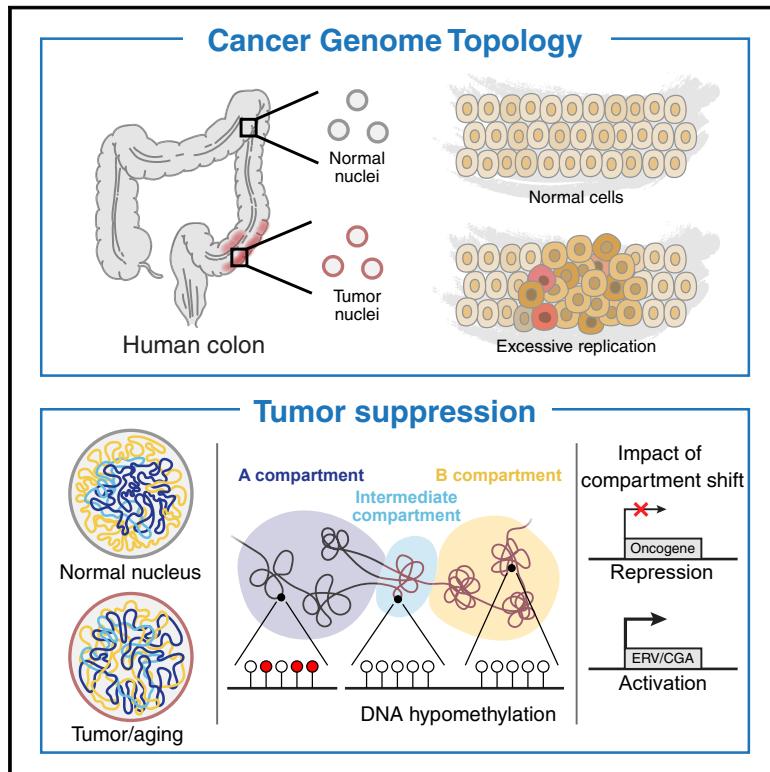


# Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer

## Graphical Abstract



## Authors

Sarah E. Johnstone, Alejandro Reyes, Yifeng Qi, ..., Bin Zhang, Martin J. Aryee, Bradley E. Bernstein

## Correspondence

aryee.martin@mgh.harvard.edu (M.J.A.), bernstein.bradley@mgh.harvard.edu (B.E.B.)

## In Brief

Integrated analyses of genome topological, epigenetic, and transcriptional features of colorectal tumors highlight substantial genome compartmental reorganization associated with tumor-suppressive rather than oncogenic transcriptional outcomes.

## Highlights

- Hierarchical layers of nuclear architecture are altered in colorectal tumors
- An intermediate genome compartment is defined in primary tissues
- Compartmental reorganization and hypomethylation occur in tumors and aging cells
- Reorganization is associated with tumor-suppressive transcriptional programs

Article

# Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer

Sarah E. Johnstone,<sup>1,2,3,10</sup> Alejandro Reyes,<sup>2,4,5,10</sup> Yifeng Qi,<sup>2,6</sup> Carmen Adriaens,<sup>1,2,3</sup> Esmat Hegazi,<sup>1,2,3</sup> Karin Pelka,<sup>2,3</sup> Jonathan H. Chen,<sup>1,2,3</sup> Luli S. Zou,<sup>2,4,5</sup> Yotam Drier,<sup>7</sup> Vivian Hecht,<sup>2</sup> Noam Shores,<sup>2</sup> Martin K. Selig,<sup>1</sup> Caleb A. Lareau,<sup>1,2,8</sup> Sowmya Iyer,<sup>1</sup> Son C. Nguyen,<sup>9</sup> Eric F. Joyce,<sup>9</sup> Nir Hacohen,<sup>2,3</sup> Rafael A. Irizarry,<sup>2,4,5</sup> Bin Zhang,<sup>2,6</sup> Martin J. Aryee,<sup>1,2,3,5,\*</sup> and Bradley E. Bernstein<sup>1,2,3,11,\*</sup>

<sup>1</sup>Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

<sup>2</sup>Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA

<sup>3</sup>Center for Cancer Research, Massachusetts General Hospital, Boston, MA 02129, USA

<sup>4</sup>Department of Data Sciences, Dana Farber Cancer Institute, Boston, MA 02215, USA

<sup>5</sup>Department of Biostatistics, Harvard School of Public Health, Boston, MA 02215, USA

<sup>6</sup>Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>7</sup>The Lautenberg Center for Immunology and Cancer Research, The Hebrew University, Jerusalem, Israel

<sup>8</sup>Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA 02215, USA

<sup>9</sup>Department of Genetics, Penn Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>10</sup>These authors contributed equally

<sup>11</sup>Lead Contact

\*Correspondence: [aryee.martin@mgh.harvard.edu](mailto:aryee.martin@mgh.harvard.edu) (M.J.A.), [bernstein.bradley@mgh.harvard.edu](mailto:bernstein.bradley@mgh.harvard.edu) (B.E.B.)

<https://doi.org/10.1016/j.cell.2020.07.030>

## SUMMARY

Widespread changes to DNA methylation and chromatin are well documented in cancer, but the fate of higher-order chromosomal structure remains obscure. Here we integrated topological maps for colon tumors and normal colons with epigenetic, transcriptional, and imaging data to characterize alterations to chromatin loops, topologically associated domains, and large-scale compartments. We found that spatial partitioning of the open and closed genome compartments is profoundly compromised in tumors. This reorganization is accompanied by compartment-specific hypomethylation and chromatin changes. Additionally, we identify a compartment at the interface between the canonical A and B compartments that is reorganized in tumors. Remarkably, similar shifts were evident in non-malignant cells that have accumulated excess divisions. Our analyses suggest that these topological changes repress stemness and invasion programs while inducing anti-tumor immunity genes and may therefore restrain malignant progression. Our findings call into question the conventional view that tumor-associated epigenomic alterations are primarily oncogenic.

## INTRODUCTION

For over a century, pathologists have observed changes in the shape, size, and chromatin texture of cancer cell nuclei (Zink et al., 2004). Nuclear features help determine cancer subtype and grade, with significant effects on prognosis and therapy. However, despite their diagnostic and clinical importance, the molecular underpinnings of these cancer-associated morphological changes remain poorly understood.

Molecular and genetic studies have documented widespread epigenetic defects in human tumors, including chromatin regulator mutations, DNA methylation changes, and altered enhancer landscapes (Baylin and Jones, 2016). Certain cancers also harbor mutations of proteins that regulate higher-order chromosomal structure (“genome topology”), including CTCF and cohesin subunits (Corces and Corces, 2016). Focal topological alterations drive oncogenic transcriptional programs in specific contexts (Flavahan et al., 2016, 2019; Hnisz et al., 2016; Klotgen et al., 2020). Despite this, genome topol-

ogy and nuclear organization in human tumors are largely uncharted.

Technological innovations, including Hi-C (Rao et al., 2014), have revealed hierarchical layers of spatial organization (Bickmore and van Steensel, 2013; Dekker and Misteli, 2015; Rowley and Corces, 2018). Chromatin loops occur when distant loci on the linear chromosome are in frequent contact and typically involve interactions between CTCF-bound sites (CTCF loops) and/or between enhancers and promoters (E-P loops). Topologically associating domains (TADs) are sub-megabase regions partitioned by boundary elements often bound by CTCF and cohesin. Finally, the genome is grossly partitioned into two large-scale compartments: an open, transcriptionally active A compartment and a compact, relatively silent B compartment.

Colorectal adenocarcinoma, the fourth most common epithelial tumor, is a well-characterized model for cancer epigenetics. These tumors exhibit profoundly altered DNA methylation landscapes, including loss of methylation from large genomic “blocks” that cover over half of the genome (Berman et al.,

2011; Hansen et al., 2011). The functional implications of hypomethylation remain obscure, but it may impact genome stability and/or transcriptional activity (Baylin and Jones, 2016). Conversely, CpG island hypermethylation occurs in a subset of colon tumors, termed the CpG island methylator phenotype (CIMP), and is associated with promoter silencing (Hinoue et al., 2012; Toyota et al., 1999). However, our understanding of these epigenomic changes and their relationship to genome topology has been hindered by a lack of data for primary tumors.

Here we mapped genome topology across a cohort of colon tumors, normal colons, and colon cancer cell lines and compared successive layers of topology. We focus on large-scale reorganization of the conventional genome compartments A and B and characterize an intermediate compartment at their interface. Remarkably, compartmental reorganization is associated with repression of stem cell, invasion, and metastasis genes and induction of genes associated with anti-tumor immunity. Our results suggest that the most profound topological alterations in tumors are actually a consequence of accumulated cell divisions and that they may have a tumor-suppressive role.

## RESULTS

### Maps of DNA Methylation, Chromatin State, and Topology in Human Tumors

To understand how nuclear architecture is altered in cancer, we profiled genome topology along with DNA methylation, chromatin modifications, and CTCF in primary colon tumors, normal colons, and colon cancer cell lines (Figure 1A). Our clinical cohort included 26 tumors and 7 normal colon tissue samples (Table S1). Our *in vitro* models included colon cancer cell lines (HCT116, SW480, RKO, and LS-174T), a line derived from normal colon (FHC), and primary fibroblasts (WI-38). Our full dataset comprises 175 libraries and 28 billion sequencing reads for Hi-C, ChIP-seq, bisulfite sequencing, chromatin immunoprecipitation sequencing (ChIP-seq) and RNA sequencing (RNA-seq).

We performed hybrid-capture bisulfite sequencing on 26 tumors, 3 normal colons, and 5 cell lines. Although only two tumors had CpG island hypermethylation (CIMP) (Figure S1A), all tumors exhibited degrees of hypomethylation across expansive genomic regions, termed “hypomethylated blocks” (Figure S1B; Berman et al., 2011; Hansen et al., 2011). We also inferred copy number variants (CNVs) for each tumor (Table S2) and controlled for CNV-related variability in further analyses by incorporating terms for copy number estimates into our linear models and verifying results in CNV-stable regions and tumors (STAR Methods).

We integrated high-resolution topological maps and epigenomics data to investigate successive layers of genome organization, from chromatin loops to TADs to large-scale genome compartments, in tissues, tumors, and cell lines.

### E-P Loops Are Associated with Oncogenic Transcriptional Programs

We began by identifying loops that could influence transcriptional states in tumors. HiChIP assays targeting the cohesin subunit SMC1 (Mumbach et al., 2016) reveal CTCF-CTCF loops, which contribute to TAD boundaries, as well as E-P loops, which are hypothesized to mediate enhancer gene activation (Stad-

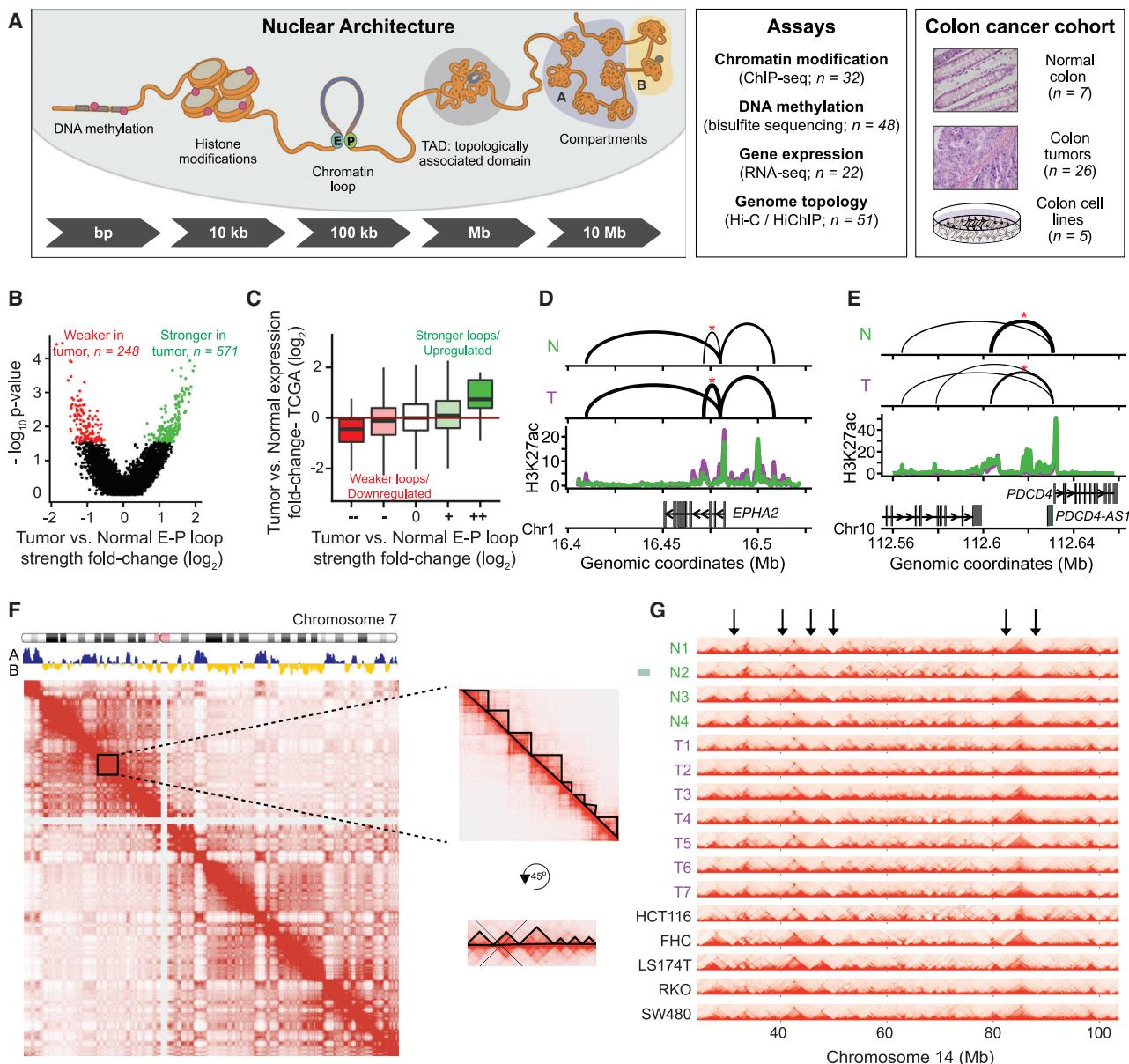
houders et al., 2019). We identified 25,125 loops in normal colon or tumors and annotated the subset that connects an histone H3 lysine 27 acetylation (H3K27ac) peak (enhancer-like) to an annotated promoter as E-P loops ( $n = 14,121$ ). Differential analysis revealed 571 E-P loops that are stronger in tumors and 248 that are weaker in tumors (Figure 1B). To relate these differential loops to transcription, we evaluated the expression of the corresponding genes in our cohort and across 521 samples from The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Network, 2012). Genes connected to E-P loops that were stronger in tumors were upregulated in tumors, whereas genes connected to E-P loops that were weaker in tumors were downregulated (Figures 1C and S2A; Table S3). This association was evident even when excluding loci subject to CNVs (Figure S2B).

Several topological alterations involved known oncogenes or tumor suppressors. For example, although the locus encoding the receptor tyrosine kinase oncogene *EPHA2* (Dunne et al., 2016) contains multiple enhancers in normal colon and tumors, the gene is connected to a strong enhancer by a tumor-associated E-P loop (Figures 1D and S2C). Accordingly, *EPHA2* is upregulated in colon tumors (Figures S2D and S2E). Conversely, the *PDCD4* tumor suppressor (Wang et al., 2017) loses an E-P loop to a distal enhancer and is downregulated in tumors (Figures 1E and S2F-S2H). The annotated E-P loops can also facilitate interpretation of single-nucleotide polymorphisms (SNPs) associated with colon cancer risk (Figures S2I-S2K; Table S4; STAR Methods).

### Topological Boundaries Are Largely Retained in Tumors

A next layer of topological organization involves TADs and their boundaries. TADs are evident in Hi-C maps as sub-megabase-scale regions with increased intra-domain interactions (Figure 1F; Dixon et al., 2015). CTCF occupies and contributes to the stability of many TAD boundaries (Rao et al., 2014; Nora et al., 2017). We used Hi-C data to assess the locations and integrity of TAD boundaries genome wide (STAR Methods). Boundary location and strength were largely concordant between tumors, normal controls, and cell lines (Figures 1G and S3A-S3D), consistent with previous studies of cell lines and non-malignant tissues (Dixon et al., 2012; Kretford et al., 2018; Nora et al., 2012; Schmitt et al., 2016). Tumors, on average, shared 92% of TAD boundaries with normal colon and 89% with the cell lines. Visual inspection revealed that discordant boundary calls were most often caused by subtle differences in strength rather than complete boundary loss.

Prior studies have shown that CTCF boundaries may be disrupted by genetic deletion or hypermethylation (Flavahan et al., 2016, 2019; Hnisz et al., 2016; Modrek et al., 2017). Consistently, we identified more than 100 TAD boundaries that gain DNA methylation and lose CTCF binding in our hypermethylated tumors (Figures S3E and S3F). The integrity of these boundaries was compromised, as evidenced by weaker “peaks” on the Hi-C contact maps (Figure S3G) and more frequent cross-boundary E-P interactions in HiChIP data (Figure S3H). However, the transcriptional consequences of these boundary losses appeared to be relatively limited (Figure S3I).



**Figure 1. Integrated Topological Maps Reveal Tumor-Specific Chromatin Loops and Stable TAD Structure**

(A) Schematic of hierarchical genome organization with indication of genomic scale (left) and summary of genome-wide assays (center) and models (right).

(B) Volcano plot presenting a differential analysis of loops between tumors and normal samples. Loops, represented as dots, with significantly stronger or weaker interactions in tumors compared with normal colon are highlighted in red and green, respectively.

(C) Boxplots depicting expression fold change ( $\log_2$ ) between tumors and normal samples (y axis) for genes engaged in enhancer-promoter (E-P) loops. Genes are stratified by change in E-P loop strength between tumors and normal colon (x axis).

(D) Genomic view of the *EPHA2* locus (~130 kb), showing SMC1 HiChIP loops (arcs) and H3K27ac enrichment for normal colon (green) and colon tumor (purple). The width of the arcs corresponds to the average loop strength summarized for the set of 2 normal and 7 colon samples. An asterisk indicates the differential loop (STAR Methods).

(E) Genomic view of the *PDCD4* locus (~100 kb) as in (D).

(F) Hi-C contact map showing pairwise contact frequencies (red heat) between genomic positions across chromosome 7 (rows, columns) in normal colon. Top: Hi-C eigenvector (PC1) based on long-range interactions demarcates compartments A (positive values, blue) and B (negative values, yellow). Right: inset with a magnified view of a representative region reveals TAD structures (highlighted by black triangles). Rotation of this inset by 45° yields a horizontal display of TAD structures (see G).

(G) Horizontal heatmaps showing local Hi-C contact patterns (red heat) across chromosome 14 for normal colon (green), colon tumors (purple), and cell lines (black). Exemplar TAD boundaries are indicated by black arrows.

In summary, topological boundaries were largely conserved across colon tumors, normal colons, and cell lines, with the exception of a relatively small set of boundaries compromised in hypermethylated tumors.

### Megabase-Scale Compartment Structure Is Reorganized in Tumors

The genome is partitioned into open A and closed B spatial compartments. TADs in the same compartment have a greater tendency to self-interact, whereas inter-compartmental interactions are disfavored (Rao et al., 2014), resulting in the characteristic checkerboard pattern of Hi-C maps (Figure 1F). We used a standard eigenvector-based method to assign compartments for each of our Hi-C datasets. In contrast to the striking conservation of TAD boundaries, compartment assignments varied between samples. For example, most cell lines were distinct from normal tissues and tumors (Figures 2A, 2B, S4A, and S4B).

Comparisons also revealed widespread differences between colon tumors and normal colon (Figure 2B). To understand these differences, we directly compared Hi-C interaction matrices. Although compartment assignments were mostly concordant between normal colon and tumors (Figure 2A), long-range interactions between compartments A and B were more frequent in tumors (Figure 2C). These differential interaction patterns were evident regardless of whether the underlying compartment assignments were derived from normal colon or tumor Hi-C data (Figure S4C) and translated to a genome-wide increase in inter-compartment interaction in tumors ( $p < 0.005$ ) (Figure S4D).

Prior studies have related compartments with nuclear positioning (Falk et al., 2019; Wang et al., 2016). Compartment B correlates with lamina-associated domains (LADs), which are located at the nuclear periphery and may be damaged in aging-related disease, senescence, and cancer (Sakthivel and Sehgal, 2016; Schreiber and Kennedy, 2013; van Steensel and Belmont, 2017). Conversely, active loci tend to localize to the nuclear interior. In principle, relative nuclear positioning of genomic loci can be inferred from Hi-C data. We therefore used a maximum entropy approach to derive topological models for normal colon and tumor nuclei (STAR Methods). Our method models the genome at 1-Mb resolution as a 3-dimensional polymer, taking into account linear constraints inherent to the DNA. We then compute an *in silico* Hi-C map for each polymer model and repeat the process iteratively until the computed map converges on the experimental Hi-C data (Figure 2D; STAR Methods).

Polymer models optimized to the normal colon Hi-C data separated compartments A and B and positioned compartment B peripherally (Figures 2E, S4E, and S4F), consistent with expectation. However, applying the same modeling approach to the colon tumor Hi-C data yielded a strikingly different result (Figures 2F and S4G). In these models, both compartments were distributed heterogeneously throughout the nucleus. We verified that these changes were not solely driven by genetic alterations by restricting analysis to stable chromosomes and a genetically stable tumor (Figures S4H and S4I). These results suggest that the asymmetric radial positioning of compartments A and B is profoundly altered in tumors.

To further investigate, we directly visualized chromatin and specific genomic loci. First, we imaged specimens by transmission electron microscopy, which revealed characteristic epithelial structures and organization. In epithelial nuclei from normal colon, electron-dense heterochromatin was juxtaposed to the nuclear membrane, consistent with peripheral lamina association, whereas characteristic light-staining euchromatin was visible throughout the nuclear interior (Figures 2G and S4K). In marked contrast, epithelial tumor nuclei had large dark-staining heterochromatin foci dispersed throughout their interiors (Figures 2H, 2I, and S4K; STAR Methods).

