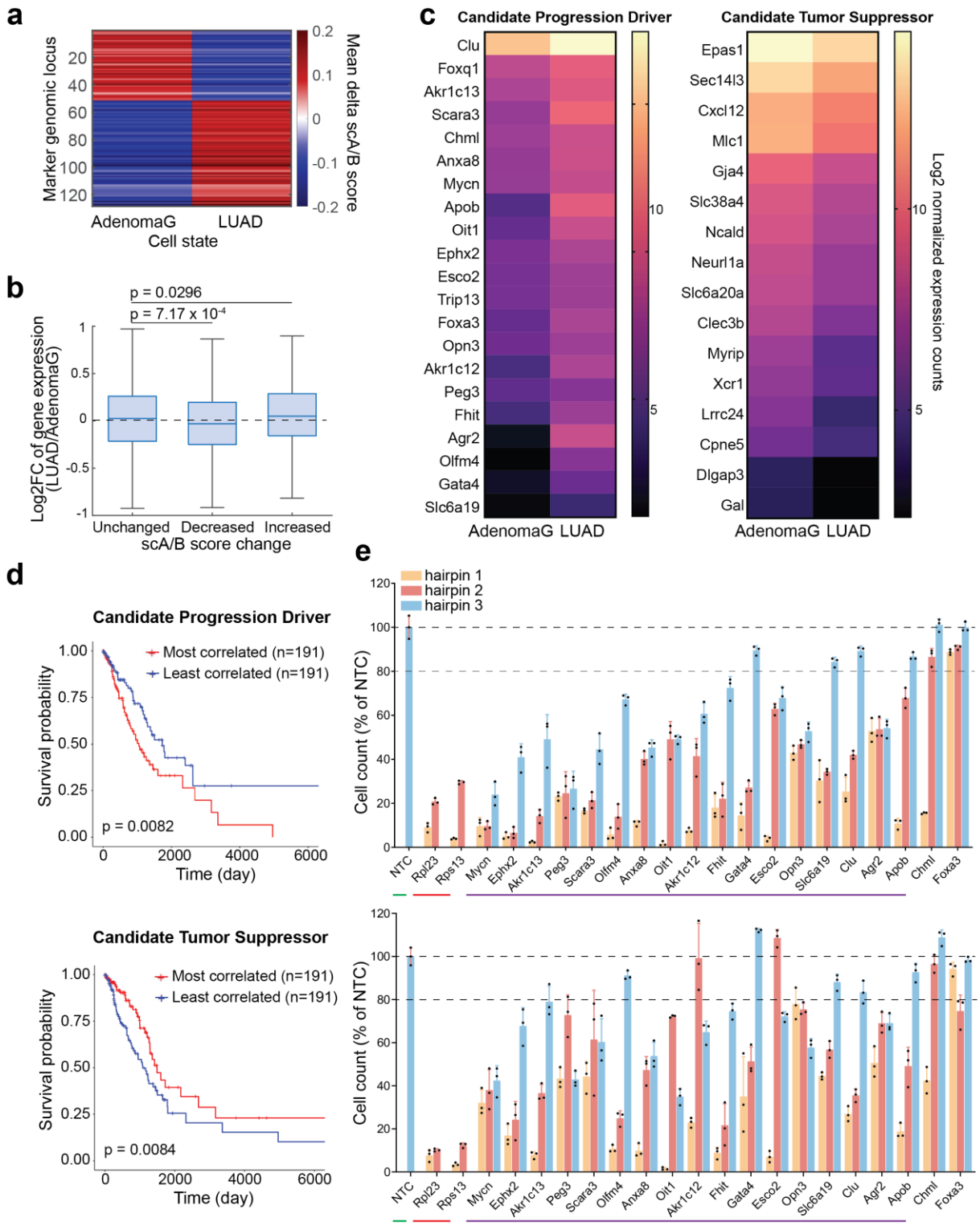


Fig. 4. The single-cell 3D genome nominates genetic dependencies during LUAD progression. a, Mean scA/B score changes in marker genomic loci between AdenomaG and

LUAD cells. **b,** Log_2 fold change of gene expression from AdenomaG to LUAD cells in genomic loci with unchanged (n = 2,574 genes), decreased (n = 705 genes) or increased (n = 390 genes) scA/B scores. The horizontal lines of each box represent the 75th percentile, median, and 25th percentile. Whiskers extend to the non-outlier maximum and non-outlier minimum. Outliers are defined as values away from the top or bottom of the box by more than 1.5 times interquartile range. p values of one-sided Wilcoxon rank-sum tests are shown. **c,** Heatmap of candidate progression driver (CPD) and candidate tumor suppressor (CTS) gene expression (log_2 normalized expression counts) comparing AdenomaG and LUAD tumors derived from *K-MADM-Trp53* mice. **d,** Kaplan-Meier survival curves comparing TCGA LUAD patients with tumors stratified by their gene expression correlation with the CPD gene signature (top/bottom 40%, n = 191 tumors per group, p = 0.0082, two-sided log-rank test, upper panel), and with the CTS gene signature (top/bottom 40%, n = 191 tumors per group, p = 0.0084, two-sided log-rank test, lower panel). **e,** Cell viability (mean ± SD, normalized to the mean of non-targeting control (NTC), n = 3 replicates per hairpin, 3 hairpins per gene) of arrayed RNAi screen targeting CPD genes in the KP (upper panel) and SA6082inf (lower panel) LUAD cell lines. NTC (green underline), positive controls (red underline), and CPD genes with significant phenotypes (at least two out of three hairpins with less than 80% of NTC cell count, purple underline) in both cell lines are indicated.

