



The hierarchical folding dynamics of topologically associating domains are closely related to transcriptional abnormalities in cancers



Guifang Du^{a,1}, Hao Li^{a,1}, Yang Ding^a, Shuai Jiang^a, Hao Hong^a, Jingbo Gan^b, Longteng Wang^b, Yuanping Yang^c, Yinyin Li^d, Xin Huang^a, Yu Sun^a, Huan Tao^a, Yaru Li^a, Xiang Xu^a, Yang Zheng^a, Junting Wang^a, Xuemei Bai^a, Kang Xu^a, Yaoshen Li^c, Qi Jiang^e, Cheng Li^b, Hebing Chen^{a,*}, Xiaochen Bo^{a,*}

^a Beijing Institute of Radiation Medicine, Beijing 100850, China

^b Center for Statistical Science, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing 100871, China

^c Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), Guangzhou 510320, China

^d Department of Liver Disease, Fifth Medical Center, Chinese People's Liberation Army General Hospital, Beijing 100069, China

^e Tongfang Cloud (Beijing) Technology Co., Ltd, Beijing 100083, China

ARTICLE INFO

Article history:

Received 1 February 2021

Received in revised form 14 March 2021

Accepted 16 March 2021

Available online 26 March 2021

Keywords:

Hierarchical TAD

Cancers

Transcriptional regulation

TH score

Hi-C

Colorectal cancer

ABSTRACT

Recent studies have shown that the three-dimensional (3D) structure of chromatin is associated with cancer progression. However, the roles of the 3D genome structure and its dynamics in cancer remains largely unknown. In this study, we investigated hierarchical topologically associating domain (TAD) structures in cancers and defined a “TAD hierarchical score (TH score)” for genes, which allowed us to assess the TAD nesting level of all genes in a simplified way. We demonstrated that the TAD nesting levels of genes in a tumor differ from those in normal tissue. Furthermore, the hierarchical TAD level dynamics were related to transcriptional changes in cancer, and some of the genes in which the hierarchical level was altered were significantly related to the prognosis of cancer patients. Overall, the results of this study suggest that the folding dynamics of TADs are closely related to transcriptional abnormalities in cancers, emphasizing that the function of hierarchical chromatin organization goes beyond simple chromatin packaging efficiency.

© 2021 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

In cancer progression, alterations in nuclear morphology are common [1,2]. In addition, nuclear features are helpful to determine the molecular subtype of tumors and are important for the diagnosis and treatment of cancer patients. Recent studies also have shown that the three-dimensional (3D) structure of chromatin is associated with cancer progression [3–10]. Moreover, cancer progression may be driven by locally abnormal promoter-enhancer interactions [11]. However, the relationship between the 3D genome and tumorigenesis and its mechanism still need to be explored.

In the nucleus, chromatin exists as a hierarchical folding structure [2,12–14]. With the development of high-throughput chromosome conformation capture (Hi-C), topologically associating domains (TADs) have been identified at the megabase level [12].

These domains are characterized by strong intra domain interaction and weak inter domain interactions. The interactions within TADs promote the 3D spatial proximity between remote genomic sites in the linear genome sequence. TADs are generally regarded as the structural and functional units of the genome, which define the regulatory pattern [15,16].

The disruption of TADs and the abnormal fusion of the TAD boundary can cause a variety of developmental disorders and diseases [14,17–19]. Many reports also indicate that the destruction of TAD boundaries in cancer cells can cause abnormal activation of oncogenes [3,8,20,21]. There are two known mechanisms of TAD destruction. First, a mutation or epigenetic inactivation can occur in a single TAD boundary, which affects gene regulation in both sides of the TAD. Second, genome rearrangement can lead to the disruption and fusion of TADs without affecting their boundaries, resulting in new regulatory domains and the abnormal activation of oncogenes [20]. Such as, in HEK-293T cells, the deletion of a TAD boundary has been demonstrated to cause the transcriptional activation of *TAL1* [8]. An *IGF2* locus tandem duplications intersecting with a TAD boundary also has been shown to cause

* Corresponding authors.

E-mail addresses: chb-1012@163.com (H. Chen), boxc@bmi.ac.cn (X. Bo).

¹ These authors contributed equally to this work.

oncogene activation in colorectal cancer [22]. However, in view of the overall conservation of TAD boundaries [4,15,23], only a few target genes related to tumorigenesis based on a survey of TAD boundary disruptions has been reported.

Recent studies have shown that TADs forming into a hierarchical structure inside. These hierarchical domains have been defined as ‘metaTADs’ [13] or ‘sub-TADs’ [24–27]. Even in a single cell, Zhuang’s lab observed TADs and a sub-TAD-like structure by multiplexed super-resolution fluorescence *in situ* hybridization imaging [28]. Quentin et al. also have observed that TADs are subdivided into discrete nanodomains by using super-resolution microscopy [28,29]. Many researchers have demonstrated that the hierarchical structures in TAD are correlated with genetic, epigenomic, and expression features [13,25,26,30]. However, the involvement of these hierarchies in biological functions and how these structures are altered in disease remains poorly understood. Because the genome topology and nuclear organization in human tumors are reported to be highly dysregulated, we were interested to determine whether the hierarchical TADs in cancer are also disrupted as well as whether they are associated with aberrant gene expression.

Here, we investigated hierarchical TAD structures in cancers. First, we used OnTAD [25], an optimized method introduced by Lin et al., to identify hierarchical TADs and to calculate the hierarchical levels of TAD boundaries. Our detailed analyses showed that although the locations of the boundaries identified by OnTAD were largely conserved across all of the tested tissues and cell lines, consistent with the previous findings [4,15,23], the topological boundary levels can surprisingly distinguish between primary colorectal tumor tissues, normal colon tissues, colorectal cancer cell lines and normal cell lines. In order to study the effect of TAD hierarchy on gene expression, we developed a method to score the TAD nesting level of each protein coding gene, denoted as the TAD hierarchical score (TH score). Using this scoring method, we determined whether there is a close relationship between the TAD hierarchical folding dynamics and transcriptional abnormalities in cancers.