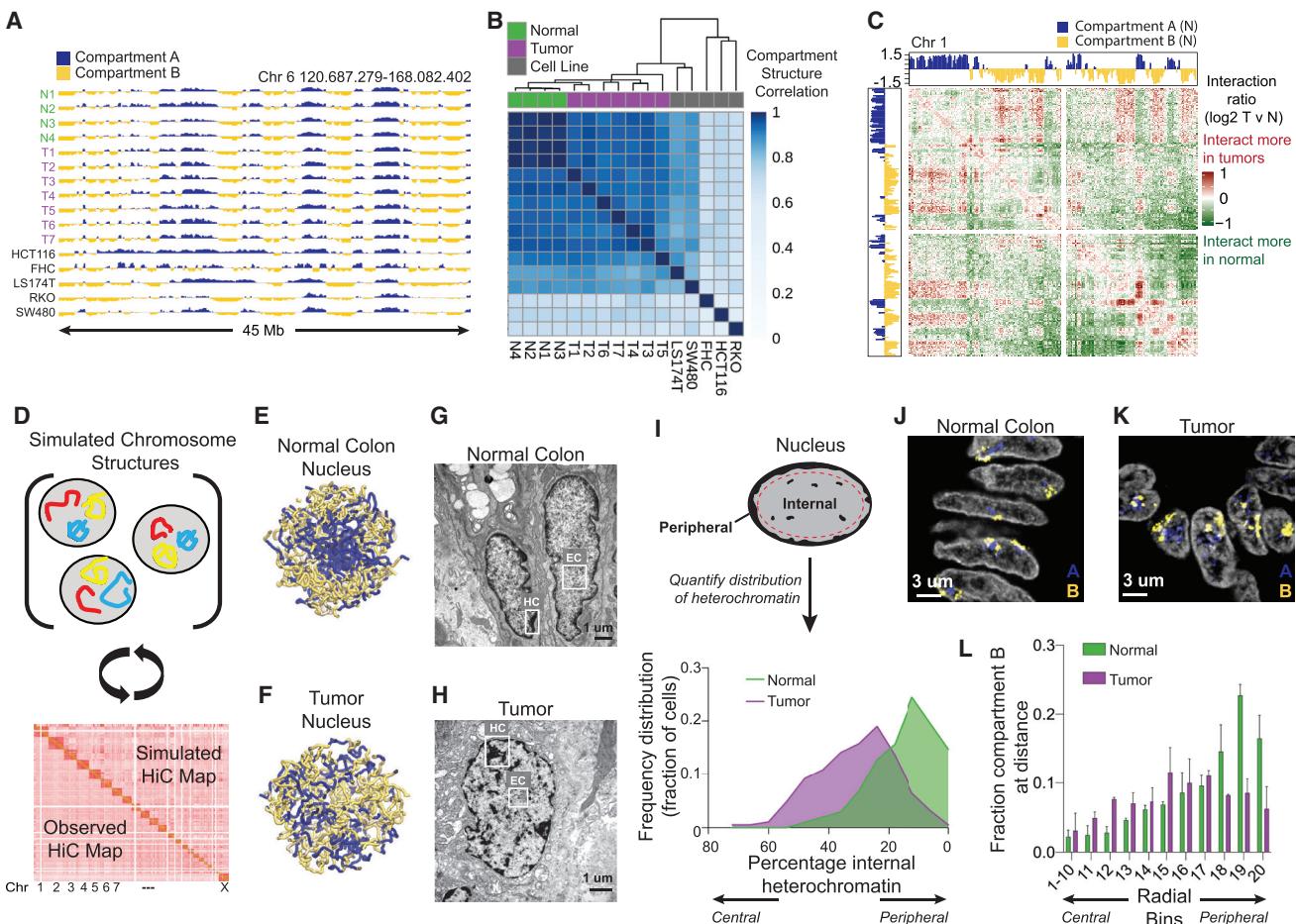
We next evaluated the positioning of specific genomic regions using DNA fluorescence *in situ* hybridization (FISH). We designed ~26,000 Oligopaint probes targeting loci on chromosome 12 that were assigned to compartment A or B according to Hi-C (Figure S4J; STAR Methods; Beliveau et al., 2012). We then labeled the respective compartments with secondary fluorescent probes and visualized them by Airyscan confocal imaging. Chromosome 12 territories were generally positioned peripherally, consistent with prior studies (Bolzer et al., 2005). To quantify radial distributions of the A and B compartments, we scored probe signals according to their intensity in 20 radial bins starting from the nuclear center in 47 nuclei from 2 normal colon samples and 82 nuclei from 2 tumors. Although the localization of fluorescence signals varied, compartment B signals were strongly skewed toward the periphery of normal colon nuclei, whereas compartment A signals were more evenly distributed (Figures 2J and 2L). In contrast, in tumor nuclei, compartment B signals lost their peripheral skew and assumed a distribution similar to compartment A (Figures 2K and 2L).

Hence, concordant analyses based on Hi-C polymer models, electron microscopy, and multi-color FISH imaging indicate profound compartmental reorganization in tumor nuclei. Spatial partitioning between compartments is compromised. Compartment B relocates from its physiologic peripheral position toward the nuclear interior.

### A Genome Compartment with Intermediate Properties

Despite widespread differences in compartmental interactions, A/B assignments were relatively consistent between tumor and normal colon. This prompted us to more closely examine the Hi-C eigenvector, which is a continuous rather than a dichotomous measure. We observed hundreds of large genomic intervals, hundreds of kilobases in size, with eigenvector values that were lower in tumors than in normal colon (Figures 3A and 3B). The majority of these intervals were assigned to compartment A in tumor and normal colon because they had positive eigenvectors. However, quantitative analysis indicated that these regions shifted their interactions toward compartment B in tumors.

When we examined the transitioning regions, we observed that they exhibited a striking loss of DNA methylation in tumors and, in fact, largely coincided with hypomethylated blocks (Figures 3A, 3B, and S5A). This was unexpected because hypomethylated blocks have been primarily associated only with compartment B (Berman et al., 2011; Fortin and Hansen, 2015; Hansen et al., 2011). We found that hypomethylated blocks, which tend to span single or consecutive TADs, covered a full

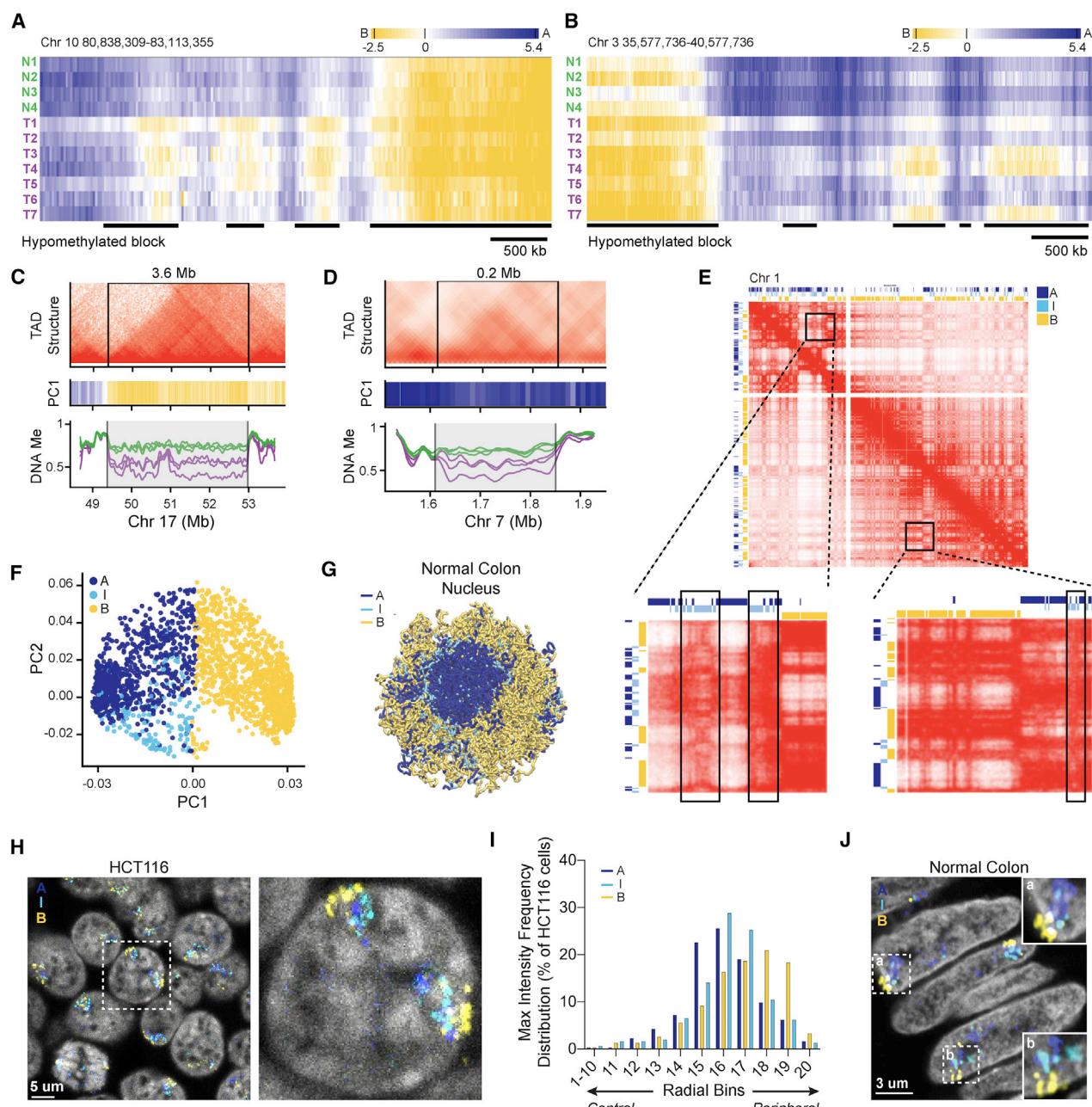


**Figure 2. Compromised Partitioning and Positioning of Genome Compartments in Tumors**

- (A) Hi-C eigenvectors (PC1) based on long-range interactions demarcate compartments A (positive values, blue) and B (negative values, yellow) across a 45-Mb region of chromosome 6. Data show eigenvectors for normal colon (green), colon tumors (purple), and cell lines (black).
- (B) Heatmap showing pairwise correlations between the first Hi-C eigenvector (blue heat) in normal colon (green), colon tumors (purple), and cell lines (gray). Samples (rows, columns) are ordered according to complete linkage hierarchical clustering (top).
- (C) Heatmap showing fold change ( $\log_2$ ) in Hi-C contact frequencies between colon tumors and normal colon across chromosome 1. Data are based on an average of normal colons ( $n = 4$ ) and tumors ( $n = 7$ ). Interactions that increase in tumors (red) or decrease in tumors (green) are evident. Top left: the Hi-C eigenvector indicates compartment assignments in normal colon (A, blue; B, yellow).
- (D) Schematic of the maximum entropy modeling approach, in which structural models of genome organization are generated and iteratively refined to improve the correlation between Hi-C maps derived from these models *in silico* and the actual experimental Hi-C data.
- (E) Whole-nucleus maximum entropy models (1-Mb resolution) for a representative normal colon sample (N1), showing compartment A in blue and compartment B in yellow.
- (F) Whole-nucleus maximum entropy models (1-Mb resolution) for a representative colon tumor sample (T1) as in (E).
- (G) Representative transmission electron microscopy (TEM) image of nuclei from normal colon epithelium, showing electron dense heterochromatin (HC) along the nuclear membrane and internally distributed euchromatin (EC). Scale bar, 1  $\mu\text{m}$ .
- (H) Representative EM image of colon tumor nucleus labeled as in panel (G). Scale bar, 1  $\mu\text{m}$ .
- (I) Top: Schematic of strategy for quantifying the internal HC in nuclei imaged by EM. Bottom: Histogram shows fraction of nuclei (y axis) that have indicated percentage of internal HC. Normal (green) reflects 102 epithelial cell nuclei from 3 normal colon specimens. Tumor (purple) reflects 184 malignant cell nuclei from 3 colon tumors. Two-sided nested t test,  $p = 0.006$ .
- (J and K) Representative image of chromosome 12 DNA FISH in nuclei from normal colon epithelial cells (J) and malignant tumor cells (K), labeling compartment A and B regions in blue and yellow, respectively.
- (L) Histogram showing a fraction of the signal from compartment B DNA FISH probes in successive radial bins from the center to the periphery of the nuclei. Normal (green) reflects 47 epithelial cell nuclei from 2 normal colon specimens. Tumor (purple) reflects 82 malignant cell nuclei from two tumors for which chromosome 12 was copy number stable (T1 and T4). Innermost bins 1–10 were aggregated.

19% of compartment A (Figures 3C, 3D, S5B, and S5C; mean methylation difference,  $>10\%$ ;  $n = 1,032$ ; mean size, 217 kb). Examination of Hi-C contact maps revealed that these noncanoni-

cal regions do not fit the typical checkerboard pattern that arises from long-range compartmental interactions (Figure 3E). Rather, they have a distinct contact pattern characterized by intermediate



**Figure 3. An Intermediate Compartment I Is Also Reorganized in Tumors**

(A) Heatmap depicting the Hi-C eigenvector used to define compartments, as in Figure 2A, except with the eigenvector values shown as heatmaps (blue, positive; yellow, negative). Data are shown for a ~2.3-Mb region (x axis) for normal colon samples (rows with green labels) and colon tumors (rows with purple labels). Hypomethylated blocks are indicated (black bars).

(B) Heatmap depicting the first eigenvector for a 5-Mb region on chromosome 3 as in (A).

(C) Hi-C contact map (top) and first eigenvector (PC1; center) for a representative tumor sample. DNA methylation levels of low-density (open-sea) CpGs (bottom) are shown for three representative normal colon samples (green) and three representative tumors (purple). The represented region on chromosome 17 contains a 3.6-Mb TAD assigned to compartment B (black square) with a hypomethylated block (gray highlight).

(D) Hi-C contact map, first eigenvector (PC1) and DNA methylation are shown for a 200 kb TAD assigned to compartment A with a hypomethylated block (data presented as in panel (C)).

(E) Hi-C contact map for chromosome 1 for a representative normal colon sample. Top left: colored bars indicate genomic regions assigned to compartment A (blue), compartment I (light blue), or compartment B (yellow). Magnified panels for representative regions highlight the intermediate long-range inter-compartmental interaction pattern typical of compartment I.

(F) First and second eigenvectors of the chromosome 1 Hi-C matrix for normal colon. Each point represents one 100-kb bin, colored by compartment.

(legend continued on next page)

interactions with both conventional compartments (Figures 3E) and preferential self-interactions (Figure S5D). This distinct contact pattern was evident in normal (Figures S5E and S5F) and tumor samples (Figures S5G and S5H).

We considered that these hypomethylated A blocks might reflect an intermediate compartment “I” that interacts with both canonical compartments at baseline and shifts toward B in tumors (Figure S5I). In support, we found that compartment I regions could be distinguished on multiple chromosomes by examining additional eigenvectors of the Hi-C matrix (Figures 3F and S5J–S5L; STAR Methods). Although the order of the declarative eigenvector varied between chromosomes, this suggested that compartment I can be distinguished from structural data alone. Furthermore, our polymer models for normal colon placed compartment I in an intermediate nuclear position between compartments A and B (Figure 3G).

To investigate further, we visualized compartment I regions by multi-color FISH. We designed a third set of ~14,500 oligonucleotide probes complementary to compartment I regions on chromosome 12 and a corresponding secondary probe with a distinct fluorophore (Figure S4J). We then used three-color FISH imaging to simultaneously localize compartment A, B, and I regions in HCT116 colon cancer cells. We selected HCT116 cells because chromosome 12 is copy number stable, and the loci targeted by our probes had similar compartment assignments as our primary tissues (Table S2; Figure S5M). We visualized and quantified the radial positioning of each compartment in 305 HCT116 nuclei. We found that compartment I is spatially intermediate between the more peripheral compartment B and the more internal compartment A (Figures 3H and 3I). We confirmed this observation in primary tissues by quantifying fluorescence FISH signals in normal colon epithelial cell nuclei (Figures 3J and S5N).

Thus, a convergence of Hi-C, methylation, polymer modeling, and imaging data support the existence of a third genomic compartment I that interacts with both conventional compartments and adopts an intermediate spatial position in the nucleus. In tumors, compartment I becomes broadly hypomethylated and shifts its interactions toward the B compartment.

### Distinct Chromatin and Transcriptional States Support the Three-Compartment Model

To investigate whether compartment I is associated with distinct histone modifications, we mapped markers of active regulatory elements (H3K27ac), elongating transcripts (histone H3 lysine 36 trimethylation/H3K36me3), constitutive heterochromatin (histone H3 lysine 9 trimethylation/H3K9me3), and facultative (histone H3 lysine 27 trimethylation/H3K27me3) heterochromatin. As expected, H3K27ac and H3K36me3 were enriched in

compartment A (Figures 4A, 4C, and 4D), whereas H3K9me3 was enriched in compartment B and relatively increased in tumors (Figures 4B–4E).

Compartment I was clearly distinguished from both conventional compartments by broad H3K27me3 enrichment (Figures 4A, 4C–4E, and S6A). H3K27me3 signal intensity was particularly pronounced in tumors and correlated with the degree of DNA hypomethylation. H3K36me3, which antagonizes H3K27me3 and has been proposed to protect against DNA hypomethylation, is depleted in compartment I (Figure 4D; Yuan et al., 2011; Zhou et al., 2018). Compartment I was also notable for relatively low transcriptional levels in normal colon and moderate gene density, both of which were in between compartments A and B (Figures 4D and S6B).

Thus, in addition to its topological features, compartment I is distinguished by its facultative heterochromatin state, modest transcriptional output, and methylation changes in tumors.

### Compartmental Changes Linked to DNA

#### Hypomethylation and Accumulated Cell Divisions

Compartments B and I largely correspond to hypomethylated blocks in tumors. Moreover, we observed a striking correlation between the extent of hypomethylation of a given region and its eigenvector: genomic loci with more extreme hypomethylation became relatively more B-like or compact (Figure 5A).

To further assess this relationship between compartmental changes and DNA hypomethylation, we treated HCT116 cells with the demethylating agent 5'-azacytidine (5-aza) for 24 h and measured methylation and topology changes by Hi-C. The treatment reduced methylation of large genomic intervals, with 56% of 100-kb windows losing more than 20% methylation (Figure S6C). Notably, we found that genomic regions with the most significant methylation loss shifted their interactions toward compartment B in a topological reorganization reminiscent of tumors (Figure 5B). This suggested that block hypomethylation may underlie the altered compartment structure in colon tumors.

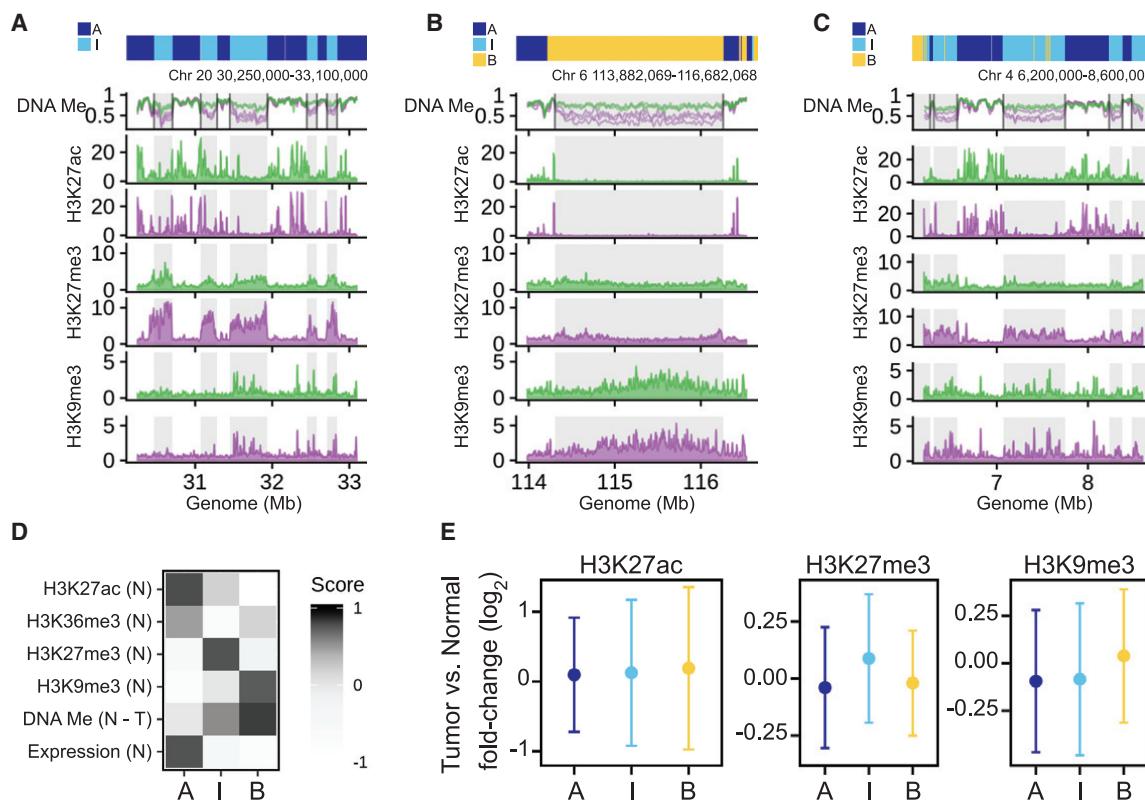
Block hypomethylation was originally described in cancer cells (Berman et al., 2011; Hansen et al., 2011; Nordor et al., 2017) but has since been recognized to be a feature of cells that have accumulated many divisions, including aging and senescent cells (Cruickshanks et al., 2013; Zhou et al., 2018). Genome topology and nuclear structure are also altered in fibroblasts passaged to replicative senescence (Criscione et al., 2016; Sati et al., 2020). We hypothesized that the compartmental shifts in tumor cells might relate to those in passaged fibroblasts, in both cases reflecting excessive replications. To test this, we passaged WI-38 fibroblasts over a 14-week course and generated DNA methylation profiles and Hi-C data for cells harvested at passages 16, 30, and 40 (STAR Methods). Late-passage cells

(G) Whole-nucleus maximum entropy model (100-kb resolution) for normal colon, showing compartments A, I, and B.

(H) Representative DNA FISH image (left) and high-magnification image (right) for HCT116 cell nuclei. Signal intensities are shown for compartment A (blue), I (light blue), and B (yellow) regions on chromosome 12.

(I) Barplot indicating the percentage of cells for which the maximum DNA FISH signal intensity for compartment A, B, or I is located at the indicated radial position for 305 HCT116 cell nuclei.

(J) Representative image of nuclei from normal colon epithelial cells. The image shows DNA FISH signal intensities of probes for compartments A, B, and I of chromosome 12. Two chromosome territories are magnified in the insets.



**Figure 4. Distinct Chromatin States Support a Three-Compartment Model**

(A–C) Plots show 3-compartment model assignments for representative regions on chromosome 20 (A), chromosome 6 (B), and chromosome 4 (C). Dark blue, compartment A; light blue, compartment I; yellow, compartment B. DNA methylation levels of low-density (open-sea) CpGs for three normal samples (green) and three tumor samples (purple) are shown below, along with ChIP-seq profiles for H3K27ac, H3K27me3, and H3K9me3 for a representative normal colon sample (green) and tumor (purple). Hypomethylated blocks are shaded in gray.

(D) Heatmap showing relative levels of H3K27ac, H3K36me3, H3K27me3, and H3K9me3 in normal colon (rows) for compartments A, I, and B (columns). DNA methylation differences between normal colon and tumors (for open-sea CpGs) and relative gene expression in normal colon (TCGA) are also shown.

(E) Plots showing mean and standard deviation of fold change ( $\log_2$ ) in enrichment of the indicated modification between tumors and normal samples. The respective plots show data for different modifications and are stratified by compartment (x axes).