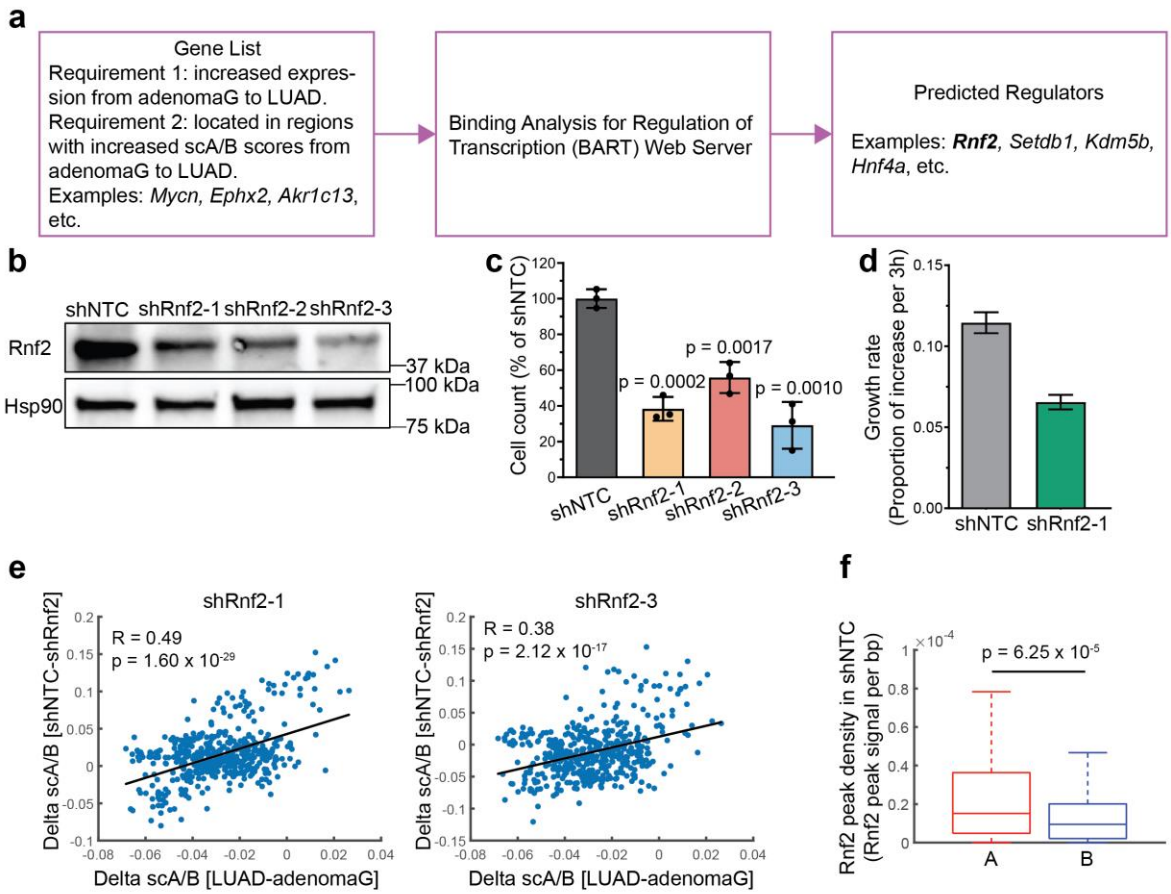


Fig.5. Rnf2 partially regulates 3D genome organization changes during the adenoma-to-LUAD transition. a, Diagram illustration of the prediction procedure that identified Rnf2 as a candidate. **b,** Western blot analysis of Rnf2 protein levels following Rnf2 knockdown with three independent hairpins in the KP cell line. Control cells were constructed with a non-targeting control shRNA sequence (shNTC). HSP90 is loading control. **c,** Cell viability (mean ± SD, normalized to mean of shNTC) following knockdown with three Rnf2 shRNAs, n = 3 replicates per hairpin. p values from two-sided two-sample t-tests are shown. **d,** Continuous assessment of cell proliferation (CellCyte) reveals that cell growth rate is reduced following Rnf2 knockdown (shRnf2-1). Error bars represent 95% confidence intervals. **e,** scA/B score changes from shRnf2 to shNTC versus those from AdenomaG to LUAD show a significant positive correlation, n = 473 target genomic regions. Black lines are fitted regression lines. Pearson correlation

coefficients and p values are shown. **f,** Rnf2 peak densities in A (n = 209) and B (n = 264) compartments in shNTC. The horizontal lines of each box from top to bottom represent the 75th percentile, median, and 25th percentile. Whiskers extend to the non-outlier maximum and non-outlier minimum. Outliers are defined as values at least 1.5 times interquartile range away from the top or bottom of the box. p value of two-sided Wilcoxon rank-sum test is shown.

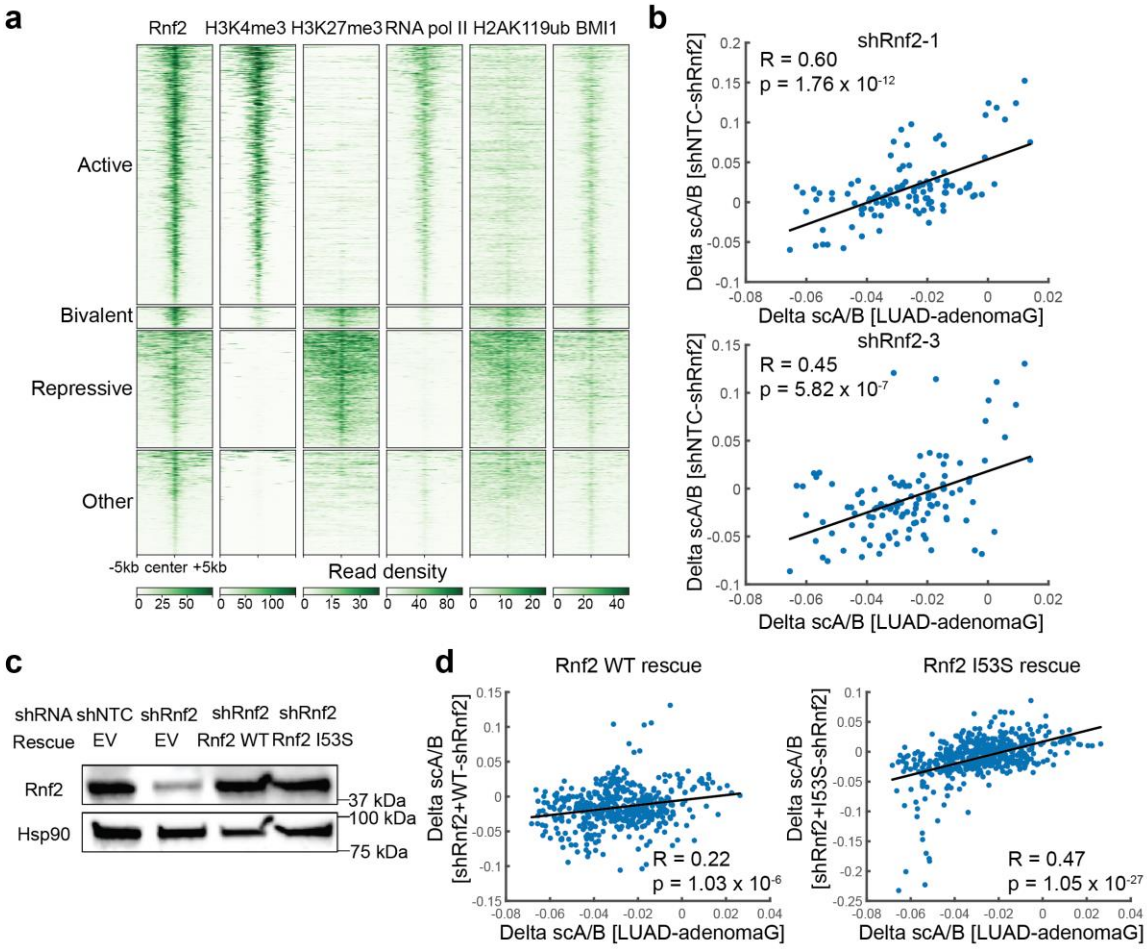
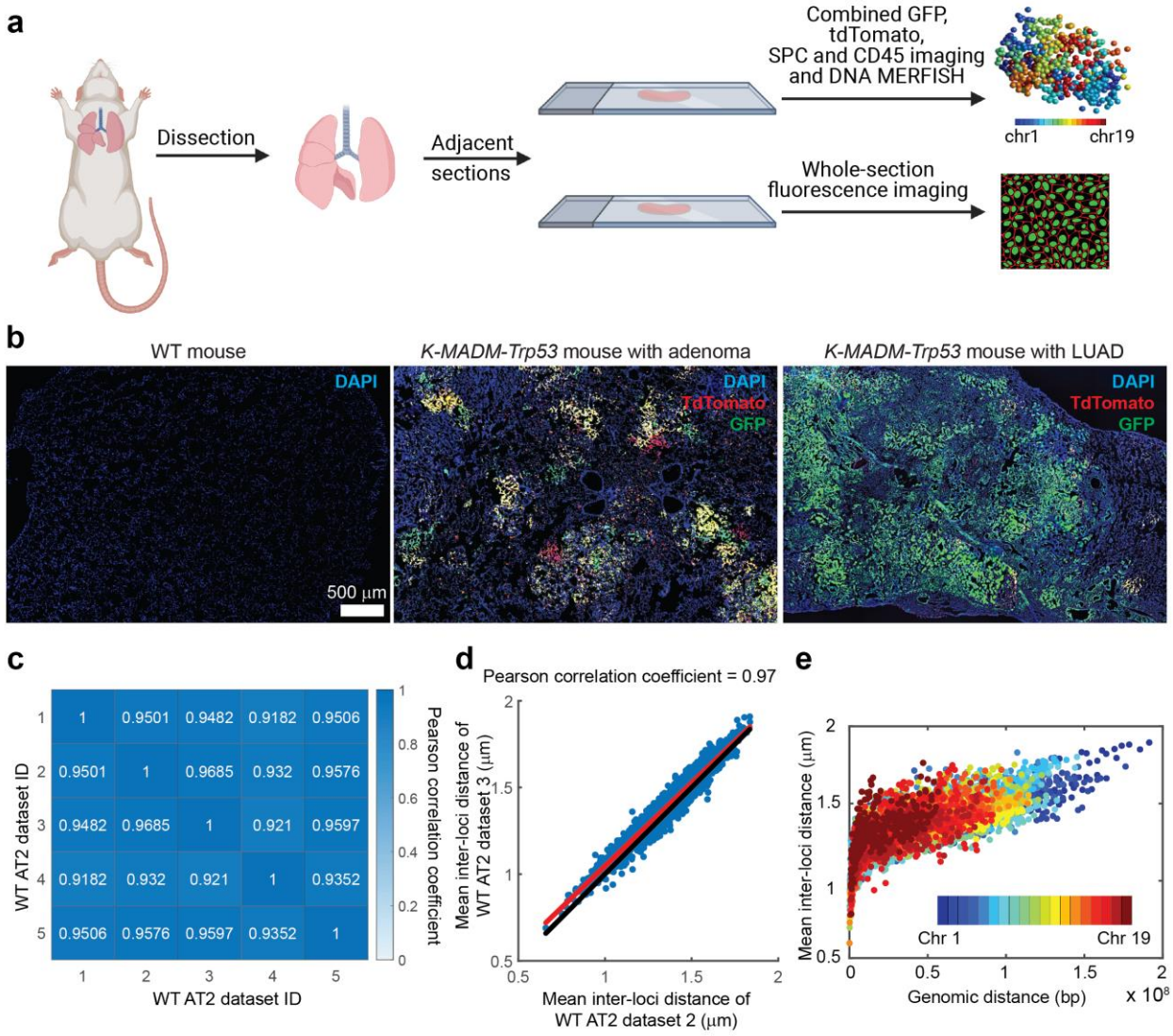


