

<https://doi.org/10.1038/s42003-025-07647-2>

# 3D genome landscape of primary and metastatic colorectal carcinoma reveals the regulatory mechanism of tumorigenic and metastatic gene expression



Xiang Xu<sup>1,2,11</sup>, Jingbo Gan<sup>3,11</sup>, Zhaoya Gao<sup>1,4,11</sup>, Ruifeng Li<sup>3,11</sup>, Dandan Huang<sup>4,5,11</sup>, Lin Lin<sup>2</sup>, Yawen Luo<sup>2</sup>, Qian Yang<sup>2</sup>, Jingxuan Xu<sup>1</sup>, Yaru Li<sup>2</sup>, Qing Fang<sup>3</sup>, Ting Peng<sup>3</sup>, Yaqi Wang<sup>3</sup>, Zihan Xu<sup>3</sup>, An Huang<sup>6</sup>, Haopeng Hong<sup>6</sup>, Fuming Lei<sup>1,4</sup>, Wensheng Huang<sup>1,4</sup>, Jianjun Leng<sup>4,7</sup>, Tingting Li<sup>3,8</sup>, Xiaochen Bo<sup>2</sup>, Hebing Chen<sup>2</sup> ✉, Cheng Li<sup>3</sup> ✉ & Jin Gu<sup>1,4,6,9,10</sup> ✉

Colorectal carcinoma (CRC) is a deadly cancer with an aggressive nature, and how CRC tumor cells manage to translocate and proliferate in a new tissue environment remains not fully understood. Recently, higher-order chromatin structures and spatial genome organization are increasingly implicated in diseases including cancer, but in-depth studies of three-dimensional genome (3D genome) of metastatic cancer are currently lacking, preventing the understanding of the roles of genome organization during metastasis. Here we perform multi-omics profiling of matched normal colon, primary tumor, lymph node metastasis, liver metastasis and normal liver tissue from CRC patients using Hi-C, ATAC-seq and RNA-seq technologies. We find that widespread alteration of 3D chromatin structure is accompanied by dysregulation of genes including *SPP1* during the tumorigenesis or metastasis of CRC. Remarkably, the hierarchy of topological associating domain (TAD) changes dynamically, which challenges the traditional view that the TAD structure between tumor and normal tissue is conservative. In addition, we define compartment stability score to measure large-scale alteration in metastatic tumors. To integrate multi-omics data and recognize candidate genes driving cancer metastasis, a pipeline is developed based on Hi-C, RNA-seq and ATAC-seq data. And three candidate genes *ARL4C*, *FLNA*, and *RGCC* are validated to be associated with CRC cell migration and invasion using in vitro knockout experiments. Overall, these data resources and results offer new insights into the involvement of 3D genome in cancer metastasis.

Colorectal carcinoma (CRC) is the third most common cancer and the second-ranked cause of cancer deaths. The incidence and mortality rates of CRC are higher in Western countries than in Asia and Africa, and are rising rapidly in developing countries<sup>1</sup>, reflecting that both genetic and

environmental factors play roles in its etiology. A total of 45% of CRC patients experience cancer metastasis. About 20% of newly diagnosed CRC patients have metastatic diseases and another 25% will develop metastasis later<sup>2</sup>. Although systemic and targeted therapies have improved the survival

<sup>1</sup>Department of Gastrointestinal Surgery, Peking University Shougang Hospital, Beijing, China. <sup>2</sup>Academy of Military Medical Sciences, Beijing, China. <sup>3</sup>Center for Bioinformatics, School of Life Sciences, Center for Statistical Science, Peking University, Beijing, China. <sup>4</sup>Center for Precision Diagnosis and Treatment of Colorectal Carcinoma and Inflammatory Diseases, Peking University Health Science Center, Beijing, China. <sup>5</sup>Department of Oncology, Peking University Shougang Hospital, Beijing, China. <sup>6</sup>Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Gastrointestinal Surgery, Peking University Cancer Hospital & Institute, Beijing, China. <sup>7</sup>Department of Hepatopancreatobiliary Surgery, Peking University Shougang Hospital, Beijing, China. <sup>8</sup>State Key Laboratory of Proteomics, Institute of Basic Medical Sciences, National Center of Biomedical Analysis, Beijing, China. <sup>9</sup>Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. <sup>10</sup>Peking University International Cancer Institute, Beijing, China. <sup>11</sup>These authors contributed equally: Xiang Xu, Jingbo Gan, Zhaoya Gao, Ruifeng Li, Dandan Huang. ✉e-mail: [chb-1012@163.com](mailto:chb-1012@163.com); [cheng\\_li@pku.edu.cn](mailto:cheng_li@pku.edu.cn); [zlgui@bjmu.edu.cn](mailto:zlgui@bjmu.edu.cn)

outcome of metastatic CRC (mCRC) patients, fewer than 20% of them currently survive beyond 5 years from diagnosis<sup>2</sup>.

Cancer metastasis is a multi-phase process where primary tumor cells invade the surrounding stroma, transit through circulatory vessels, and develop into metastatic lesions at distant sites<sup>3</sup>. Although the initial phases are thought to involve epithelial-mesenchymal transition that increases the motility and invasiveness of tumor cells, less is known about how circulating tumor cells manage to survive and proliferate in a new tissue environment<sup>3</sup>. As primary and metastatic tumors mostly share similar mutational landscape and driver genes<sup>4</sup>, non-genetic mechanisms have been suggested for metastatic progression, such as amplification of oncogenic pathways previously activated in primary tumors<sup>5</sup>, activation of developmental epigenetic programs<sup>6</sup>, and global alteration of chromatin accessibility<sup>7</sup>.

Higher-order chromatin structures and spatial genome organization are increasingly implicated in various diseases<sup>8</sup>. In recent years, new sequencing technologies such as ATAC-seq<sup>9</sup> and high-throughput chromosome conformation capture (Hi-C)<sup>10</sup> have been developed to investigate genome structures and organization. Chromatin is entangled in histones to form nucleosomes, and the degree to which nucleosomes are stacked together determines the accessibility of chromatin<sup>9</sup>. A/B compartments are two states of chromatin obtained after principal component analysis (PCA) of Hi-C data. Compartment A is usually enriched in active chromatin, while compartment B is usually enriched in silent chromatin<sup>10</sup>. At the megabase scale, chromatin folds to form mutually insulating spheres, named topologically associated domains (TADs). The interaction frequency within TADs is significantly higher than that between TADs, thereby mediating relatively independent elements regulation within TADs<sup>11,12</sup>. Finally, the CTCF/cohesin protein pulls two sites that are linearly far apart to form a spatially adjacent loop which often represents the interaction between enhancer and promoter (E-P loop)<sup>13</sup>.

Genetic mutations in the germline or somatic genome, such as point mutations and structural variations, were reported to markedly change chromatin folding maps in the cancer genomes. And gene expression was significantly changed in the recombinant regions<sup>14–17</sup>. The current 3D genome studies of clinical cancer are gradually covering a variety of cancer types, revealing the regulatory mechanism of the 3D genome in cancer. Structural variations in acute myeloid leukemia (AML) induce enhancer-hijacking and silencer-hijacking events and further lead to 3D genome organization and gene expression alteration<sup>18</sup>. Johnstone et al. identified a compartment at the interface between the canonical A and B compartments that is reorganized in colon tumors<sup>19</sup>. Subtype-specific transcription in muscle-invasive bladder cancers is accompanied by specific open chromatin, epigenomic marks, and 3D genome organizations<sup>20</sup>. Zhu et al. traced the evolution and malignant transformation of colorectal cancer by generating high-resolution chromatin conformation maps of 33 colon samples spanning different stages of early neoplastic growth and revealed a substantial progressive loss of genome-wide cis-regulatory connectivity at early malignancy stages, correlating with nonlinear gene regulation effects<sup>21</sup>.

Although these studies discover A/B compartment and loop differences between samples from patients and healthy donors, individual heterogeneity brings challenges to comparative analysis of the 3D genome dynamics during tumor development. Additionally, the changes in the three-dimensional structure of chromatin in cancer tissues at the TAD scale have not been fully studied. Due to the limited size and strict experimental requirements of clinical samples, there are few nuclear architecture studies of solid metastatic tumors. Recently, Ren et al. used normal human pancreatic epithelial cell lines (HPNE), primary pancreatic cancer cell lines (PANC-1), and metastatic pancreatic cancer cell lines (Capan-1) to study multiscale 3D epigenome reprogramming during pancreatic cancer metastasis<sup>22</sup>. The nucleus of cancer cells often displays abnormalities in many aspects, including nucleolus<sup>23</sup>, nuclear size, shape, and chromatin texture<sup>24</sup>. However, the changes and function of three-dimensional chromatin structure in clinical metastatic samples remain for further research.

Here we performed multi-omics profiling of matched normal colon, primary tumor, and liver metastasis from 13 CRC patients using Hi-C,

ATAC-seq, and RNA-seq technologies. We integrated the quality-controlled 3D genome, chromatin accessibility and transcriptome data to explore the roles of genome organization in CRC.

## Results

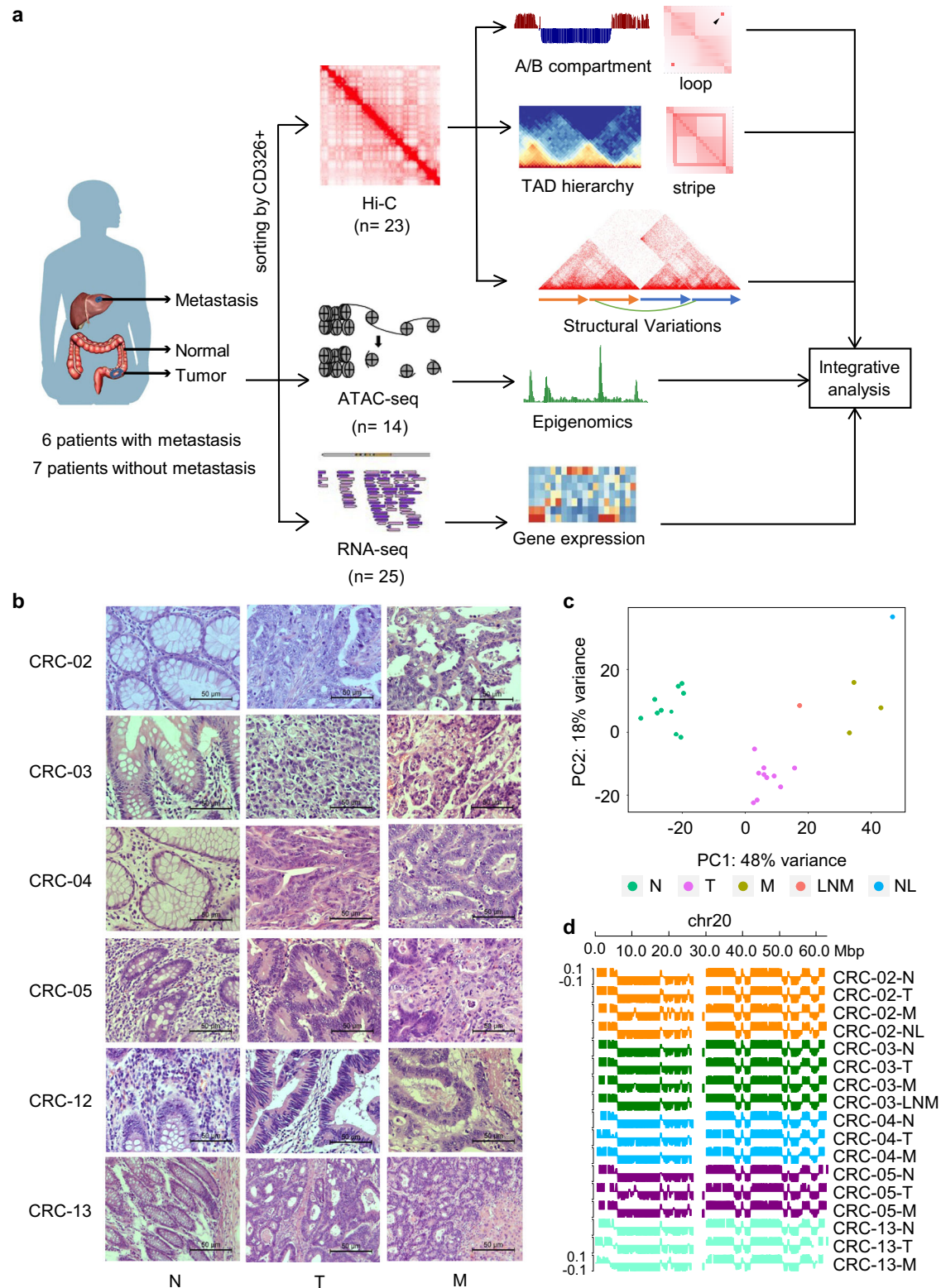
### Comprehensive maps of chromatin topology, gene expression, and chromatin accessibility in matched normal tissues, primary tumors, and metastatic CRC

To investigate the dynamic 3D genome landscape of primary and mCRC, we collected matched normal colon and colorectal tumor samples from seven metastasis-free CRC patients, and matched normal colon, colorectal tumor, liver metastasis, lymph node metastasis, and normal liver samples from 6 metastasis CRC patients treated at Peking University Shougang Hospital (Fig. 1a). We conducted a follow-up survey on all patients for 6 months. According to the definition of synchronous metastasis, we chose metastasis CRC patients who have liver metastasis when diagnosed with CRC. We named patients who had not experienced liver metastasis within 6 months after surgery as metastasis-free CRC patients. We arranged the clinical information of patients (Table S1).