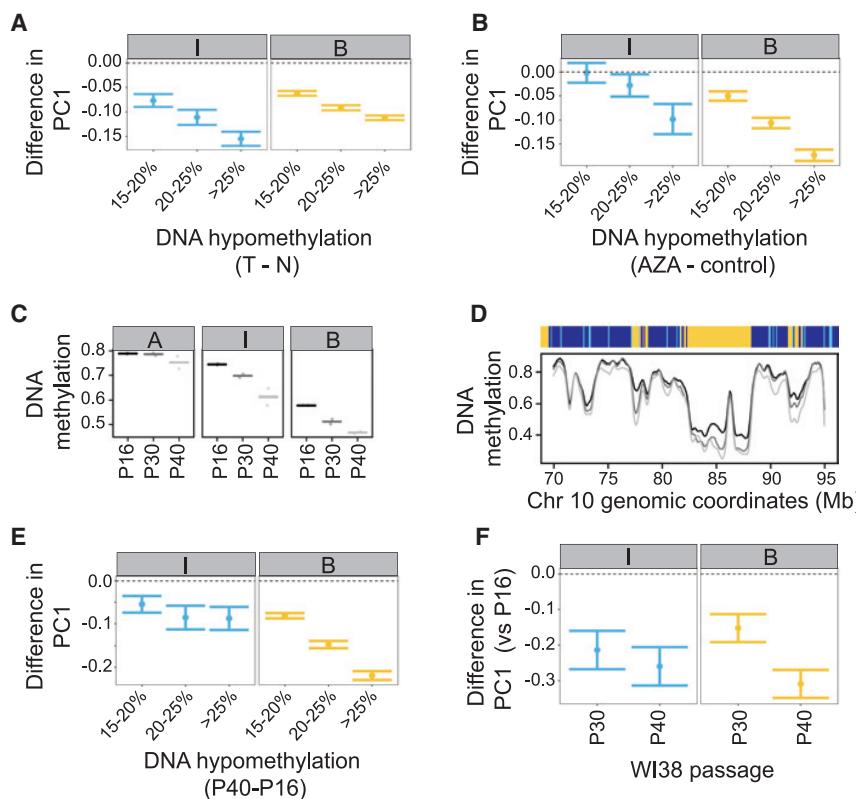
continued to replicate and had not yet progressed to replicative senescence. Comparison of early and late passage data confirmed progressive hypomethylation of compartments B and I with passage (Figures 5C and 5D). Moreover, late-passage hypomethylation was accompanied by topological changes analogous to tumors: the most hypomethylated regions had reduced eigenvector values, suggestive of compaction (Figure 5E). Importantly, intermediate-passage fibroblasts exhibited intermediate degrees of hypomethylation and structural reorganization (Figures 5C and 5F).

The indication that compartmental hypomethylation and topological shifts arise gradually as cells accumulate divisions prompted us to examine methylation in colonic adenomas. These pre-malignant lesions are entirely submitted for diagnostic evaluation, precluding assessment of topology, but their DNA methylation can be profiled from paraffin sections. Assessment of a published cohort of adenomas confirmed that compartmental hypomethylation was evident in these pre-malignant lesions and more severe in higher-grade cases

(Figure S6D; Fan et al., 2020). Taken together, these results suggest that compartmental shifts in colorectal tumors closely relate to DNA hypomethylation and may arise gradually over the course of proliferation. Thus, the most profound topological changes in tumors likely reflect their accumulated cell divisions rather than specific oncogenic programs.

### Transcriptional Consequences of Compartmental Reorganization

We next considered the transcriptional consequences of compartment B/I reorganization. We were struck that the overall transcriptional activity in these compartments was actually reduced in tumors, despite loss of DNA methylation and relocation of compartment B from the nuclear periphery. Considering genes in compartments B and I with detectable expression (transcripts per million/TPM > 0.1), we found that 3-fold more were downregulated than upregulated (Figures S6E and S6F). In contrast, compartment A genes did not exhibit such an imbalance.



**Figure 5. Compartmental Reorganization Is Closely Associated with DNA Hypomethylation and Accumulated Divisions**

(A) Association between Hi-C eigenvector (PC1) and hypomethylation of 100-kb windows in tumor samples. Data are stratified by compartment and extent of hypomethylation. Points and horizontal bars represent point estimates and 95% confidence intervals from a linear regression model.

(B) Data visualized as in (A), showing the association between Hi-C eigenvector and hypomethylation for HCT116 cells treated with 5'-aza versus DMSO.

(C) Plot showing DNA methylation for WI-38 cells at passages 16 (black), 30 (dark gray), and 40 (light gray). Data are stratified by compartment. Bars represent the average of two replicates (dots).

(D) Plot showing DNA methylation (for open-sea CpGs) for a representative region on chromosome 10. Traces represent methylation values for WI-38 fibroblasts at passages 16 (black,  $n = 2$ ), 30 (dark gray,  $n = 2$ ), and 40 (light gray,  $n = 2$ ). Compartment assignments for WI-38 are shown at the top (A, blue; I, light blue; B, yellow).

(E) Data visualized as in (A), showing the association between Hi-C eigenvector and hypomethylation for late (passage 40 [P40]) versus early (P16) WI-38 cells.

(F) Plot showing change over time in Hi-C eigenvector for 100-kb windows that show more than 20% hypomethylation in late-passage (P40) versus early-passage (P16) WI-38 fibroblasts.

We reasoned that gene silencing in the reorganized compartments might be sustained (or enhanced) by their repressive chromatin states. Indeed, compartment I gained widespread H3K27me3 in tumors (Figure 4E), and analysis of published data showed that many compartment I genes were upregulated in colon cancer-initiating cells treated with the Ezh2 inhibitor (Figure S6G; Lima-Fernandes et al., 2019). Compartment B is broadly covered by the repressive H3K9me3 mark, and a subset of its promoters is silenced by focal hypermethylation within the hypomethylated blocks (Figures 4D and S6H). Thus, alternate epigenetic mechanisms may actually further repress compartments B and I upon reorganization.

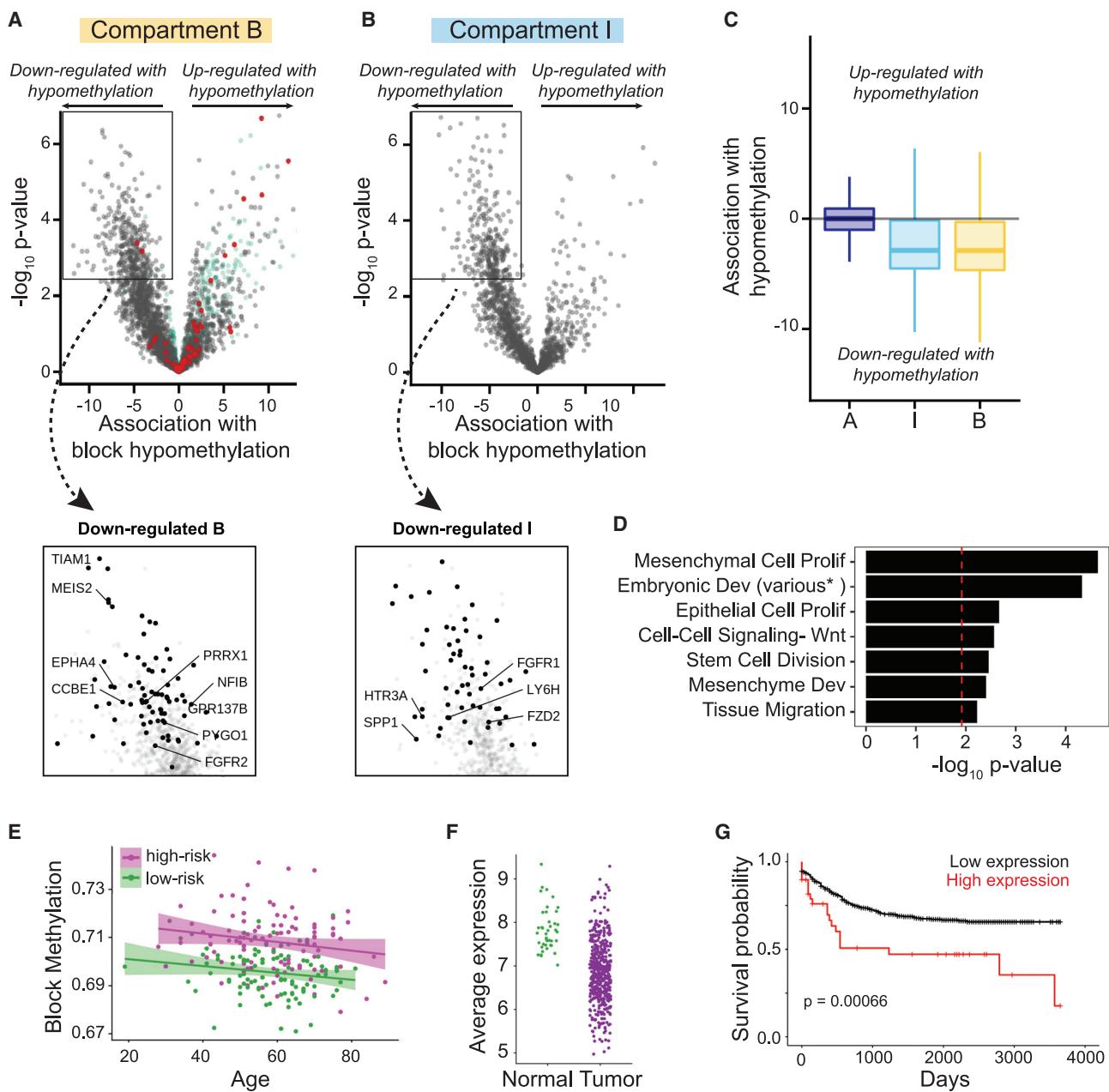
We therefore sought to identify specific genes that were deregulated by this process. Because reorganization correlated quantitatively with hypomethylation (Figure 5A), we reasoned that methylation could be a surrogate for compartmental changes. We collated methylation and RNA-seq data for 239 colorectal tumors (TCGA) and used a correlation metric to identify genes in compartments B and I whose expression was consistently altered in association with block hypomethylation (Figures 6A, 6B, and S7A; STAR Methods). Notably, genes in compartment A showed no consistent transcriptional change in association with hypomethylation (Figure 6C).

Correlation analysis highlighted two gene sets (Figures 6A and 6B). A small set of genes was upregulated with block hypomethylation and included cancer germline antigens (CGAs) and endogenous retroviruses (ERVs). De-repression of CGAs and ERVs has been described in colon tumors and linked to viral

mimicry and immunogenicity (Gibbs and Whitehurst, 2018; Rooney et al., 2015; Roulois et al., 2015).

A much larger set of genes in compartments B and I were downregulated with block hypomethylation (Figures 6A and 6B). Downregulated genes in compartment B were marked by H3K9me3 and/or promoter methylation in tumors, whereas those in compartment I were enriched for H3K27me3 (Figure S7B). We curated a list of robustly downregulated genes in these compartments (Table S5). To focus on malignant cell-intrinsic expression, we controlled for stromal content and excluded genes with high expression in immune cells or other non-epithelial cell types (STAR Methods). Remarkably, the resulting list of 146 genes was highly enriched for functions related to mesenchymal development, stem cell proliferation, and Wnt signaling (Figure 6D). Further analysis highlighted specific genes with established roles in colorectal cancer progression, Wnt signaling, epithelial-mesenchymal transition (EMT), invasion, and metastasis (e.g., CCBE1, EPHA4, FGFR1, FGFR2, FZD2, GPR137B, MEIS2, NFIB, PRRX1, PYGO1, SPP1, and TIA1; Table S6; Koveitypour et al., 2019; Nguyen et al., 2020). In contrast, only a few of the downregulated genes were nominally associated with tumor-suppressive functions.

Our analyses suggest that compartmental reorganization drives induction of CGAs and ERVs, which are associated with anti-tumor immunity, and repression of genes with functions in Wnt signaling, EMT, invasion, and metastasis. Hence, the most profound topological alterations evident in tumors are actually associated with tumor-suppressive transcriptional programs.



**Figure 6. Compartmental Reorganization Linked to Tumor-Suppressive Transcriptional Programs**

(A) Volcano plot depicting the association (x axis) between expression and block hypomethylation for genes in compartment B, computed across tumors in the TCGA colorectal cohort with a purity of more than 40%. The y axis indicates the significance of the association. Genes (points) plotted toward the top right are upregulated in association with block hypomethylation. They are highly enriched for cancer germline antigens (CGAs; red) and ERV elements (green). Genes plotted toward the top left are downregulated in association with hypomethylation. A magnified panel (below) highlights high-confidence downregulated genes after excluding genes expressed in non-epithelial cell types (black; [STAR Methods](#)). Genes related to EMT, Wnt signaling, invasion, and metastasis are labeled.

(B) Data presented as in (A) for compartment I genes.

(C) Boxplot showing association between expression change and DNA block hypomethylation for genes in compartments A, B, and I. Negative values indicate repression in association with DNA hypomethylation, and positive values indicate upregulation in association with hypomethylation.

(D) Functional gene set annotations enriched (false discovery rate [FDR] < 20%) among 146 high-confidence downregulated B/I genes (from the insets in A and B). \*, enriched annotations included multiple overlapping sets related to embryonic development ([Table S6](#)).

(E) Plot showing average block methylation levels for normal colon biopsies (y axis) as a function of donor age (x axis). Each point represents one sample from a low-risk (green) or high-risk (magenta) donor ([Wang et al., 2020](#)). Linear regression fit (lines) and 95% confidence intervals (shades) are indicated.

(legend continued on next page)

## Compartment-Specific Epigenetic Changes Restrain Tumor Progression

Our collective findings suggested that accumulation of excess cell divisions leads to compartmental shifts that enact tumor-suppressive transcriptional programs. They led us to hypothesize that compartmental reorganization hinders malignant progression. To test this, we examined whether the compartmental shifts were predictive of disease risk and outcome.

First, we considered a recent survey of DNA methylation in 206 normal colon biopsies, stratified into low- and high-risk groups according to whether the donor had a concurrent colorectal tumor elsewhere in the colon (Wang et al., 2020). Examination of these data confirmed that compartments B and I became progressively hypomethylated with increasing donor age (Figure S7C). To test whether hypomethylation was protective, we compared low- and high-risk groups. We found that compartments B and I were significantly less hypomethylated in normal colon biopsies from the high-risk group, consistent with our hypothesis that the compartment shift is tumor suppressive (Figure 6E).

Second, we examined two clinical cohorts of colorectal tumors (Marisa et al., 2013; The Cancer Genome Atlas Network, 2012). Although these tumors have presumably overcome impediments posed by compartment shifts, we reasoned that the associated transcriptional changes should nonetheless hinder their progression and correlate with favorable patient outcomes. Here we focused on the 146 genes in compartments B and I that were robustly downregulated with block hypomethylation (Table S5). As expected, these genes were expressed at substantially lower levels in tumors (Figure 6F). However, there was considerable tumor-to-tumor variability. Remarkably, we found that this set of genes was highly enriched for poor prognosis markers in a cohort of 566 colon tumors (Figure S7D;  $p = 2.4 \times 10^{-10}$ ; Marisa et al., 2013). A risk score constructed from their average expression was a strong predictor of shorter recurrence-free survival (RFS) (Figure 6G;  $p = 0.0007$ ). This prognostic association was also validated in a second cohort of 443 tumors from TCGA ( $p = 0.03$ ).

Survival difference was evident even after controlling for microsatellite instability (MSI), BRAF mutations, and clinical stage ( $p = 0.025$ ) (Figure S7E). It was evident even in node-negative stage II tumors ( $p = 0.039$ ), which is significant, given the clinical challenge associated with the uncertain course of these intermediate-stage tumors (Figure S7F) (Fotheringham et al., 2019). The gene set was also associated with metastases ( $p = 0.018$ ), consistent with its functional annotations and supportive of its clinical significance.

Finally, we considered whether compartmental reorganization could be a general tumor-suppressive mechanism. We examined methylation and expression data for cohorts spanning 10 epithelial tumor types (ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, 2020). We confirmed that block

hypomethylation correlated with reduced gene expression in all 10 cancers (Figure S7G). We next collated genes in blocks that were significantly downregulated in association with hypomethylation in each cohort. The resulting sets were highly overlapping and include a shared set of 367 genes that were commonly downregulated with hypomethylation in at least 7 of the 10 tumor types (Table S7). Remarkably, these shared genes were enriched for annotated oncogenes ( $p = 0.002$ ), suggesting that compartmental reorganization may also hinder the development and progression of other epithelial tumor types.

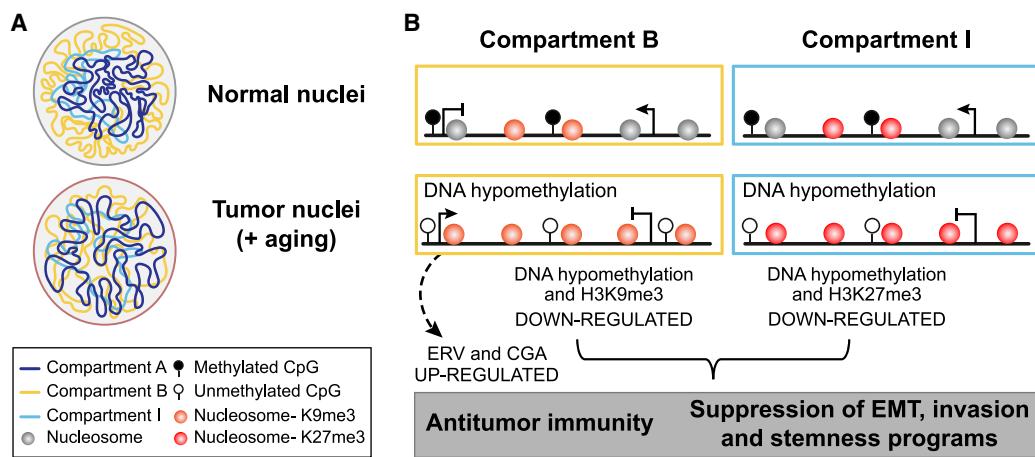
In conclusion, we document profound compartmental shifts in tumors and other cells that have accumulated many divisions (Figures 7A and 7B). This reorganization is associated with widespread transcriptional changes, including repression of EMT, invasion, metastasis, and stemness programs. Further analysis of methylation and expression data for normal colon biopsies and tumor cohorts supports the hypothesis that the compartmental shifts and associated transcriptional programs restrain tumor progression.

## DISCUSSION

We presented a systematic integration of genome topology, methylation, and chromatin state in colorectal cancer. Our data and analyses parse multiple organizational layers, from E-P loops to TADs to compartmental structures. In particular, they revealed three principles of large-scale compartmental organization. First, topology data for primary tissues uncovered a structurally distinct intermediate compartment I. Second, comparison of tumors and normal colon revealed widespread changes in spatial partitioning, nuclear positioning, and epigenetic states of compartments B and I that appear to be shared by tumor, aging, and other excessively replicated cells. Third, these compartmental changes correlate with and may promote tumor-suppressive expression programs associated with reduced cancer risk and better prognosis. Although tumor-associated epigenetic changes are typically construed to be oncogenic, our findings suggest that these most profound topological alterations actually restrain malignant progression.

Evidence of compartment I emerged from our analysis of primary epithelial tissue. Compartment I resides at the interface between the A and B compartments and engages in promiscuous long-range interactions with both conventional compartments. Polymer models and FISH imaging data indicate that compartment I regions occupy intermediate radial positions in nuclei. Compartment I is also distinguished epigenetically by broad H3K27me3 and robust block hypomethylation in tumors. Compartment I is distinct from previously described sub-compartments, showing the highest overlap (47%) with B1 (Rao et al., 2014). It may relate to nuclear foci documented in high-resolution imaging studies (Boettiger et al., 2016; Rowley and Corces, 2018; Xu et al., 2018). For example, Boettiger et al. (2016)

(F) Plots showing average expression of the high-confidence downregulated B/I genes (from the insets in A and B) in clinical specimens. Each point corresponds to a different sample from a cohort of colorectal tumors and normal colons (Cancer Genome Atlas Network, 2012). The y axis represents log<sub>2</sub>-normalized counts. (G) Kaplan-Meier curve depicting survival outcomes of patients stratified by their average tumor expression of the high-confidence downregulated compartment B/I genes.



**Figure 7. Compartment Shifts in Excessively Replicated Cells Restrain Malignant Progression**

The schematic depicts compartment shifts and proposed functional consequences.

(A) In normal nuclei, compartments A, B, and I are robustly partitioned and spatially segregated. In tumor or aging cells that have accumulated excess divisions, compartmental organization is compromised, and compartment-specific epigenetic states are altered.

(B) Exemplar loci are shown for each compartment in normal (top) or pathologic (bottom) states. Hypomethylation of compartment B induces ERVs and CGAs, which promote anti-tumor immunity. Repressive chromatin in compartments B and I downregulates genes associated with EMT, invasion, and stemness.

visualized a compartment enriched for Polycomb-associated marks and developmental genes in *Drosophila*.

The coherent organization of compartments A, B, and I in normal colon was profoundly distorted in tumors. We observed a breakdown of partitioning between compartments A and B, whereas compartment I shifted its interactions toward the closed B compartment. The aberrations appear to be closely related to nuclear architecture. Concordant polymer models, electron microscopy, and FISH imaging data indicate that compartment B loses its tight association with the periphery and shifts toward the nuclear interior. Falk et al. (2019) found previously that the radial asymmetry of compartments A and B is inverted in rod photoreceptors. However, compartmental partitioning was largely maintained in photoreceptors, in contrast to the overall disorganization of compartmental structure in tumors.

The breakdown of compartment structure was closely tied to pervasive changes to methylation and chromatin state. Compartments B and I acquired near-uniform hypomethylation in tumors and became further enriched for their characteristic chromatin states, H3K9me3 and H3K27me3, a pattern that has also been described for hypomethylated loci in breast cancer cell lines (Hon et al., 2012). The reorganized compartments might relate to phase-separated condensates and nuclear foci thought to play wide-ranging roles in gene and genome regulation (Larson et al., 2017; Strom et al., 2017). Notably, high-resolution analyses have identified senescence-associated heterochromatin foci with a central density of H3K9me3-marked heterochromatin surrounded by a ring of H3K27me3 (Chandra and Narita, 2013; Sati et al., 2020), features consistent with our three-compartment model in tumors.

Compartmental reorganization appears to be tightly linked to DNA hypomethylation and proliferative history. Hypomethylated blocks in tumors correspond to compartments B and I, with the

most severely hypomethylated loci undergoing more negative eigenvector shifts, indicative of compaction. Methylation loss may be causal because demethylating agents directly induce topological changes. Although block hypomethylation was initially described in cancer, it is increasingly recognized to be a common feature of cells that have accumulated excess divisions, including aging and senescent cells (Berman et al., 2011; Cruickshanks et al., 2013; Nordor et al., 2017; Timp et al., 2014). Indeed, compartments B and I become progressively hypomethylated and structurally reorganized in passaged fibroblasts. Moreover, colon adenomas exhibit intermediate compartmental hypomethylation, consistent with a replicative history between normal colon and tumors. Although future studies are needed to examine the topology of these and other pre-malignant lesions, our findings suggest that compartmental shifts are not a consequence of malignancy but, rather, arise progressively as cells accumulate divisions.