Fig.6. Rnf2 partially regulates 3D genome organization changes via a ubiquitin ligase-independent activity. a, CUT&RUN read density heatmaps of Rnf2, H3K4me3, H3K27me3, RNA polymerase II with phosphorylated S5 modification, H2AK119ub, and BMI1 in shNTC KP cells. Rnf2 peak regions [-5k, +5k] of all target genomic regions are shown and are categorized as active (H3K4me3+, H3K27me3-), bivalent (H3K4me3+, H3K27me3+), repressive (H3K4me3-, H3K27me3+), or other (H3K4me3-, H3K27me3-) based on chromatin marks. **b,** scA/B score changes from shRnf2 to shNTC versus those from AdenomaG to LUAD show a stronger positive correlation using target genomic regions with only active Rnf2 peaks, n = 113 target genomic regions. **c**, Western blot analysis of Rnf2 protein levels following rescue of shRNA knockdown (shRnf2-3) with stable transduction of Rnf2 WT, ubiquitin-ligase dead Rnf2

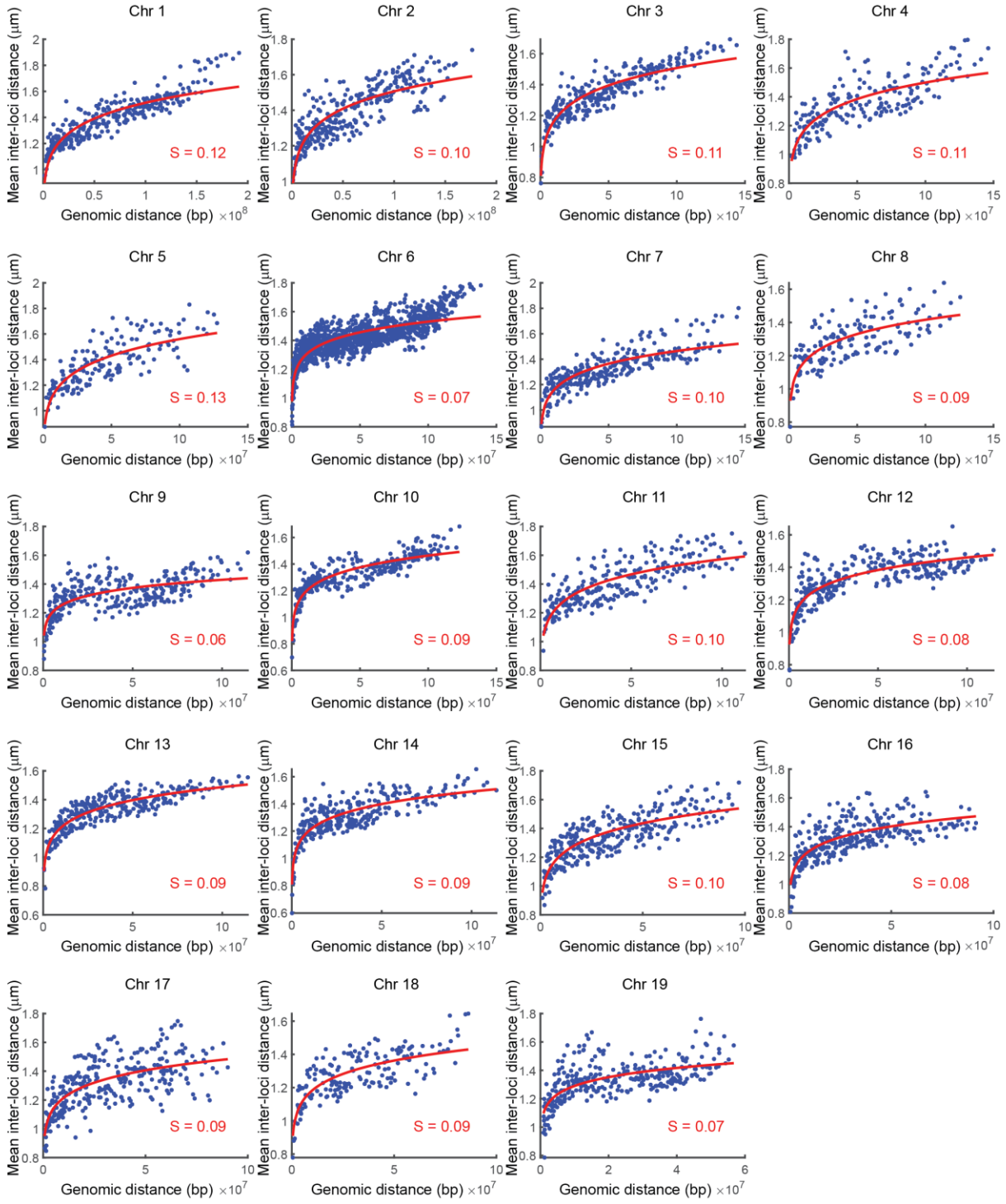
I53S mutant, or empty vector (EV) control. HSP90 is loading control. **d**, scA/B score changes from shRnf2-3 to shRnf2-3+WT Rnf2 (left) and to shRnf2-3+I53S Rnf2 (right) versus those from AdenomaG to LUAD show a stronger positive correlation with expression of a catalytically dead mutant of Rnf2, n = 473 target genomic regions. In **b, d,** the black lines are fitted regression lines. Pearson correlation coefficients and p values are shown.

Extended Data Figures



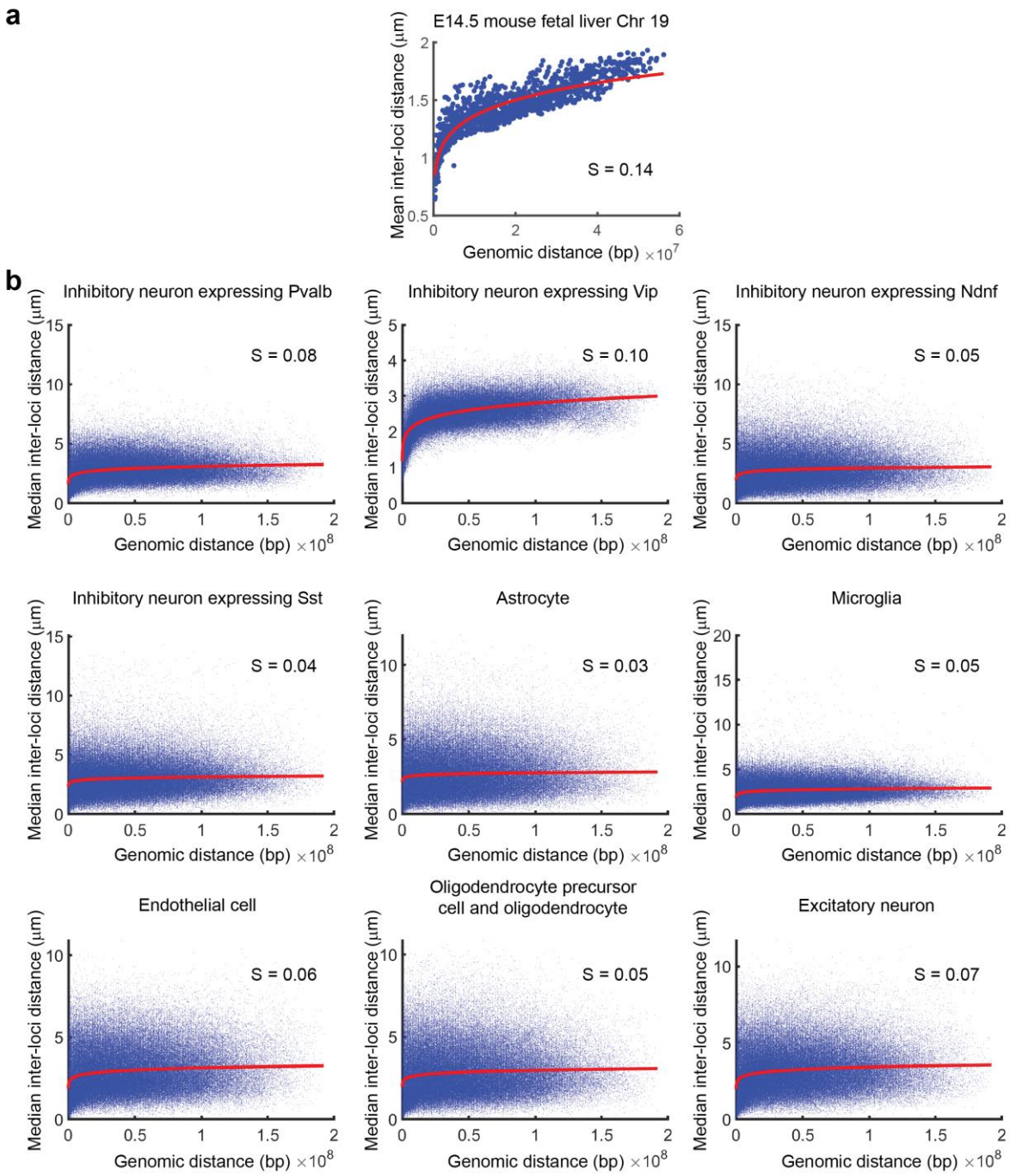
Extended Data Fig. 1. Validation of the genome-scale chromatin tracing strategy to visualize 3D genome organization *in vivo*. a, Schematic illustration of the experimental procedure. **b,** Whole-section fluorescence images of a wild-type (WT) mouse lung, a *K-MADM-Trp53* lung with adenomas and a *K-MADM-Trp53* lung with a LUAD. **c,** Pearson correlation coefficients of mean inter-loci distances between WT AT2 datasets. **d,** Mean inter-loci distances of WT AT2 dataset 3 versus mean inter-loci distances of WT AT2 dataset 2. The red line is a fitted linear regression line. The black line is the y = x line. **e,** Mean inter-loci spatial distance