Then, all samples were digested into single-cell suspension immediately after collection to construct Hi-C, RNA-seq, and ATAC-seq libraries for sequencing (“Methods”). Notably, the experimental conditions for clinical solid tissues are strict. For example, too few cells can result in lower Hi-C resolution, too long experimental time will lead to RNA degradation, and too low cell activity may reduce the enrichment of chromatin accessibility signal. Therefore, the quality control of multi-omics data for solid tissue samples is important.

Rigorous quality control was applied to our sequencing data. We screened qualified Hi-C data through quality control of sequencing quality (Table S2), alignment rate (Fig. S1a, b), fragment pairs distribution (Fig. S1c, d), valid interaction distribution (Fig. S2a, b), and distance vs. Hi-C counts (Fig. S2c, d). We also performed quality control on RNA-seq data by checking sequencing quality (Table S3), housekeeping gene coverage (Fig. S3a), read duplication (Fig. S3b), transcript integrity (Fig. S3c), and RPKM saturation (Fig. S3d). Further, we checked the ATAC-seq data (Table S4; Fig. S4) and observed enriched ATAC-seq signals at transcription start sites (TSS). A significant correlation was observed between chromatin accessibility of gene promoter and gene expression in our data (Fig. S5a). We found that increased promoter accessibility of cancer-related genes *MYC* and *NOTCH1* is consistent with the upregulation of gene expression (Fig. S5b, c). Therefore, integrating ATAC-seq and gene expression profiles discovered global chromatin accessibility alteration coordinated with gene expression changes, and nominated candidate genes dysregulated in mCRC.

Overall, we provided high-quality maps of chromatin topology, gene expression, and chromatin accessibility in matched normal tissues, primary tumors, and metastatic tumors (Fig. S6). The hematoxylin-eosin staining of matched normal colon, primary tumor, and liver metastatic tumor reflects the phenotypic difference of cancer tissue types and proves the correctness of our sampling (Fig. 1b). All hematoxylin-eosin staining slides were double-checked by pathological experts, and they ensured the expected cells (normal, primary or metastatic tumor cells) were present in the obtained tissue. Next, PCA was applied to analyze RNA-seq and Hi-C data. We showed that transcriptome data of normal colon, colorectal tumor, liver metastasis, lymph node metastasis, and normal liver samples were grouped into separate clusters (Fig. 1c). Notably, previous studies based on cell lines have shown that normal cells are more similar to metastatic cancer cells than primary cancer cells by clustering transcriptome and three-dimensional genome data<sup>22</sup>. However, in our clinical data, the cell types are sorted according to the order of normal colon, colorectal tumor, lymph node metastasis, liver metastasis, and normal liver. The data of adjacent cell types are more similar, which is consistent with the development process of cancer cells. We also found that liver metastasis is more similar to the normal liver than other tissue types in the transcriptome, which may be one of the prerequisites for cancer metastasis. To check the robustness of our result, we



**Fig. 1 | Genome organization and quality control in CRC samples. a** Experimental design and data analysis workflow of this study. **b** Hematoxylin-eosin staining of matched normal colorectum, primary tumor, and liver metastatic tumor from six patients with a scale bar of 50  $\mu$ m. **c** PCA of RNA-seq data in normal colon, colorectal

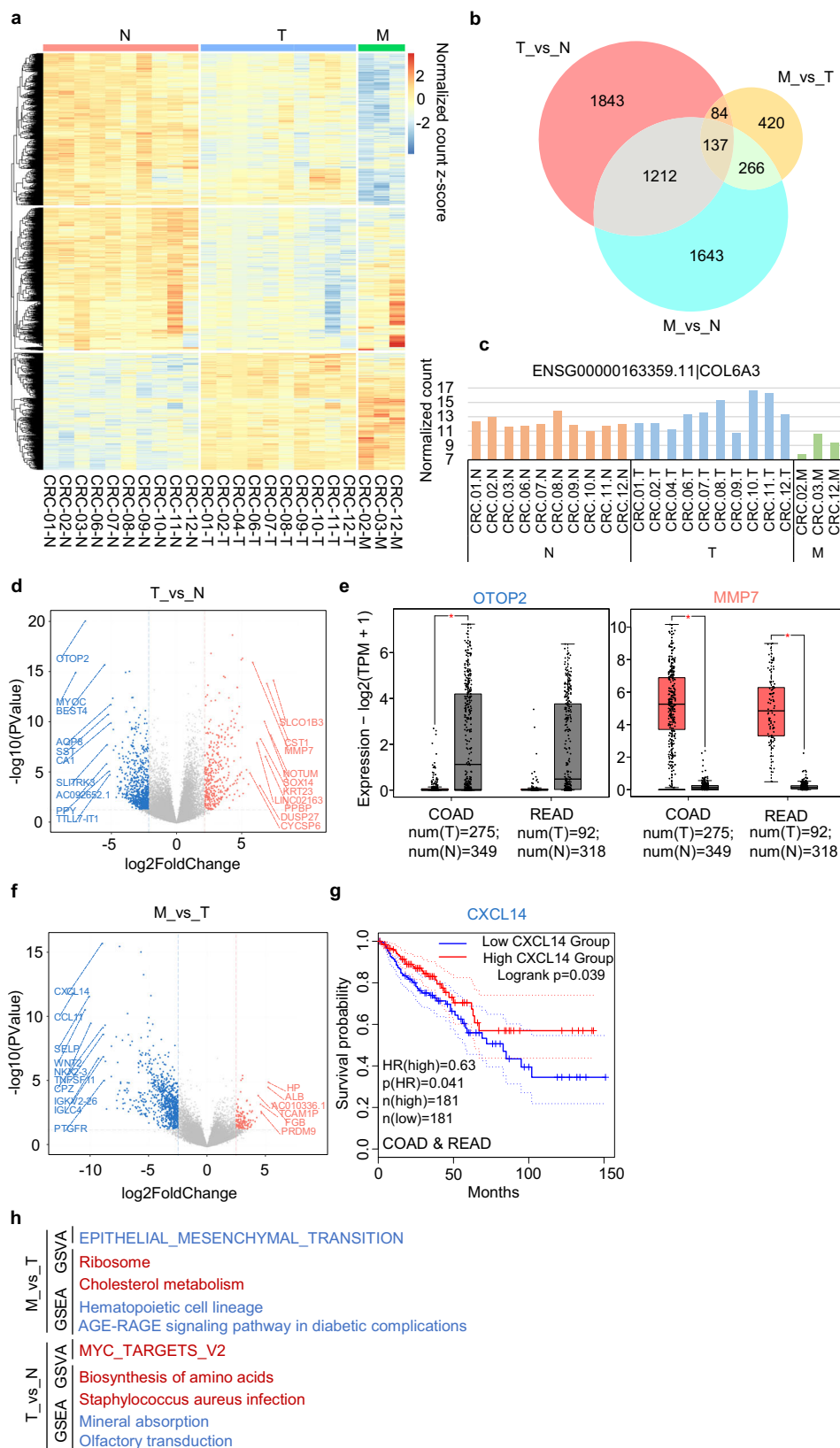
tumor, liver metastasis, lymph node metastasis, and normal liver samples. N: normal colorectum, T: primary tumor, M: liver metastasis, LNM: lymph node metastasis, NL: normal liver. **d** PCA1 snapshot of Hi-C data.

collected published RNA-seq data of matched normal colon, primary tumor, lymph node metastasis, liver metastasis, and normal liver tissues<sup>25–27</sup>, and we used a consistent pipeline for data quality control and PCA (Fig. S7). The high-quality published data achieved similar PCA result with our data.

According to the first principal component (PC1) track of Hi-C data, we confirmed the similarity of large-scale Hi-C features and specific differences in different tissue types (Fig. 1d). These high-quality clinical data provide a good basis for our research.



**Fig. 2 | Dynamic transcription patterns in the tumorigenesis and metastasis of colorectal carcinoma.** **a** Clustering heatmap of significantly differentially expressed genes. Gene expression count was normalized with z-score. The threshold for significant differences in gene expression is  $p$  value  $< 0.1$ . **b** Venn diagram of significantly differentially expressed genes among different comparison groups. **c** Normalized count of COL6A3 gene expression in different samples. **d** Differentially expressed genes in primary tumor vs. normal colorectum. Red dots: significantly higher expression gene in the primary tumor, blue dots: significantly lower expression gene in the primary tumor. **e** OTOP2 and MMP7 gene expression in colorectal tumor and normal colorectum. The red filling represents the gene expression level of colon or rectal cancer, while the gray filling represents the gene expression level of normal colon and rectum. The red asterisk indicates that the absolute value of  $\log_2$  (fold change) is greater than 1 and the  $p$  value is less than 0.01. **f** Differentially expressed genes in metastatic tumor vs. primary tumor. **g** Survival curve of CRC patients with different CXCL14 gene expression. **h** GSEA and GSVA enriched KEGG pathway in different comparison groups. The red font represents increased pathways, and the blue font represents decreased pathways.



## Primary tumor-specific and metastatic tumor-specific gene expression patterns

To find out the dynamic changes of the transcriptome during tumorigenesis and metastasis, we first identified differentially expressed genes in different types of samples (Fig. 2a). We obtained 5605 genes that were significantly changed in at least one of the pairwise comparisons, and 137 genes that were

significantly changed in each of the pairwise comparisons of three different sample types (Figs. 2b and S8a). To understand what role these dynamically expressed genes play in the tumorigenesis and metastasis of CRC, we selected *COL6A3*, *RP11-708H21.4*, and *ENO1* as examples. The patterns of upregulation and down-regulation of these genes in different samples differ from one another, but they are related to the prognosis of CRC patients

(Figs. 2c and S8b, c). The *P* value of *COL6A3* is the lowest among eight hub genes associated with poor prognosis in CRC, and in silico analysis of cell type-specific gene expression and *COL6A3* knockout experiments indicate the clinical relevance of *COL6A3* in the development of CRC<sup>28</sup>. Down-regulation of long non-coding RNA *RP11-708H21.4* is associated with poor prognosis for CRC and promotes tumorigenesis through regulating AKT/mTOR pathway<sup>29</sup>. In addition, the expression level of *ENO1* may serve as a valuable diagnostic biomarker, as it is associated with patient outcomes, such as cancer survival and prognosis<sup>30–32</sup>. The highly dynamic changes of these genes among different samples reflect the complexity and dynamics of gene regulatory networks related to cancer courses.