We therefore propose that the compartmental reorganization reflects a fundamental epigenetic process primed by excess cell divisions, a perspective that enabled us to interpret attendant transcriptional changes. Compartmental hypomethylation was associated with repression of compartment B and I genes, likely as a consequence of repressive epigenetic states that arise in the hypomethylated compartments. Repressed genes were enriched for oncogenic functions related to EMT, invasion, and Wnt signaling, leading us to speculate that the topological shifts present a barrier to tumorigenesis. Indeed, prior studies have shown that cultured cells undergoing EMT remodel large heterochromatin domains (McDonald et al., 2011). Although most compartment B and I genes were downregulated, CGAs and ERVs with pro-immunity functions were induced and could complement restriction of stemness and invasion programs to restrain malignant progression in aging colonic epithelium or pre-malignant lesions.

Final support for a proposed tumor-suppressive role emerged from our analysis of clinical cohorts. First, methylation profiles for normal colon biopsies revealed that age-associated compartment B and I hypomethylation was associated with reduced colorectal cancer risk (Wang et al., 2020), consistent with our model and with a recent report relating morphological changes in uninvolved colonic nuclei to tumor risk (Gladstein et al., 2018). Second, examination of colorectal tumor cohorts revealed that a transcriptional signature of compartmental shift was predictive of patient outcome and likelihood of metastasis. Finally, a pan-cancer analysis suggested that the compartmental shifts and tumor-suppressive effects may be generalizable to other epithelial cancers. Future studies of these pervasive architectural changes and their functional significance in cancer and aging could inform new strategies for early detection, patient stratification, and therapeutic intervention.

## STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

- Lead Contact
- Materials Availability
- Data and Code Availability

- EXPERIMENTAL MODEL AND SUBJECT DETAILS

- Human tumor specimens
- Cell lines

- METHOD DETAILS

- Tissue dissociation and crosslinking
- Hybrid selection bisulfite sequencing
- Whole genome bisulfite library preparation
- ChIP-seq
- Hi-C
- HiChIP
- RNA-seq
- Electron microscopy
- DNA-FISH
- Treatment with 5-azacytidine

- QUANTIFICATION AND STATISTICAL ANALYSIS

- DNA methylation data preprocessing and quantification
- Cancer-associated DNA methylation alterations in normal colon tissue
- Hi-C analysis
- Association of compartmental organization with DNA hypomethylation
- HiChIP analysis
- Copy number variant analysis
- SNP Analysis
- Gene expression analysis
- Pan-cancer analysis of gene expression associated to hypomethylation
- Exclusion of genes of likely non-tumor cell-origin
- Survival analysis
- ChIP-seq analysis

- DNA polymer modeling
- Energy function
- Parameter optimization
- Molecular dynamics simulation details
- Radial density profile
- Electron Microscopy Analysis
- DNA-FISH analysis

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cell.2020.07.030>.

## ACKNOWLEDGMENTS

We thank Rianne Boursiquot for assistance with sequencing; Mohammed Miri for assistance with clinical samples; Elizabeth Gaskell, Volker Hovestadt, Christine Eyerl and Ryan Corcoran, Angela Shih, and Omer Yilmaz for thoughtful discussions; and Leslie Gaffney for graphic support. S.E.J. is supported by NIH T32CA009216. A.R. and R.A.I. are supported by NIH R01GM083084 and R01HG005220. J.H.C. is supported by NIH 1T32CA207021-01. M.J.A. is supported by a Broad Institute Merkin Fellowship. B.E.B. is the Bernard and Mildred Kayden Endowed MGH Research Institute Chair and an American Cancer Society Research Professor. This research was supported by the National Cancer Institute (DP1CA216873) and the Starr Cancer Consortium. This paper is dedicated to the memory of Yaw Adu Kuffour.

## AUTHOR CONTRIBUTIONS

Conception and Experimental Design, S.E.J., A.R., M.J.A., and B.E.B.; Methodology and Data Acquisition, S.E.J., A.R., C.A., E.H., K.P., J.H.C., and B.E.B.; Analysis and Interpretation of Data, S.E.J., A.R., Y.Q., C.A., L.S.Z., Y.D., V.H., N.S., M.K.S., C.L., S.I., S.C.N., E.F.J., N.H., R.A.I., B.Z., M.J.A., and B.E.B. Manuscript Writing and Revision, S.E.J., A.R., M.J.A., and B.E.B.

## DECLARATION OF INTERESTS

N.H. is an equity holder of BioNTech and a consultant for Related Sciences. M.J.A. declares outside interest in Excelsior Genomics. B.E.B. declares outside interests in Fulcrum Therapeutics, 1CellBio, HiFiBio, Arsenal Biosciences, Cell Signaling Technologies, BioMillenia, and Nohla Therapeutics.

Received: July 8, 2019

Revised: May 4, 2020

Accepted: July 20, 2020

Published: August 24, 2020

## REFERENCES

- Baylin, S.B., and Jones, P.A. (2016). Epigenetic Determinants of Cancer. *Cold Spring Harb. Perspect. Biol.* 8, a019505.
- Beliveau, B.J., Joyce, E.F., Apostolopoulos, N., Yilmaz, F., Fonseka, C.Y., McCole, R.B., Chang, Y., Li, J.B., Senaratne, T.N., Williams, B.R., et al. (2012). Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. USA* 109, 21301–21306.
- Berman, B.P., Weisenberger, D.J., Aman, J.F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C.P.E., van Dijk, C.M., Tollenaar, R.A.E.M., et al. (2011). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* 44, 40–46.
- Bickmore, W.A., and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. *Cell* 152, 1270–1284.
- Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C.-T., and Zhuang, X. (2016). Super-

- resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529, 418–422.
- Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Müller, S., Eils, R., Cremer, C., Speicher, M.R., and Cremer, T. (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* 3, e157.
- Calon, A., Lonardo, E., Berenguer-Llergo, A., Espinet, E., Hernando-Mombina, X., Iglesias, M., Sevillano, M., Palomo-Ponce, S., Tauriello, D.V.F., Byrom, D., et al. (2015). Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.* 47, 320–329.
- Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330–337.
- Cao, Y., Chen, Z., Chen, X., Ai, D., Chen, G., McDermott, J., Huang, Y., Guo, X., and Han, J.-D.J. (2019). Accurate loop calling for 3D genomic data with cLoops. *Bioinformatics* 36, 666–675.
- Chandra, T., and Narita, M. (2013). High-order chromatin structure and the epigenome in SAHFs. *Nucleus* 4, 23–28.
- Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbio-links: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 44, e71.
- Collado-Torres, L., Nellore, A., Kammers, K., Ellis, S.E., Taub, M.A., Hansen, K.D., Jaffe, A.E., Langmead, B., and Leek, J.T. (2017). Reproducible RNA-seq analysis using recount2. *Nat. Biotechnol.* 35, 319–321.
- Corces, M.R., and Corces, V.G. (2016). The three-dimensional cancer genome. *Curr. Opin. Genet. Dev.* 36, 1–7.
- Criscione, S.W., De Cecco, M., Siranosian, B., Zhang, Y., Kreiling, J.A., Sedivy, J.M., and Neretti, N. (2016). Reorganization of chromosome architecture in replicative cellular senescence. *Sci. Adv.* 2, e1500882.
- Cruickshanks, H.A., McBryan, T., Nelson, D.M., Vanderkraats, N.D., Shah, P.P., van Tuyn, J., Singh Rai, T., Brock, C., Donahue, G., Duncan, D.S., et al. (2013). Senescent cells harbour features of the cancer epigenome. *Nat. Cell Biol.* 15, 1495–1506.
- Dekker, J., and Misteli, T. (2015). Long-Range Chromatin Interactions. *Cold Spring Harb. Perspect. Biol.* 7, a019356.
- Di Piero, M., Zhang, B., Aiden, E.L., Wolynes, P.G., and Onuchic, J.N. (2016). Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci. USA* 113, 12168–12173.
- Dias-Santagata, D., Akhavanfard, S., David, S.S., Vernovsky, K., Kuhlmann, G., Boisvert, S.L., Stubbs, H., McDermott, U., Settleman, J., Kwak, E.L., et al. (2010). Rapid targeted mutational analysis of human tumours: a clinical platform to guide personalized cancer medicine. *EMBO Mol. Med.* 2, 146–158.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380.
- Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature* 518, 331–336.
- Dunne, P.D., Dasgupta, S., Blayney, J.K., McArt, D.G., Redmond, K.L., Weir, J.A., Bradley, C.A., Sasazuki, T., Shirasawa, S., Wang, T., et al. (2016). EphA2 Expression Is a Key Driver of Migration and Invasion and a Poor Prognostic Marker in Colorectal Cancer. *Clin. Cancer Res.* 22, 230–242.
- Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* 3, 99–101.
- Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B.R., Leonhardt, H., Joffe, B., Dekker, J., Fudenberg, G., Solovei, I., et al. (2019). Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* 570, 395–399.
- Fan, J., Li, J., Guo, S., Tao, C., Zhang, H., Wang, W., Zhang, Y., Zhang, D., Ding, S., and Zeng, C. (2020). Genome-wide DNA methylation profiles of low- and high-grade adenoma reveals potential biomarkers for early detection of colorectal carcinoma. *Clin. Epigenetics* 12, 56.
- Flavahan, W.A., Drier, Y., Liu, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suvà, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529, 110–114.
- Flavahan, W.A., Drier, Y., Johnstone, S.E., Hemming, M.L., Tarjan, D.R., Hergazi, E., Shareef, S.J., Javed, N.M., Raut, C.P., Eschle, B.K., et al. (2019). Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature* 575, 229–233.
- Fortin, J.-P., and Hansen, K.D. (2015). Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol.* 16, 180.
- Fotheringham, S., Mozolowski, G.A., Murray, E.M.A., and Kerr, D.J. (2019). Challenges and solutions in patient treatment strategies for stage II colon cancer. *Gastroenterol. Rep. (Oxf.)* 7, 151–161.
- Gibbs, Z.A., and Whitehurst, A.W. (2018). Emerging Contributions of Cancer/Testis Antigens to Neoplastic Behaviors. *Trends Cancer* 4, 701–712.
- Gladstein, S., Damania, D., Almassalha, L.M., Smith, L.T., Gupta, V., Subramanian, H., Rex, D.K., Roy, H.K., and Backman, V. (2018). Correlating colorectal cancer risk with field carcinogenesis progression using partial wave spectroscopic microscopy. *Cancer Med.* 7, 2109–2120.
- Gusnanto, A., Wood, H.M., Pawitan, Y., Rabbits, P., and Berri, S. (2012). Correcting for cancer genome size and tumour cell content enables better estimation of copy number alterations from next-generation sequence data. *Bioinformatics* 28, 40–47.
- Hansen, K.D., Timp, W., Bravo, H.C., Sabuncyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D., et al. (2011). Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.* 43, 768–775.
- Hansen, K.D., Langmead, B., and Irizarry, R.A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol.* 13, R83.
- Hinoue, T., Weisenberger, D.J., Lange, C.P.E., Shen, H., Byun, H.M., Van Den Berg, D., Malik, S., Pan, F., Noushmehr, H., van Dijk, C.M., et al. (2012). Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res.* 22, 271–282.
- Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.-L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., et al. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* 351, 1454–1458.
- Hon, G.C., Hawkins, R.D., Caballero, O.L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L.E., et al. (2012). Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* 22, 246–258.
- Huang, J., Gusnanto, A., O'Sullivan, K., Staaf, J., Borg, A., and Pawitan, Y. (2007). Robust smooth segmentation approach for array CGH data analysis. *Bioinformatics* 23, 2463–2469.
- Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* 12, 115–121.
- ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020). Pan-cancer analysis of whole genomes. *Nature* 578, 82–93.
- Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* 9, 999–1003.
- Kloetgen, A., Thandapani, P., Ntziachristos, P., Ghebrechristos, Y., Nomikou, S., Lazaris, C., Chen, X., Hu, H., Bakogianni, S., Wang, J., et al. (2020). Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat. Genet.* 52, 388–400.
- Koveitpour, Z., Panahi, F., Vakilian, M., Peymani, M., Seyed Forootan, F., Nasr Esfahani, M.H., and Ghaedi, K. (2019). Signaling pathways involved in colorectal cancer progression. *Cell Biosci.* 9, 97.

- Krefting, J., Andrade-Navarro, M.A., and Ibn-Salem, J. (2018). Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol.* 16, 87.
- Lareau, C.A., and Aryee, M.J. (2018a). diffloop: a computational framework for identifying and analyzing differential DNA loops from sequencing data. *Bioinformatics* 34, 672–674.
- Lareau, C.A., and Aryee, M.J. (2018b). hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat. Methods* 15, 155–156.
- Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1 $\alpha$  suggests a role for phase separation in heterochromatin. *Nature* 547, 236–240.
- Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15, R29.
- Lawrence, M., Huber, W., Pagès, H., Aboyou, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9, e1003118.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Liau, B.B., Sievers, C., Donohue, L.K., Gillespie, S.M., Flavahan, W.A., Miller, T.E., Venetecher, A.S., Hebert, C.H., Carey, C.D., Rodig, S.J., et al. (2017). Adaptive Chromatin Remodeling Drives Glioblastoma Stem Cell Plasticity and Drug Tolerance. *Cell Stem Cell* 20, 233–246.e7.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.
- Lima-Fernandes, E., Murison, A., da Silva Medina, T., Wang, Y., Ma, A., Leung, C., Luciani, G.M., Haynes, J., Pollett, A., Zeller, C., et al. (2019). Targeting bivalency de-represses Indian Hedgehog and inhibits self-renewal of colorectal cancer-initiating cells. *Nat. Commun.* 10, 1436.
- Liu, Y., Sun, J., and Zhao, M. (2017). ONGene: A literature-based database for human oncogenes. *J. Genet. Genomics* 44, 119–121.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Love, M.I., Anders, S., Kim, V., and Huber, W. (2015). RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Res.* 4, 1070.
- Lun, A.T.L., and Smyth, G.K. (2016). csaaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* 44, e45.
- MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., et al. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 45 (D1), D896–D901.
- Marisa, L., de Reyniès, A., Duval, A., Selvès, J., Gaub, M.P., Vescovo, L., Etienne-Grimaldi, M.-C., Schiappa, R., Guenot, D., Ayadi, M., et al. (2013). Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* 10, e1001453.
- McDonald, O.G., Wu, H., Timp, W., Doi, A., and Feinberg, A.P. (2011). Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition. *Nat. Struct. Mol. Biol.* 18, 867–874.
- McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D., et al. (2018). Cell-Profiler 3.0: Next-generation image processing for biology. *PLoS Biol.* 16, e2005970.
- Modrek, A.S., Golub, D., Khan, T., Bready, D., Prado, J., Bowman, C., Deng, J., Zhang, G., Rocha, P.P., Raviram, R., et al. (2017). Low-Grade Astrocytoma Mutations in IDH1, P53, and ATRX Cooperate to Block Differentiation of Human Neural Stem Cells via Repression of SOX2. *Cell Rep.* 21, 1267–1280.
- Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* 13, 919–922.
- Nguyen, L.H., Goel, A., and Chung, D.C. (2020). Pathways of Colorectal Carcinogenesis. *Gastroenterology* 158, 291–302.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385.
- Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169, 930–944.e22.
- Nordor, A.V., Nehar-Belaid, D., Richon, S., Klatzmann, D., Bellet, D., Dangles-Marie, V., Fournier, T., and Aryee, M.J. (2017). The early pregnancy placenta foreshadows DNA methylation alterations of solid tumors. *Epigenetics* 12, 793–803.
- Olshen, A.B., Venkatraman, E.S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5, 557–572.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419.
- Peltekova, V.D., Lemire, M., Qazi, A.M., Zaidi, S.H.E., Trinh, Q.M., Bielecki, R., Rogers, M., Hodgson, L., Wang, M., D'Souza, D.J.A., et al. (2014). Identification of genes expressed by immune cells of the colon that are regulated by colorectal cancer-associated variants. *Int. J. Cancer* 134, 2330–2341.
- Plimpton, S. (1995). Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* 117, 1–19.
- Qi, Y., Reyes, A., Johnstone, S.E., Aryee, M.J., Bernstein, B.E., and Zhang, B. (2020). Data-driven polymer model for mechanistic exploration of diploid genome organization. *bioRxiv*. <https://doi.org/10.1101/2020.02.27.968735>.
- Qi, Y., and Zhang, B. (2019). Predicting three-dimensional genome organization with chromatin states. *PLoS Comput. Biol.* 15, e1007024.
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160, 48–61.
- Rosin, L.F., Nguyen, S.C., and Joyce, E.F. (2018). Condensin II drives large-scale folding and spatial partitioning of interphase chromosomes in Drosophila nuclei. *PLoS Genet.* 14, e1007393.
- Roulois, D., Loo Yau, H., Singhania, R., Wang, Y., Danesh, A., Shen, S.Y., Han, H., Liang, G., Jones, P.A., Pugh, T.J., et al. (2015). DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell* 162, 961–973.
- Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* 19, 789–800.
- Sakthivel, K.M., and Sehgal, P. (2016). A novel role of lamins from genetic disease to cancer biomarkers. *Oncol. Rev.* 10, 309.
- Sati, S., Bonev, B., Szabo, Q., Jost, D., Bensadoun, P., Serra, F., Loubiere, V., Papadopoulos, G.L., Rivera-Mulia, J.-C., Fritsch, L., et al. (2020). 4D Genome Rewiring during Oncogene-Induced and Replicative Senescence. *Mol. Cell* 78, 522–538.e9.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682.

- Schmitt, A.D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C.L., Li, Y., Lin, S., Lin, Y., Barr, C.L., and Ren, B. (2016). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* 17, 2042–2059.
- Schreiber, K.H., and Kennedy, B.K. (2013). When lamins go bad: nuclear structure and disease. *Cell* 152, 1365–1375.
- Serra, F., Baù, D., Goodstadt, M., Castillo, D., Filion, G.J., and Martí-Renom, M.A. (2017). Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput. Biol.* 13, e1005665.
- Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* 16, 259.
- Shav-Tal, Y. (2013). *Imaging Gene Expression* (Springer).
- Smit, A., Hubley, R., and Green, P. (2015). RepeatMasker Open-4.0. 2013–2015 (RepeatMasker). <http://www.repeatmasker.org>.
- Stadhouders, R., Filion, G.J., and Graf, T. (2019). Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* 569, 345–354.
- Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. *Nature* 547, 241–245.
- Teng, M., and Irizarry, R.A. (2017). Accounting for GC-content bias reduces systematic errors and batch effects in ChIP-seq data. *Genome Res.* 27, 1930–1938.
- Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al.; Cancer Genome Atlas Research Network (2019). The Immune Landscape of Cancer. *Immunity* 51, 411–412.
- Timp, W., Bravo, H.C., McDonald, O.G., Goggins, M., Umbrecht, C., Zeiger, M., Feinberg, A.P., and Irizarry, R.A. (2014). Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med.* 6, 61.
- Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., and Issa, J.P. (1999). CpG island methylator phenotype in colorectal cancer. *Proc. Natl. Acad. Sci. USA* 96, 8681–8686.
- van Steensel, B., and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* 169, 780–791.
- Wang, S., Su, J.H., Beliveau, B.J., Bintu, B., Moffitt, J.R., Wu, C.T., and Zhuang, X. (2016). Spatial organization of chromatin domains and compartments in single chromosomes. *Science* 353, 598–602.
- Wang, Q., Zhu, J., Wang, Y.-W., Dai, Y., Wang, Y.-L., Wang, C., Liu, J., Baker, A., Colburn, N.H., and Yang, H.-S. (2017). Tumor suppressor Pcd4 attenuates Sin1 translation to inhibit invasion in colon carcinoma. *Oncogene* 36, 6225–6234.
- Wang, T., Maden, S.K., Luebeck, G.E., Li, C.I., Newcomb, P.A., Ulrich, C.M., Joo, J.E., Buchanan, D.D., Milne, R.L., Soutey, M.C., et al. (2020). Dysfunctional epigenetic aging of the normal colon and colorectal cancer risk. *Clin. Epigenetics* 12, 5.
- Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* 10, 232.
- Xiong, K., and Ma, J. (2019). Revealing Hi-C subcompartments by imputing inter-chromosomal chromatin interactions. *Nat. Commun.* 10, 5069.
- Xu, Y., Hu, B., Choi, A.J., Gopalan, B., Lee, B.H., Kalady, M.F., Church, J.M., and Ting, A.H. (2012). Unique DNA methylome profiles in CpG island methylator phenotype colon cancers. *Genome Res.* 22, 283–291.
- Xu, J., Ma, H., Jin, J., Uttam, S., Fu, R., Huang, Y., and Liu, Y. (2018). Super-Resolution Imaging of Higher-Order Chromatin Structures at Different Epigenomic States in Single Mammalian Cells. *Cell Rep.* 24, 873–882.
- Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* 286, 7983–7989.
- Zhang, B., and Wolynes, P.G. (2015). Topology, structures, and energy landscapes of human chromosomes. *Proc. Natl. Acad. Sci. USA* 112, 6062–6067.
- Zhang, B., and Wolynes, P.G. (2016). Shape Transitions and Chiral Symmetry Breaking in the Energy Landscape of the Mitotic Chromosome. *Phys. Rev. Lett.* 116, 248101.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.
- Zhou, W., Dinh, H.Q., Ramjan, Z., Weisenberger, D.J., Nicolet, C.M., Shen, H., Laird, P.W., and Berman, B.P. (2018). DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.* 50, 591–602.
- Zink, D., Fischer, A.H., and Nickerson, J.A. (2004). Nuclear structure in cancer cells. *Nat. Rev. Cancer* 4, 677–687.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CTCF	Cell signaling	Cat# 3418; RRID: AB_2086791
SMC1	Bethyl	Cat# A300-055A; RRID AB_2192467
H3K27ac	Active Motif	Cat# 39133; RRID AB_2561016
H3K9me3	Abcam	Cat# 8898; RRID AB_306848
H3K27me3	Cell Signaling	Cat# 9733; RRID AB_2616029
H3K36me3	Abcam	Cat# 9050; RRID AB_306966
Biological Samples		
See Table S1 for a list of patient samples included in the study.		N/A
Chemicals, Peptides, and Recombinant Proteins		
Azacitidine (5-azacytidine)	Selleckchem	S1782
Triethanolamine	Sigma	90279
Acetic Anhydride	Sigma	320102
Formamide	Sigma	47671
CX1723 Supelco Citric Acid, Anhydrous	Sigma	CX1723
Thermo ScientificRNase A, DNase and protease-free (10 mg/mL)	ThermoScientific	EN0531
Dextran sulfate sodium salt from Leuconostoc spp.	Sigma	D8906
Invitrogen DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride)	ThermoFisher	D1306
SlowFade Gold Antifade Mountant	Invitrogen	S36936
Tissue-Tek® O.C.T. Compound	VWR/Sakura	25608-930
Oligo Pool, 160mers	Twist Biosciences	
Taq polymerase	Thermo Fisher	18038042
Maxima H Minus Reverse Transcriptase (200 U/μL)	Thermo Fisher	EP0751
Critical Commercial Assays		
Nextera DNA Library Prep Kit (for HiChIP)	Illumina	20018704
NEXTFLEX bisulfite library prep kit	Perkin Elmer	NOVA-5119-01
HiScribe T7 High Yield RNA Synthesis Kit	NEB	E2040S
Kapa HiFi Hotstart PCR Kit	Roche	#KK2502
Deposited Data		
Imaging data	This paper	Mendeley data: <a href="https://dx.doi.org/10.17632/6k4hfw76.1">https://dx.doi.org/10.17632/6k4hfw76.1</a>
Raw and processed sequencing data	This paper	GEO: GSE133928
Experimental Models: Cell Lines		
HCT116	ATCC	CCL-247
SW480	ATCC	CCL-228
LS 174T	ATCC	CL-188
RKO	ATCC	CRL-2577
FHC	ATCC	CRL-1831
WI38	Coriell	AG06814-N
Oligonucleotides		
Chr 12 Probe F- for DNA FISH probe preparation	IDT	CGGTCCCCGTCCGAGGTATAC
Chr 12 Probe R- for DNA FISH probe preparation	IDT	TCCAATACGCACCGATCGAG

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
chr12_A_all (Secondary 1 Binding Site) F	IDT	CACCGACGTCGCATAGAACGGAAGAGCGTGTG GACAGCCGGTTCGGTCGTTTC
chr12_A_all (Secondary 1 Binding Site) R	IDT	TAATACGACTCACTATAAGGGCGGTCCCCTCCG AGGTATAC
chr12_B_all (Secondary 5 Binding Site) F	IDT	TAGCGCAGGAGGTCCACGACGTGCAAGGGTG TTCGTTACCGCGCGTTGAAG
chr12_B_all (Secondary 5 Binding Site) R	IDT	TAATACGACTCACTATAAGGGCGGTCCCCTCCG AGGTATAC
chr12_I_all (Secondary 6 Binding Site) F	IDT	CACACGCTCTCGTCTGGCCGTGGTCGATCA GCGATCTGCGCATGTTAAC
chr12_I_all (Secondary 6 Binding Site) R	IDT	TAATACGACTCACTATAAGGGCGGTCCCCTCCG AGGTATAC
Secondary 1- Alexa488	IDT	ACACACGCTCTTCCGTTCTATGCGACGTGCGGTGA
Secondary 1- Alexa647	IDT	ACACACGCTCTTCCGTTCTATGCGACGTGCGGTGA
Secondary 5- Atto565	IDT	ACACCCCTTGACGTGCTGGACCTCTGCGCTA
Secondary 6- Alexa647	IDT	TGATCGACCACGCCAAGACGGAGAGCGTGTG
Software and Algorithms		
Juicebox	Durand et al., 2016	<a href="http://aidenlab.org/juicebox/">http://aidenlab.org/juicebox/</a>
Cell Profiler	McQuin et al., 2018	version 3.1.9
FIJI	Schindelin et al., 2012	version 2.0.0-rc-69/1.52p
HiC-Pro	Servant et al., 2015	version 2.10.0
Bioconductor	Huber et al., 2015	release 3.11
Code supporting this study	This paper	<a href="https://github.com/aryeelab/colon-dna-topology">https://github.com/aryeelab/colon-dna-topology</a>
OligoMiner scripts	Beliveau et al., 2012	<a href="https://github.com/beliveau-lab/OligoMiner">https://github.com/beliveau-lab/OligoMiner</a>

## RESOURCE AVAILABILITY

### Lead Contact

Further information and requests for resources and reagents should be directed to the Lead Contact, Bradley Bernstein ([Bernstein.Bradley@mgh.harvard.edu](mailto:Bradley@mgh.harvard.edu)).