2. Materials and methods

2.1. Hi-C data sets

Hi-C data of 13 primary colorectal tumors, 9 normal colons, 4 colorectal cancer cell lines, a normal colon cell line and a primary fibroblast line were obtained from Johnstone et al., binned at 40 kb [4]. Hi-C data of the cell lines K562, GM12878, IMR90, HMEC, HUVEC, KBM7 and NHEK were obtained from Rao et al. [12]. Hi-C data of the cell lines SK-N-MC, CAK12, PANC1, NCIH460, T47D, G401, RPMI7951, LNCaP and SKMEL5 were obtained from the Dekker lab of the Encyclopedia of DNA Elements (ENCODE) project [31]. Hi-C data of the cell lines RWPE1, 22RV1 and C42B were obtained from Farnham lab of ENCODE. Hi-C data of the cell lines RPMI8226 and U266 were obtained from Pengze et al. [32] (details in Supplementary Tables 1, 2, and 3).

2.2. Identification of hierarchical TADs and calculation of the TAD boundary level

We obtained hierarchical TADs using OnTAD software [25] from Hi-C matrix data. In brief, OnTAD first uses an adaptive local minimum search algorithm to identify candidate TAD boundaries. Then, OnTAD assembles TADs from the candidate boundaries using a recursive algorithm. Because the structure of <3 bins is too small to form a domain, we set the minimum size of TAD to 3 bins for 40 kb matrix Hi-C data. Because TADs are known to be smaller than a few megabytes, the maximum size of TAD was set to 50 bins. We

used 0.1 as the value of “penalty” in our analyses to select positive TADs, as lower penalty score would result in false positive TADs. These parameters were constants across the analyses for all the samples, in order to ensure comparability.

The methods for calculating the TAD boundary level in our study were the same as those reported previously [25]. In detail, if a boundary was shared by no more than one TAD on each side, the boundary was classified as level 1; if a boundary was shared by no more than two TADs on each side, the boundary was classified as level 2; by analogy, if there were no more than five TADs on each sides of a boundary, then it could be upgraded to level 5. For instance, a boundary should be classified as level 4 if it was shared by three TADs to its left and four TADs to its right.

2.3. TH score and differential gene analysis

We defined a “TH score” to assess the TAD nesting level of genes. Specifically, the TH score for gene_{*i*} is defined as $s_i = \frac{1}{l_i} \sum_{j=1}^{n_i} l_{ij} \times h_{ij}$, where l_i is the length of gene_{*i*}, l_{ij} ($j = 1, \dots, n_i$) is the length of segments obtained by partitioning the gene locus with the TAD nesting level and h_{ij} is the TAD nesting level in the corresponding segment (Fig. 3A). We calculated “TH score” for each gene using bedtools software with “coverage” command on two given BED files describing gene regions and TAD regions (see Supplementary data for more details).

We then calculated the differential TH score (Δ TH score = TH_{can} - TH_{normal}) of each gene between normal and tumor tissues and calculated the statistical significance of the Δ TH score using Mann-Whitney *U* test. Genes with Δ TH score > 0 and $p < 0.05$ was considered significantly upregulated and genes with Δ TH score < 0 and $p < 0.05$ was considered significantly downregulated.

2.4. Clustering analysis

The clustering analyses were performed and visualized using the “scipy.cluster.hierarchy” python library [33]. In practice, the method parameter was set to “Ward”, which is suitable for quantitative variables and “optimal_ordering” parameter was set to “False”.

2.5. RNA sequencing (RNA-seq) data processing

We downloaded the processed RNA-seq data (counts and fragments per kilobase of transcript per million mapped reads; FPKM) from the Cancer Genome Atlas (TCGA) project at the University of California at Santa Cruz Genomic Data Commons Hub (<https://gdc.xenahubs.net>). Then, we performed differential gene expression analysis using DESeq2 [34]. Absolute of log₂ fold change > 1, adjusted p-value < 0.05 were used as cutoff for differentially expressed genes.

The association between genes and colorectal cancer survival was analyzed using patients’ information from the TCGA database and the “survminer” R package, and $p < 0.05$ (unadjusted p value) was considered significant.

2.6. Kyoto Encyclopedia of genes and genomes (KEGG) pathway and gene ontology (GO) enrichment analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 [35] was used to perform KEGG pathway and GO enrichment analysis. The terms with a p value < 0.05 were reported.

3. Results

3.1. The landscape of hierarchical TADs in normal colons, colorectal tumors, and cell lines

To investigate the hierarchical TAD structures in cancer cells, we analyzed a published Hi-C data set that included 13 colorectal tumor tissues, 9 normal colon tissues, 4 colorectal cancer cell lines (SW480, RKO, HCT116 and LS-174T), 1 normal colon derived cell line (FHC) and 1 primary fibroblast cell line (Wi38) [4]. First, we used OnTAD [25] to identify the hierarchical TADs (Fig. 1A). We found 5644 (SD = 197) TADs per sample, on average, and 4571 (SD = 115) boundaries. We also calculated the hierarchical levels of the TAD boundaries, which were defined as the maximum number of TADs that share a boundary on either its left or right side (Fig. 1B). As shown in Fig. 1C–D, hierarchical structures were common in all tissue and cell line samples analyzed. The numbers of TADs and TAD boundaries decreased at each consecutive level.