To comprehensively analyze which pathways are enriched by dynamically changing genes and regulate tissue type-specific gene expression networks in the development of cancer, we analyzed the differences in the expression and pathways of primary tumors vs. normal tissues and metastatic tumors vs. primary tumors respectively. Firstly, we identified the top genes with significant differences in expression among different samples (Fig. 2d, f). Then we analyzed the expression of these differentially expressed genes in public CRC transcriptome data from the Cancer Genome Atlas (TCGA)<sup>33</sup> with consistent results (Fig. 2e). Because of the poor prognosis and low survival rate beyond 5 years, we speculated that the differentially expressed genes in metastatic tumors vs. primary tumors were correlated to patient survival. We chose the most significantly differential expressed gene *CXCL14* for survival analysis. As expected, the *CXCL14* was significantly associated with survival in colorectal patients (Fig. 2g). The chemokine *CXCL14* has been reported to play an important role in the progression of many malignancies, and tumoral *CXCL14* expression levels were significantly correlated with TNM (Tumor-node-metastasis) stage, histodifferentiation, and tumor size<sup>34</sup>. The gene set enrichment analysis (GSEA)<sup>35</sup> and gene set variation analysis (GSVA)<sup>36</sup> showed enrichment pathways of differentially expressed genes among different samples (Figs. 2h and S8d, e).

Here we performed transcriptome profiling of matched samples from both metastatic and non-mCRC patients. Therefore, we can compare the difference in gene expression patterns between metastatic and non-metastatic patients in normal tissues and primary tumors, so as to explore the driving factors of cancer metastasis. By analyzing the different expression patterns of metastatic and non-metastatic patients, we found that the differential expressed genes in normal tissues are enriched in intestinal immune-related pathways, and the differential expressed genes in primary tumors mainly affect the metastasis-related pathways (Fig. S9), which revealed that the difference between different tissues in patients with metastasis and non-metastasis might be related to the occurrence of tumor metastasis.

### Reorganization of A/B compartment is associated with the differential gene expression specific to primary and metastatic colorectal carcinoma

The dynamic change of three-dimensional chromatin structure provides a new perspective for the change of gene expression regulation in CRC metastasis. Previous studies have revealed the alteration of higher-order genome structures in colorectal cell lines and primary tumors, in which translocations induce genome reorganization and compartmental changes that may restrain malignant progression<sup>19</sup>. To further explore the 3D genome of metastatic tumors as well as its potential relationship with the progression of CRC, we first inspected the A/B compartment landscape of the samples, which correspond to transcriptionally active and inactive genome regions, respectively<sup>10</sup>. We showed the dynamic A/B compartment switch during the development of individual CRC (Fig. 3a). The proportion of compartments that switched A/B status between normal and tumor samples ranged from 4.73% to 18.99% and the proportion between primary and metastatic tumors ranged from 6.35% to 21.62% in individual patients (Figs. 3b and S10a, b). Interestingly, the greatest switches were observed in CRC-04, CRC-05, and CRC-13, all of which developed metastasis. To clarify the association between individual heterogeneity in the A/B compartment and tumor metastasis, we defined the compartment stability score for comparative research of A/B compartment switch events among different

groups (1). A significantly higher compartment stability score was observed in patients without metastasis (Fig. 3c).

$$\text{Compartment stability score} = \frac{\text{Stable compartment}}{\text{Stable compartment} + \text{Switched compartment}} \quad (1)$$

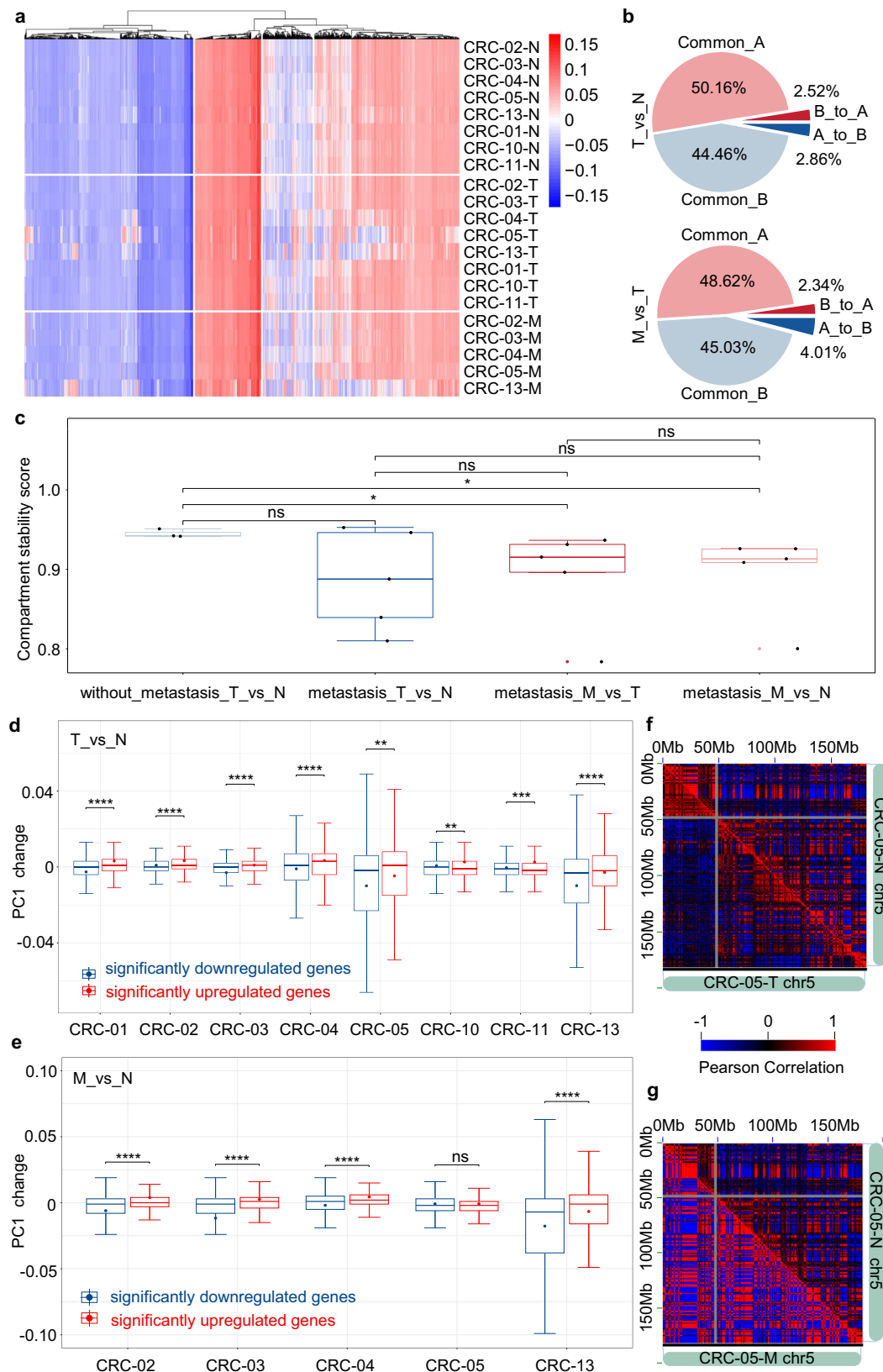
Since the compartment stability score is a novel concept in our research, we then tested this concept on published clinical Hi-C data. We first tested the performance of the compartment stability score in measuring A/B compartment difference on CRC<sup>19</sup> and AML<sup>18</sup> samples (Fig. S11). We noticed that the compartment stability score could well quantify the difference in A/B compartment between tumor and normal samples. Moreover, the tumor samples always show more heterogeneity than normal samples, so we tested whether this phenomenon could be reflected in the compartment stability score. As expected, we showed that the compartment stability score of normal sample pairs was significantly higher than that of tumor sample pairs (Fig. S12). Overall, the compartment stability score is robust and stable in measuring A/B compartment difference and evaluating Hi-C data heterogeneity within datasets.

We next asked whether the compartment switching is relevant to the gene expression change in tumorigenesis or metastasis of CRC. Since the A/B compartment switch events are highly heterogeneous among CRC patients, we first focused on the PC1 changes of significant differentially expressed genes. As expected, the PC1 value of upregulated genes increased more significantly than that of downregulated genes (Fig. 3d, e). This suggested that gene expression changes during the tumorigenesis and metastasis of CRC are correlated with PC1 value. However, we noticed an exception in comparison between CRC-05-M and CRC-05-N. The PC1 change did not show a significant difference between downregulated and upregulated genes. To explore the particularity of A/B compartment pattern of CRC-05-M, we showed the Pearson correlation heatmap of CRC-05-N, CRC-05-T, and CRC-05-M (Fig. 3f, g). The significant enhancement of A/B compartment pattern was observed in CRC-05-M Pearson correlation heatmap. We suspected that the great genome-wide alterations in CRC-05-M might influence the correlation analysis between gene expression and PC1 values. Although the pre-surgical carcinoembryonic antigen of CRC-05 patients is extremely high (129.33, with an average of 9.595 for other patients), the mechanism of great A/B compartment alterations remains unknown.

To further demonstrate the impact of A/B compartment changes on gene expression, we divided the switch mode of A/B compartment into four categories: B compartment remains unchanged (B-B), A compartment switches to B compartment (A-B), B compartment switches to A compartment (B-A), A compartment remains unchanged (A-A). By comparing the gene expression changes of different types of A/B compartment switching, we observed significant gene expression differences (Fig. S13a). The genes converted from B compartment to A compartment tend to be upregulated in expression, while the genes converted from A compartment to B compartment tend to be downregulated in expression. During the tumorigenesis and metastasis of CRC, the switching of compartment structure was consistent with the change of expression, suggesting that the change of compartment structure was one of the factors affecting the tumorigenesis and metastasis of cancer. Although the gene expression is affected by the compartment structure, the compartment is a structure at the megabase scale. The expression of some genes is not consistent with the switching of A/B compartment (Fig. S13b). We reasoned that the gene expression regulation will also be affected by more detailed regulatory units such as TAD and loop.

### TAD hierarchy and strength reorganized in tumorigenesis and metastasis of CRC

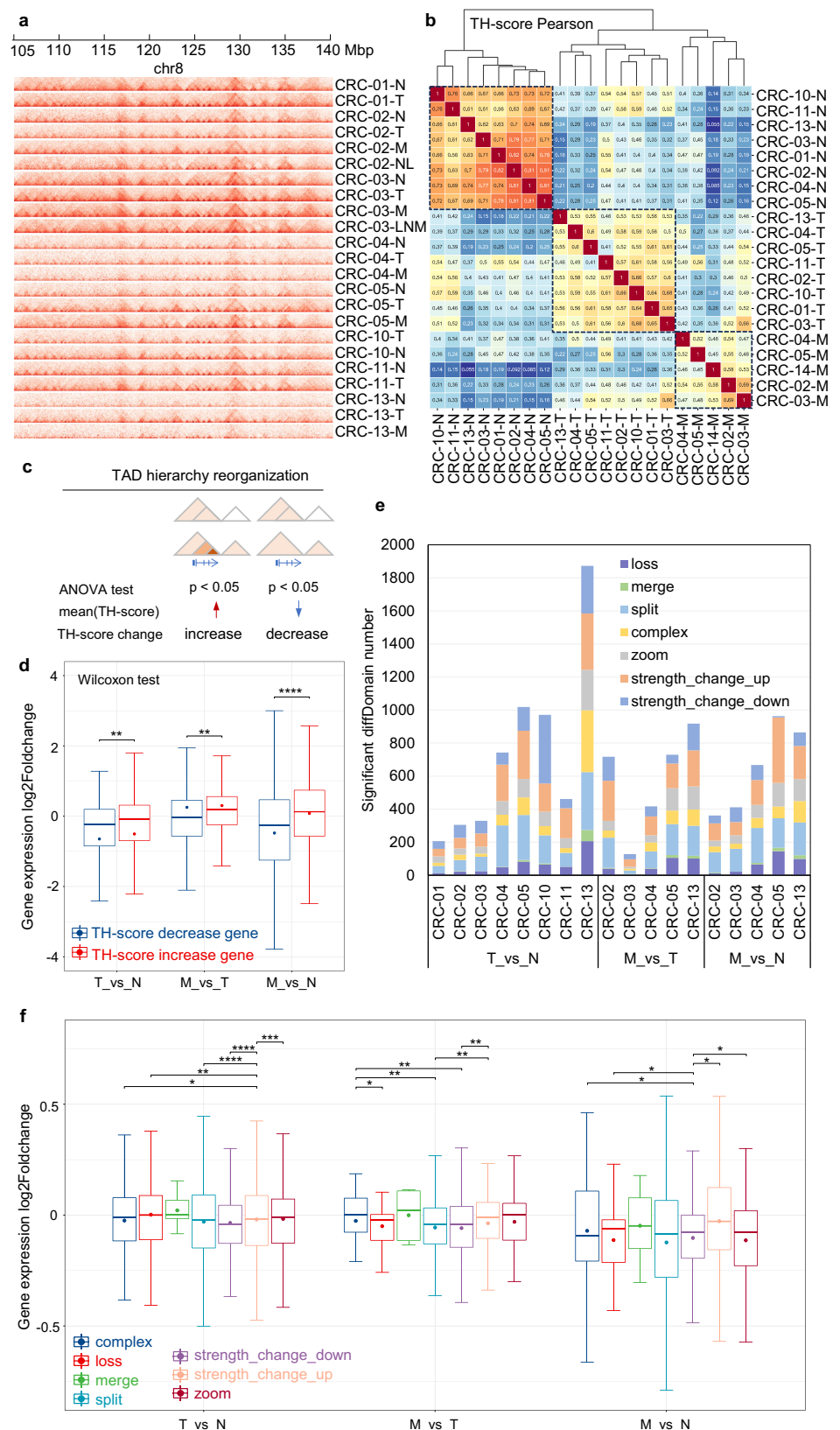
TADs are spatially self-interacting genomic regions, which physically restrict the interactions between promoters and regulatory elements within