### Materials Availability

No unique reagents were generated for this study.

### Data and Code Availability

All next generation sequencing data generated in the study were deposited at dbGaP and the Gene Expression Omnibus (GEO): GSE133928. Original data including all raw microscopy images were deposited at Mendeley Data: <https://dx.doi.org/10.17632/6k4hjkfw76.1>. Code supporting the study is deposited at Github: <https://github.com/aryeelab/colon-dna-topology>.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Human tumor specimens

Tumors included in this cohort are either collected as part of a Massachusetts General Hospital (MGH) Pathology discarded tissue tumor banking protocol or on a consented protocol. All samples were acquired with Institutional Review Board approval (2012P001475/PHS). For the discarded tissue cohort, the tissue was collected from colonic adenocarcinomas at the time of surgical resection at MGH prior to 2015 and saved as part of a de-identified tissue bank. As a de-identified cohort, no clinical data is available. A second set of tumors came from patients at the MGH who were consented preoperatively to take part in the study. For a subset of these patients, normal colon was taken from adjacent normal areas in resection specimens from consented patients. Both normal and tumor tissue from the consented cohort was snap frozen. A summary of the tissue (normal and tumor) samples is provided in [Table S1](#). Tumor genotyping was based on the SNaPshot assay, performed at Massachusetts General Hospital ([Dias-Santagata et al., 2010](#)). Our clinical cohort included 26 tumors, and 7 normal colon tissue samples ([Table S1](#)).

### Cell lines

Our *in vitro* models included colon cancer cell lines (HCT116, SW480, RKO, LS-174T), a line derived from normal fetal colonic epithelium (FHC) and a primary fibroblast line (WI-38). Colon cell lines were purchased from ATCC: HCT116 (CCL-247), SW480 (CCL-228), LS174-T(CL-188), RKO (CRL-2577) and FHC (CRL-1831). The primary fibroblast line WI38 was obtained from Coriell (AG06814-N). HCT116 and SW480 were grown in McCoy's 5A medium (GIBCO 16600082), 10% FBS and 0.5% pen-strep (GIBCO 10378016). LS174T and RKO were grown in EMEM (ATCC 30-2003), 10% FBS and pen-strep (GIBCO 10378016). FHC was cultured as per ATCC in DMEM/F12 media (ATCC 30-2006), 25mM HEPES, 10 ng/ml cholera toxin, 0.005 mg/ml insulin, 0.005 mg/ml transferrin, 100 ng/ml hydrocortisone, 20 ng/ml recombinant EGF (Thermo Fisher PHG0311) and 10% FBS. WI38 was cultivated in EMEM with 15% FBS and passaged serially (approximately twice weekly) for 14 weeks.

### METHOD DETAILS

#### Tissue dissociation and crosslinking

Tumor and normal colon samples were diced on dry ice into small (< 1mm) pieces and resuspended in ice cold PBS (GIBCO 10010-023). When crosslinking, formaldehyde was added to 1% and tissue was rotated at room temperature (RT) for 15 minutes. Glycine (2.5 M) was added to quench the formaldehyde and samples were rotated for an additional 5 minutes at RT. For cell lines, cells were pelleted, crosslinked in 1% formaldehyde for 10 minutes at 37 degrees and quenched with 2.5M glycine. All crosslinked samples were washed in ice cold PBS with protease inhibitors, pelleted and flash frozen in liquid nitrogen.

#### Hybrid selection bisulfite sequencing

Total genomic DNA was isolated using the DNAeasy Blood & Tissue Kit (QIAGEN) and sheared using the Covaris LE220. Ampure XP beads (Agencourt) were then used to size select gDNA fragments within 150-320bp, and sheared distribution was verified via Bio-Analyzer (Agilent) prior to continuation. Sheared genomic DNA (1 ug) was then generated into a sequencing library via end repair, 3' A base tailing (KAPA Hyper Prep Kit #KK8502) and sequencing adaptor ligation (Roche SeqCap Epi Enrichment System). Post-ligation clean-up was performed using Ampure XP beads. Following bead clean-up, DNA library was then bisulfite converted using the EZ DNA Methylation-Lightning Kit (Zymo Research) and amplified via PCR using KAPA HiFi U+ HotStart ReadyMix (KAPA #KK2800). Equal concentration of each bisulfite-converted library was then combined in sets of either three or four libraries/per pool along with SeqCap Epi universal & indexing oligos and bisulfite capture enhancer (SeqCap Epi Accessory Kit). Each pool was subsequently lyophilized using TOMY Micro-Vac (MV100) and resuspended in hybridization buffer (SeqCap Epi Hybridization and Wash Kit) prior to being hybridized to SeqCap Epi Probe Pool (Roche) for 72 hours at 47C in a thermocycler with a heated lid at 57C. Following the 72-hour incubation, captured bisulfite-converted libraries were recovered (SeqCap Pure Capture Bead Kit) at 47C in a thermocycler for 45 minutes, with intermediate vortexing every 15 minutes. Capture beads were then washed (SeqCap Hybridization and Wash Kit) in a 47C water bath at room temperature, respectively. Captured bisulfite-converted libraries were amplified via PCR (SeqCap Epi Accessory Kit). Libraries were sequenced with 10% PhiX spike-in as 100-base paired end reads on the HiSeq2500 in rapid run mode.

#### Whole genome bisulfite library preparation

Total genomic DNA was isolated using the DNAeasy Blood & Tissue Kit (QIAGEN) and for each sample, one microgram of DNA was sheared using the Covaris LE220. Libraries were prepared according to manufacturer's instructions using the NEXTFLEX bisulfite library prep kit (NOVA-5119-01). Libraries were sequenced on the Illumina NextSeq500 instrument.

#### ChIP-seq

We generated chromatin state maps (H3K27ac, H3K36me3, H3K9me3 and H3K27me3) and binding profiles for the CTCF insulator protein by ChIP-seq. ChIP-seq was performed as described previously ([Liau et al., 2017](#)). In brief, crosslinked cells were lysed and DNA was sheared to between 400 and 2,000 base pair fragments. Antibodies were as follows: CTCF (Cell signaling #3418), H3K27ac (Active Motif #39133), H3K9me3 (Abcam #8898), H3K27me3 (Cell Signaling #97335) and H3K36me3 (Abcam #9050). ChIP DNA was used to generate sequencing libraries by end repair (End-It DNA repair kit, Epicenter), 3' A base overhang addition via Klenow fragment (NEB), and ligation of barcoded sequencing adapters. Barcoded fragments were amplified via PCR. Libraries were sequenced as 38-base paired-end reads on an Illumina NextSeq500 instrument.

#### Hi-C

Hi-C maps of chromosome topology were initially generated for a cohort of 7 primary tumors, 4 normal colon tissue samples and 5 cell lines (cohort 1; [Table S1](#)). We then confirmed our results by acquiring Hi-C data for a validation cohort of 5 tumors and 3 normal colon samples (cohort 2; [Table S1](#)).

*In situ* Hi-C was performed as described previously ([Rao et al., 2014](#)). In brief, crosslinked cells or tumor were thawed on ice in Hi-C lysis buffer. Tissue samples were mechanically disrupted with the Biomasher tissue grinder (Kimble Chase). Tissue and cell line samples were permeabilized in 0.5% SDS at 37 degrees, quenched with Triton X-100 and chromatin was digested with 100-200U MboI at 37 degrees overnight. Nuclei were then pelleted, ends were marked with biotin-14-dATP (ThermoFisher 19524016) and chromatin was ligated for 5 hours by T4 DNA ligase (M0202). Samples were treated with proteinase K at 55 degrees for 30 minutes

and cross-links were reversed at 68 degrees overnight. DNA was ethanol precipitated and sheared on a Covaris LE220. DNA was cleaned up via AMPure XP beads (Beckman Coulter, A63881) and quantified by Qubit dsDNA High Sensitivity Assay (Life Technologies, Q32854). Samples were bound to Dynabeads MyOne Streptavidin T1 beads (Life technologies, 65602) and washed. End repair, dATP attachment and adaptor ligation was performed. Final PCR amplification was performed using barcoded sequencing primers and PCR. Libraries were purified using AMPure XP beads and sequenced on either a NextSeq500 (150 cycle kit), HiSeq2500 (high output; 200 cycle kit) or NovaSeq S4 (200 cycles).

### HiChIP

We acquired SMC1 HiChIP data for cohort 1 ([Table S1](#)). HiChIP was performed as described previously ([Mumbach et al., 2016](#)). Briefly, crosslinked samples were lysed in Hi-C lysis buffer and chromatin was permeabilized in 0.5% SDS at 63°C for 10 minutes. Chromatin was digested with MboI for 2 hours at 37°C. Overhangs were filled in and marked with Biotin-dATP (ThermoFisher 19524016), and ends were ligated with T4 DNA Ligase (NEB M0202) for 4 hours at room temperature. Nuclei were pelleted and lysed and chromatin was sheared on the Covaris E220 with the following conditions: Fill level 5, Duty Cycle 5, PIP 140, Cycles/burst 200, Time 4 minutes. Samples were clarified, diluted in ChIP dilution buffer and precleared with Protein G beads (Invitrogen 11205D) for 1 hour at 4°C. Samples were cleared on magnet and supernatant was added to antibody. Chromatin was ChIP'd overnight at 4°C with rotation. Protein G beads were added and incubated for 2 hours at 4°C with rotation. Beads were washed with low salt, high salt and LiCl buffer. Sample was eluted in ChIP elution buffer, treated with Proteinase K and crosslinks were reversed. DNA was purified using the Zymo clean & concentrate kit (DCC-100). Streptavidin M280 beads (Invitrogen 11205D) were washed and resuspended in 2x biotin binding buffer and DNA was bound for 15 minutes at room temperature. Beads were washed and Tn5 fragmentation was carried out as per Mumbach et al., 2016, with dilutions of Tn5 to process low input samples. Libraries were amplified using the Nextera DNA Library Prep kit (Illumina). Material was cleaned up using Ampure XP beads. Libraries were sequenced on NextSeq500 (150 cycle kit) or the HiSeq2500 (high output; 200 cycle kit).

### RNA-seq

Whole RNA was extracted using the QIAGEN RNeasy kit according to the manufacturer's protocol. For RNA-seq library preparation, Poly(A)<sup>+</sup> RNA was enriched using magnetic oligo(dT)-beads (Life Technologies) and then ligated to RNA adaptors for sequencing. RNA-seq was performed with two biological replicates per colon cancer line and in singlicate for tumor samples. Libraries were sequenced as 38-base paired-end reads on an Illumina NextSeq500 instrument.

### Electron microscopy

Fresh tissue biopsies were placed directly into EM fixative (2.5% glutaraldehyde, 2.0% paraformaldehyde, 0.025 calcium chloride in a 0.1M sodium cacodylate buffer, pH 7.4) and allowed to fix for 3 hours at room temperature or ON at 4°C. Further processing was done in an EMS (Electron Microscopy Sciences) Lynx II automatic tissue processor. Briefly, tissues were post-fixed with osmium tetroxide, dehydrated in a series of ethanol solutions, en block stained in the 70% ethanol step with uranyl acetate, further dehydrated in 100% ethanol and propylene oxide. Tissues were infiltrated in a series of propylene oxide, Epon mixtures and embedded in pure Epon. The Epon blocks were polymerized overnight in a 60°C oven. One micron sections were cut using glass knives and stained with toluidine blue. Representative areas were chosen by light microscopy. Thin sections were cut using an LKB ultramicrotome and diamond knife. The sections were stained with Sato's lead stain and examined with a FEI Mogagni transmission electron microscope. Images were captured with an AMT (Advanced Microscopy Techniques) 2K digital CCD camera.

### DNA-FISH

#### Generation of DNA-FISH probes

Based on evidence of CNV stability in tumors with available biological material, chromosome 12 was chosen for validation experiments. Only the p-arm of chromosome 12 was considered to avoid measuring arm-related differences in nuclear positioning. For compartments A and B, consecutive regions of at least 300Kb with absolute PC1 values larger than 0.5 in normal colon were identified (see methods section of eigenvector decomposition of Hi-C matrices). For these consecutive segments, the 100Kb regions at the middle of each segment were selected. For compartment I, a set of 100Kb regions was selected so that their linear distances to the B candidate regions was equivalent to the distance between the A candidate regions and the B candidate regions. Furthermore, the candidate regions were screened to have the same compartment-specific characteristics in HCT116 cells. Then, oligopaint libraries were designed using the Oligominer pipeline ([Beliveau et al., 2012](#)). Specifically, candidate regions were mined for probes using the length requirement of 80 nucleotides of homology, melting temperature range of 47–80°C, and default settings for the remaining Oligominer parameters. Resulting oligos yielded an average probe density of 4.6 probes/kb. An oligo pool (Twist Bioscience) was synthesized such that all probes targeting A, B, or I regions could be created in aggregate. Single-stranded probes were produced using PCR, T7 RNA synthesis, and reverse transcription as described previously ([Rosin et al., 2018](#); [Shav-Tal, 2013](#)).

#### DNA-FISH on tissue

For DNA-FISH, tissue samples were fixed in 4% paraformaldehyde (FisherScientific, 15710) for 2 hours, washed in 1X PBS and soaked for 2x 10 minutes in 0.5M NH4Cl (26.6g/L in PBS, Sigma-Aldrich, 213330). The samples were cryoprotected by overnight incubation in 30% w/v sucrose (Sigma-Aldrich, S0389) in PBS at 4°C, nutating. The next morning, the sucrose was replaced with

30% sucrose, 50% OCT (Tissue-Tek<sup>\*</sup> O.C.T. Compound, Sakura, 25608-930) for two hours at 4°C, nutating. They were embedded in plastic peel-away molds in OCT, frozen on a cold block in liquid nitrogen, and stored in the freezer until sectioning. 5 µm sections were sectioned using a cryostat, collected on Superfrost Plus microscope slides (FisherSci 22-037-246) and stored at –20°C.

Slides were thawed at room temperature for 30' and rehydrated in 10 mM sodium citrate for 5'. Next, the slides were incubated for 10' in 10 mM sodium citrate at 80°C and allowed to cool down at room temperature for ~30'. They were washed twice at 2X SSC, 5' each, and transferred to 50% formamide (Sigma Aldrich, F9037) in 2X SSCT (SSC + 0.1% Tween-20) for at least 1h. Following denaturation, the tissue was dehydrated in an ethanol row (70%, 95%, 100%, three minutes each) and air-dried for at least 90 minutes. They were then acetylated with acetic anhydride as follows: dried slides were equilibrated in 0.1M fresh ethanolamine (Sigma Aldrich, 90279) in dH<sub>2</sub>O, pH 8.0 for 10 minutes, then transferred to 0.25% v/v acetic anhydride (Sigma Aldrich, 320102) in 0.1M ethanolamine for 5' followed by washing for 10' in 2X SSCT. During the previous steps, probes against compartments A, B and I were prepared by adding 1 µL of 100 µM dNTPs (Life Technologies, R1121) and 50 pmol probes each per slide, speed vacuuming on high for 20' (or until the liquid had evaporated), and resuspending the probes in 5.25 µL per sample. This probe-dNTP mix was then added to the hybridization mixture containing a final concentration of 50% formamide, 10 µg RNaseA (Thermo Scientific EN0531) and 1x Dextran Sulfate Mix (10% Dextran Sulfate D8906, Sigma Aldrich, 2xSSC, 0.1% Tween-20). Next, the slides were mounted with probe mix, covered with a glass coverslip and rubber cement (Staples, EPI231) and incubated on a hot block at 42°C for 1-3 hours to allow infiltration with the probe. Then, the slides were heat-shocked to denature the section and probe at 85°C for 7 minutes and incubated overnight (16h+) at 37°C in a temperature-controlled oven. The next day, the coverslips were removed from the sections and washed at 60°C for 15 minutes in pre-warmed 2X SSCT, at room temperature for 10' in 2X SSCT, at room temperature for 10' in 0.2X SSCT, and transferred to 2X SSC while secondary mix was prepared consisting of 10% formamide, 1x Dextran Sulfate Mix and 10 pmol of secondary probes labeled with Cy3 (B compartment), Cy5 (A or I compartment) and A488 (A compartment) in dH<sub>2</sub>O. The sections were incubated with secondary mix in a humid chamber at room temperature in the dark for at least two hours, followed by the same washes as before. In the last step with 2X SSC, the nuclei were counterstained with DAPI (Sigma-Aldrich, D1305, 5 µg/mL final concentration), briefly washed in 2X SSC and mounted using SlowFade Gold Antifade Mountant (Invitrogen, S36936).

#### DNA-FISH on HCT116 cells

Cells were grown on glass coverslips, fixed in 4% PFA for 15 minutes, and processed as described above with the following adjustments: sodium citrate and acetic anhydride treatments were omitted; instead, the cells were permeabilized after fixation using 0.5% Triton X-100 for 15' followed by a 5' wash in PBS and alcohol denaturation. Prior to probe mix addition, the samples were re-equilibrated using 2X SSCT + formamide (50% v/v 4x SSCT + 50% v/v formamide) for 1h at 37°C. The probes were then immediately denatured at 80°C for 5' and incubated overnight at 37°C. The DNA-FISH stainings were imaged on a Zeiss LSM800 confocal microscope with Airyscan settings and a 63x oil objective.

#### Treatment with 5-azacytidine

5-azacytidine was obtained from SelleckChem (S1782) and HCT116 cells were cultured in the presence of 5 µM or 1:2000 DMSO control. Cells were plated in media containing 5-aza or DMSO and harvested at 24 h.

## QUANTIFICATION AND STATISTICAL ANALYSIS

#### DNA methylation data preprocessing and quantification

BSMAP version 2.74 was used both to map the sequenced reads to the reference genome (hg19) and to calculate the methylated fraction for each CpG across the genome (Xi and Li, 2009). Data was further analyzed within the framework of the bsseq Bioconductor package (Hansen et al., 2012).

In order to avoid potential biases introduced by the DNA capture enrichment step, genomic coordinates of hypomethylated blocks were defined based on differentially methylated regions (DMRs) using previously published whole-genome bisulfite sequencing data (Hansen et al., 2011). DMRs with methylation differences larger than 10% between tumors and normal samples were considered. Consecutive DMRs smaller than 500Kb bins were merged if the genomic distance between the DMRs was smaller than 10% of the width of the individual DMRs. Only merged DMRs larger than 100Kb were considered for further analysis in order to match the resolution of the compartment calls of the Hi-C data. To quantify block-level methylation in our samples, we used only methylation at open-sea CpGs, i.e., those located >2kb from CpG islands. We confirmed that hypomethylated blocks were recapitulated in our capture bisulfite sequencing data, displaying methylation differences both between tumor samples and normal samples, and between regions inside a block and the flanking genomic regions.

For each sample, the degree of CIMP hypermethylation was assessed by measuring average methylation at CIMP-specific methylated islands (Xu et al., 2012). Samples with the highest methylation levels at these sites were labeled as CIMP. A clinical genotyping assay (Dias-Santagata et al., 2010) revealed BRAF mutations in the two CIMP tumors, consistent with reported associations (Hinoue et al., 2012). TCGA methylation data was used to verify that the methylation levels of our CIMP samples were comparable to the methylation levels of CIMP tumors in TCGA. TCGA DNA methylation data was downloaded using TCGAbiolinks (Colaprico et al., 2016). Comparison of CpG island hypermethylation and block hypomethylation across our samples and a larger cohort of colon tumors from the Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Network, 2012) confirmed that CIMP and block hypomethylation are indeed independent features.