It has been reported that the locations of TAD boundaries are highly conserved across different cell types and tissues [4,15,23]. First, we compared pairwise similarities across all of the samples. The tumor tissues had 85.32% of their TAD boundaries overlapping with those of the normal colon tissues and 84.08% of their TAD boundaries overlapping with those of the cell lines, on average. The locations of the topological boundaries were similar across the tissues and cell lines, especially for tissue samples (Fig. 1E–F). Additionally, genome-wide systematic comparison analysis across samples showed that the hierarchical clustering of samples according to the pairwise similarity scores could not distinguish cancer samples and normal samples (Fig. S1A). These results are consistent with those of previous studies, thus demonstrating the reliability of our identified results.

3.2. The hierarchical level of the topological boundary alone can distinguish normal colons, colorectal tumors, and cell lines

Next, we investigated the levels of the TAD boundaries in detail. The normal and tumor tissues were well-separated by the clustering with hierarchical TAD boundary levels (Fig. 2A). As shown in the dendrogram of Fig. 2A, all the colorectal tumor tissues were clustered into one cluster, and all the normal tissues were clustered into the other cluster. Similar results were observed for all other cell lines, such as Wi38, HCT116, FHC. It could also be observed from the heat map that the same type of samples has high Pearson correlation coefficient (colored in dark red) in gene scores. These results indicated that the hierarchical levels of the topological boundaries were tissue- and cell type-specific, although the locations of the topological boundaries were similar across the tissues and cell lines (Fig. 1E–F). This analysis performed better than a similar analysis of the mRNA levels (Fig. 2B, and Fig. S2A–B), suggesting changes in the folding dynamics of TADs during tumorigenesis. It is noteworthy that the levels were less homogeneous in the colorectal tumor tissues (marked by dashed yellow lines in Fig. 2A) than in the others (marked by solid yellow lines in Fig. 2A), indicating that the hierarchical levels of topological boundary could reflect the inter-heterogeneity of tumor tissues. TAD boundaries have been reported to be enriched with actively epigenetic signals and expressed genes. Therefore, we studied the genes located at the boundary of the TADs. It was found that compare to the low level boundaries (level 1), hub boundaries (level 3+) were more enriched with essential genes, which were proved differentially expressed in normal and tumor tissues (Fig. 2C) [36]. This result may be related to the higher transcription level of the genes at the hub boundaries (level 3+). The number of samples sharing one boundary increased along with the boundary

levels, indicated that compared with other boundary levels, a high-level TAD boundary was more conserved among the 42 investigated samples (Fig. 2D). This finding suggests that an abnormality of a high-level TAD boundary indicates a greater likelihood of leading to a transcriptional regulation-related disorder or disease.

3.3. The TAD nesting level of genes in tumors is different from that in normal tissues

To study the effect of the TADs hierarchy on gene regulation, we defined a “TH score” to assess the TAD nesting level of genes (Fig. 3A–B). We found that at least 93.5% of the genes in each sample were located in TAD structures (TH score > 0, Fig. 3C), and the TH scores of genes in tumor tissues were statistically different compared to those in normal tissues. In tumor tissues, the number of genes with a high TH score (2 ~ 3 and 3+) was statistically significantly greater than that in normal colon tissues; by contrast, the number of low-TH-score genes (0 ~ 1) in tumor tissues was statistically significantly less than that in normal colon tissues (Fig. 3C). The same pattern could be found in cell lines (Fig. S3). We speculated that these results may be due to two reasons: 1) more frequent hierarchical folding of TADs in single cancer cells; 2) aggregation reflection in bulk cells caused by intra-tumor heterogeneity. Moreover, the hierarchical clustering of genes by the TH score also showed that the TAD nesting level of genes could distinguish colorectal tumor and normal tissues (Fig. 3D), with a better performance than that based on the mRNA-level (Fig. 2B, and Fig. S2A–B). Moreover, as shown in the dendrograms of Figs. 2A and 3D, all the cancer cell lines were clustered together by TH score-based clustering, however, the clustering with hierarchical TAD boundary levels were not able to do this, suggesting that TH score-based clustering performed better than the clustering with hierarchical TAD boundary levels. Like the hierarchical levels of the topological boundaries, the TAD nesting levels of genes were also less homogeneous in the colorectal tumor tissues (marked by dashed yellow lines in Fig. 3D) than in the others (marked by solid yellow lines in Fig. 3D), suggesting that the analysis of hierarchical TADs might be helpful for molecular typing of colorectal cancer. A rank-sum test analysis showed that compared to the normal samples, there were 2646 genes with statistically significantly upregulated TH scores and 1201 genes with downregulated scores in the tumor samples (Fig. 3E). The fractions of these genes whose TH scores were statistically upregulated or downregulated in tumor tissues were increased along with the levels of the TAD boundaries (Fig. 3F). Remarkably, these genes included 35 (17.5% of 200) epithelial mesenchymal transition (EMT)-associated genes, which play pivotal roles in colorectal cancer progression (e.g., *FOXC2*, *MCM7*, *COL7A1*, *TGM2*, *FSTL1*, *EFEMP2*, *MXRA5*, *SPP1*, *SPOCK1*, *TNFRSF12A*, *FBLN1*, *CD59* and *FGF2*, Fig. 3E) [37].

Our analyses suggest that the hierarchical TAD level of genes can distinguish between colorectal tumors, normal colons, and cell lines. Hence, the most profound hierarchical folding dynamics of TADs in tumors might be associated with tumor progression. The hierarchical levels of TADs reflected the heterogeneity of tumor tissues, making it a prospective tool to screen potential targets for the diagnosis and treatment of colorectal cancer.