**Fig. 3 | Dynamic A/B compartment switch in the tumorigenesis and metastasis of colorectal carcinoma. a** Clustering heatmap of Hi-C PCA1 value in different samples. **b** Proportional distribution of A/B compartment switch in CRC-03 patient during the tumorigenesis and metastasis of colorectal carcinoma. **c** Compartment stability score in tumorigenesis or metastasis of patients with or without metastasis.  $*p \leq 0.05$ , ns: no significance, Wilcoxon-test. **d** Box plot of PC1 changes of significantly downregulated and upregulated genes in colorectal cancer tumorigenesis.  $*p \leq 0.05$ ,

$**p \leq 0.01$ ,  $***p \leq 0.001$ ,  $****p \leq 0.0001$ , ns: no significance, *t*-test. **e** Box plot of PC1 changes of significantly downregulated and upregulated genes in colorectal carcinoma metastasis.  $*p \leq 0.05$ ,  $**p \leq 0.01$ ,  $***p \leq 0.001$ ,  $****p \leq 0.0001$ , ns: no significance, *t*-test. **f** Pearson correlation heatmap of chr5 in CRC-05-T and CRC-05-N samples. **g** Pearson correlation heatmap of chr5 in CRC-05-M and CRC-05-N samples.

**Fig. 4 | TAD hierarchy reorganization in the tumorigenesis and metastasis of colorectal carcinoma.** **a** Snapshot of Hi-C heatmap in different samples. **b** Pearson correlation clustering heatmap based on TH-score. **c** Schematic diagram of TAD hierarchy reorganization classification based on TH-score. ANOVA test was performed to identify significantly differential TH-score genes. **d** Box plot of gene expression log2Foldchange between TH-score decrease and increase genes. Wilcoxon test,  $*p \leq 0.05$ ,  $**p \leq 0.01$ ,  $***p \leq 0.001$ ,  $****p \leq 0.0001$ . **e** Number of significant reorganized TADs using diffDomain. **f** Box plot of gene expression log2Foldchange located in different types of reorganized TADs. Wilcoxon test,  $*p \leq 0.05$ ,  $**p \leq 0.01$ ,  $***p \leq 0.001$ ,  $****p \leq 0.0001$ .



each TAD<sup>11</sup>. Disruption of TAD boundaries by structural variations results in gene expression dysregulation in developmental diseases and cancer<sup>16,37,38</sup>. However, topological boundaries were largely conserved across colon tumors, normal colons, and cell lines<sup>19,22</sup>. The similarity of TAD structure among normal tissue, primary tumor, and metastatic samples was also found in our data (Fig. 4a). Interestingly, we have noticed the differences in

the internal substructures of TAD among different cell types, with subTADs nested within larger TADs, this structure was reported as TAD hierarchy in previous studies<sup>39</sup>. An et al. reported that gene expression of higher-level TAD was significantly higher than that of lower-level TAD and our previously developed TH-score algorithm was defined to represent the TAD level of each gene<sup>39,40</sup>. We further explored whether this change in TAD



hierarchy was related to gene expression regulation. The TAD levels of all genes were first evaluated with TH-score. Based on TH-score and ANOVA test ( $p$  value < 0.05), we obtained genes with significant hierarchical TAD differences and performed Pearson correlation clustering (Fig. 4b). The heatmap showed good clustering of cell types based on TH-score, and the heterogeneity increased with the progression of CRC tumorigenesis and metastasis. Then significant differential TH-score genes were divided into two categories according to the changes in levels among different types of samples (Fig. 4c). The change in TAD levels was observed to be consistent with the change in gene expression (Fig. 4d), indicating that the TAD hierarchy also affected gene expression in tumorigenesis and metastasis of CRC.

To further classify TAD reorganization types, we used the diffDomain<sup>41</sup> to analyze the TAD domain difference between normal colon and primary tumor, between primary tumor and metastasis tumor, and between normal colon and metastasis tumor. All TAD reorganization events were classified into loss, merge, split, complex, zoom, strength\_change\_up, and strength\_change\_down (Fig. 4e). We next explored whether different TAD reorganization types were related to different gene expression change patterns. By overlapping significantly differentially expressed genes with reorganized TAD region, we found that gene expression in strength\_change\_up domain was significantly higher than that in strength\_change\_down domain (Fig. 4f). This suggested that TAD strength is related to gene expression during the tumorigenesis and metastasis of CRC. Overall, recombination of TAD hierarchy and TAD intensity occurs during CRC progression, which is associated with differential gene expression.

### Enhanced *SPP1* loop correlates with *SPP1* upregulation during the progression of CRC

Chromatin loops are formed by architectural proteins including CTCF, YY1, cohesin complex, and transcription factors through a loop extrusion process<sup>42,43</sup>. They bring promoters and enhancers into spatial proximity and often correlate with gene transcription in biological and pathological processes<sup>44,45</sup>. To determine whether the progression of CRC is accompanied by dynamic chromatin loops, we first detected the differential loops for each patient using Juicer hiccupdiff<sup>46</sup>. We observed that the change in loop strength has a consistent correlation with the differential expression of loop anchor genes (Fig. S14). However, we did not find loops with consistent strength changes occurred in all patients. To identify the consistent loop changes during the tumorigenesis and metastasis of CRC, we merged our Hi-C data that belong to each of the normal tissue, primary tumor, and metastasis tumor categories to improve the data resolution for detecting loops. The average length of the loops was 225 kb (Fig. S15a). We identified 2657 enhanced loops and 649 weakened loops in primary tumors relative to normal tissue (Fig. S15b), and 477 enhanced loops and 2745 weakened loops in metastasis tumors relative to primary tumors (Fig. 5a). Genes whose TSS are located in the anchors of the enhanced loops (hereinafter referred to as loop-related genes.) were significantly upregulated (Figs. 5b and S13c). The enhanced loop-related genes in primary tumors were enriched in the cell cycle and proliferation pathways (Fig. S15d). In contrast, the enhanced loop-related genes in metastasis tumors were enriched in pathways of stem cell population maintenance and negative regulation of adhesion (Fig. 5c). These results suggest that the dynamic loops during CRC progression are involved in tumorigenesis and metastasis through regulating gene expression.

Among the enhanced loops, 25 loops (Table S5) were shared between the normal vs. primary tumor and the primary tumor vs. metastasis tumor comparisons. We inspected the genes whose TSS is located within the anchor regions of these loops for their potential roles in CRC tumorigenesis or metastasis (Table S5). Then we explored whether these gene expressions changed in CRC tumorigenesis or metastasis (Fig. 5d). We noticed that the mRNA expression of *SPP1* gene was significantly upregulated in both CRC tumorigenesis and metastasis. The cancer metastasis-related gene *SPP1*<sup>47</sup> (alias: *OPN*) is located at the anchor of an enhanced loop, which connects

the *SPP1* promoter to a region with strong enhancer signals (Fig. 5e, f). *SPP1* encodes a secreted phosphorylated glycoprotein and high expression of *SPP1* is related to tumor progression in gastric cancer and esophageal adenocarcinoma<sup>48,49</sup>. The *SPP1* loop was weak in normal tissues, increased in primary tumors, and further enhanced in metastasis tumors (Fig. 5e, f). This loop was also enhanced in public Hi-C data of CRC relative to normal tissue (Fig. S15e). Concomitant with the increased loop strength, the *SPP1* expression was continuously upregulated during the progression of metastasis CRC (Fig. S15f).

To confirm the clinical relevance of *SPP1* upregulation, we performed *SPP1* immunohistochemical staining of the samples from patients CRC-01, CRC-02, CRC-03, CRC-04, CRC-10, and CRC-13. The expression of *SPP1* protein was very low in normal tissues, increased in tumors, and further enhanced in metastasis CRC (Figs. 5g and S15g). Survival analysis showed that higher expression of *SPP1* is associated with poorer prognosis of CRC patients (Fig. 5h) and patients with adrenocortical carcinoma, liver hepatocellular carcinoma, and lung adenocarcinoma (Fig. S15h). Taken together, dynamic chromatin loops during CRC progression and metastasis are associated with the transcription of cancer-related genes, which may have clinical consequences.