### Cancer-associated DNA methylation alterations in normal colon tissue

We obtained pre-processed Illumina EPIC and 450k methylation array beta values data from the colon tissue samples described in Wang et al., 2020 (NCBI GEO GSE132804 Series Matrix files). Analysis was restricted to probes common to the two platforms. We computed a block hypomethylation score for each sample as the average DNA methylation of open-sea CpGs (i.e., those > 2kb from a CpG island) within hypomethylated blocks, using the same block region definitions used elsewhere in the study for tumor versus normal comparisons. We verified that the low and high risk group samples were spread across the two platforms (EPIC: 52 high risk, 57 low risk; 450k: 49 high risk, 48 low risk), and that results were consistent when analysis was performed for each platform separately.

### Hi-C analysis

We confirmed data quality by assessing the fraction of cis-long range contacts for each library. Each map contained an average of 320 million contacts, for an average resolution of 10 Kb. We used these data to derive TAD and compartment structures.

Data were controlled for quality, mapped to the reference genome (hg19) and converted into interaction matrices using HiC-Pro v2.10.0 ([Servant et al., 2015](#)) using pipeline code available at [https://github.com/aryeelab/topology\\_tools](https://github.com/aryeelab/topology_tools). Within sample normalization was performed using the Iterative Correction and Eigenvector decomposition (ICE) method ([Imakaev et al., 2012](#)). For each chromosome in each sample, compartments were called using the standard PCA method ([Lieberman-Aiden et al., 2009](#)). Briefly, the interaction matrix  $X = x_{ij}$  was transformed into an observed over expected (O/E) matrix by dividing each element of the matrix by the expected interaction frequency for a given distance from the diagonal  $k = i - j$ , defined as the mean of the values  $x_{ij}$  with the same value of  $k$ . A correlation matrix was generated by estimating the pairwise correlation coefficients of all the rows of the O/E matrix. Then, an eigendecomposition was performed on the correlation matrix, and the sign of the first eigenvector was used to assign compartment labels. We used expression data and GC content to flip the sign of the eigenvector such that values larger than 0 correspond to open (A) regions and values smaller than 0 correspond to closed (B) regions.

In addition to using the eigenvector (PC1) metric, we also directly quantified the tendency of each region to interact with other regions in either the A or B compartments. We calculated the “A/B interaction ratio,” defined for each 100kb genomic window as the ratio of interaction frequency with the A versus B compartments using the O/E matrix. Specifically, we calculate  $\log_2(\text{mean O/E interaction frequency with A regions}) - \log_2(\text{mean O/E interaction frequency with B regions})$ . For each comparison, A and B were defined in the ‘baseline’ condition: normal colon tissue, HCT116 cells treated with DMSO, or early passage fibroblasts (passage 16). The same A and B definitions were used for all samples within a comparison. We confirmed that assessing the relationship between compartmental change and hypomethylation gave consistent results when using A/B interaction ratio instead of eigenvector as the outcome (See “[Association of compartmental organization with DNA hypomethylation](#)”).

In primary tissues, compartment I was defined as those genomic regions with a positive value of the first eigenvector that were within a block of DNA hypomethylation (defined by comparing tumors versus normal). We verified that compartment I could be identified based only on Hi-C data. To do this, we plotted the first 5 eigenvectors of the matrix decomposition used to define compartments A and B. The first eigenvector separates compartment A and B. Compartment I could be identified using the second eigenvector on several chromosomes. In other chromosomes, compartment I was often separated by lower eigenvectors, and became more evident when we applied the matrix decomposition method to individual chromosome arms. For IMR90 cells, compartment I was assigned as in primary tissues using DNA methylation differences between proliferating and senescent cells. For HCT116, 100Kb genomic bins were labeled as I if two consecutive genomic bins had a positive value on the eigenvector decomposition described ([Lieberman-Aiden et al., 2009](#)), and if their open sea CpG DNA methylation values were equal or less than 80%. We confirmed that this approach enriched for regions that were consistent with our definition of compartment I, i.e., that had intermediate A/B contact Hi-C patterns and that were enriched for the H3K27me3 chromatin mark compared to compartments A and B.

Sub-compartment structures defined by Rao et al., 2014 were called using SNIPER ([Xiong and Ma, 2019](#)). SNIPER uses Hi-C data and sub-compartment calls to train a model that is able to learn the interaction patterns of each sub-compartment. These trained models can be used to define sub-compartments from Hi-C data that are distinct from the training Hi-C data. To call sub-compartments in colon tissue, we used the pre-trained models provided by SNIPER that were trained using 5% of GM12878 Hi-C data ([Rao et al., 2014](#)), as these data were of a comparable coverage to our individual HiC samples.

Insulation scores were calculated by defining two windows of length  $l$ , one upstream and one downstream of a given genomic position. For each chromosome, the log (base 2) ratio of the sum of interaction counts within each window and the interaction counts between the two windows was calculated at each position. These log ratios were further transformed into z-scores by subtracting the median and dividing by the median absolute deviation. In order to capture boundaries of TADs of varying sizes, including nested TAD structures, the z-scores were calculated genome-wide for different values of  $l$ , specifically  $l = 200\text{Kb}$ ,  $l = 400\text{Kb}$  and  $l = 800\text{Kb}$ . Boundaries were called for genomic positions where the z-scores were larger than the 90% quantile of a standard normal distribution in at least one resolution in one sample. Three strategies were used to assess the stability of TADs across samples: (1) we calculated the correlation coefficients of the insulation scores, (2) for each sample, we defined the position of TAD boundaries in that sample and calculated the percentage of TAD boundaries where the insulation scores of other samples were also in a local minima and (3) we plotted metaplots of the insulation scores to visually inspect conservation of TAD boundaries.

Juicebox ([Durand et al., 2016](#)) was used for exploratory visualization of the Hi-C data.

### Association of compartmental organization with DNA hypomethylation

We used linear mixed effects models to assess the association between open-sea CpG hypomethylation and Hi-C eigenvector in three settings: 1) tumor versus normal colon, 2) HCT116 cells treated with DMSO (control) or 5-azacytidine, and 3) aging WI38 fibroblasts sampled at passage 16, 30 and 40. For each 100kb window in each sample we computed DNA hypomethylation relative to a baseline condition (normal colon, HCT116+DMSO, or WI-38 passage 16). In each case, we fit linear mixed effects model to I and B separately with eigenvector as the outcome, hypomethylation bin as a fixed effect and random effect terms for sample and genomic window. The intercept corresponds to regions with minimal (< 15%) hypomethylation, and the coefficient estimates shown in Figure 5 represent the mean eigenvector change relative to these methylation-stable regions. An additional model was used to assess PC1 change over time in WI-38 fibroblasts for 100kb windows that show > 20% hypomethylation in late passage (P40) versus early passage (P16) WI-38 fibroblasts. This model uses passage number as the fixed effect instead of hypomethylation.

### HiChIP analysis

Data were controlled for quality, mapped to the reference genome (hg19) and converted into interaction matrices using the HiC-Pro pipeline (Servant et al., 2015). Chromatin loops were called using the hicchipper 0.7.3 with the parameter to use user-defined peaks (Lareau and Aryee, 2018b). hicchipper was run for each HiChIP sample using the union of CTCF and H3K27ac peaks as the predefined peak set. The hicchipper pipeline defines potential loop anchors by extending peaks by a fixed window (i.e., 500bp) to account for uncertainty in the peak calling and merging peaks whose genomic distance is below 500bp. These extended peaks are overlapped with restriction fragments and are further extended to the edges of the restriction fragments they overlap. For each pair of potential loop anchors, hicchipper counts the number of valid contact pairs (defined by HiC-Pro) that support their 3D interaction. Running this step for each sample results in a matrix  $Z_{ij} = z_{ij}$  where each column is a sample  $j$  and each row  $i$  represents a pair of loop anchors (i.e., a loop), and  $z_{ij}$  is the number of valid read pairs that supports a loop  $i$  in sample  $j$ . To distinguish between random background contacts from contacts due to DNA looping, hicchipper runs the mango background correction model on the sum of loop counts across samples (i.e.,  $\sum_j z_{ij}$ ). The mango correction consists of modeling the counts using a binomial distribution to estimate the probability of

observing the counts between two genomic loci given its genomic distance. The resulting p values are corrected for multiple testing. Loops with a q-value smaller than 0.1, with at least 4 valid contacts in two or more samples and with at least 20 counts across all samples were considered high-confidence loops and were considered for further analysis. Significant loops were annotated as enhancer-promoter loops if one of the anchors overlapped an H3K27ac peak (enhancer-like) and the other anchor overlapped the promoter of a gene.

To assess the robustness of our results to the loop calling algorithm, we repeated our analyses using a different loop calling algorithm, cLoops (Cao et al., 2019). We ran cLoops version 0.92 using the parameters “-hic -eps 2500,5000,7500,10000 -minPts 3,5,10 -j -s -w” for each sample and used the union of the per-sample significant loops as our final set of loops. The global trends described in the main text were robust to the loop calling algorithm.

The software tool diffloop was used to test for differential looping (Lareau and Aryee, 2018a). diffloop uses the statistical engine of the edgeR package (Robinson et al., 2010), where the matrix of loop counts  $Z_{ij}$  is modeled using generalized linear models (GLM) of the negative binomial distribution:

$$Z_{ij} \sim NB(\mu_{ij}, \alpha_i)$$

$$\mu_{ij} = n_{ij}z_{ij}$$

where  $\mu_{ij}$  is the fitted mean and  $\alpha_i$  is the dispersion estimate, which is estimated using the common dispersion method from edgeR.  $n_{ij}$  are normalization factors that account both for library size and for copy number differences between samples. Specifically, we overlapped the genomic coordinates of the loop anchors with the copy number estimates of each sample (see Section “Copy number variant analysis” for details) to generate a matrix of copy number estimates (loops  $i$  times sample  $j$ ). This matrix of copy numbers was row-centered. The resulting matrix was multiplied by the library size factors estimated by edgeR and the resulting values defined  $n_{ij}$ , which were introduced as offsets when fitting the GLMs. To test for differential looping between conditions, a GLM was fitted for each gene and a likelihood ratio test was used to calculate p values for each loop and the Benjamini-Hochberg was used to correct for multiple testing.

### Copy number variant analysis

The coverage of our DNA methylation sequencing assays was used to assess copy number alterations in tumors. For each sample, the number of sequencing reads was counted for non-overlapping fixed-width genomic regions of 40Kb and CNAnorm was used to infer copy number alterations. CNAnorm inputs two vectors of read counts,  $w_{kl}$  and  $y_{kl}$ , which represent read counts for genomic region  $l$  of chromosome  $k$  for a tumor sample and a matching normal sample, respectively. CNAnorm performs the following steps to determine copy number alterations. First, the ratio of the read counts is calculated,  $r_{kl} = \frac{w_{kl}}{y_{kl}}$ . This ratio is normalized using a loess-based method that defines  $r_{kl}^{norm}$ , which removes technical dependencies between  $r_{kl}$  and GC content. Second, to remove random

error variability,  $r_{kl}^{norm}$  is smoothed throughout the genome (Huang et al., 2007). Third,  $r_{kl}^{norm}$  values are transformed so that the most common genomic regions are centered to ratio one. Then, CNAnorm normalizes for tumor purity by shrinking  $r_{kl}^{norm}$  so that the modes of the distribution fit values resulting from copy number alteration processes (deletion = 0, deletion of one chromosome copy = 0.5, no CNV = 1, amplification of one copy = 1.5, etc). Finally, a circular binary segmentation algorithm is used to define regions with copy number alterations (Olshen et al., 2004).

Since not all our tumor samples contained a matching normal sample, we used the normal sample with the highest coverage as the reference sample for all our tumors. We used copy number calls to verify that the epigenetic differences between tumors and normal were not only driven by copy number alterations by introducing the estimated CNV as offsets in the statistical models when doing inferences and making sure that the epigenetic differences were present after we masked genomic regions with CNVs.

### SNP Analysis

We identified 28 colon cancer risk SNPs (MacArthur et al., 2017) that coincide with E-P loop enhancer anchors and assigned target genes based on the corresponding promoter contact (Table S4). These looping data confirmed predicted targets, including risk SNPs previously associated with COLCA1/2 or TERC expression (Figures S2I–S2K; Peltekova et al., 2014). 5 of the 28 risk SNPs were associated with distal genes rather than the nearest promoter (Table S4).

### Gene expression analysis

#### Download and processing of transcriptome data

RNA-seq processed data from TCGA was downloaded programmatically using TCGAbiolinks (Colaprico et al., 2016) and recount2 (Collado-Torres et al., 2017). Gene count normalization was done using DESeq2 (Love et al., 2014) and statistical inference was done using limma-voom (Law et al., 2014). Genomic ranges operations were done using GenomicRanges infrastructure (Lawrence et al., 2013). RepeatMasker was used to extract the genomic coordinates of repeat elements (Smit et al., 2015). For each repeat element in the human genome, we calculated the total coverage for each TCGA sample and followed the default recount2 pipeline to estimate scaled read counts. Our in-house RNA-seq data was quantified using salmon (Patro et al., 2017) and differential expression analyses were done as described in Bioconductor's RNA-seq workflow (Love et al., 2015).

### Pan-cancer analysis of gene expression associated to hypomethylation

To test for association between block level (open sea) DNA methylation and gene expression, we defined a block score as the mean open sea CpG DNA methylation across hypomethylation blocks for each TCGA sample. To minimize confounding correlations driven by stromal cell fractions, only tumors with stromal cell fractions smaller than 60% were considered for further analysis (Thorsson et al., 2019). Furthermore, the remaining TCGA tumors were grouped into 5 equally-sized bins according to their stromal cell fraction content. A limma-voom model was fitted using as predictors the block scores and the stromal cell fraction bins (i.e., categorical labels) as a blocking factor. This strategy enabled the identification of genes correlated with hypomethylation, while adjusting for potential confounding by stromal cell content. Using this pipeline, we defined genes associated to block hypomethylation in each tumor type. We tested for oncogene enrichment using Fisher's exact test using published oncogene annotations (Liu et al., 2017).

### Exclusion of genes of likely non-tumor cell-origin

For downstream analyses, we sought to exclude from analyses those genes whose expression patterns could be driven by differences in stromal or immune cell composition by two filters: 1) We downloaded gene expression data from cell populations purified using fluorescence activated cell sorting (Calon et al., 2015) and excluded genes where the average expression level was higher (> 2 fold change) in fibroblast cells compared to epithelial cells. 2) We downloaded GTEx blood RNA-Seq data from the recount2 project (Collado-Torres et al., 2017) and scaled toward a target read count of 1,000,000 reads using the scale\_counts function of the recount R/Bioconductor package. We excluded genes where the median scaled count in blood samples was greater than 10.

### Survival analysis

Preprocessed Affymetrix U133Plus2 microarray gene expression data (NCBI GEO GSE39582) was downloaded using the curate-dCRCDATA R/Bioconductor package (Marisa et al., 2013). Eight of 566 samples with low average pairwise correlation (< 0.9) with other samples were excluded in a QC filtering step. Gene expression values were Z-score transformed.

We computed a gene expression risk score as the average normalized expression level of the 146 genes that are downregulated with hypomethylation in the I and B compartments for samples from Marisa et al. (2013) and the TCGA COAD cohort (Marisa et al., 2013). For Marisa et al. (2013), we restricted to the 145 genes where the corresponding gene symbol was present. We defined a "high" score as those values that were 2sd (robustly estimated with the R 'mad' function) above the median for the Marisa et al. (2013) study, and 1 sd above the median for the TCGA cohort. We constructed Kaplan-Meier survival curves using available data for recurrence-free survival (Marisa et al., 2013) and overall survival (Cancer Genome Atlas Network, 2012). To assess the association of expression risk score with survival outcomes while adjusting for known risk factors we fit a Cox Proportional Hazards model to data from the Marisa et al. study, using gene expression risk score, clinical stage, BRAF mutation status and MSI status as predictors.

### ChIP-seq analysis

Reads were mapped to the reference genome (hg19) using bwa version 0.7.12 (Li and Durbin, 2009). CTCF peaks were called using GCAPC with default parameters (Teng and Irizarry, 2017) and H3K27ac peaks were called using MACS (Zhang et al., 2008). The data revealed the expected punctate peaks of the enhancer-associated mark, H3K27ac, as well as broader regions of the repressive modifications, H3K9me3 and H3K27me3. CTCF binding sites were highly enriched for the CTCF binding motif (OR = 19.04;  $p < 10^{-15}$ ). For differential CTCF peak analysis, the union of peaks that were detected in at least two samples were considered and reads were counted for each peak in each sample. Differential CTCF binding was inferred using DESeq2, introducing offsets to the generalized linear model to normalize for library size (Love et al., 2014), copy number differences, and non-linear trends (Lun and Smyth, 2016). Copy number estimates for each genomic region were obtained using CNAnorm (Gusnanto et al., 2012). To account for technical experimental variation, the loadings of the first principal component were introduced as a covariate on the generalized linear model. CTCF sites were considered lost if they had a q-value smaller than 0.1 and a methylation difference larger than 20%. To evaluate the effects of methylation of CTCF sites in chromatin looping, we performed aggregate peak analyses based on a comprehensive list of loops annotated on the human genome (Rao et al., 2014). Lost CTCF peaks were assigned to loops if they overlapped with loop's bidirectional CTCF motifs. To assign genes upregulated upon TAD boundary disruption, only peaks within 50Kb of a TAD boundary were considered. For each gene near a disrupted TAD boundary, a linear model was fitted on TCGA data using gene expression as a response variable and the average methylation at the lost CTCF site as predictors. Genes were selected if they had a significant positive (at a false discovery rate of 15%) association between their expression and the methylation levels at the corresponding TAD boundary CTCF.

### DNA polymer modeling

To visualize genome organization in normal tissues and primary tumors, we applied a computational approach to derive polymer models and 3D structures for each sample. As illustrated in Figure 2D, these models were fine-tuned to ensure that simulated in-silico contact maps reproduce the corresponding experimental results. We followed the same protocol to parameterize the models for different samples independently, and any difference between simulated structures, if observed, should reflect the alterations in genome organization detected by Hi-C.

The polymer models explicitly consider each individual chromosome as a string of beads for structural representation. Only one copy of the two homologs was included since the Hi-C data do not provide allele-specific contacts. We studied the genome organization at two resolutions that represent each bead as either 1Mb or 100kb long genomic segments. The 1Mb-resolution model is computationally efficient and allows us to explore the genome organization in multiple samples. On the other hand, the 100kb-resolution model provides a more detailed representation of the genome and will enable us to characterize the spatial localization of compartment I, introduced in the main text.

A key innovation of the polymer modeling approach is its use of an ensemble of structures, instead of a single, unique conformation, to reproduce Hi-C data. The ensemble of structures is assumed to follow a Boltzmann distribution with a potential energy function  $U_{ME}(\mathbf{r})$ , the expression of which can be derived following the maximum entropy principle (Qi et al., 2020; Qi and Zhang, 2019; Zhang and Wolynes, 2015, 2016) and is provided below. Parameters of the polymer model are solely encoded in the energy function, and their values were determined iteratively such that the simulated structures reproduce Hi-C contact maps. Molecular dynamics simulations were carried out to collect structures consistent with the energy function and the Boltzmann distribution.

### Energy function

The potential energy function for the genome adopts the following form:

$$U_{ME}(\mathbf{r}) = \sum_I [U(\mathbf{r}_I) + U_{ideal}(\mathbf{r}_I)] + U_{compt}(\mathbf{r}), \quad [1]$$

where  $\mathbf{r}$  represents the 3D conformation of the entire genome.  $I$  indexes over different chromosomes and  $\mathbf{r}_I$  corresponds to the conformation of chromosome  $I$ . By definition,  $\mathbf{r} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{23}\}$ .  $U(\mathbf{r}_I)$  and  $U_{ideal}(\mathbf{r}_I)$  are generic potentials shared by all chromosomes, and  $U_{compt}(\mathbf{r})$  describes compartment-type-specific interactions within the same chromosome and between different chromosomes.

Specifically,  $U(\mathbf{r}_I)$  is the energy function for a confined homopolymer and consists of four terms,  $U_{bond}$ ,  $U_{angle}$ ,  $U_{sc}$  and  $U_c$ .  $U_{bond}$  is the bonding potential between neighboring beads.  $U_{angle}$  is the angular potential applied among every three neighboring beads to define the persistence length of the polymer.  $U_{sc}$  is a soft-core potential applied to all the non-bonded pairs to enforce the excluded volume effect among genomic loci.  $U_c$  models a spherical boundary and is introduced to mimic the confinement effect applied by the nuclear envelop onto the chromosomes. The radii of the spherical confinement is chosen to ensure a volume fraction of 0.1. Explicit expressions for  $U(\mathbf{r}_I)$  can be found in Zhang and Wolynes, 2015 and Qi and Zhang, 2019.

$U_{ideal}(\mathbf{r}_I)$  is introduced to reproduce the power law decay of the contact probability as a function of genomic separation for each chromosome (Di Pierro et al., 2016; Qi and Zhang, 2019). It describes the tendency for chromosomes to collapse and form territories in addition to what has been enforced by the confinement potential  $U_c$ .  $U_{ideal}(\mathbf{r}_I)$  is defined as

$$U_{ideal}(\mathbf{r}_I) = \sum_{i,j \in I} \alpha_{ideal}(|j-i|) f(r_{ij}), \quad [2]$$

where  $f(r_{ij})$  determines the contact probability of a genomic pair with a spatial distance of  $r_{ij}$ , and  $i, j$  index over all pairs of non-bonded chromatin beads from chromosome  $I$ . Following Qi and Zhang (2019), we define  $f(r)$  as

$$f(r) = \begin{cases} \frac{1}{2}[1 + \tanh(\sigma(r_c - r))], & \text{if } r \leq r_c \\ \frac{1}{2}\left(\frac{r_c}{r}\right)^4, & \text{if } r > r_c \end{cases} \quad [3]$$

where  $r_c = 2.0$  and  $\sigma = 2.0$ .  $\alpha_{ideal}(|j-i|)$  measures of the strength of the contact at a given genomic separation  $|j-i|$  and its value can be determined from Hi-C data as detailed below. It contributes to a total of  $N - 1$  parameters, where  $N$  is the number of beads for the longest chromosome (chromosome 1, 249Mb).

For the 1Mb-resolution model,  $U_{compt}(\mathbf{r})$  is defined as

$$U_{compt}(\mathbf{r}) = \sum_I \sum_{i,j} \alpha_{intra}(C'_i, C'_j) f(r_{ij}) + \sum_{I,J} \sum_{i \in I, j \in J} \alpha_{inter}(C'_i, C'_j) f(r_{ij}), \quad [4]$$

where  $I$  and  $J$  index over different chromosomes and  $i$  and  $j$  index over non-bonded pairs of chromatin beads.  $C'_i$  denotes the compartment type for bead  $i$  from chromosome  $I$  and can be either  $A$  or  $B$ . We used different parameters  $\alpha_{intra}$  and  $\alpha_{inter}$  for intra- and inter-chromosome interactions to account for the presence of different molecular players that organize the genome at various lengthscales (Qi and Zhang, 2019). This potential contributes a total of 6 parameters to the model.

Therefore, for the 1Mb-resolution model, the total number of parameters is  $N - 1 + 6 = 254$ .