3.4. The hierarchical TAD level of genes is related to abnormal transcription and prognosis in colorectal cancer

Based on the above analyses, we identified four gene patterns for the TH score (upregulated, downregulated, stable high (TH score > 3), and stable low (TH score < 1)) (Fig. 4A). By performing KEGG enrichment analysis, we found that the genes with different TH scores in tumor tissues and normal tissues were enriched in tumor related signaling pathways (Fig. 4B, Table S4). Using the col-

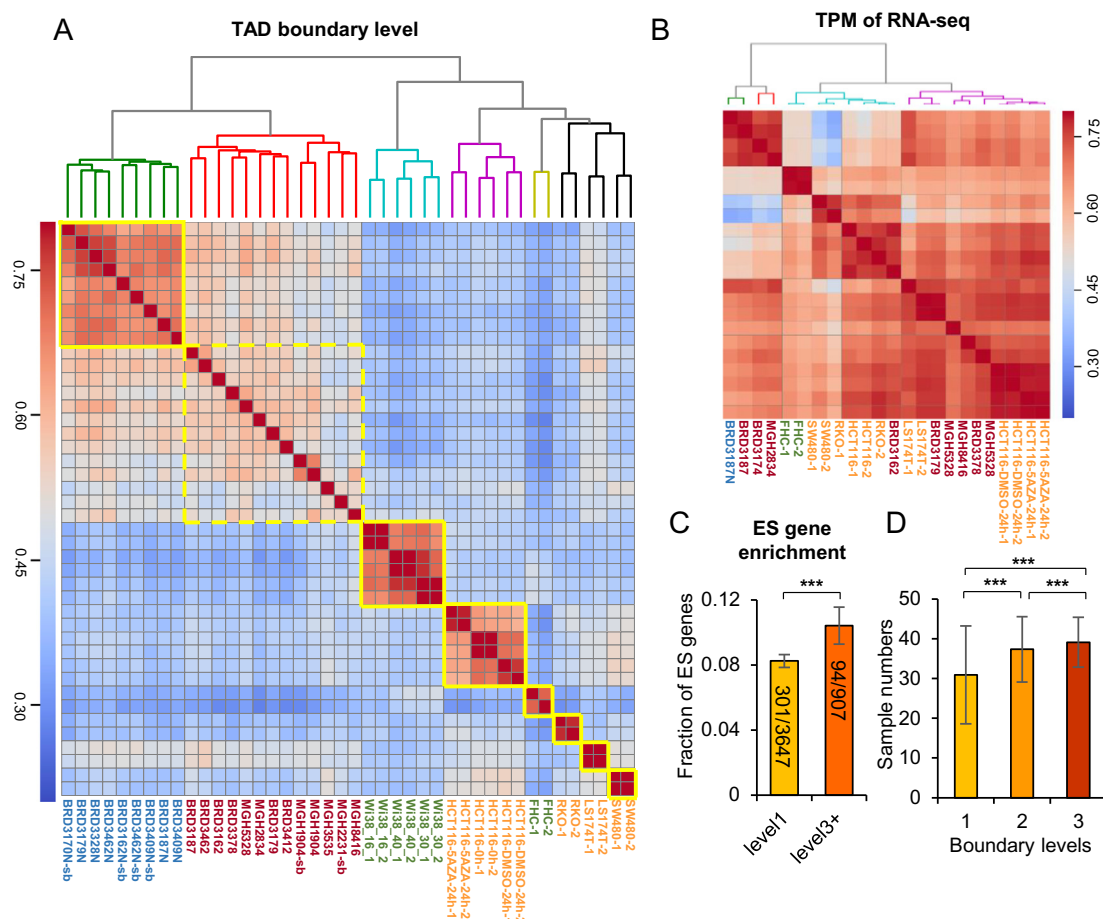


Fig. 2. Topological boundary levels can distinguish normal colon tissues, colon tumor tissues, and cell lines. (A) Heatmap showing the pairwise Pearson correlation coefficient between the boundary levels (coolwarm heat) in colon tumor tissues (sample ID colored in red), normal colon tissues (sample ID colored in blue), cancer cell lines (sample ID colored in orange) and cell lines (sample ID colored in green). The order of samples (rows, columns) is consistent with Ward's linkage hierarchical clustering (top). (B) Heatmap showing the pairwise Pearson correlation coefficient between the TPM of genes (coolwarm heat) in all samples. (C) Barplot showing the enrichment of essential genes located at level 1 or level 3 + boundaries. The fraction of genes located within each specific boundary is presented in the respective column. *** indicates $P < 0.001$ using the t -test. (D) Barplot showing the number of samples sharing one boundary in different boundary levels. *** indicates $P < 0.001$ using the t -test. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

orectal cancer RNA-seq data in TCGA, we defined active genes as those with FPKM > 5 in all samples. Then, for the stable-high and stable-low sets, we computed the fraction of active genes. The results demonstrated that the active genes were more enriched in the stable high gene set (Fig. 4C), which is in line with previous studies [25]. Furthermore, we found that the genes with the most significantly increased TH scores (Δ TH score (cancer – normal) > 0 and $p < 0.01$) were significantly enriched with respect to the genes having up-regulated mRNA levels (Fig. 4D) and that the genes with the most significantly decreased TH scores (Δ TH score (cancer – normal) < 0 and $p < 0.01$) were significantly enriched with respect to the genes having down-regulated mRNA levels (Fig. 4E), suggesting that the abnormal transcription of these genes might be related to changes in the TAD nesting level.