versus genomic distance for all pairs of genomic loci on each autosome in AT2 cells. Different pseudo-colors represent different autosomes. n = 6,039 intra-chromosomal inter-loci pairs in **d-e**. n = 5,146 WT AT2 cells in **d-e**.


Chr 1
S = 0.12
Chr 2
S = 0.10
Chr 3
S = 0.11
Chr 4
S = 0.11
Chr 5
S = 0.13
Chr 6
S = 0.07
Chr 7
S = 0.10
Chr 8
S = 0.09
Chr 9
S = 0.06
Chr 10
S = 0.09
Chr 11
S = 0.10
Chr 12
S = 0.08
Chr 13
S = 0.09
Chr 14
S = 0.09
Chr 15
S = 0.10
Chr 16
S = 0.08
Chr 17
S = 0.09
Chr 18
S = 0.09
Chr 19
S = 0.07
Mean inter-loci distance (μm)
Genomic distance (bp) ×10^8
Genomic distance (bp) ×10^7

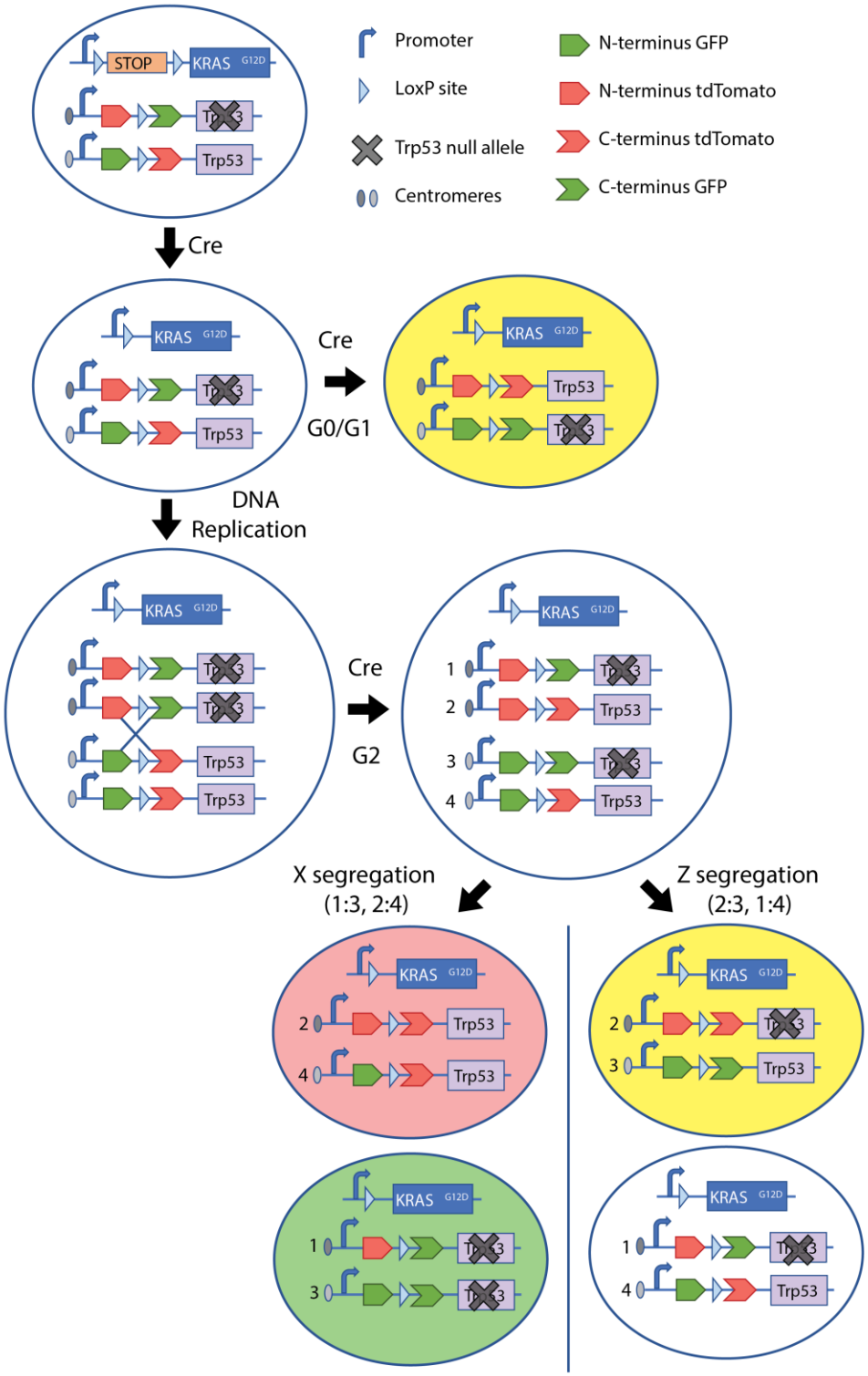
Extended Data Fig. 2. Power-law scaling of mean spatial distance versus genomic distance between pairs of genomic loci on each autosome. The red lines are fitted power-law functions. S is the scaling factor of the power-law function.