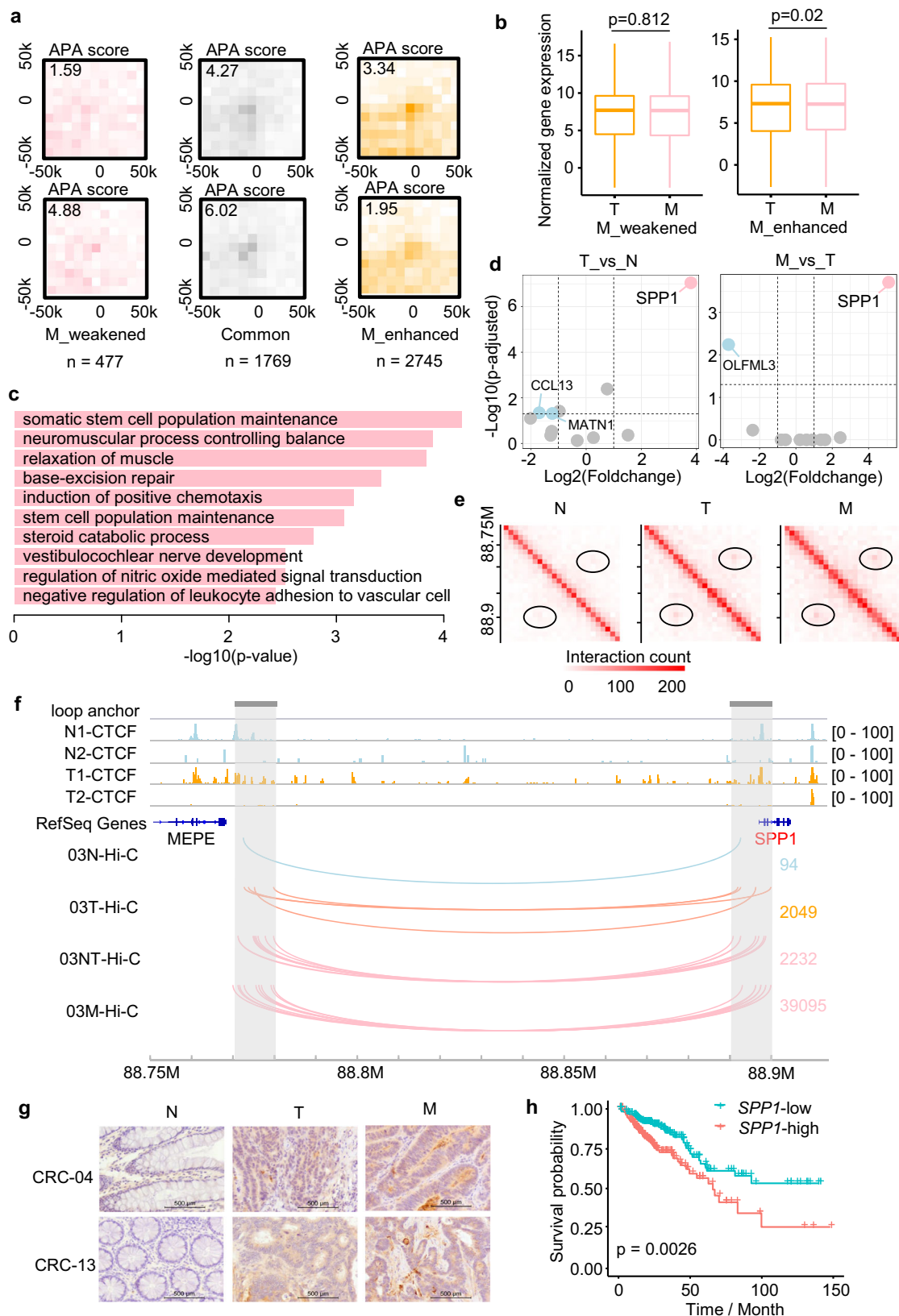
To facilitate the usage of our generated data resources, we developed a filtering pipeline to prioritize genes that are potentially affected by the 3D genome during the tumorigenesis and metastasis of CRC (“Methods”, Fig. S16a). We classified these genes by their putative function and searched their relevant literature on involvement in cancer, and provided the information as a reference (Supplementary Data 1). To verify the biological function of the selected 3D genome-related genes, we selected 3 genes (*ARL4C*, *FLNA*, and *RGCC*) which are predicted to be the 3D genome-related oncogenes during the metastasis. Our filtering pipeline shows that these genes were significantly upregulated in metastasis samples and related to the survival period of CRC patients, while no significant epigenome difference and direct report in CRC metastasis. Therefore, we knocked down these genes by siRNAs in HCT116 cells (Fig. S16b), and evaluated the changes of cell migration and invasion by transwell assay, respectively (Fig. S16c, d). Results show that cell migration and invasion are significantly decreased after knocking down the genes (Fig. S16e, f), especially in the group of *ARL4C* and *FLNA* knockdown. These data demonstrated that those 3D genome-related genes are able to promote the migration and invasion properties of CRC cells.

### Chromatin structural variations in the progression and metastasis of CRC are associated with altered chromatin accessibility and spatial interactions

Chromatin structural variations play an important role in tumorigenesis and metastasis, and are often regarded as decisive diagnostic characterization and therapeutic targets of cancer<sup>50–53</sup>. It has been reported that structural variations can lead to the alteration of chromatin structure, including the formation of new topologically associated domains<sup>54</sup>, induced enhancer-hijacking events<sup>55</sup>, and silencer-hijacking events<sup>18</sup>. To explore the dynamic changes of structural variations in the progression and metastasis of CRC, we used the EagleC<sup>56</sup> to identify structural variations (Fig. 6a, Supplementary Data 2) and gene fusion events (Fig. 6b, Supplementary Data 3).

To explore the impact of structural variations on chromatin accessibility and spatial conformation alteration, we aggregated the ATAC-seq and insulation score profile in structural variation breakpoints flanking regions (Figs. 6c and S17a). We noticed that structural variations were consistent with high chromatin accessibility and low insulation, especially for translocations and genes located in SV flanking regions were evaluated using TH-score to quantify TAD hierarchy. We noticed that the TH-score of tumor-specific SV flanking genes was significantly higher than that of normal-specific SV flanking genes (Fig. 6d). This suggested that structure variation breakpoints tended to form new TAD boundaries and increase the TAD hierarchy of SV flanking genes. For example, we focused on an obvious translocation of chromosome 5 and comprehensively analyzed the A/B





compartment, TAD level, and gene expression in the translocation region (Fig. 6e). We found A/B compartment was switched and the TAD hierarchy was remodeled in chromatin translocation region. A new TAD boundary is formed at the structure variation breakpoint, and the genes located in the breakpoint flanking region are settled in the new subTAD structure,

resulting in the increase of the TAD hierarchy of these genes. These structural alterations jointly affect the differential expression of *CEP120*, *CSNK1G3*, *ZNF608*, *RGAMD2B*, *ALDH7A1*, and other genes around the translocation region. *CEP120* was reported to promote the expansion of the centrosome and the progress of gastric cancer through the ubiquitination of

**Fig. 5 | Enhanced SPP1 loop correlates with SPP1 upregulation during the progression of CRC. a** Heatmaps of representative enhanced loops (left), common loops (middle), and weakened loops (right) comparing primary tumors (T) and mCRC (M). Top: the loops in the primary tumor, bottom: the loops in liver metastasis. The aggregated peak analysis (APA) scores measure the average loop strength. **b** Boxplots of average expression levels of the genes located in the anchors of enhanced and weakened loops (T vs. M) (Significance is based on paired one-tailed Wilcoxon test; the number of genes located in the anchors of weakened and enhanced loops are 1424 and 233, respectively). **c** Representative of most significant gene ontology terms enriched in genes within metastasis-enhanced loops relative to primary tumors. **d** The differential expression volcano plot of genes located in enhanced loop anchors. Red dots: significantly upregulated genes, blue dots: significantly downregulated genes, gray dots: not significantly differentially expressed

*PLK4* mediated by *USP54*<sup>57</sup>. Cyclic RNA (circCSNK1G3) formed by the transcription of oncogene *CSNK1G3* is highly expressed in renal cell carcinoma. Abnormal circCSNK1G3 promotes the growth and metastasis of cancer<sup>58</sup>. Furthermore, the *ALDH7A1* gene is related to the development and prognosis of many kinds of cancer, and its overexpression promotes the metastasis of cancer cells<sup>59</sup>. We also observed upregulation of this gene in our metastatic cancer samples. Therefore, chromatin translocation is associated with chromatin structural remodeling and disorder of gene regulation.

To explore the potential impact of structural variation on chromatin structure of clinical CRC-related genes, we showed the following genes with structural variations (Fig. S17b). First, *ZMIZ1* is overexpressed in a significant percentage of human breast, ovarian, and colon cancers, and ectopic expression of *ZMIZ1* induces cutaneous squamous cell malignancies in a mouse model of cancer<sup>60</sup>. We found chromatin translocation around *ZMIZ1* gene occurred in primary CRC samples and enhanced in metastasis, which may activate the expression of *ZMIZ1* truncated transcript. Next, the *TTC28* gene in CRC has been found to cause L1 element reverse transposon to change the gene expression of the insertion site, which is related to tumorigenesis and development of CRC<sup>61</sup>. We also found a gradually strengthened chromatin translocation signal with the development of cancer around the *TTC28* gene site. Third, the *CLN3* gene is overexpressed in CRC cell lines and tissues, and the overexpression of this gene is related to anti-apoptosis and promotes the growth and proliferation of cancer cells<sup>62</sup>. In our Hi-C data of CRC patients, we found specific chromatin translocation around *CLN3* gene site in metastatic samples, which may drive CRC deterioration and further metastasis. Last, *ABCG1* gene is highly expressed in CRC cells with rapid metastasis and aggregation. Inhibition of *ABCG1* gene can reduce cell aggregation and tumor growth lowering cellular aggregation and tumoroid growth as well as hypoxia-inducible factor 1 $\alpha$  in cancer cells around the central necrotic areas in tumors<sup>63</sup>. We also found significant chromatin translocation around *ABCG1* only in mCRC samples. The chromatin translocations around these CRC-related genes are closely related to the tumorigenesis and metastasis of CRC.

## Discussion

Cancer cell metastasis is a major cause of cancer recurrence, lethality, and poor prognosis<sup>64</sup>. Metastatic cancers differ from primary cancers at the genomic and transcriptional levels, with the genome of metastatic cancers being more unstable and mutated than that of primary cancers<sup>65</sup>. Chromatin structure bridges genomic variation with transcriptional regulation. At present, no clinical cancer metastasis Hi-C data is available. Here, we selected highly mCRC to build comprehensive maps of chromatin topology, gene expression, and chromatin state in matched normal tissues, primary cancers, and metastasis. We provided high-quality and multi-omics data for studying the mechanism of cancer metastasis from the perspective of the 3D genome.

By clustering the gene expression of normal colon, primary cancer, lymph node metastasis, liver metastasis, and normal liver samples of CRC, we found dynamic and continuous alteration of gene expression with the tumorigenesis and metastasis of CRC, which might be related to the “seed and soil” hypothesis proposed by Stephen Paget<sup>66</sup>. The dynamic gene

genes. Horizontal dashed line: p-adjusted value equal to 0.05, Vertical dashed lines: log2Foldchange value equal to  $-1$  and  $1$ . **e** The interaction heatmap of the SPP1 loop in merged normal tissue (N), primary tumors (T) and mCRC (M). **f** The epigenomic and interaction of SPP1 loop. Top: genomic features of chromosome region containing the SPP1 loop; bottom: chromatin interactions in the normal tissue, primary tumor, and two metastatic samples of CRC-03. The increased interactions of the SPP1 promoter with nearby enhancers are accompanied with SPP1 expression increase during tumorigenesis and metastasis. Public CTCF and H3K27ac data are from Johnstone’s research. **g** Immunofluorescence staining of the SPP1 protein level in normal tissue, primary tumor, and liver metastasis of patients CRC-04 and CRC-13 with a scale bar of 500  $\mu$ m. **h** Survival curves of TCGA primary COAD patients stratified by SPP1 expression level. Significance is based on a log-rank test.

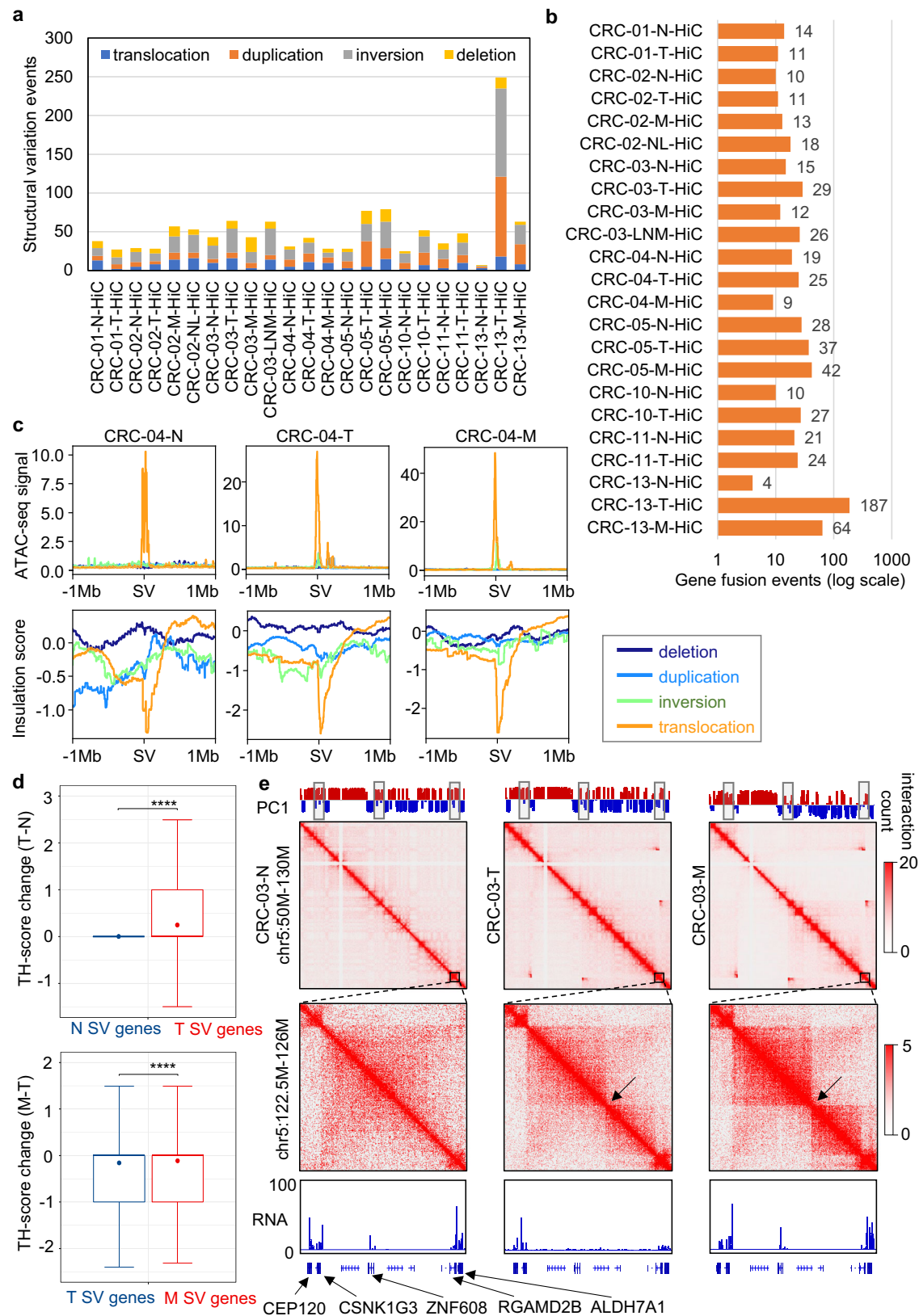
expression supports the presence of dynamic changes in the metabolism of metastasizing cells, contributing to their ability to successfully transition through the changing microenvironments of the metastatic cascade<sup>67</sup>. Compared with normal colon, differentially expressed genes in primary cancer served as tumor markers, and differentially expressed genes in metastatic cancer correlated with patient survival cycle and prognosis in our research. Moreover, compared with the normal colon of patients without metastasis, the differentially expressed genes in the normal colon of patients with metastasis are mainly enriched in the intestinal immune-related pathways. This is consistent with the finding that activating a collaborative innate-adaptive immune response controls metastasis<sup>68</sup>. We showed that CRC patients with and without metastasis have expression differences in both the precancerous lesions and the primary cancer.