Based on differential gene analysis, we constructed a pipeline (Fig. 5A) and screened 141 genes whose TH score and mRNA level were significantly upregulated (Table S5). Among these, there were 13 genes whose high mRNA expression levels were statistically significantly associated with poor prognosis of colorectal cancer. On the contrary, there were 78 genes (Table S5) whose TH score and mRNA level were significantly downregulated. Among these, there were 6 genes whose high mRNA expression levels were statistically significantly associated with a good prognosis of colorectal cancer (Fig. 5A). In particular, *C5orf46*, *SPP1*, *IBSP*, *EDAR*, and

oxysterol-binding protein-like protein 3 (*OSBPL3*) were statistically significantly up-regulated in the tumor tissue in terms of the TAD level and the mRNA level (Fig. 5B and C), and their high expression levels were statistically significantly associated with a poor prognosis (Fig. 5D and S4A–D). Meanwhile, *CA2*, *PPARGC1B*, *FOXD2*, *CLCA1*, and *NRAP* were significantly down-regulated in the tumor tissue in terms of the TAD level and the mRNA level (Fig. 5E–F), and their high expression levels were statistically significantly associated with a good prognosis (Fig. 5G, S4E–H). The upregulation of *OSBPL3* has been reported to be involved in the promotion of colorectal cancer progression by activating the Ras signaling pathway [38]. As shown in the Hi-C heat maps in Fig. 5H, the TAD nesting level of *OSBPL3* was significantly higher in the tumor tissue than in normal colon tissue. In addition, *PPARGC1B* is known to play a crucial role in multiple metabolic processes, and the down-regulation of *PPARGC1B* by miR-21 has been reported to promote non-small-cell lung cancer growth. The TAD nesting level of *PPARGC1B* was significantly lower in the tumor tissue than in the normal colon tissue. Hence, we suggest that the hierarchical folding dynamics of TADs at the gene loci might cause the alternation of local interactions and be an important reason for the abnormal transcription of these two genes, which is closely linked to cancer prognosis.

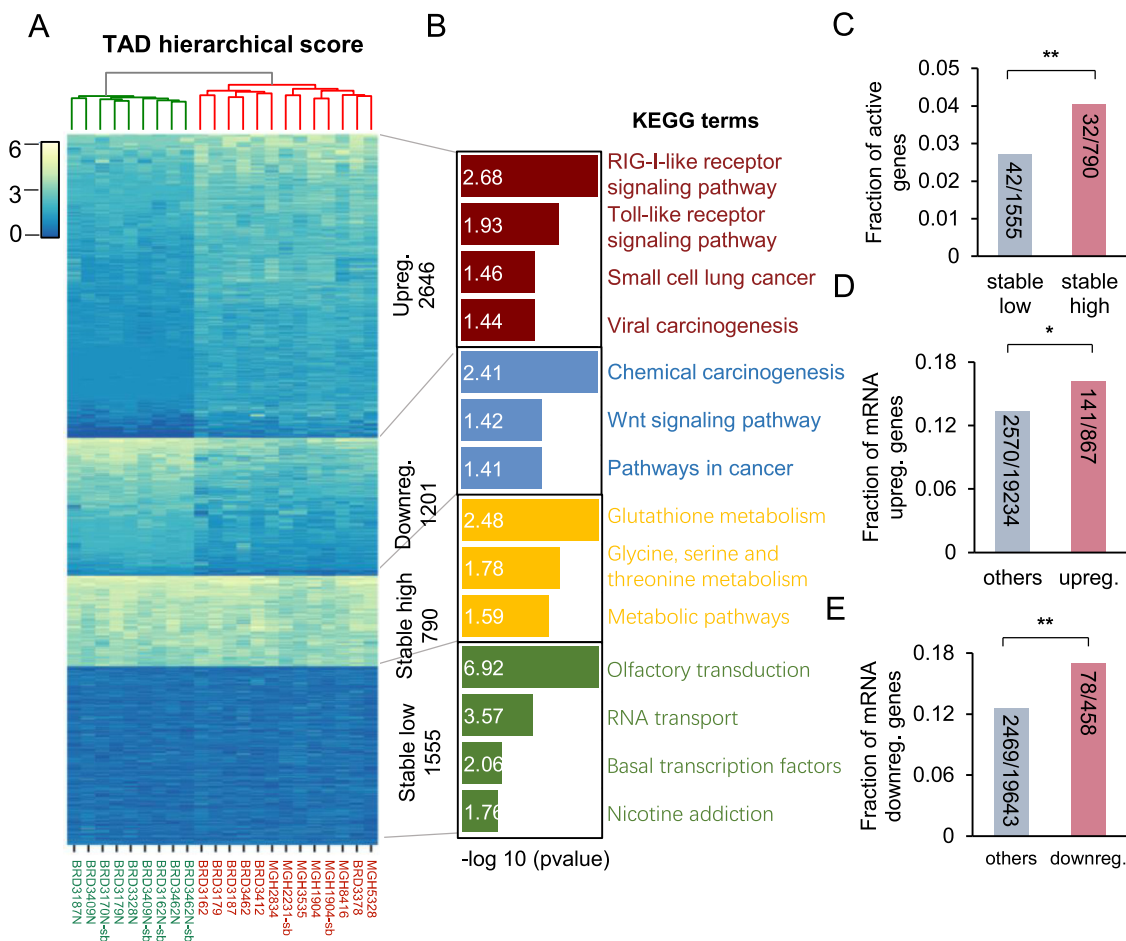


Fig. 4. The hierarchical folding of TADs is related to transcriptional changes in colon cancer. (A) Heatmap showing the TH scores of 4 patterns (upregulated, downregulated, stable high and stable low) genes (viridis color scale) for each tissue sample. The order of samples (columns) is consistent with Ward's linkage hierarchical clustering (top). The genes (rows) are ordered by average TH score of normal samples from high to low. (B) Selected KEGG terms enriched in genes sets of these four patterns. (C) Enrichment of active genes (FPKM > 5 in all samples) in TH score-stable high and stable low patterns. (D) and (E) Enrichment of mRNA-upregulated (downregulated) genes in TH score-upregulated (downregulated) genes. * indicates $P < 0.05$ using Fisher's exact test. ** indicates $P < 0.01$ using Fisher's exact test.

Moreover, in order to learn about the hierarchical folding dynamics of TADs in pan-cancers, we also investigated the TH scores in other types of cancer cell lines and normal cell lines. Encouragingly, the TH score of genes could distinguish cancer and normal cell lines to a certain extent (Fig. S5A). The genes with different TH scores in cancer and normal cell lines were enriched in tumor-related signaling pathways (Fig. S5B and C).