For the 100kb-resolution model, we further separated the intra-chromosome potential into intra- and inter-TAD interactions depending on whether the pair of beads are within the same topologically associating domain (TAD) or not. Specifically,

$$U_{compt}(\mathbf{r}) = \sum_I \sum_{i,j} \left[ \alpha_{intra}^{TAD}(C'_i, C'_j) \delta_{T'_i, T'_j} + \alpha_{inter}^{TAD}(C'_i, C'_j) (1 - \delta_{T'_i, T'_j}) \right] f(r_{ij}) + \sum_{I,J} \sum_{i \in I, j \in J} \alpha_{inter}(C'_i, C'_j) f(r_{ij}), \quad [5]$$

where  $T'_i$  denotes the TAD index for bead  $i$  from chromosome  $I$ .  $\delta_{T'_i, T'_j}$  is the Kronecker delta function and equals to 1 if  $T'_i = T'_j$  and 0 otherwise. The positions of TAD boundaries were determined from experimental Hi-C data using the software TADbit (Serra et al., 2017). Here  $C'_i$  can adopt three values:  $A$ ,  $B$  and  $I$ . Therefore,  $U_{compt}(\mathbf{r})$  contributes 18 parameters to the model. The total number of parameters for the 100kb-resolution model is thus  $2492 + 18 = 2510$ .

### Parameter optimization

Parameters in the above energy function can be derived using the iterative algorithm introduced in our previous works (Qi and Zhang, 2019). In particular, parameters  $\alpha_{ideal}(|j-i|)$ ,  $\alpha_{intra}(C'_i, C'_j)$  and  $\alpha_{inter}(C'_i, C'_j)$  are tuned to ensure that the following ensemble averages determined with simulated genome conformation matches corresponding experimental constraints calculated using Hi-C data.

$$\left\langle \sum_I \sum_{i,j} f(r_{ij}) \delta_{|j-i|, s} \right\rangle = \sum_I \sum_{i,j} f_{ij}^{exp} \delta_{|j-i|, s}, \text{ for } s = 1, \dots, N - 1 \quad [6]$$

$$\left\langle \sum_I \sum_{i,j} f(r_{ij}) \delta_{C'_i, c_1} \delta_{C'_j, c_2} \right\rangle = \sum_I \sum_{i,j} f_{ij}^{exp} \delta_{C'_i, c_1} \delta_{C'_j, c_2}, \text{ for } (c_1, c_2) \in \{(A, A), (A, B), (B, B)\}$$

$$\left\langle \sum_{I,J} \sum_{i \in I, j \in J} f(r_{ij}) \delta_{C'_i, c_1} \delta_{C'_j, c_2} \right\rangle = \sum_{I,J} \sum_{i \in I, j \in J} f_{ij}^{exp} \delta_{C'_i, c_1} \delta_{C'_j, c_2}, \text{ for } (c_1, c_2) \in \{(A, A), (A, B), (B, B)\}$$

In the above equations, the Kronecker delta function  $\delta_{C'_i, c_1}$  equals to 1 if  $C'_i = c_1$  and 0 otherwise.  $\delta_{C'_j, c_2}$  is similarly defined.  $f_{ij}^{exp}$  is the contact probability between the pair of genomic segments  $i$  and  $j$  determined from Hi-C.  $U_{ME}(\mathbf{r})$  can be shown as the least biased potential to reproduce these experimental constraints following the maximum entropy principle.

The constraints used to parameterize the 100kb-resolution model can be similarly defined.

### Molecular dynamics simulation details

The software package LAMMPS (Plimpton, 1995) was used to carry out molecular dynamics simulations with reduced units and collect ensembles of genome organization. Simulations were maintained at a constant temperature  $T = 1.0$  via the Langevin dynamics with a damping coefficient  $\gamma = 10.0$  and a time step of  $dt = 0.01$ .

To generate an initial configuration for these simulations, we first placed all the chromosomes consecutively on a cubic lattice with an edge length of  $0.9R/\sqrt{3}$ , where  $R$  is the radii of the spherical confinement introduced to ensure a volume fraction of 0.1. This configuration was subsequently equilibrated along a 100,000-step-long simulation under the potential  $\sum_i U(\mathbf{r}_i)$  to relax both the topology and energy of the polymer structures. The last configuration from this equilibration trajectory was then used to initialize our whole genome simulations. We note that the long sampling time used in our simulations ensures their convergence. Therefore, all the results presented in the manuscript are independent of this initial configuration.

Parameters of the whole genome models were determined iteratively. We initialized the first iteration of these simulations using the equilibrated configuration mentioned above. All subsequent simulations were initialized using the end configurations from the previous iteration. During each iteration, we carried out six independent ten-million-time-step-long simulations for the 1Mb-resolution model and ten independent two-million-time-step-long simulations for the 100kb-resolution model. Genome conformations were saved at every 2000 timesteps to calculate the ensemble averages. A total of 10 iterations were performed for the 1Mb-resolution model to reach an error of less than 5%. We define the error as  $\varepsilon = \sum |f_i^{sim} - f_i^{exp}| / \sum f_i^{exp}$ , where  $f_i^{exp}$  are the experimental constraints defined in [Equation 6](#) and  $f_i^{sim}$  are the corresponding ensemble averages determined from computer simulation. We used 35 iterations for the 100kb-resolution model to reach an error of less than 15%.

With the converged parameters, we performed additional six independent twenty-million-time-step-long simulations for the 1Mb-resolution model and ten independent four-million-time-step-long simulations for the 100kb-resolution model. A total of 60,000 and 20,000 structures were collected for the 1Mb- and 100kb-resolution model respectively to perform all the analysis presented in the main text.

### Radial density profile

The compartment-specific radial density functions were calculated with the following expression

$$\rho(r) = \frac{n(r)}{4\pi r^2 \Delta r \cdot N}$$

where  $r$  is the spatial distance from the nuclear center.  $n(r)$  is the number of genomic loci of a given compartment type found in the spherical shell from  $r$  to  $r + \Delta r$ , and the angular brackets indicates an ensemble average over all the simulated genome structures.  $N$  is the total number of genomic loci of that given compartment type.

### Electron Microscopy Analysis

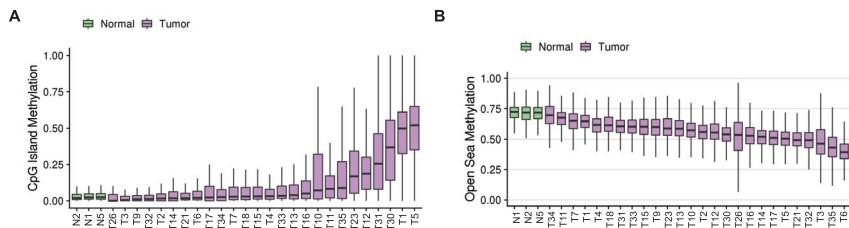
Quantifications were performed using Fiji version 2.0.0-rc-69/1.52p by thresholding the 8-bit EM images. The outer border of each nucleus was delineated using the freehand tool and its area and % positive for the heterochromatin threshold was measured by set measurements > limit to Threshold. Next, the same nucleus was measured again with a freehand selection on the inner border of the nucleus, excluding the peripheral heterochromatin (the chromatin touching the nuclear membrane). The percentage internal heterochromatin was calculated by dividing the internal heterochromatin by the area of the internal measurement and multiplied by the total nuclear area and plotted as the percentage of internal over the percentage total heterochromatin. Independent data from three samples with  $N = 19, 37, 46$  nuclei for normal tissues and  $N = 19, 95, 71$  nuclei for tumor tissues were obtained. Visualization of the data and statistical testing were performed using Prism 8 version 8.4.2. The data was imported as a nested dataset and statistical significance was tested using a nested, unpaired, two-sided t test with alpha = 0.05. Data were represented using a frequency distribution plot with percentage internal heterochromatin using bin size 6% on the x axis and the percentage of cells in each of those bins on the y axis.

### DNA-FISH analysis

To calculate redistribution of the A and B compartment in primary tissues, the nuclei for 2D images were manually curated to delineate intact single tumor and colon epithelial nuclei in FIJI version 2.0.0-rc-69/1.52p ( $N = 2$  tumors and 2 normal samples). This step was necessary to avoid generating data from poorly oriented and non-tumor and -colon epithelial nuclei such as immune cells and fibroblasts. As the latter cells have strikingly different nuclear morphologies, we could easily exclude them by visual inspection. Next, the pictures were loaded into Cellprofiler version 3.1.9, the nuclei and compartment spots segmented, and the original channel images masked on the identified DNA-FISH spots. The radial intensity distribution of the masked images was calculated in the nuclei using 20 scaled bins per cell. To determine redistribution of the B compartment in copy number stable tumors toward the nuclear interior, the Fraction at Distance of each masked image bin was plotted in Prism Version 8.4.2. Because the chromosome territory of chromosome 12 is in general peripherally located, bin 1-10 were summed together for their visualization.

Radial distribution of DNA-FISH in cells was quantified using Cellprofiler version 3.1.9 2 ([McQuin et al., 2018](#)), segmenting the nuclei, followed by fill holes and exclusion of cells touching the edge of the image. Next, the radial distribution of each channel (A compartment (A488); B compartment (Cy3) and I compartment (Cy5) was measured using the module MeasureObjectIntensityDistribution using each nucleus as the center of the points. To obtain distributions for each cell's radial bin in which the maximum of the signal was located, we multiplied the Fraction at Distance for each bin and channel with each Mean Fraction value, and gave a value of 1 to the bin containing the highest value. The counts for each cell and channel were then plotted using Prism Version 8.4.2. Representative images and insets in all panels were generated using FIJI version 2.0.0-rc-69/1.52p.

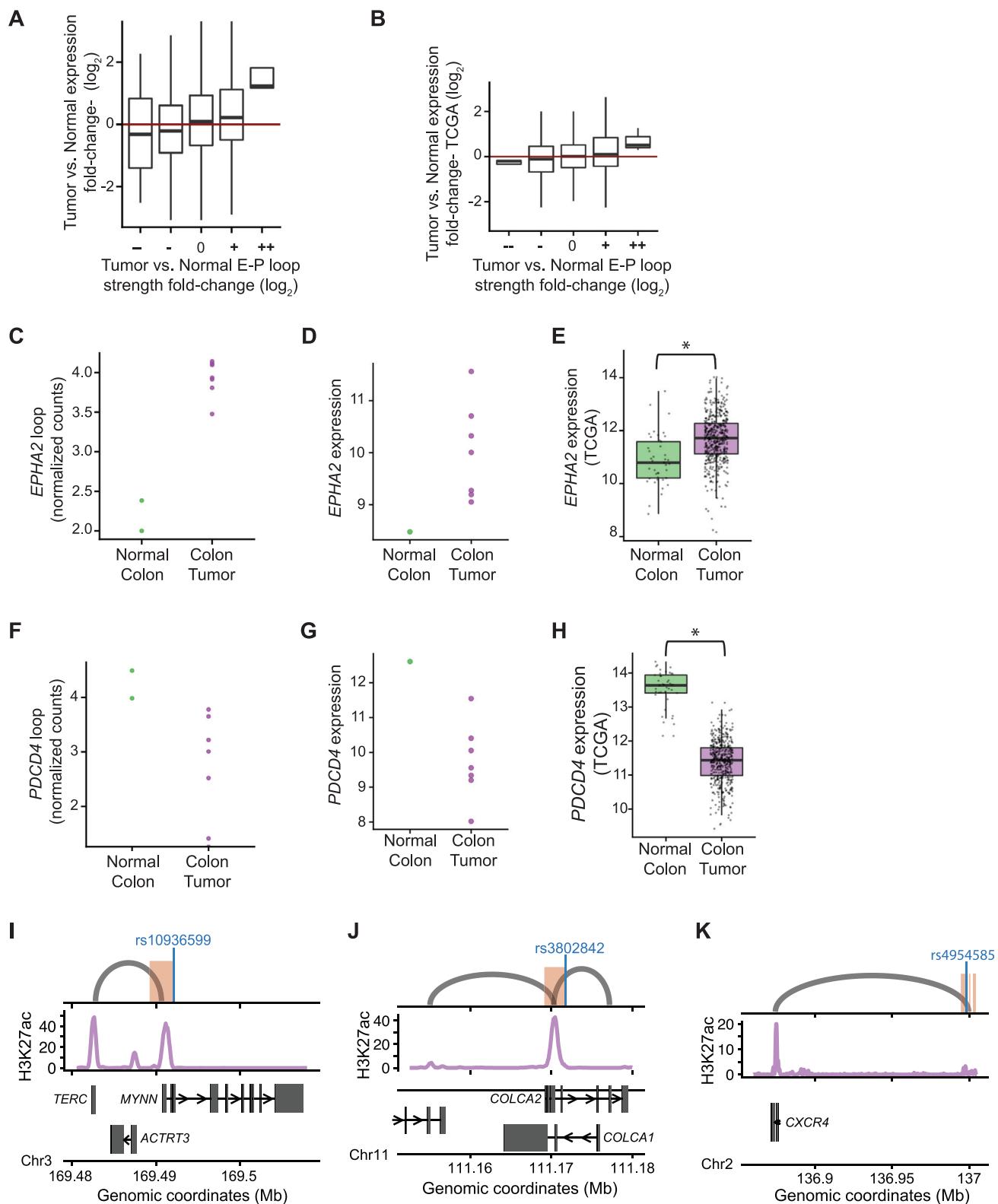
## Supplemental Figures



**Figure S1. CpG Island and Block methylation, Related to Figure 1**

(A) Boxplot representation of the distribution of DNA methylation values across CpG islands for each tumor sample (x axis). Tumor samples are shown in purple and normal samples are shown in green.

(B) For each sample (x axis), boxplot representation of the distribution of DNA methylation values across open sea CpGs in hypomethylated blocks. Tumor samples are shown in purple and normal samples are shown in green.

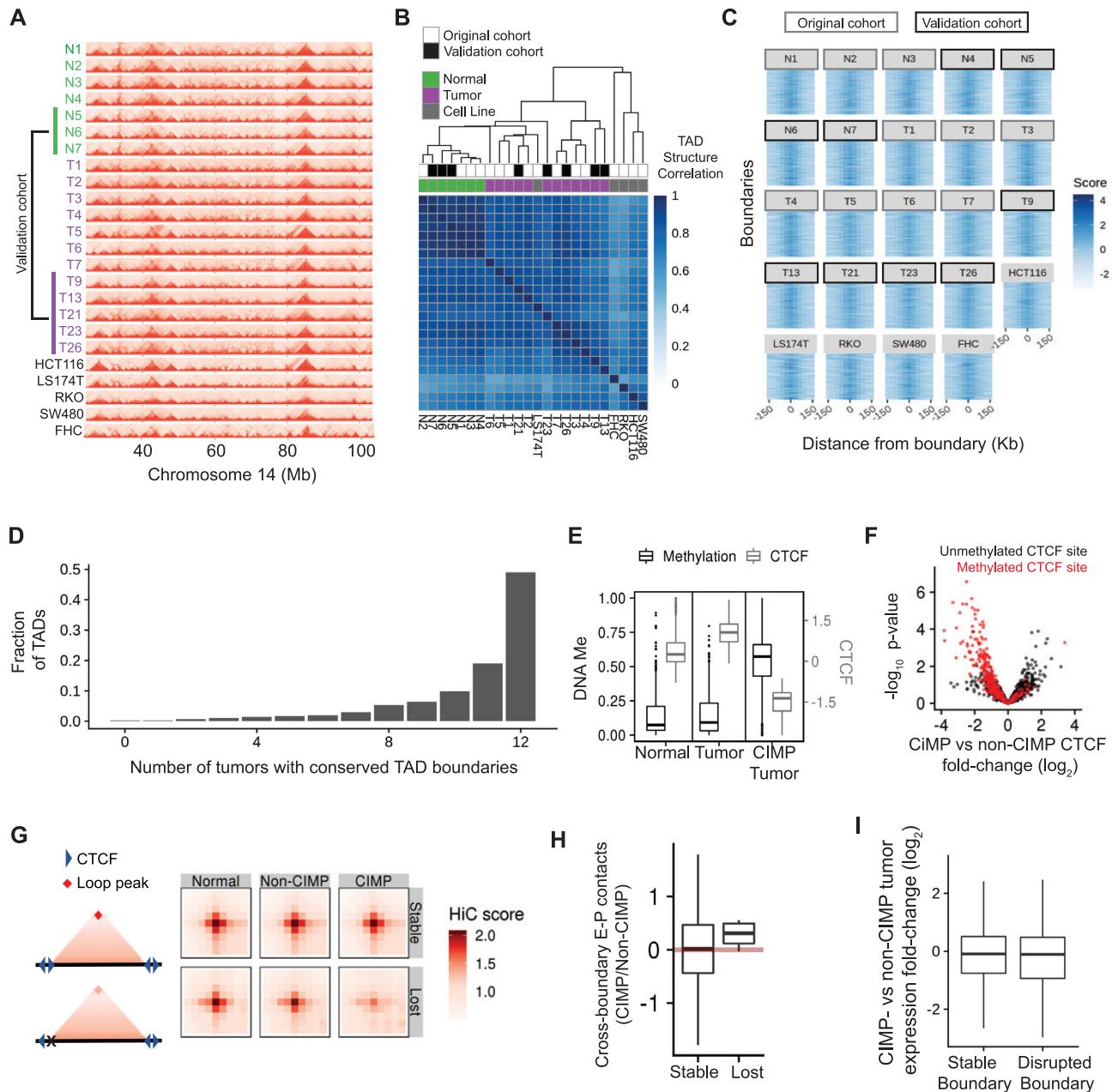


(legend on next page)

---

**Figure S2. Chromatin Loops, Related to Figure 1**

- (A) Boxplots depict expression fold-change ( $\log_2$ ) between tumors and normal samples (y axis) for genes engaged in enhancer-promoter (E-P) loops. Genes are stratified by change in E-P loop strength between tumors and normal (x axis). This plot is equivalent to [Figure 1C](#), but uses RNA-seq data from our cohort instead of TCGA data.
- (B) Boxplot representation of the distribution of expression log (base 2) fold-changes between tumors and normal stratified by the strength of loop strength fold changes (x axis). This plot is equivalent to [Figure 1C](#), but includes only copy number stable loci.
- (C) Dot plot shows normalized (EPHA2 E-P) loop counts for each normal (green) and tumor (purple) sample. Counts are shown for the differential loop highlighted in [Figure 2D](#).
- (D) Dot plot shows EPHA2 expression levels for normal colon (green) and tumor (purple). Each point represents an RNA-seq sample from our cohort.
- (E) Boxplots depict EPHA2 expression in 41 normal colon samples and 480 colon tumors from TCGA.
- (F) Normalized (PDCD4 E-P) loop counts are shown for each normal (green) and tumor (purple) sample. Counts are shown for the differential loop highlighted in [Figure 2E](#).
- (G) Dot plot shows PDCD4 expression levels for normal colons (green) and tumors (purple). Each point represents an RNA-seq sample from our cohort.
- (H) Boxplots depict PDCD4 expression in 41 normal colon samples and 480 colon tumors from TCGA.
- (I-K) Genomic views of the TERC locus (F), COLCA1/2 locus (G) and CXCR4 locus (H). Upper panels show SMC1 HiChIP loops as gray arcs; middle panels show H3K27ac signal in colon tumors in purple. Colon cancer risk SNP positions are indicated by blue lines. H3K27ac peaks with coincident loop anchors are indicated in orange.



**Figure S3. Topologically Associated Domains, Related to Figure 1**

(A) Horizontal heatmaps show local Hi-C contact patterns (red heat) across chromosome 14 for normal colon (green), colon tumors (purple) and cell lines (black). Validation cohort is highlighted (vertical bars, left).

(B) Heatmap shows pairwise correlations between genome wide TAD boundary scores (blue heat) in normal colons (green), colon tumors (purple) and cell lines (gray). These samples (rows, columns) are ordered according to a complete linkage hierarchical clustering (top). Original cohort is indicated by white squares and validation cohort is in black squares.

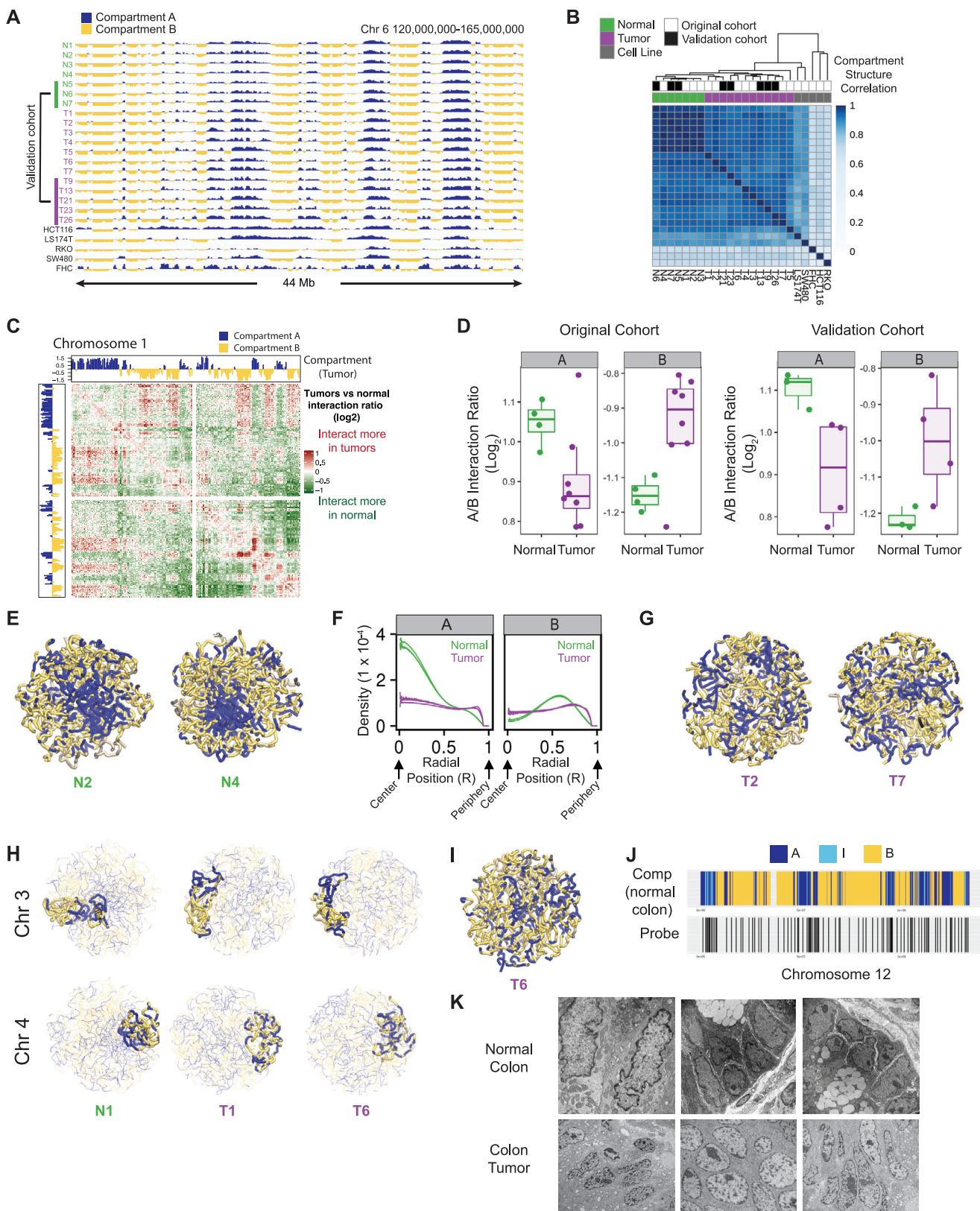
(C) Metaplot of the 40Kb-resolution boundary scores are shown for each Hi-C sample. Each panel represents one Hi-C sample, each row of the heatmap represent one TAD boundary, and the columns represent the genomic position relative to the TAD boundary. Original cohort indicated by gray outlines and validation cohort is indicated by black outlines.