In the study of A/B compartment of CRC, we showed that A/B compartment switched during the tumorigenesis and metastasis of CRC and significantly affected gene expression. The A/B compartment with the larger absolute value of the first principal component is relatively stable, and the smaller one is more prone to A/B compartment conversion, which is consistent with the previous three-dimensional genome study of CRC. Johnstone et al. defined an intermediate genome compartment (I compartment) with higher convertibility, largely coinciding with hypomethylated blocks<sup>19</sup>. The switch from either A to B or B to A compartment was also found in our study, and the pattern of A/B compartment switch differed in different patients. We also noticed that A/B compartment is relatively stable in patients without metastasis, which may be due to the more unstable genome of metastatic tumors.

In the regulation of chromatin structure, TAD boundaries are conservative. Therefore, few changes in TAD structure were observed in the previous three-dimensional genome study of cancer, despite changes in gene expression within TADs<sup>14,18,22</sup>. From the TAD hierarchy perspective, we discovered the dynamic changes in TAD internal interactions and regulated gene expression during the development of CRC.

We identified and validated the dynamic changes in the SPP1 loop at the loop scale, which are consistent with the expression of the gene and protein. Identified the impact of loop structural changes on gene enrichment pathways. Based on the chromatin structure screening algorithm, the role of *ARL4C*, *FLNA*, and *RGCC* in the migration and invasion of cancer cell migration was found and verified in vitro.

Frequent structural variation of chromatin is a key feature of cancer cell genome<sup>69</sup>. Structural variation can induce dramatic changes in chromatin organization, thus creating specific signatures that are noticeable by visual inspection of interaction maps<sup>70</sup>. We found that structural variations were consistent with high chromatin accessibility and low insulation, especially for translocations. Moreover, structure variation breakpoints tended to form new TAD boundaries and increase the TAD hierarchy of SV flanking genes. We found structural variation of chromatin at multiple cancer metastasis-related gene sites in our Hi-C data. Overall, recognition of chromatin structural variation based on Hi-C requires low sequencing depth, while it can achieve high recognition accuracy and reflect spatial information. This approach builds a bridge between structural variation and regulatory function.



**Fig. 6 | Chromatin structure variations in the tumorigenesis and metastasis of colorectal carcinoma. a** The distribution of structural variation events called by EagleC. **b** The distribution of gene fusion events mediated by chromatin translocation events. **c** ATAC-seq signal and Insulation score enrichment in SV flanking

region. **d** TH-score change in SV flanking region. \*\*\*\* $p \leq 0.0001$ ,  $t$ -test. **e** A/B compartment, TAD and RNA-seq track in structural variation region between chr5:61600000-63100000 and chr5:121500000-124500000, between chr5:87300000-90500000 and chr5:124500000-13100000.

Genomic instability and high mutation are important characteristics of cancer<sup>71</sup>, but the mechanism of correlation between this instability and cancer occurrence and metastasis has not yet been fully revealed. Our study provides a three-dimensional genome map of clinical metastatic samples of CRC, providing a new perspective for understanding the dynamic changes in colorectal metastasis. Due to the lack of lineage differentiation information, our research has not been able to identify cancer-driving genes. The differential expression of a large number of genes in cancer cells may be related to the trend of more suitable cell migration and invasion in the development and evolution of cancer. We checked the *in vitro* effects of ARL4C, FLNA, and RGCC on cancer metastasis. The recent rapid development of lineage tracking and single-cell Hi-C technology could provide new tools for screening cancer-driving genes.

The 3D genome is a promising direction for studying cancer metastasis, not only because the 3D genome is closer to the upstream of expression regulation, but also because the rules of 3D genome regulation are conserved in normal, primary, and metastatic tumors. This provides a new idea for future research on stable anticancer drug targets. At present, the quantity and quality of clinical solid tissue sequencing data remain challenging. Although our sampling is random and we did not specifically select subgroup patients based on age, gender, or other variables, we could not completely exclude possible confounding factors. We advocate more clinical cancer metastasis samples, strict data quality control, and *in vivo* knockout experiments in future research.

## Methods

### Experimental design and subject information

Patients diagnosed with CRC and willing to participate in the present study were prospectively selected. Thirteen patients who underwent radical resection of CRC or synchronous resection of colorectal primary tumor and liver metastases between April 2018 and September 2018 at Peking University Shougang Hospital were included in the study. The inclusion criteria were as follows: (1) patients are diagnosed with colorectal adenocarcinoma; (2) primary CRC or CRC with liver metastases; and (3) no preoperative chemoradiotherapy, interventional radiology, or radio frequency. Patients with hereditary CRC or multiple primary CRC were excluded. Clinicopathological characteristics included gender, age, tumor histologic type, tumor size, lymph node status, histologic grade, and TNM staging (Table S1). The extent of disease (assessed by chest computed tomography, abdominal computed tomography, rectal or liver magnetic resonance imaging) was recorded. The application to access patient tissue was approved by the Peking University Shougang Hospital Ethics Committee, and informed consent was obtained from all the patients. All ethical regulations relevant to human research participants were followed.

### Biospecimen collection and processing

To ensure the quality of biological samples, we obtained patients' samples during surgery and stored them in a 1 mL DMEM medium containing 5% FBS on ice for short-term transportation. The samples are immediately transferred to the laboratory for procession. Firstly, fat, fibers, blood vessels, and other tissues are separated under a micromanipulator to reduce the contamination of non-tumor cells. Then transfer the cleaned sample into a 1.5 mL EP tube, cut it into a homogenized shape with scissors, mix each tube with 1.5% trypsin (PBS), place it at 37 °C for 1.5 h, and mix it gently every 15 min. After obtaining the single-cell suspension, we filter it once with a 70 µm filter, wash it twice with PBS solution, and then sort the tumor cells by CD326 MicroBeads (MACS, 130-061-101). Finally, the cells are dissolved in PBS solution immediately to construct sequencing libraries. The processing of biological samples is completed within 3 h to ensure the freshness of the sample for sequencing.

### RNA-seq library construction

We used PureLink™ RNA Mini Kit (Invitrogen, 12183018A) to extract total RNA from ~10<sup>6</sup> cells. The experimental procedure was performed

according to the manufacturer's protocol. The total RNA concentration was measured using a Qubit® 3.0 fluorometer (Invitrogen).

RNA-seq library construction was performed according to the manufacture protocol provided by NEBNext® Ultra™ RNA Library Prep Kit (NEB, #E7530). In brief, Poly(A) mRNA was isolated by NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, #E7490) from 0.2 to 1 µg of total RNA. Then the mRNA was fragmented to an average insert size of 200–400 bp at 95 °C for 15 min. Following fragmentation, random primers and reverse transcriptase (NEB, #E7530) were added to reverse transcribe the RNA fragments into single-stranded cDNA. The single-stranded cDNA was then converted into double-stranded DNA. A portion of these cDNA fragments have 5'-phosphorylated dA-tailed ends, so end repairment and adapter ligation were subsequently performed. USER enzyme (NEB #E7350) was added to the ligation mixture to cut cDNA for the reason that circular cDNA was formed after adapters (NEB #E7350) in the NEB kit were ligated to the cDNA. The linear cDNA was purified by AMPure XP beads (Beckman Coulter, A63881) and amplified via 14 PCR cycles to generate the sequencing library. After quantification by Qubit® 3.0 fluorometer (Invitrogen) and fragment size inspection by Agilent Bioanalyzer 2100 (Agilent Technologies), the libraries were sequenced using the Illumina HiSeq X Ten platform (Illumina Inc.).

### ATAC-seq library construction

The ATAC-seq library was constructed using the established method<sup>72</sup>. Briefly, 5 W cells were first separated into a clean 1.5 mL EP tube and centrifuged at 500 rpm, 4 °C, and the supernatant was discarded. 50 µl of ice-cold ATAC-seq lysate was added and pipetted to the cells, which were then left on ice for 5 min to lyse. The lysed cells were centrifuged at 1000 rpm, 4 °C to obtain a precipitate. Tn5 enzyme and buffer (Vazyme, TD501) were added, and the mixture was allowed to react at 37 °C for 30 min. Then, the DNA was purified using Qiagen MinElute PCR Purification Kit immediately. Finally, the sequencing library was constructed using Custom Nextera PCR Primer 1 and 2 (Vazyme, TD202) for 8 cycles of PCR amplification, purified using Qubit® 3.0 fluorometer (Invitrogen) and quality checked by Agilent Bioanalyzer 2100 (Agilent Technologies). The library was sequenced on the Illumina NovaSeq 6000 platform (Illumina Inc.).

### Hi-C library construction

*In situ*, Hi-C library was constructed according to the previously established method<sup>13</sup>. Briefly, ~5 × 10<sup>6</sup> cells were fixed with 1% formaldehyde for 10 min at room temperature and then quenched with 2.5 M glycine. The fixed cells were lysed on Hi-C lysis buffer for 30 min on ice, followed by Mbo I (NEB, R0147) digestion overnight. On the next day, the sample was treated at 62 °C for 10 min to inactivate Mbo I, and then biotin-14-dATP (Life Technologies, 19524-016) was added to label the digested ends. Then, we added ligation master mix (NEB, B0202 & M0202) to ligate proximal DNA ends overnight. On the third day, we used Covaris LE220 (Covaris) to break the end-ligated DNA fragments into a suitable length (300–500 bp), followed by size selection with AMPure XP beads (Beckman Coulter, A63881). Dynabeads MyOne Streptavidin T1 beads (Life Technologies, 65602) were used to pull down the biotin-labeled fragments, which represent spatial chromatin interactions. Finally, the pulled-down fragment was purified by AMPure XP beads (Beckman Coulter, A63881), and the sequencing library was constructed using Illumina primers and protocols (Illumina: [https://support.illumina.com.cn/content/dam/illumina-support/documents/documentation/chemistry\\_documentation/samplepreps\\_truseq/truseq-stranded-mrna-workflow/truseq-stranded-mrna-workflow-reference-1000000040498-00.pdf](https://support.illumina.com.cn/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseq-stranded-mrna-workflow/truseq-stranded-mrna-workflow-reference-1000000040498-00.pdf)). Specifically, twelve PCR cycles were used to amplify the DNA, which was then purified and the library concentration was measured using Qubit® 3.0 fluorometer (Invitrogen). The Agilent Bioanalyzer 2100 (Agilent Technologies) analyzed the library fragment distribution. If both concentration and length distribution passed the quality control, the library was then sequenced using the Illumina HiSeq X Ten platform (Illumina Inc.).