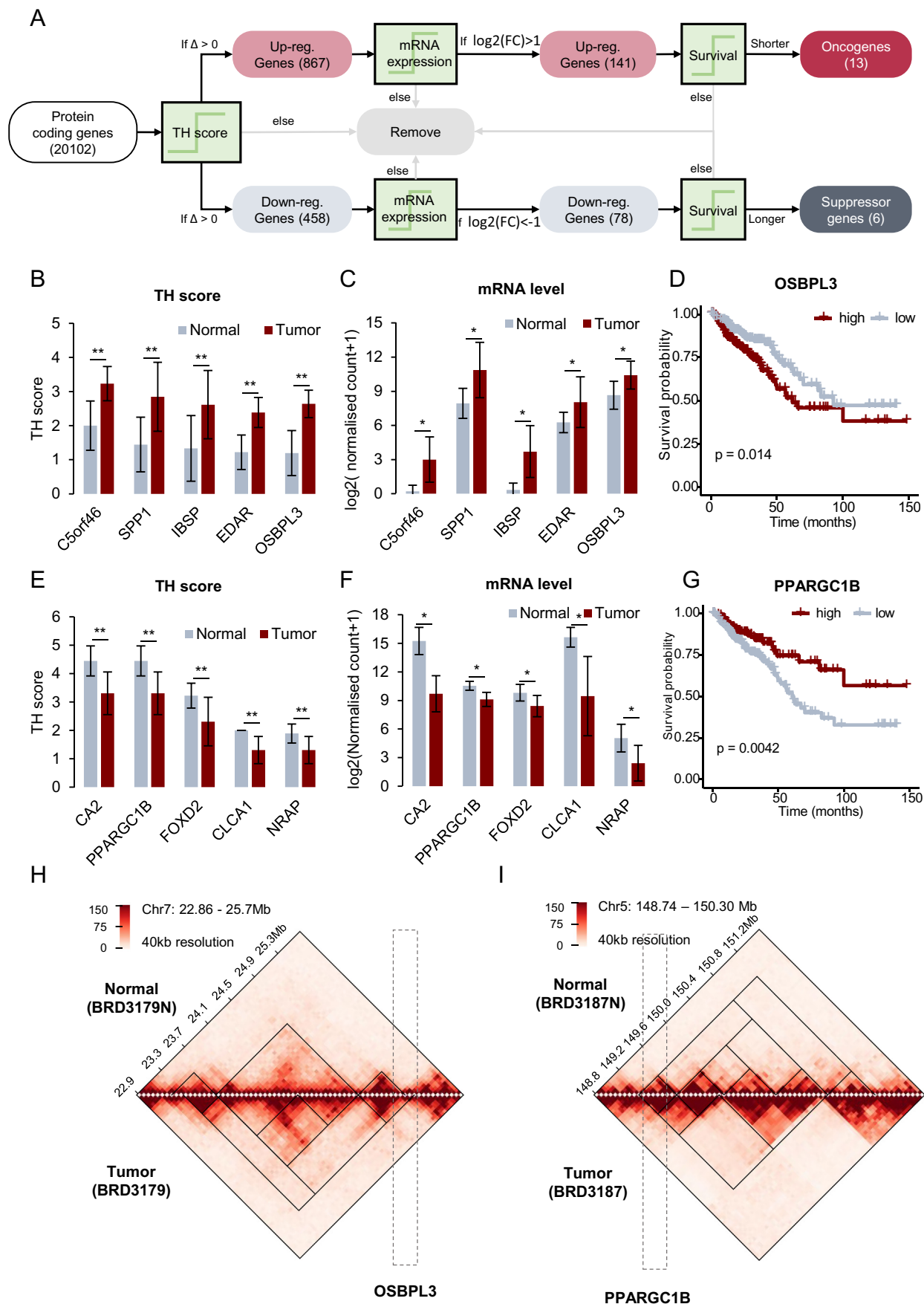
4. Discussion

Overall, this study investigated hierarchical TADs structures in colorectal cancer and introduced the “TH score” to assess the TAD nesting level of specific genes. The TH score performed well in distinguishing cancer and normal samples and reflected the heterogeneity of tumor tissues to a great extent. We also found that the hierarchical TAD level of genes was related to transcriptional changes in colorectal cancer and that some of the genes with a hierarchical level change were significantly related to the patient's prognosis. Based on these facts, we suggest that TAD hierarchical changes frequently occur in cancers, leading to significant changes in gene expression that can possibly drive tumor progression.

It should be noted that the gene TH score performs well in the clustering of normal and tumor tissue samples and could reflect the heterogeneity of tumors. Possible reasons for this are as fol-

lows: 1) Although the basic locations of TAD were stable, the hierarchical folding of TAD was dynamic and could reflect the inter-heterogeneity of tumor tissues. 2) Unlike focusing only on the genes located at TAD boundaries, the TH score allowed consideration of all the protein-coding genes. 3) Compared with the transcriptional level, the chromatin structure could reflect differences in larger regions of the genome, including non-coding region.

The mechanisms of the hierarchical TAD dynamics of tumors in terms can be discussed of two aspects: 1) The hierarchical structure maybe reflect the overlaying of different interaction maps of bulk-cell Hi-C data. The intra-tumor heterogeneity is one of the features of malignant tumors, i.e., there are many different genotypes or subtypes of cells in the same tumor, which is caused by clonal variation and micro environmental influences. Thus, the components of tumor tissues are more complex than those of normal colon tissue, which contributed to the more hierarchical TAD level (Fig. 6A). 2) Hierarchical TAD folding does indeed exist in single cells (Fig. 6B). Although the basic locations of the topological boundaries were stable across normal and tumor tissues, the hierarchical folding of the TADs was tissue- and cell type-specific. Therefore, modifications of in TAD-level folding in the process of tumorigenesis leads to a change of the TAD nesting level of the gene, which is closely related to gene transcription. This hypothesis emphasizes that the function of hierarchical chromatin organization goes beyond simple chromatin packaging efficiency.