(D) Bar plot of TAD boundary conservation analysis using the approach by Schmitt et al., 2016. Data are summarized over both the original and validation cohorts. Plot shows the number of tumors where TAD boundaries are called at the same location as TADs in colon normal tissues (x axis). The y axis shows the fraction of normal colon TAD boundaries. The overall conservation of TAD boundaries is similar to what has been described across different tissue types (Schmitt et al., 2016).

(E) Boxplots depict DNA methylation (black) and CTCF binding (gray) for CTCF binding sites that are differential between CIMP and non-CIMP tumors. Data shown for normal colon, non-CIMP tumors and CIMP tumors.

(legend continued on next page)

(F) Volcano plot shows differential analysis of CTCF binding sites (points) between CIMP and non-CIMP tumors (points to the upper left represent CTCF binding sites lost in CIMP tumors). Sites that are hypermethylated in CIMP tumors relative to non-CIMP samples are highlighted (red; methylation difference > 15%).  
(G) Left: Cartoon schematic of Hi-C heatmap shows a strong loop peak corresponding to an interaction between two CTCF bound loop anchors flanking a TAD (top panel). This theoretical CTCF-CTCF loop interaction is weakened in a sample with reduced CTCF binding at one or both anchors (bottom). Right: Heatmaps show actual Hi-C signals aggregated over CTCF-CTCF loops, revealing interaction peaks (i.e., averaged signal for the pixels corresponding to the tops of the TAD triangles illustrated at left). Top: Heatmaps aggregate signals for loops whose CTCF anchors are stable in normal colon, non-CIMP tumors and CIMP tumors (top). Bottom: Heatmaps aggregate signals for loops whose CTCF anchors are lost in CIMP tumors. These loop anchor interactions are weakened in CIMP tumors.  
(H) Boxplots depict fold-change ( $\log_2$ ) in E-P loop strength between tumors and normal. Loops crossing TAD boundaries are shown. Loops are stratified according to whether the TAD boundary that they span loses CTCF binding and gains methylation in CIMP tumors (lost) or whether it retains CTCF (stable).  
(I) Boxplots depict expression fold-change ( $\log_2$ ) between CIMP and non-CIMP tumors stratified by whether the genes are located in a disrupted TAD or not.

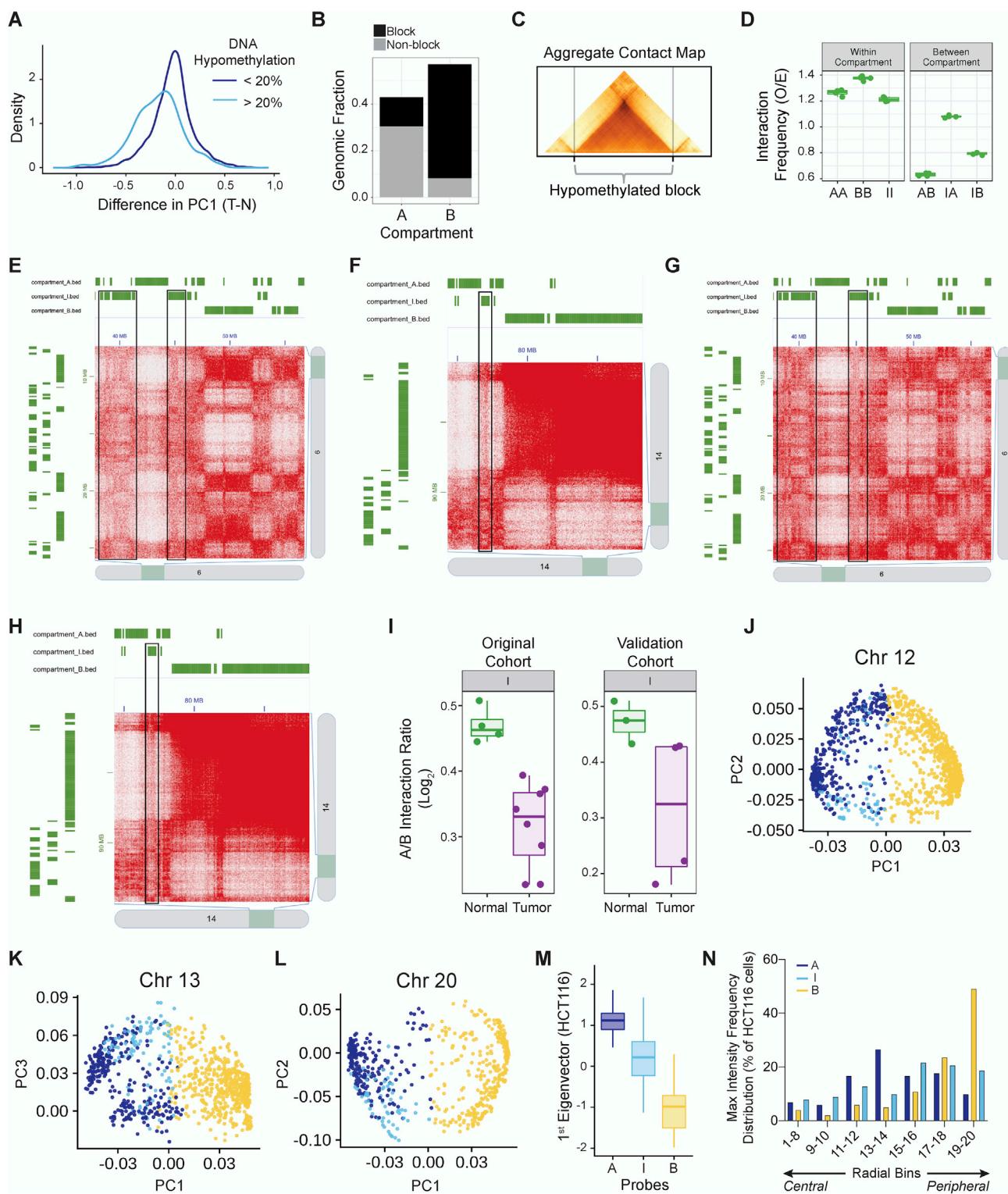


(legend on next page)

---

**Figure S4. Compartment Reorganization, Related to Figure 2**

- (A) Hi-C eigenvectors (PC1) based on long-range interactions demarcate compartments A (positive values, blue) and B (negative values, yellow) across a 45 Mb region of chromosome 6. Data show Hi-C eigenvectors for normal colon (green), colon tumors (purple) and cell lines (black). Validation cohort is indicated (vertical bars, left).
- (B) Heatmap shows pairwise correlations between the Hi-C eigenvector (blue heat) in normal colon (green), colon tumors (purple) and cell lines (gray). Samples (rows, columns) are ordered according to a complete linkage hierarchical clustering (top). Original cohort is indicated by white squares and validation cohort by black squares.
- (C) Heatmap shows fold-change ( $\log_2$ ) in Hi-C contact frequencies between colon tumors and normal colon across chromosome 1. Data are based on an average of normal colons ( $n = 4$ ) and tumors ( $n = 7$ ). Interactions that increase in tumors (red) or decrease in tumors (green) are evident. Top, left: Hi-C eigenvector indicates compartment assignments in colon tumor (A = blue, B = yellow).
- (D) Plot shows average ratio of interactions with the A versus B compartments (y axis), summarized for compartment A and B loci for original (left) and validation (right) cohorts. Each point represents the average for 100 kb windows per sample for normal colons (green) or tumors (purple). Statistical significance was computed by a two-sided Wilcoxon rank sum test comparing tumor versus normal absolute A/B ratio values (original cohort:  $p = 0.004$ ; validation cohort:  $p = 0.005$ ).
- (E) Whole nucleus maximum entropy models (1 Mb resolution) for two normal colon samples showing compartment A in blue and compartment B in yellow.
- (F) Density plots depict radial distributions of compartment A and compartment B regions in the maximum entropy models derived for normal colons (green, as in panel (E)) or colon tumors (purple, panel (G)) (0 = interior of the nucleus, 1 = periphery of the nucleus). Each line corresponds to a maximum entropy model derived for one experimental Hi-C dataset (2 colon tumors and 2 normal colons).
- (G) Whole nucleus maximum entropy models (1 Mb resolution) for two colon tumors showing compartment A in blue and compartment B in yellow.
- (H) Maximum entropy models for the whole genome for normal colons and colon tumors, highlighting copy number stable chromosomes 3 (upper) and 4 (lower). Compartment A is colored in blue and compartment B is colored in yellow.
- (I) Maximum entropy model for whole genome for a copy-number stable tumor (T6) sample showing compartment A colored in blue and compartment B colored in yellow.
- (J) Genomic view of chromosome 12 shows compartment assignment from normal colon (top) and DNA-FISH probe distribution (bottom; black bars).
- (K) Representative transmission electron microscopy (EM) images of nuclei from normal colon epithelium (top row) and colon tumors (bottom row).

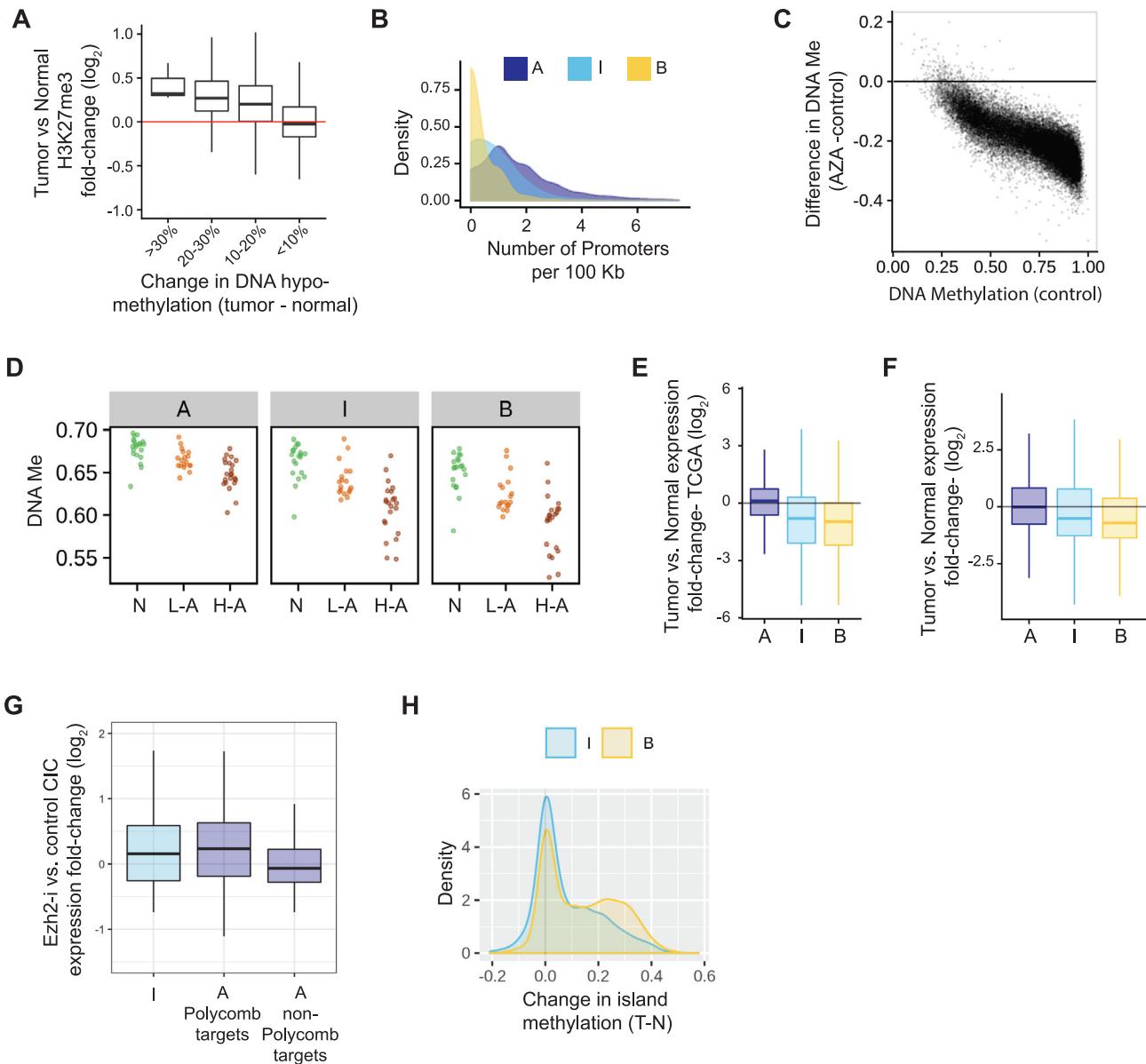
**Figure S5. Intermediate Compartment I, Related to Figure 3**

(A) Density plot shows Hi-C eigenvector difference between tumor and normal for 100 kb windows with  $PC1 > 0$  in normal colon tissue. Regions are stratified based on degree of tumor hypomethylation.

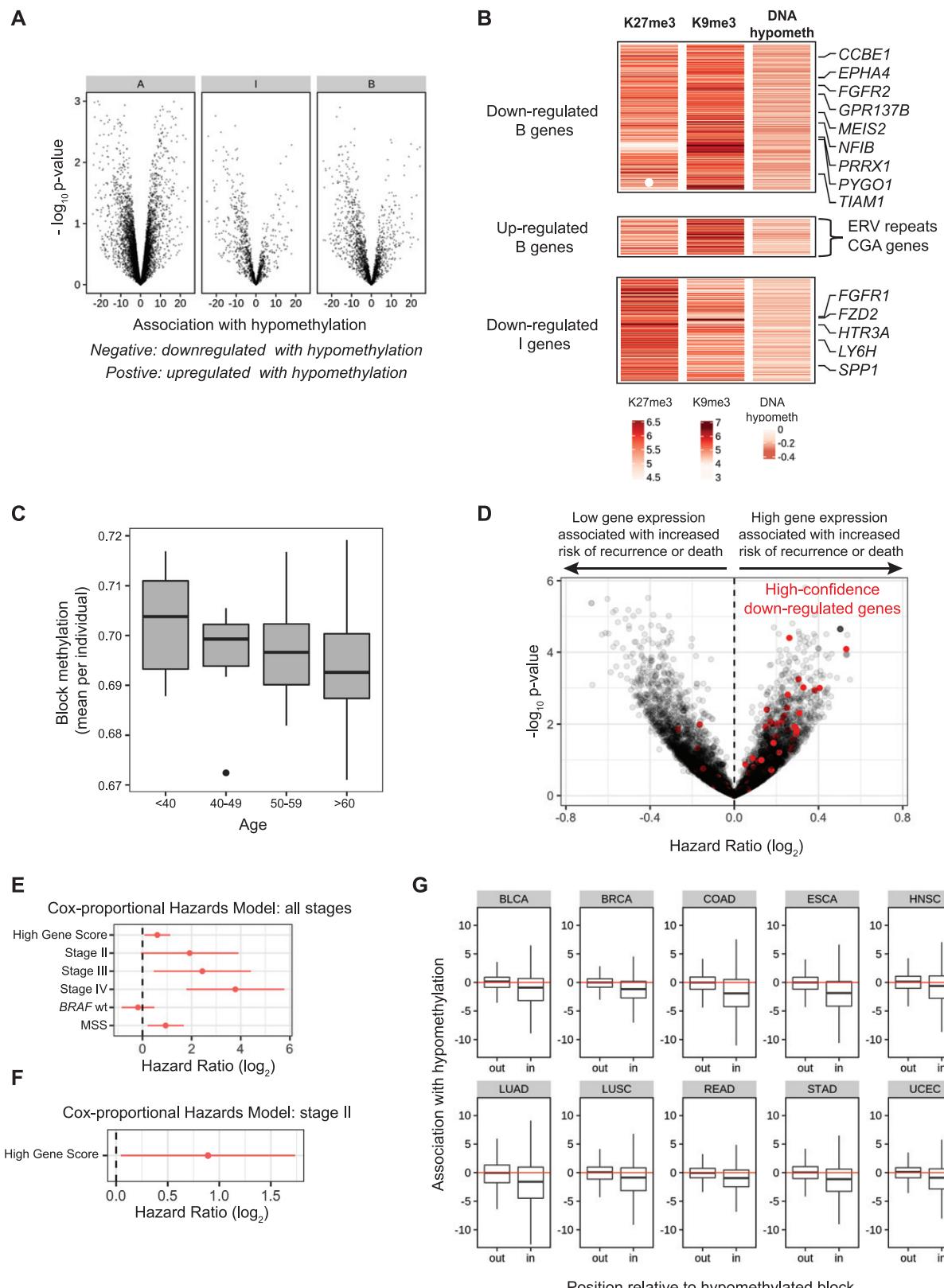
(B) Barplot showing the fraction of the genome (y axis) assigned to each compartment (x axis). Black bar represents the fraction of the genome that is block hypomethylated in tumors with respect to normal.

(legend continued on next page)

- 
- (C) Aggregated contact map shows Hi-C signal averaged over all hypomethylated blocks across normal and tumor samples. The x axis shows genomic positions relative to hypomethylated blocks. The edges of hypomethylated blocks correspond to TAD boundaries.
- (D) Plots show average frequency of Hi-C contacts for pairwise interactions that occur within the same genomic compartment (left), and between different compartments (right). Data are shown for four normal colon samples (dots). Compartment I regions have inter-compartment interactions with both A and B regions.
- (E-F) Hi-C contact map of observed versus expected interactions in normal colon for two representative regions across chromosomes 6 (E) and 14 (F). Compartment designations are shown for both rows and columns.
- (G-H) Hi-C contact map of observed versus expected interactions in colon tumors for two representative regions across chromosomes 6 (G) and 14 (H). Compartment designations are shown for both rows and columns.
- (I) Plot shows average ratio of interactions with the A versus B compartments (y axis), summarized for compartment I. Each point represents the average of 100 kb windows for normal colons (green) and tumors (purple). Shown for original (left) and validation (right) cohorts.
- (J-L) Scatterplots of first and second (J and L) or first and third (K) eigenvectors for chromosomes 12 (I), 13 (J) and 20 (K) resulting from the eigenvector decomposition method to define compartments. Data are shown for the aggregated normal colon Hi-C matrices. Each point represents one 100 kb bin and is colored by compartment (A: dark blue; I: light blue; B: yellow).
- (M) Boxplot shows the distribution of PC1 values resulting from the eigenvector decomposition of the HCT116 Hi-C matrix. Data are shown for the 100Kb-bins that overlap with our DNA-FISH probes. Probes for the respective compartments have the expected distributions of PC1 values.
- (N) Barplot indicates the percent of cells for which the maximum DNA-FISH signal intensity for compartments A, B or I is located at the indicated radial position for 102 normal colon nuclei.

**Figure S6. Related to Figure 4**

- (A) Boxplots show H3K27me3 fold-change (y axis) between tumors and normal colon as a function of DNA hypomethylation in tumors (x axis).
- (B) Histogram shows the distribution of gene density, measured as number of promoters per 100Kb, for each genomic compartment.
- (C) For each 100Kb genomic bin, change in DNA methylation upon 5-aza treatment (y axis) is plotted against baseline (DMSO) methylation level. Regions showing high levels of methylation in the control (DMSO) sample showed the greatest loss of methylation upon treatment with 5-aza.
- (D) Plots show average DNA methylation for a cohort of normal (N) colon samples and low- (L-A) and high-grade adenomas (H-A) (Fan et al., 2020). Points represent individual samples. Data are stratified by compartment. Points represent individual samples. Data are stratified by compartment.
- (E) Boxplot depicts fold-change ( $\log_2$ ) in expression for genes in compartments A, I or B between tumor and normal colon using TCGA gene expression data. Genes in compartments I and B are downregulated in tumors.
- (F) Boxplot depicts fold-change ( $\log_2$ ) in expression for genes in compartments A, I or B between tumor and normal colon. This panel is equivalent to (E), but the values reflect RNA-seq data from our cohort rather than TCGA.
- (G) Boxplot representation of gene expression fold-changes in cancer initiating cells treated with EZH2 inhibitor, relative to control (Lima-Fernandes et al., 2019). Data are shown for genes in compartment I, known EZH2 targets in compartment A and expressed genes in compartment A that are not EZH2 targets.
- (H) Density plot shows the distribution of methylation differences in CpG islands in tumors, relative to normal colons (x axis). Compartment I is depicted in light blue and compartment B is depicted in yellow. Although both compartments are globally hypomethylated in open sea regions, a relatively larger subset of CpG islands gains methylation in compartment B.



(legend on next page)

---

**Figure S7. Related to Figure 6**

(A) Volcano plot depicts the association between expression and block hypomethylation for all genes in compartments A, I and B. A positive association on the x axis indicates upregulation with hypomethylation and a negative association indicates downregulation with hypomethylation. This panel is equivalent to Figures 6A and 6B, but the values were calculated using RNA-seq data from our cohort rather than TCGA data.

(B) Heatmap shows H3K27me3 and H3K9me3 ChIP-seq signal in tumors (first two columns), and the extent of DNA hypomethylation in tumors relative to normal colon (third column). Each row of the heatmap represents a gene. Three groups of genes are shown: Genes in compartment B that are downregulated with hypomethylation, genes in compartment B that are upregulated with hypomethylation (including CGA genes and ERV repeats), and genes in compartment I that are downregulated with hypomethylation.

(C) Boxplot shows open sea methylation levels for compartments B and I (mean per individual) for a cohort of normal colon samples stratified by age (x axis) (Wang et al., 2020).

(D) Volcano plot shows  $\log_2$  hazard ratio from a Cox proportional hazards model of 7,694 variable genes with expression sd > 0.5 from previously published cohort (Marisa et al., 2013). The set of genes in compartments B and I that are downregulated with hypomethylation are shown in red. A positive log hazard ratio indicates higher expression is associated with increased risk of recurrence or death.

(E) Coefficient estimates and 95% confidence intervals for association with recurrence free survival from a Cox Proportional Hazards model. The model includes the 146 gene high risk indicator and the other indicated variables (Data from Marisa et al., 2013).

(F) Coefficient estimates and 95% confidence intervals for association with recurrence free survival from a Cox Proportional Hazards model fit only to samples from patients with Stage II disease (Marisa et al., 2013).

(G) Boxplots show associations between gene expression and block hypomethylation for 10 epithelial tumor types. Positive values indicate upregulation with hypomethylation. Separate boxes show data for genes outside ('out') or inside ('in') a hypomethylated block. Data shown for 10 TCGA cohorts: bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), esophageal squamous cell carcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), rectal adenocarcinoma (READ), stomach adenocarcinoma (STAD), uterine corpus endometrial cancer (UCEC) (The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, 2020).