## Cell lines and cell culture

The human CRC cell lines HCT116 were purchased from the American Type Culture Collection (Manassas, VA, USA), which were cultured in RPMI 1640 medium (GIBCO, Carlsbad, CA) supplemented with 10% fetal bovine serum (FBS; GIBCO, Carlsbad, CA), 100 U/mL penicillin (HyClone), and 100 µg/mL streptomycin (HyClone) at 37 °C in an atmosphere of 5% CO<sub>2</sub>. The cultured HCT116 cells were tested by MycAway-Color Kit (Yeasten, Cat. No. 40611) according to the manufacturer's protocol and the result showed there was no contamination of mycoplasma.

## RNA interference and transfection

siRNAs targeting human ARL4C, FLNA, RGCC, and negative control siRNA were designed and synthesized by Gx-health (Suzhou, China). Transfection of three mixed siRNAs (8 pM) was carried out using LipofectamineRNAiMAX (Invitrogen) according to the manufacturer's recommendations. The target sequences for siRNAs are listed in Table S6.

## RNA extraction and RT-PCR analyses

Total RNA from cells was extracted using TRIZOL reagent (Invitrogen). For RT-PCR, RNA was reverse transcribed to cDNA by using a Reverse Transcription Kit (Takara, Beijing, China). RT-PCR was performed using Power SYBR Green (Takara, Beijing, China). The 2<sup>-ΔCT</sup> was calculated for every sample and normalized to GAPDH. The primer sequences are listed in Table S6.

## Cell migration and invasion assay

The migration and invasion of CRC cells were evaluated by Transwell assay (8-µm pore; BD Biosciences). 4 × 10<sup>4</sup> cells in serum-free medium were placed into the upper chamber of an insert for migration assays and invasion assays with 0.2% Matrigel (Corning, New York, USA). Medium containing 10% FBS was added to the lower chamber. After 36–48 h of incubation, the cells that had migrated or invaded through the membrane were fixed in 4% paraformaldehyde and stained by DAPI (4',6-diamidino-2-phenylindole) and counted under microscope (Olympus) in five different fields. The results represent the average cell number in three wells per cell line.

## RNA-seq data analysis

RNA-seq reads were first aligned to the hg19 reference genome with Hisat2 (v2.2.1, parameters: --dta-cufflinks)<sup>73</sup>. The aligned files were converted to binary files using samtools (v1.11, parameters: view -h -bs)<sup>74</sup>. The binary files were further sorted and indexed using samtools (v1.11, with default parameters). The sorted files were then performed quality control using RSeQC (v4.0.0, reference genome: hg19)<sup>75</sup>. The sorted files were then used for transcript assembly, transcript quantification, and expression matrix extraction using StringTie (v2.2.1, parameters: -B -e)<sup>76</sup>. The batch effects in gene expression matrix were removed and differentially expressed genes were screened by DESeq2 (v1.34.0, parameter: padj < 0.1)<sup>77</sup>. Gene set enrichment and variation analysis were performed by R package RNAseqStat2 (v0.1.0.9993). The differential expression and survival effects of genes in TCGA and GTEx databases were verified by GEPIA2<sup>78</sup>.

The filtering pipeline and parameters for high-quality RNA-seq data were as follows. To ensure transcript integrity and avoid significant RNA degradation, we used transcript integrity number (Tin > 50) and gene body coverage skewness (cutoff > -2) to choose high-quality RNA-seq data. All RNA-seq data with low quality were excluded from our research (Fig. S6). Public RNA-seq data were also quality-controlled with the same pipeline. Data obtained from SRA database with accession number SRR13234223, SRR13234224, SRR13234225, SRR13234228, SRR13234229, SRR13234230, SRR13234222, SRR975580, SRR975560 and SRR975562 reached standard.

## ATAC-seq data analysis

ATAC-seq reads were first trimmed with TrimGalore (v3.4, parameters: --phred33 --fastqc --stringency 3 --length 20 -e 0.1). Next, the trimmed fastq

files were aligned to the hg19 reference genome with Bowtie2 (v2.3.5.1, parameters: --very-sensitive-local --no-unal --no-mixed --no-discordant --phred33 -I 10 -X 700)<sup>79</sup>. The aligned files were indexed using samtools (v1.11, with default parameters)<sup>74</sup>. The aligned files were performed quality control using ATACseqQC (v1.16.0)<sup>80</sup>. MACS2 (v2.2.7.1)<sup>81</sup> was used for peak calling with the following parameters: --nomodel, --shift -100, --extsize 200.

## Hi-C data analysis

Hi-C reads were first trimmed with TrimGalore (v3.4, parameters: --phred33 --fastqc --stringency 3 --length 20 -e 0.1). Next, the trimmed fastq files were processed using the HiC-Pro software (v2.11.4, parameters: genome\_fragment = hg19haeiii.bed, ligation\_site = GGCC)<sup>82</sup>. To eliminate the impact of sequencing depth, the output valid data of different tissues in each patient were downsampled to the same amount. The distance vs. Hi-C counts plots and Hi-C data heatmaps were generated using HiCExplorer (v2.2.1.1, command: hicPlotDistVsCounts)<sup>83</sup>. Then, A/B compartments were recognized by R package HiTC (v1.36.0, resolution: 500 kb)<sup>84</sup>. TAD hierarchical structure was identified using OnTAD (v1.2, parameters: resolution 40 kb, -maxsz 50, -minsz 1, -penalty 0.100, -lsize, -ldiff1.960000)<sup>39</sup>. The gene TAD hierarchical score (TH-score)<sup>40</sup> was calculated using the coverage command of bedtools (v2.30.0, parameter: -mean)<sup>85</sup>. Insulation score was calculated using matrix2insulation.pl (v1.0.0, parameters: -is 500000 -ids 200000 -im mean -bmoe 3 -nt 0.1 -v)<sup>86</sup>. TAD reorganization events were detected using diffDomain<sup>41</sup> at 40 kb resolution. Chromatin loops were called using HICCUPS (parameter: -r 1000, -p 2, -f 0.1)<sup>46</sup>. We performed aggregated peak analysis (APA) to assess the quality of the detected loops. The interaction signals surrounding a group of chromatin loops were pooled and represented by a 10 × 10 bin matrix by dividing the observed contact number by the expected contact number. In brief, the APA scores were determined by dividing the center square value by the median value of the nine squares at the lower-right corner. To identify differential loops, we merged all loops across the sample categories to generate a combined loop list and calculated APA scores for each loop in all the sample categories. APA scores were pairwise compared and fold change FC ≥ 1.5 was used as the cutoff for differential loops. Differential loops for each patient were detected using Juicer software with hiccupdiff command. The visualization of Hi-C data was performed using HiCExplorer (v2.2.1.1)<sup>83</sup> and Juicer (v1.11.08)<sup>46</sup> software.

## Filtering pipeline for candidate 3D genome-related genes

A filtering pipeline was developed based on multi-omics data.

Firstly, differential A/B compartment, TAD, TAD score, and loop anchor regions in primary and metastatic tumors were merged. Genes whose TSS falls on the merged chromatin region were selected for further filtering. Detailed differential definitions are as follows. All chromatin bins were first assigned according to PC1 value. The label of bin equals to 1 when the PC1 value of bin is greater than zero. The label of bin equals to 0 when the PC1 value of bin equals to zero. The label of bin equals to -1 when the PC1 value of bin is less than zero. Then sum of tumor labels minus all controlled normal colon labels. If the subtraction result of one bin is greater than 8, this bin is classified as BtoA and differential compartment region. And if the subtraction result of one bin is less than -8, this bin is classified as AtoB and differential compartment region. Differential TADs are defined as at least one of two TAD boundaries that are differential. So, we merged all TAD boundaries as candidate differential boundaries, then if one TAD boundary is recognized as boundary in greater than 4 normal samples and recognized as boundary in less than 2 normal samples, this TAD boundary is classified as loss and differential boundary. If one TAD boundary is recognized as boundary in less than 2 normal samples and recognized as boundary in greater than 5 normal samples, this TAD boundary is classified as gain and differential boundary. TAD score was defined as interaction within TAD divided by the sum of interaction between TAD region and whole genome. For TADs with unchanged TAD boundaries in tumor and normal samples,

the significance of TAD score between tumor and normal samples was calculated using DEseq2 (v1.34.0, parameter:  $\text{padj} < 0.05$ ). Loops were first filtered with  $\text{FDR} < 0.1$  and distance of anchors greater than 4 kb. For all loop anchor candidates, if the number of times a loop anchor appears in tumor samples minus the number of times the loop anchor appears in normal samples greater than 3 or less than  $-3$ , this loop anchor is defined as differential loop anchor.

Secondly, the genes located in chromatin structure differential regions intersected with differentially expressed genes. The significance of differentially expressed genes was called using DEseq2 (v1.34.0, parameter:  $\text{padj} < 0.05$ ).

Next, to eliminate the effect of epigenetic modification, we filtered out genes with significant differences in H3K27ac or H3K27me3. H3K27ac and H3K27me3 with significant differences were calculated based on the signal of H3K27ac and H3K27me3 peaks. We filtered out all differential peaks with  $\text{padj} < 0.01$  and  $|\log_2(\text{foldchange})| > 1$ .

At last, the remained genes were intersected with genes in TCGA that significantly affected the survival period of CRC patients. The significance of the survival period was analyzed using R package survminer (v0.4.9, parameter:  $p < 0.1$ ).

### Chromatin structural variations analysis

Chromatin structural variation breakpoints were identified using EagleC<sup>56</sup> (v0.1.8) with command predictSV (parameters: --hic-5k, --hic-10k, --hic-50k, -g hg19, --balance-type Raw, --prob-cutoff-5k 0.8, --prob-cutoff-10k 0.8, --prob-cutoff-50k 0.8). The output 5K\_combined files were used for data analysis. Gene fusion events were annotated using EagleC (v0.1.8) with command annotate-gene-fusion (parameters: --buff-size 10000 --skip-rows 1 --ensembl-release 75 --species human).

Chromatin translocations were visualized using Hi-Cbreakfinder<sup>87</sup>.

### Statistics and reproducibility

Comparisons of mean differences between two groups of samples were performed by the *t*-test for unpaired data. Details of performed statistical analyses are reported in the figure legends and Methods. Data are presented as mean  $\pm$  SD. *P* value  $< 0.05$  was considered statistically significant. All samples were included when comparing the differences among groups, which determined the sample sizes. At least three biological replicates can be used for the significance test. To ensure reproducibility, all the information regarding participants and experimental design can be found in the “Methods”.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Raw sequenced data and processed data generated from this study are deposited at the BIGD database under accession number [HRA000417](https://bigd.org/HRA000417). Published RNA-seq raw data of matched normal colon, colorectal tumor, liver metastasis, lymph node metastasis, and normal liver were downloaded from the GEO database with accession numbers GSE50760, GSE92914, and GSE162960. Hi-C data of AML and CRC were downloaded from the GEO database with accession numbers GSE152135 and GSE133928. Public CTCF ChIP-seq data were downloaded from the GEO database with accession numbers GSM3930219 and GSM3930229. Public H3K27ac data were downloaded from the GEO database with accession numbers GSM3930220 and GSM3930230, public H3K27me3 data were downloaded from the GEO database with accession numbers GSM3930231 and GSM3930252. Uncropped and unedited images of cells migration and invasion evaluated by transwell assay are provided in Fig. S18.

### Code availability

Custom code used to analyze data in this study is available at GitHub ([https://github.com/XiangXuCode/CRC\\_metastasis\\_code](https://github.com/XiangXuCode/CRC_metastasis_code)) and Zenodo<sup>88</sup>.