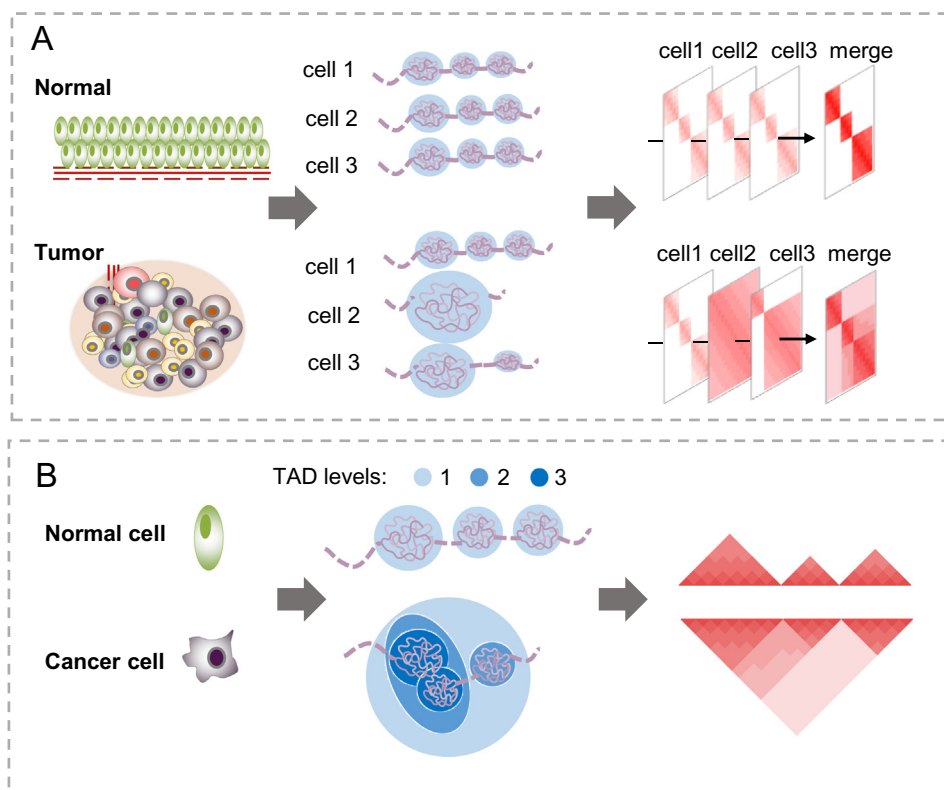


Fig. 6. A proposed model for dynamic TAD hierarchical folding in cancers.

In view of its excellent performance using the Hi-C data of colorectal cancer, the TH score may be a potential diagnostic tool for identifying cancer and have wider application in addressing other biological problems. Furthermore, future studies of hierarchical TAD structures may contribute to the early diagnosis of cancer and the development of new treatment strategies. The ability of researchers to predict the broader consequences of disease-related local genomic abnormalities will also improve, as our understanding of the mechanism and function of TAD structures increases. Looking into the future, there are many obstacles to be solved both in terms of experiments and ethics. However, insights into the formation and abnormal mechanism of TAD will undoubtedly bring new ideas for the treatment of cancer.

Funding

This research was supported by the National Natural Science Foundation of China 31801112 (awards to HC), the Beijing Nova

Program of Science and Technology Z191100001119064 (awards to HC), the National Natural Science Foundation of China 61873276 and 31900488 (awarded to XB and HL), and the Beijing Natural Science Foundation 5204040 (awards to HL).

CRediT authorship contribution statement

Guifang Du: Investigation, Writing - original draft preparation. **Hao Li:** Investigation. **Yang Ding:** Methodology. **Shuai Jiang:** Writing - review & editing. **Hao Hong:** Resources. **Jingbo Gan:** Investigation. **Longteng Wang:** Software. **Yuanping Yang:** Investigation. **Yinyin Li:** Investigation. **Xin Huang:** Investigation. **Yu Sun:** Investigation. **Huan Tao:** Investigation. **Yaru Li:** Investigation. **Xiang Xu:** Investigation. **Yang Zheng:** Investigation. **Junting Wang:** Writing - original draft. **Xuemei Bai:** Writing - original draft. **Kang Xu:** Investigation. **Yaoshen Li:** Writing - original draft preparation. **Qi Jiang:** Writing - original draft. **Cheng Li:** Writing - review & editing. **Hebing Chen:** Writing - review & editing, Supervision, Project administration. **Xiaochen Bo:** Writing - review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank Dr. Yang Chen for comments on early drafts of the manuscript.

Fig. 5. Genes with different TH scores in cancer tissues vs. normal tissues are significantly related to tumor survival. (A) Schematic overview of the oncogenes and suppressor genes related to the hierarchical TAD screening process. (B) and (E) The TH scores of 5 significantly upregulated (downregulated) genes. (C) and (F) The mRNA expression levels of 5 genes with significantly upregulated (downregulated) TH score from the TCGA database. * indicates $P < 0.05$. The bar plot is based on 471 colon cancer samples (marked in red) and 41 normal samples (marked in gray). (D) and (G) Kaplan-Meier curves displaying relapse-free survival for 471 patients with colon cancer based on *OSBPL3* (*PPARGC18*) gene expression. High expression of *OSBPL3* (*PPARGC18*) are shown in red and low expression are shown in gray. P values were derived from the log rank test. (H) and (I) Hi-C heatmaps of regions (chr7: 22.86–25.7 Mb; and chr5: 148.74–150.30 Mb) containing *OSBPL3* and *PPARGC18* in normal and tumor tissue samples. In all matrices, all of the identified hierarchical TADs are shown as black lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2021.03.018>.