5



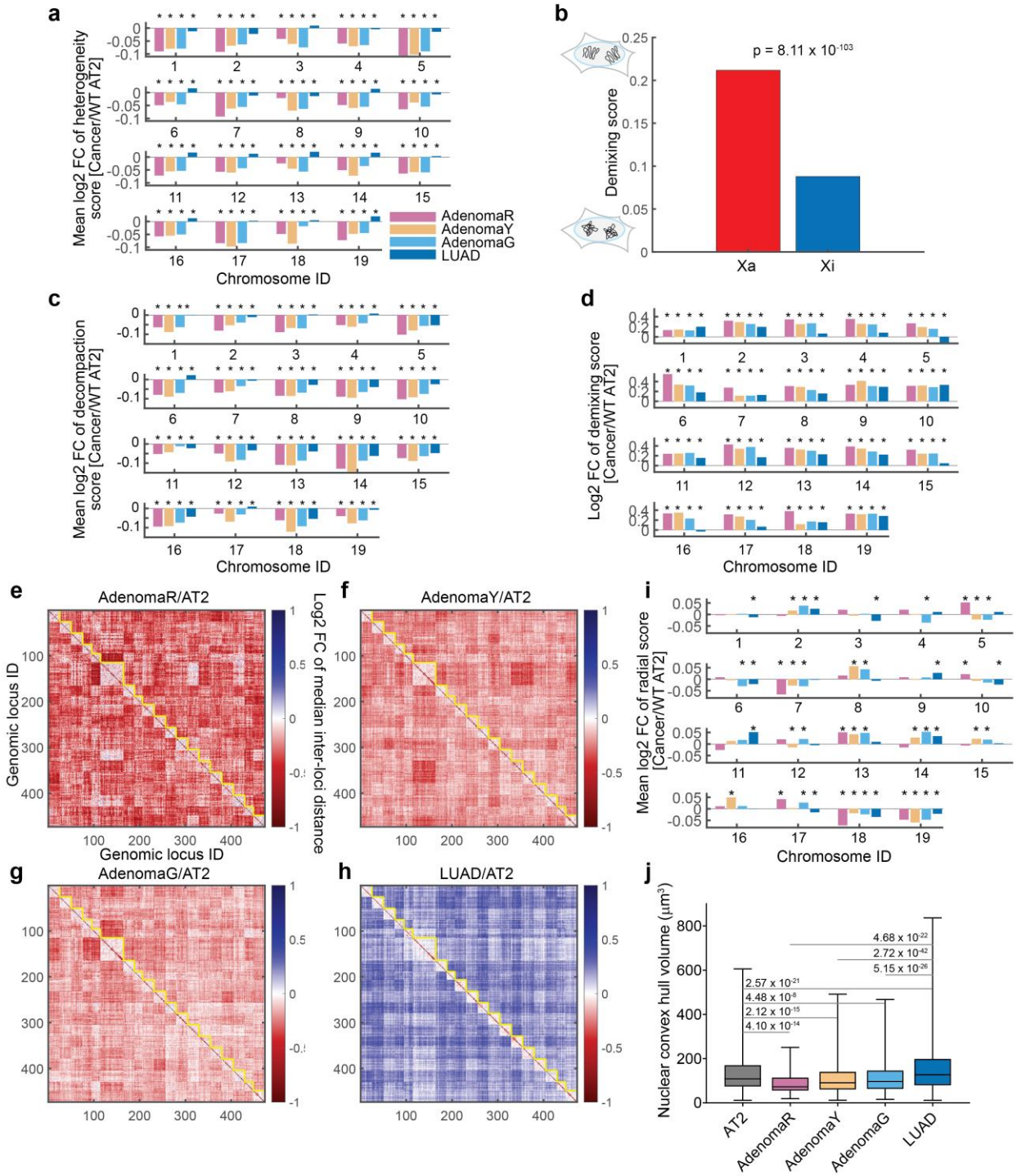
Extended Data Fig. 3. Power-law scaling of mean/median inter-loci spatial distance versus genomic distance on mouse chromosomes from other tissue types. a, Power-law scaling of Chr 19 in E14.5 mouse fetal liver. Data were re-analyzed from Liu et al. *Nat. Commun.* (2020)[23]. The lines are fitted power-law functions. S is the scaling factor. **b,** Power-law scaling of all 20

mouse chromosomes (Chr 1 - Chr 19, Chr X) in the mouse brain cortex. Data were re-analyzed from Takei et al. *Science*. (2021)[22].



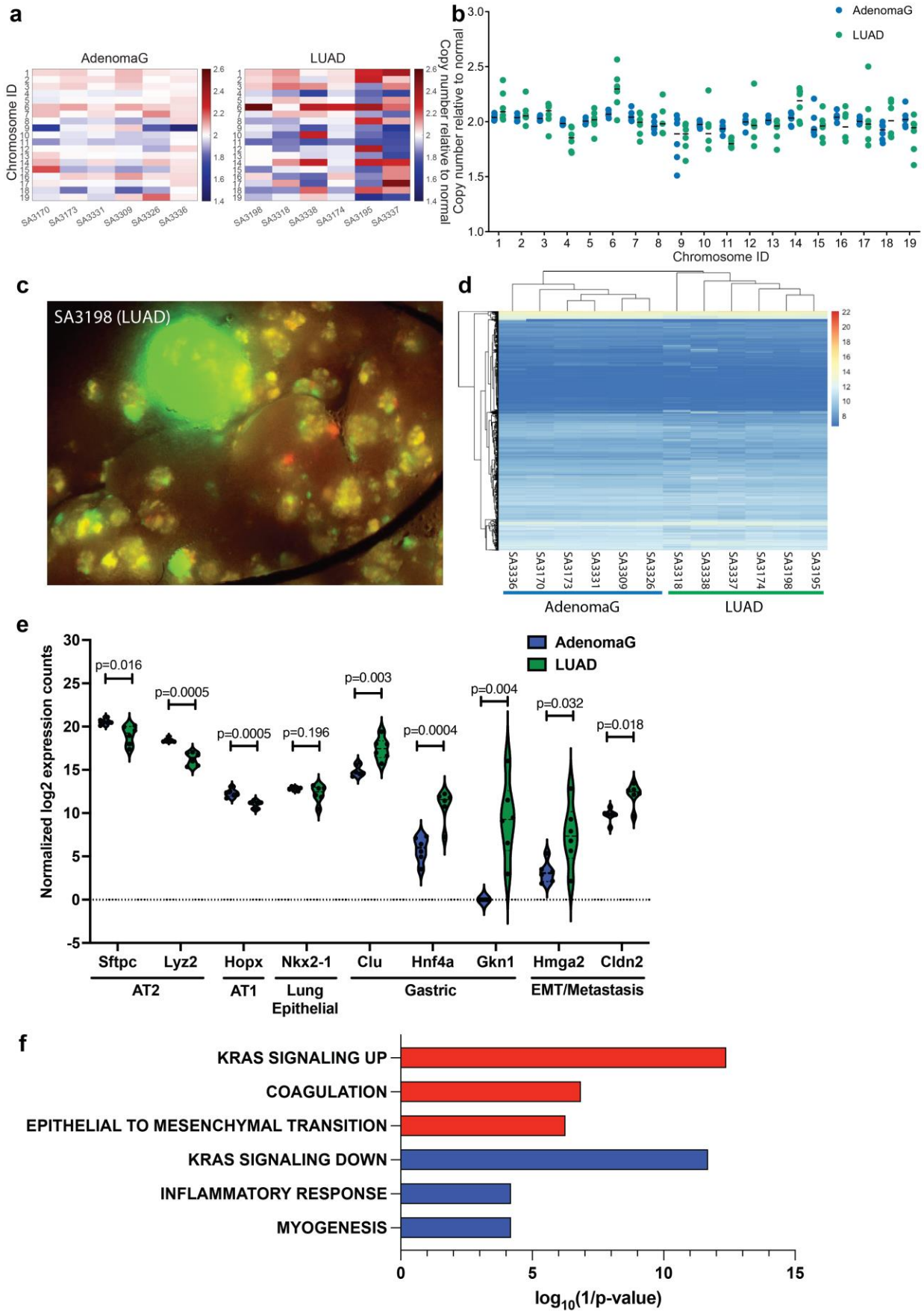
Extended Data Fig. 4. Schematic of MADM system for generation of fluorescently labelled lung tumor subclones of defined *Trp53* genotype. Efficient Cre-mediated *intra*-chromosomal

recombination deletes a STOP cassette to induce oncogenic Kras^{G12D} expression and initiate lung tumor (adenoma) formation. Thereafter, less efficient Cre-mediate inter-chromosomal recombination using MADM leads to unambiguous labeling of tdTomato+ (red) *Trp53*$^{+/+}$, tdTomato+/GFP+ (yellow) *Trp53*$^{+/-}$, GFP+ (green) *Trp53*$^{-/-}$, and tdTomato-/GFP- (colorless) subclones. Sibling green *Trp53*$^{-/-}$ and red *Trp53*$^{+/+}$ cells are generated from a single Cre-mediated mitotic recombination event followed by X-segregation in G2 (G2-X).