Received: 22 June 2024; Accepted: 31 January 2025;

Published online: 04 March 2025

### References

1. Arnold, M. et al. Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66**, 683–691 (2017).
2. Biller, L. H. & Schrag, D. Diagnosis and treatment of metastatic colorectal cancer: a review. *JAMA* **325**, 669–685 (2021).
3. Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging biological principles of metastasis. *Cell* **168**, 670–691 (2017).
4. Birkbak, N. J. & McGranahan, N. Cancer genome evolutionary trajectories in metastasis. *Cancer Cell* **37**, 8–19 (2020).
5. Vanharanta, S. & Massague, J. Origins of metastatic traits. *Cancer Cell* **24**, 410–421 (2013).
6. Pomerantz, M. M. et al. Prostate cancer reactivates developmental epigenomic programs during metastatic progression. *Nat. Genet.* **52**, 790–799 (2020).
7. Denny, S. K. et al. Nf1b promotes metastasis through a widespread increase in chromatin accessibility. *Cell* **166**, 328–342 (2016).
8. Li, R. et al. 3D genome and its disorganization in diseases. *Cell Biol. Toxicol.* **34**, 351–365 (2018).
9. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
10. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
11. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
12. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
13. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
14. Akdemir, K. C. et al. Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat. Genet.* **52**, 294–305 (2020).
15. Akdemir, K. C. et al. Somatic mutation distributions in cancer genomes vary with three-dimensional chromatin structure. *Nat. Genet.* **52**, 1178–1188 (2020).
16. Flavahan, W. A. et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).
17. Wu, P. et al. 3D genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. *Nat. Commun.* **8**, 1937 (2017).
18. Xu, J. et al. Subtype-specific 3D genome alteration in acute myeloid leukaemia. *Nature* **611**, 387–398 (2022).
19. Johnstone, S. E. et al. Large-scale topological changes restrain malignant progression in colorectal cancer. *Cell* **182**, 1474–1489.e1423 (2020).
20. Iyyanki, T. et al. Subtype-associated epigenomic landscape and 3D genome structure in bladder cancer. *Genome Biol.* **22**, 105 (2021).
21. Zhu, Y. et al. Global loss of promoter-enhancer connectivity and rebalancing of gene expression during early colorectal cancer carcinogenesis. *Nat. Cancer* **5**, 1697–1712 (2024).
22. Ren, B. et al. High-resolution Hi-C maps highlight multiscale 3D epigenome reprogramming during pancreatic cancer metastasis. *J. Hematol. Oncol.* **14**, 120 (2021).
23. Salvetti, A. & Greco, A. Viruses and the nucleolus: the fatal attraction. *Biochim. Biophys. Acta* **1842**, 840–847 (2014).
24. Zink, D., Fischer, A. H. & Nickerson, J. A. Nuclear structure in cancer cells. *Nat. Rev. Cancer* **4**, 677–687 (2004).
25. Kim, S. K. et al. A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. *Mol. Oncol.* **8**, 1653–1666 (2014).

26. Ma, Y. S. et al. Proteogenomic characterization and comprehensive integrative genomic analysis of human colorectal cancer liver metastasis. *Mol. Cancer* **17**, 139 (2018).
27. Menck, K. et al. High-throughput profiling of colorectal cancer liver metastases reveals intra- and inter-patient heterogeneity in the EGFR and WNT pathways associated with clinical outcome. *Cancers* **14**, 2084 (2022).
28. Liu, W., Li, L., Ye, H., Tao, H. & He, H. Role of COL6A3 in colorectal cancer. *Oncol. Rep.* **39**, 2527–2536 (2018).
29. Sun, L. et al. Down-regulation of long non-coding RNA RP11-708H21.4 is associated with poor prognosis for colorectal cancer and promotes tumorigenesis through regulating AKT/mTOR pathway. *Oncotarget* **8**, 27929–27942 (2017).
30. Takashima, M. et al. Overexpression of alpha enolase in hepatitis C virus-related hepatocellular carcinoma: association with tumor progression as determined by proteomic analysis. *Proteomics* **5**, 1686–1692 (2005).
31. Tomaino, B. et al. Circulating autoantibodies to phosphorylated  $\alpha$ -enolase are a hallmark of pancreatic cancer. *J. Proteome Res.* **10**, 105–112 (2011).
32. Tsai, S. T. et al. ENO1, a potential prognostic head and neck cancer marker, promotes transformation partly via chemokine CCL20 induction. *Eur. J. Cancer* **46**, 1712–1723 (2010).
33. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
34. Zeng, J. et al. Chemokine CXCL14 is associated with prognosis in patients with colorectal carcinoma after curative resection. *J. Transl. Med.* **11**, 6 (2013).
35. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
36. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7 (2013).
37. Hnisz, D. et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).
38. Lupiáñez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
39. An, L. et al. OnTAD: hierarchical domain structure reveals the divergence of activity among TADs and boundaries. *Genome Biol.* **20**, 282 (2019).
40. Du, G. et al. The hierarchical folding dynamics of topologically associating domains are closely related to transcriptional abnormalities in cancers. *Comput. Struct. Biotechnol. J.* **19**, 1684–1693 (2021).
41. Hua, D. et al. DiffDomain enables identification of structurally reorganized topologically associating domains. *Nat. Commun.* **15**, 502 (2024).
42. Weintraub, A. S. et al. YY1 is a structural regulator of enhancer-promoter loops. *Cell* **171**, 1573–1588.e1528 (2017).
43. Stadhouders, R., Filion, G. J. & Graf, T. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569**, 345–354 (2019).
44. Kubo, N. et al. Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat. Struct. Mol. Biol.* **28**, 152–161 (2021).
45. Zhou, Q. et al. ZNF143 mediates CTCF-bound promoter-enhancer loops required for murine hematopoietic stem and progenitor cell function. *Nat. Commun.* **12**, 43 (2021).
46. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
47. Denhardt, D. T. & Chambers, A. F. Overcoming obstacles to metastasis-defenses against host defenses: osteopontin (OPN) as a shield against attack by cytotoxic host cells. *J. Cell. Biochem.* **56**, 48–51 (1994).
48. Zhuo, C. et al. Elevated THBS2, COL1A2, and SPP1 expression levels as predictors of gastric cancer prognosis. *Cell. Physiol. Biochem.* **40**, 1316–1324 (2016).
49. Lin, J. et al. Osteopontin (OPN/SPP1) isoforms collectively enhance tumor cell invasion and dissemination in esophageal adenocarcinoma. *Oncotarget* **6**, 22239–22257 (2015).
50. Mitelman, F., Johansson, B. & Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* **7**, 233–245 (2007).
51. Roukos, V., Burman, B. & Misteli, T. The cellular etiology of chromosome translocations. *Curr. Opin. Cell Biol.* **25**, 357–364 (2013).
52. Lieber, M. R. Mechanisms of human lymphoid chromosomal translocations. *Nat. Rev. Cancer* **16**, 387–398 (2016).
53. Roukos, V. & Misteli, T. The biogenesis of chromosome translocations. *Nat. Cell Biol.* **16**, 293–300 (2014).
54. Franke, M. et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).
55. Ooi, W. F. et al. Integrated paired-end enhancer profiling and whole-genome sequencing reveals recurrent CCNE1 and IGF2 enhancer hijacking in primary gastric adenocarcinoma. *Gut* **69**, 1039–1052 (2020).
56. Wang, X., Luan, Y. & Yue, F. EagleC: a deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps. *Sci. Adv.* **8**, eabn9215 (2022).
57. Zhang, C. et al. Centrosomal protein 120 promotes centrosome amplification and gastric cancer progression via USP54-mediated deubiquitination of PLK4. *iScience* **26**, 105745 (2023).
58. Li, W. et al. CircCSNK1G3 up-regulates miR-181b to promote growth and metastasis via TIMP3-mediated epithelial to mesenchymal transitions in renal cell carcinoma. *J. Cell. Mol. Med.* **26**, 1729–1741 (2022).
59. Lu, H. J. et al. The impact of ALDH7A1 variants in oral cancer development and prognosis. *Aging* **14**, 4556–4571 (2022).
60. Rogers, L. M., Riordan, J. D., Swick, B. L., Meyerholz, D. K. & Dupuy, A. J. Ectopic expression of Zmiz1 induces cutaneous squamous cell malignancies in a mouse model of cancer. *J. Investig. Dermatol.* **133**, 1863–1869 (2013).
61. Pitkäänen, E. et al. Frequent L1 retrotranspositions originating from TTC28 in colorectal cancer. *Oncotarget* **5**, 853–859 (2014).
62. Rylova, S. N. et al. The CLN3 gene is a novel molecular target for cancer drug discovery. *Cancer Res.* **62**, 801–808 (2002).
63. Namba, Y. et al. Depletion of lipid efflux pump ABCG1 triggers the intracellular accumulation of extracellular vesicles and reduces aggregation and tumorigenesis of metastatic cancer cells. *Front. Oncol.* **8**, 376 (2018).
64. Steeg, P. S. Targeting metastasis. *Nat. Rev. Cancer* **16**, 201–218 (2016).
65. Nguyen, B. et al. Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* **185**, 563–575.e511 (2022).
66. Paget, S. The distribution of secondary growths in cancer of the breast. 1889. *Cancer Metastasis Rev.* **8**, 98–101 (1989).
67. Bergers, G. & Fendt, S. M. The metabolism of cancer cells during metastasis. *Nat. Rev. Cancer* **21**, 162–180 (2021).
68. Sun, L. et al. Activating a collaborative innate-adaptive immune response to control metastasis. *Cancer Cell* **39**, 1361–1374.e1369 (2021).
69. Li, Y. et al. Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112–121 (2020).
70. Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467 (2018).
71. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).



72. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21–29 (2015).
73. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
74. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
75. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
76. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
77. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
78. Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **47**, W556–W560 (2019).
79. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
80. Ou, J. et al. ATACseqQC: a bioconductor package for post-alignment quality assessment of ATAC-seq data. *BMC Genom.* **19**, 169 (2018).
81. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
82. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
83. Wolff, J. et al. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **48**, W177–W184 (2020).
84. Servant, N. et al. HiTC: exploration of high-throughput ‘C’ experiments. *Bioinformatics* **28**, 2843–2844 (2012).
85. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
86. Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
87. Dixon, J. R. et al. Integrative detection and analysis of structural variation in cancer genomes. *Nat. Genet.* **50**, 1388–1398 (2018).
88. Xu, X. The code for metastatic colorectal carcinoma multi-omic data analysis. Zenodo, <https://doi.org/10.5281/zenodo.14676683> (2025).

## Acknowledgements

Z.G., D.H., and J.Gu were supported by the National Natural Science Foundation of China (82073223) and Beijing Natural Science Foundation (Z171100001017087). J.Gan, R.L., and C.L. were supported by the National Natural Science Foundation of China (32025006, 32288102, 31871266) and the National Key Research and Development Program of China (2021YFA1100300). X.X. and H.C. were supported by the National Natural Science Foundation of China (62173338) and the Beijing Nova Program of Science and Technology (Z191100001119064). Part of the data analysis was performed on the High-Performance Computing Platform of the Center for Life Sciences, Peking University.

## Author contributions

J.Gu, C.L., and H.C. designed and supervised the study. X.X. and J.Gan performed data analysis and wrote the manuscript. Z.G. acquired patient informed consents and clinical samples. J.Gan and R.L. performed sample processing and constructed sequencing libraries. D.H. collected clinical information of patients and performed immunohistochemistry imaging. L.L., Y.Luo, Q.Y., J. X., Y.Li, Q.F, T.P., Y.W., and Z.X. participated in data analysis and interpretation. A.H. and H.H. participated in immunohistochemistry imaging. F.L., W.H., J.L., T.L., and X.B. provided constructive guidance and detailed inspection. All authors edited and approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains

supplementary material available at

<https://doi.org/10.1038/s42003-025-07647-2>.

**Correspondence** and requests for materials should be addressed to Hebing Chen, Cheng Li or Jin Gu.

**Peer review information** *Communications Biology* thanks Motoki Takaku and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editors: Aylin Bircan, Christina Karlsson Rosenthal.

**Reprints and permissions information** is available at

<http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025, corrected publication 2025