References

- [1] Fritz AJ, Ghule PN, Boyd JR, Tye CE, Page NA, Hong D, et al. Intranuclear and higher-order chromatin organization of the major histone gene cluster in breast cancer. *J Cell Physiol* 2018;233(2):1278–90.
- [2] Chen X, Ke Y, Wu K, Zhao H, Sun Y, Gao L, et al. Key role for CTCF in establishing chromatin structure in human embryos. *Nature* 2019;576(7786):306–10.
- [3] Kloetgen A, Thandapani P, Ntziachristos P, Ghebrechristos Y, Nomikou S, Lazaris C, et al. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat Genet* 2020;52(4):388–400.
- [4] Johnstone SE, Reyes A, Qi Y, Adriaens C, Hegazi E, Pelka K, et al. Large-scale topological changes restrain malignant progression in colorectal cancer. *Cell* 2020;182(6):1474–1489.e23.
- [5] Flavahan WA, Drier Y, Johnstone SE, Hemming ML, Tarjan DR, Hegazi E, et al. Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature* 2019;575(7781):229–33.
- [6] Orlando G, Law PJ, Cornish AJ, Dobbins SE, Chubb D, Broderick P, et al. Promoter capture Hi-C-based identification of recurrent noncoding mutations in colorectal cancer. *Nat Genet* 2018;50(10):1375–80.
- [7] Taberlay PC, Achinger-Kawecka J, Lun ATL, Buske FA, Sabir K, Gould CM, et al. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res* 2016;26(6):719–31.
- [8] Hnisz D, Weintraub AS, Day DS, Valton A-L, Bak RO, Li CH, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* 2016;351(6280):1454–8.
- [9] Flavahan WA, Drier Y, Liao BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 2016;529(7584):110–4.
- [10] Achinger-Kawecka J, Taberlay PC, Clark SJ. Alterations in three-dimensional organization of the cancer genome and epigenome. *Cold Spring Harb Symp Quant Biol* 2016;81:41–51.
- [11] Dekker J, Mirny L. The 3D genome as moderator of chromosomal communication. *Cell* 2016;164:1110–21.
- [12] Rao SP, Huntley M, Durand N, Stamenova E, Bochkov I, Robinson J, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014;159(7):1665–80.
- [13] Fraser J et al. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol Syst Biol* 2015;11:852.
- [14] Krumm A, Duan Z. Understanding the 3D genome: emerging impacts on human disease. *Semin Cell Dev Biol* 2019;90:62–77.
- [15] Krefting J, Andrade-Navarro MA, Ibn-Salem J. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol* 2018;16:87.
- [16] Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* 2015;518(7539):331–6.
- [17] Lupiáñez Darío G, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 2015;161(5):1012–25.
- [18] Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* 2016;538(7624):265–9.
- [19] Zhang X, Zhang Y, Zhu X, Purmann C, Haney MS, Ward T, et al. Local and global chromatin interactions are altered by large genomic deletions associated with human brain development. *Nat Commun* 2018;9(1). <https://doi.org/10.1038/s41467-018-07766-x>.
- [20] Valton A-L, Dekker J. TAD disruption as oncogenic driver. *Curr Opin Genet Dev* 2016;36:34–40.
- [21] Dixon JR et al. Integrative detection and analysis of structural variation in cancer genomes. *Nat Genet* 2018;50:1388–98.(Dixon, 2018 #265)
- [22] Weischenfeldt J, Dubash T, Drainas AP, Mardin BR, Chen Y, Stütz AM, et al. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat Genet* 2017;49(1):65–74.
- [23] Schmitt A, Hu M, Jung I, Xu Z, Qiu Y, Tan C, et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep* 2016;17(8):2042–59.
- [24] Norton HK, Emerson DJ, Huang H, Kim J, Titus KR, Gu S, et al. Detecting hierarchical genome folding with network modularity. *Nat Methods* 2018;15(2):119–22.
- [25] An L et al. OnTAD: hierarchical domain structure reveals the divergence of activity among TADs and boundaries. *Genome Biol* 2019;20:282.
- [26] Soler-Vila P et al. Hierarchical chromatin organization detected by TADpole. *Nucleic Acids Res* 2020;48:e39.
- [27] Kumar V, Leclerc S, Taniguchi Y. BHI-Cect: a top-down algorithm for identifying the multi-scale hierarchical structure of chromosomes. *Nucleic Acids Res* 2020;48:e26.
- [28] Bintu B, Mateo LJ, Su J-H, Sinnott-Armstrong NA, Parker M, Kinrot S, et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* 2018;362(6413):eaau1783. <https://doi.org/10.1126/science.aau1783>.
- [29] Szabo Q, Donjon A, Jerković I, Papadopoulos GL, Cheutin T, Bonev B, et al. Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat Genet* 2020;52(11):1151–7.
- [30] Cresswell KG, Stansfield JC, Dozmorov MG. SpectralTAD: an R package for defining a hierarchy of topologically associated domains using spectral clustering. *BMC Bioinf* 2020;21:319.
- [31] Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
- [32] Wu P, Li T, Li R, Jia L, Zhu P, Liu Y, et al. 3D genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. *Nat Commun* 2017;8(1). <https://doi.org/10.1038/s41467-017-01793-w>.
- [33] Virtanen P et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17:261–72.
- [34] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
- [35] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44–57.
- [36] Chen H et al. New insights on human essential genes based on integrated analysis and the construction of the HEGIAP web-based platform. *Brief Bioinform* 2020;21:1397–410.
- [37] Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27(12):1739–40.
- [38] Jiao H-I, Weng B-S, Yan S-S, Lin Z-m, Wang S-Y, Chen X-P, et al. Upregulation of OSBPL3 by HIF1A promotes colorectal cancer progression through activation of RAS signaling pathway. *Cell Death Dis* 2020;11(7). <https://doi.org/10.1038/s41419-020-02793-3>.