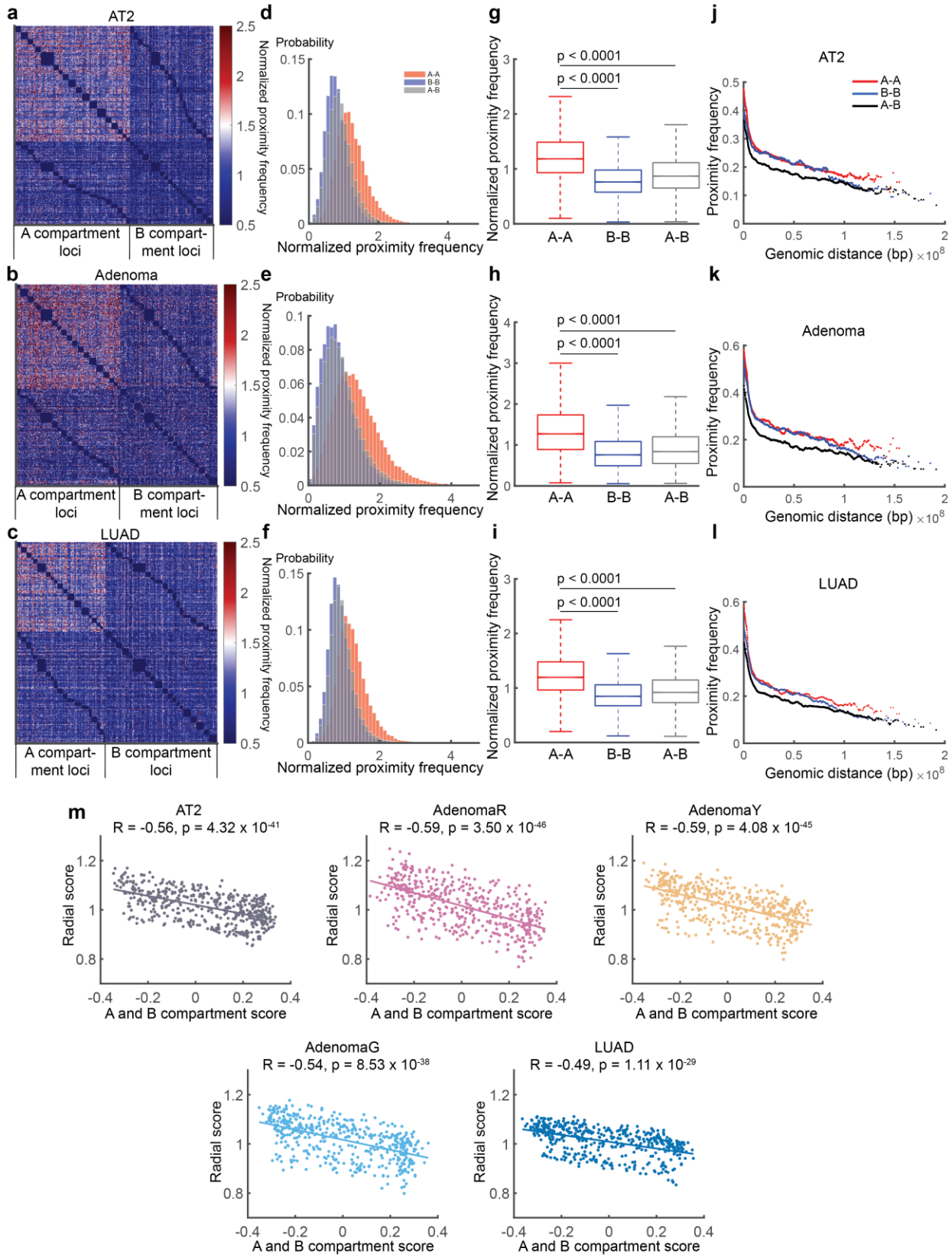


Extended Data Fig. 5. Changes of cancer-state-specific 3D genome features comparing each cancer state to the AT2 state. a, Mean $\log_2$ fold change of heterogeneity scores of each chromosome, comparing each cancer state to WT AT2 cell state. * indicates FDR < 0.05, two-sided Wilcoxon signed-rank test. **b,** Demixing scores (standard deviation of normalized mean

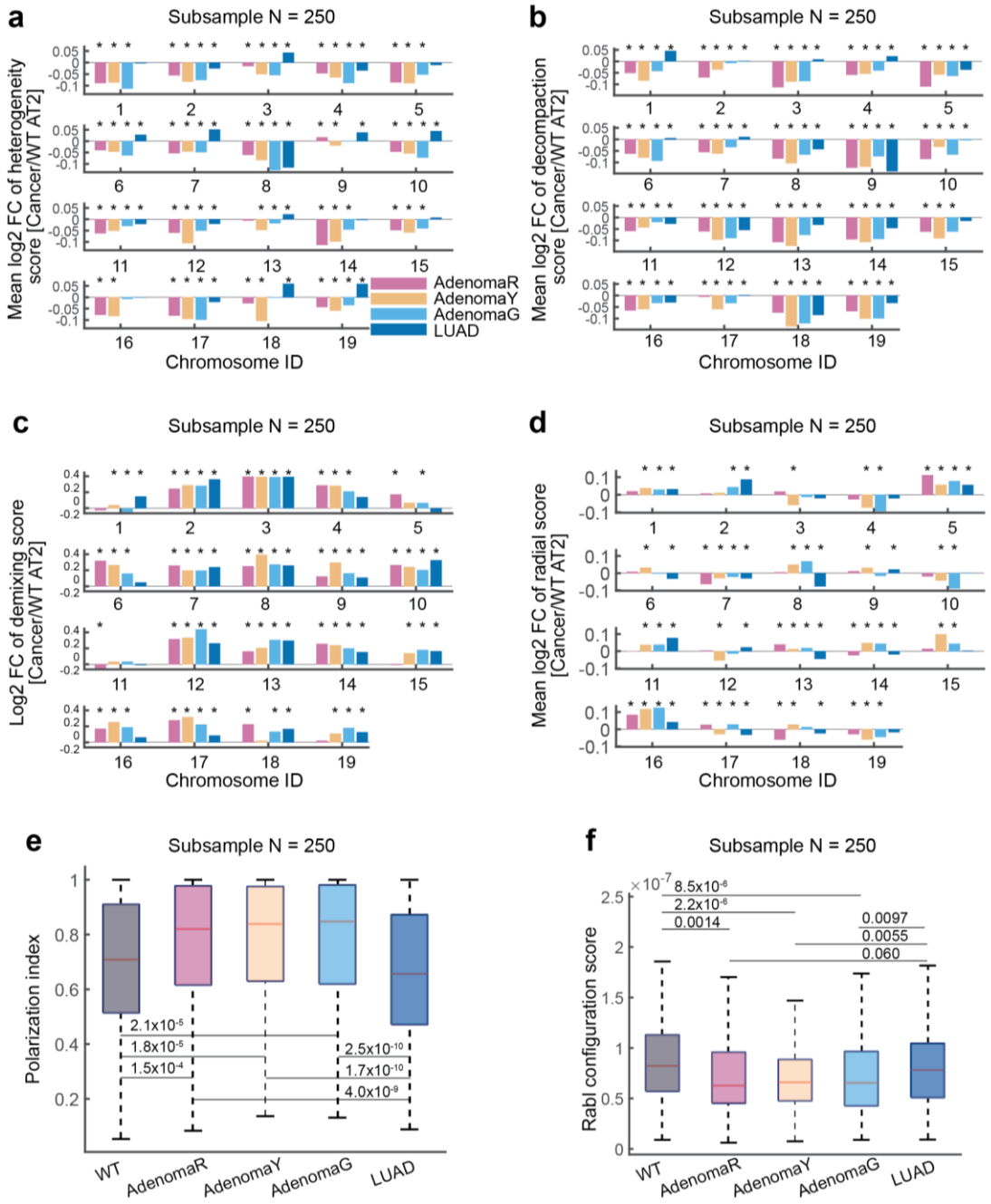
inter-loci distances) of the active X chromosomes (Xa, n = 95) and inactive X chromosomes (Xi, n = 95) in human IMR90 cells show a reduction (increased intermixing) in Xi, as previously described using other analyses[16]. p value from two-sided Levene's test is shown. **c,** Mean $\log_2$ fold change of decompaction scores of each chromosome, comparing each cancer state to WT AT2 cell state. * indicates FDR < 0.05, two-sided Wilcoxon signed-rank test. **d,** Log_2 fold change of the demixing score of each chromosome, comparing each cancer state to WT AT2 cell state. * indicates FDR < 0.05, two-sided Levene's test. **e-h**, Log2 fold changes of median inter-loci distances for adenoma red (AdenomaR) (**e**), adenoma yellow (AdenomaY) (**f**), adenoma green (AdenomaG) (**g**), and LUAD (**h**) relative to AT2 cells. Yellow lines highlight the boundaries of chromosomes. **i,** Mean log2 fold change of radial scores of each chromosome, comparing each cancer state to the AT2 cell state. * indicates FDR < 0.05, two-sided Wilcoxon signed-rank test. **j,** Nuclear convex hull volume of WT AT2, AdenomaR, AdenomaY, AdenomaG, and LUAD cells. p values of two-sided Wilcoxon rank-sum test are shown. The horizontal lines of each box from top to bottom represent the 75th percentile, median, and 25th percentile. Whiskers extend to the maximum and minimum values. Cell numbers in **a, c-j** are the same as in **Fig. 2**.


a
AdenomaG
LUAD
Chromosome ID
SA3170
SA3173
SA3331
SA3309
SA3326
SA3336
SA3198
SA3318
SA3338
SA3174
SA3195
SA3337
b
AdenomaG
LUAD
Copy number relative to normal
Chromosome ID
c
SA3198 (LUAD)
d
AdenomaG
LUAD
e
AdenomaG
LUAD
Normalized log2 expression counts
p=0.016
p=0.0005
p=0.0005
p=0.196
p=0.003
p=0.0004
p=0.004
p=0.032
p=0.018
Sftpc
Lyz2
Hopx
Nkx2-1
Clu
Hnf4a
Gkn1
Hmga2
Cldn2
AT2
AT1
Lung Epithelial
Gastric
EMT/Metastasis
f
KRAS SIGNALING UP
COAGULATION
EPITHELIAL TO MESENCHYMAL TRANSITION
KRAS SIGNALING DOWN
INFLAMMATORY RESPONSE
MYOGENESIS
log10(1/p-value)

Extended Data Fig. 6. Copy number variation (CNV) and RNA-seq analyses of lung tumors from *K-MADM-Trp53* mice. a, Heatmap of mean CNVs of each autosome across AdenomaG and LUAD samples (n = 6 tumors per group). Classification of AdenomaG and LUAD is based on parallel gene expression analysis by RNA-seq on the same tumors (**d-e**). **b,** Distribution of copy numbers of each autosome in AdenomaG and LUAD tumors relative to paired normal. Each dot represents a single tumor. Black lines represent the median values. **c,** Representative large GFP+ (green) tumor dissected under fluorescence microscopy for RNA-seq and Exome-seq analyses. **d-e**, Unsupervised hierarchical clustering of all expressed genes (**d**) segregates dissected green tumors into two clusters (n = 6 tumors per cluster) defining AdenomaG and LUAD cells based on the expression of previously described markers of histologic progression[37] (**e**), including loss of AT2/AT1 genes (*Stfpc*, *Lyz2*, and *Hopx*) and acquisition of genes associated with gastric differentiation, epithelial-to-mesenchymal transition (EMT), and metastasis (*Clu*, *Hnf4a*, *Gkn1*, *Hmga2*, *Cldn2*). Log2 normalized expression counts are shown. The horizontal lines in each violin represent the 75th percentile, median, and 25th percentile. p values of two-tailed Wilcoxon rank-sum test are displayed. **f,** Gene set enrichment analysis using the MSigDB Hallmarks (H1) shows the top 3 enriched gene sets for upregulated (red, log2 fold change>2 and FDR<0.05) and downregulated (blue, log2 fold change<-1 and FDR<0.05) genes in LUAD compared to AdenomaG. Log10(1/p value) is plotted (one-sided hypergeometric test).

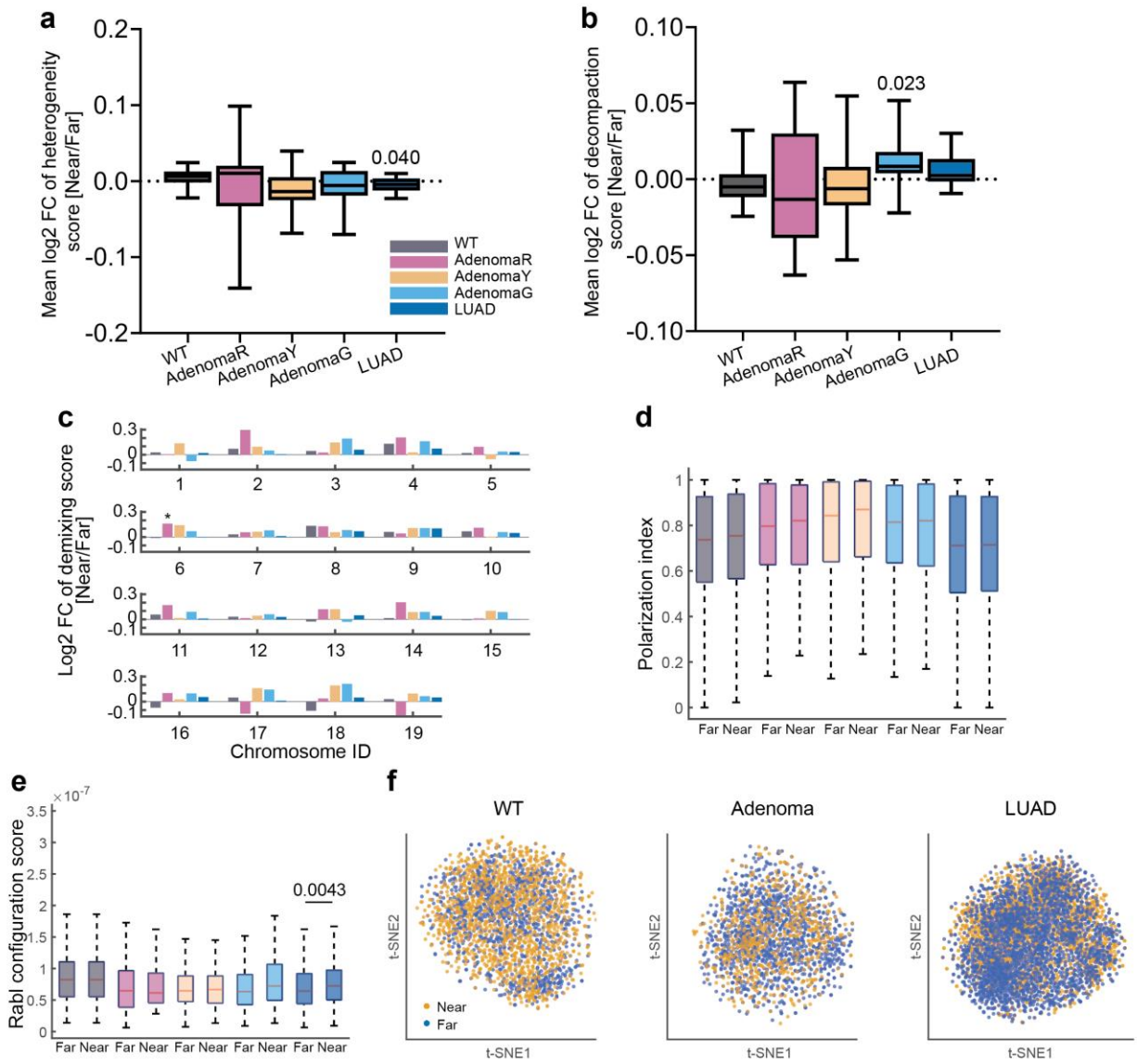

a
AT2
A compartment loci
B compartment loci
b
Adenoma
c
LUAD
A compartment loci
B compartment loci
d
e
f
Probability
Normalized proximity frequency
A-A
B-B
A-B
g
h
i
p < 0.0001
p < 0.0001
Normalized proximity frequency
A-A
B-B
A-B
j
k
l
Proximity frequency
Genomic distance (bp) ×10^8
m
AT2
R = -0.56, p = 4.32 x 10^-41
AdenomaR
R = -0.59, p = 3.50 x 10^-46
AdenomaY
R = -0.59, p = 4.08 x 10^-45
AdenomaG
R = -0.54, p = 8.53 x 10^-38
LUAD
R = -0.49, p = 1.11 x 10^-29
Radial score
A and B compartment score

Extended Data Fig. 7. Changes of cancer-state-independent 3D genome features during lung cancer progression. (**a, b, c**) Normalized trans-chromosomal proximity frequency between genomic loci in AT2 (**a**), adenoma (**b**), and LUAD (**c**) cells. The proximity frequency between each pair of trans-chromosomal genomic regions were normalized to the mean proximity frequency of all loci pairs of the two corresponding chromosomes. A cutoff distance of 800 nm was used for defining proximity. The genomic regions were re-ordered so that A compartment loci were grouped separately from B compartment loci. (**d, e, f**) Distribution of normalized trans-chromosomal proximity frequencies of pairs of A loci (A-A), pairs of B loci (B-B) and pairs of A and B loci (A-B) in AT2 (**d**), adenoma (**e**) and LUAD (**f**) cells. (**g, h, i**) Normalized trans-chromosomal proximity frequencies of A-A, B-B and A-B loci pairs in AT2 (**g**), adenoma (**h**) and LUAD (**i**) cells. In **g-i**, the horizontal lines of each box from top to bottom represent the 75th percentile, the median and the 25th percentile. Whiskers extend to the non-outlier maximum and non-outlier minimum. Outliers are defined as values at least 1.5 times interquartile range away from the top or bottom of the box. p values from two-sided Wilcoxon rank-sum test are shown. (**j-l**) The proximity frequency between each pair of *cis*-chromosomal genomic regions as a function of their genomic distances in AT2 (**j**), adenoma (**k**), and LUAD (**l**) cells. **m,** Radial scores versus A-B compartment scores of genomic loci in the five cell states. The lines are fitted linear regression lines. Correlation coefficients (R) and p values of the correlations are shown. Cell numbers for all panels are the same as in **Fig.2**.



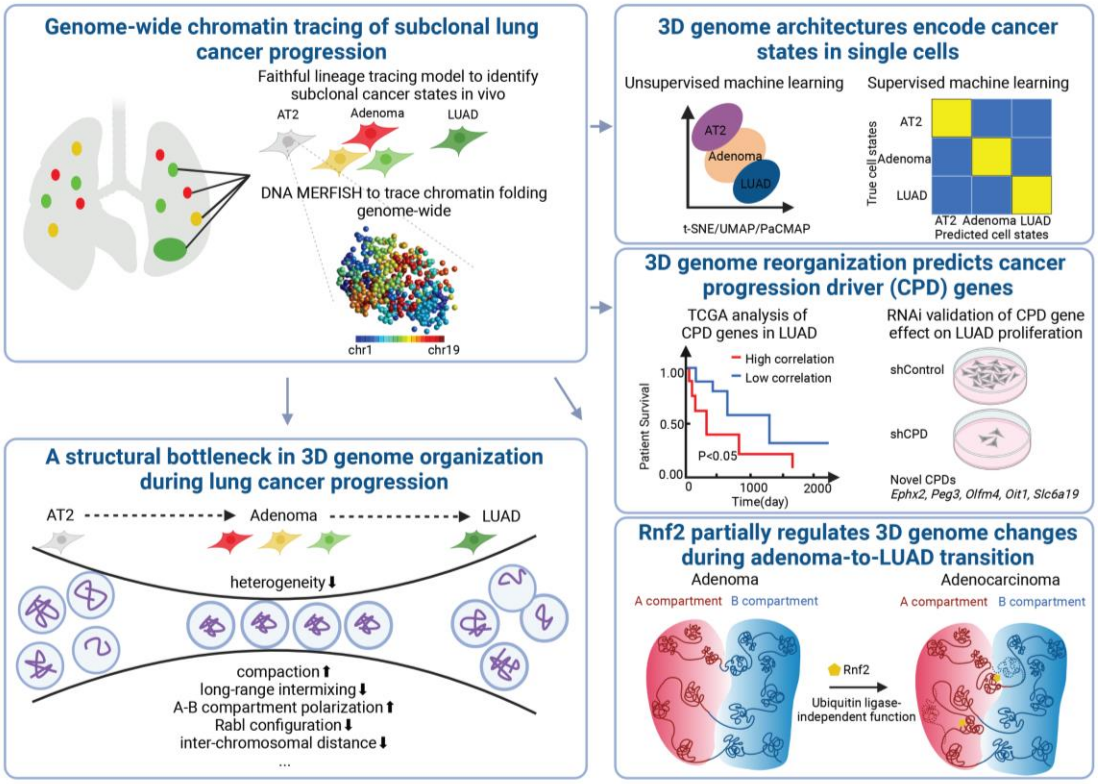
Extended Data Fig. 8. Subsampling analysis of chromatin folding changes during cancer progression. a-d, Mean $\log_2$ fold change of heterogeneity (**a**), decompaction (**b**), demixing (**c**), and radial (**d**) scores of each chromosome, comparing each cancer state to the WT AT2 cell state in a randomly subsampled population of 250 cells per cell state. * indicates FDR < 0.05 of either two-sided Wilcoxon signed-rank tests (**a, b, d**) or two-sided Levene's test (**c**). **e,** Distribution of

polarization indices of A and B compartments in individual chromosome copies in the five cell states with a randomly subsampled population of 250 cells per cell state. p values of two-sided Wilcoxon rank-sum tests are shown. **f,** Distribution of Rabl configuration scores of single cells in the five cell states with a randomly subsampled population of 250 cells per cell state. p values of two-sided Wilcoxon rank-sum tests are shown. In all panels, if a cell state has fewer than 250 cells, all cells of that state were analyzed. In **e** and **f**, the horizontal lines of each box from top to bottom represent the 75th percentile, median, and 25th percentile. Whiskers extend to the non-outlier maximum and non-outlier minimum. Outliers are defined as values at least 1.5 times interquartile range away from the top or bottom of the box.



Extended Data Fig. 9. 3D genome organization in cancer cells is largely independent of spatial proximity to immune cells. a-c, Mean log2 fold change of heterogeneity (**a**), decompaction (**b**), and demixing (**c**) scores of each chromosome, comparing AT2/cancer cells near (less than 5 μm) versus far (more than 5 μm) from CD45+ immune cells. The horizontal lines of each box from top to bottom represent the 75th percentile, median, and 25th percentile. Whiskers extend to the maximum and minimum values. p values (**a, b**) or FDR (**c**) with significance (p/FDR < 0.05) from two-sided Wilcoxon signed-rank test (**a, b**) or two-sided

Levene's test (**c**) are displayed. **d,** Distribution of polarization indices of A-B compartments in AT2/cancer cells near or far from immune cells. Two-sided Wilcoxon rank-sum test yielded no significant p values ($p < 0.05$). **e,** Distribution of Rabl configuration scores for AT2/cancer cells near or far from immune cells. p value with significance ($p < 0.05$) from two-sided Wilcoxon rank-sum test is displayed. In **d** and **e**, the horizontal lines of each box from top to bottom represent the 75th percentile, median, and 25th percentile. Whiskers extend to the non-outlier maximum and non-outlier minimum. Outliers are defined as values at least 1.5 times interquartile range away from the top or bottom of the box. **f,** t-SNE plots of single-cell 3D chromatin conformations in AT2, adenoma, and LUAD cells show no distinct clusters based on spatial proximity to immune cells.



Extended Data Fig. 10. Schematic illustration of the experimental approach and major findings. During lung cancer progression from AT2 to preinvasive adenoma to LUAD, a 3D genomic structural bottleneck exists at the preinvasive adenoma stage, indicating an early stringent selection of the 3D genome conformation. The structural bottleneck manifests in 3D genome conformation heterogeneity, compaction, long-range intermixing, A-B compartment polarization, Rabl configuration, etc. Single-cell 3D genome organization encodes different cancer cell states and predicts cancer progression drivers and genetic vulnerabilities. Finally, Rnf2 – through a ubiquitin ligase-independent function – partially reorganizes A-B compartmentalization during the adenoma-to-LUAD transition. The cartoons were created with BioRender.com.

Captions for Supplementary Tables

Supplementary Table 1. Template oligonucleotide probe libraries for genome-wide chromatin tracing

Supplementary Table 2. Codebook for all target genomic loci for genome-wide chromatin tracing.

The first column indicates which chromosome each target TAD is located in. The second and third column indicate the start coordinate and end coordinate of the target TAD. The fourth column indicates the HW2 binary barcode for each TAD. Each barcode consists of 100 bits, with 2 bits assigned as '1' and all the other 98 bits as '0'.

Supplementary Table 3. Oligo sequences for adapters, common readout probes, blocking oligos and primers.

The file contains four spreadsheets – 'Adapters', 'Common readout oligos', 'Blocking oligos' and 'Primers'. 'Adapters' contains the sequences of all 100 oligos that can bind to the primary probes and correspond to the 100 bits in the codebook. 'Common readout oligos' contains the sequences of dye-conjugated readout probes. 'Blocking oligos' contains the dye-free oligos that have the same sequences as the common readout oligos to block any unbound sites of the previous adapters. 'Primers' contains the forward and reverse primer sequences for the template probe library amplification and primary probe synthesis.

Supplementary Table 4. RNA-seq analysis of lung tumors from *K-MADM-Trp53* mice.

Differential expression analysis (DESeq2) of large green tumors derived from *K-MADM-Trp53* mice and classified as AdenomaG and LUAD based on marker gene expression. For each gene, $\log_2$ normalized expression counts (DESeq2) for each sample, average expression counts across all samples, $\log_2$ fold change (LUAD vs. AdenomaG), p-value, and FDR (padj) are shown.

Supplementary Table 5. List of candidate progression driver, candidate tumor suppressor and candidate tumor initiation genes.

Lists of candidate progression drivers, candidate tumor suppressors, and candidate tumor initiation genes. References are listed for genes with prior literature evidence for a functional role in tumor initiation, progression, or suppression.

Supplementary Table 6. shRNA sequences for lentiviral transduced stable cell lines to knock down the candidate genes.

Sequences for all the shRNAs for the candidate genes, positive controls, and negative controls are included. The first column contains the gene names. The second column contains the shRNA sequences. The third column contains the hairpin numbers.