

Regulation Network of Colorectal Cancer Specific Enhancers in Progression of Colorectal Cancer

Bohan Chen¹, Yiping Ma¹, Jinfang Bi¹, Wenbin Wang¹, Anshun He¹, Guangsong Su¹, Zhongfang Zhao¹, Jiandang Shi¹, Lei Zhang^{1,*}

affiliation and address

¹State Key Laboratory of Medicinal Chemical Biology and College of Life Sciences, Nankai University, 94 Weijin Road, 300071 Tianjin, China

* Corresponding authors:

Lei Zhang:

Tel: +86 22 23503617; Fax: +86 22 23503617; Email: joyleizhang@nankai.edu.cn

Abstract

Enhancers regulate multiple genes through higher-order chromatin structure and further affect cancer progression. Epigenetic changes in cancer cells activate several cancer specific enhancers that are silenced in normal cells. These cancer specific enhancers are potential therapeutic targets of cancer. However, functions and regulation network of colorectal cancer specific enhancers are still unknown. Here in this study, we profile colorectal cancer specific enhancers and reveal the regulation network of these enhancers by analysis of HiChIP, Hi-C and RNA-seq data. We propose the regulation network of colorectal cancer specific enhancers plays important role in progression of colorectal cancer.

Key words: Colorectal cancer-specific enhancer, long-range interaction, Hi-C, HiChIP, TAD, loop

1. Introduction

Colorectal cancer is the world's fourth most deadly cancer [1]. Previous studies on colorectal cancer indicated sequential accumulation of genetic mutations and chromosomal instability are the important reasons for the occurrence and development of colorectal cancer [2]. Typical genomic events in the initiation of colorectal cancer are *APC* mutation and followed by *RAS* activation or function loss of *TP53* [1]. Besides, *MAPK*, *WNT*, *PI3K*, TGF- β signaling pathway are reported play important roles in initiation and progression of colorectal cancer [2, 3]. Chromosomal changes including deletion, amplifications and translocations induced alterations of genes' expressions and functions that could further initiate colorectal cancer. Studies on higher-order chromatin structures revealed long-range gene regulation networks are important in modulating biological process in cells [4-6]. Further studies on gene regulation networks could help unveil mechanism of occurrence and progression of colorectal cancer.

Gene regulatory elements like enhancers or silencers could regulate multiple target genes through higher-order chromatin structures [7, 8] and function in gene regulation networks in cancer cells. Mutations in gene regulatory elements and changes of higher-order chromatin structures are related to gene transcriptions and cancer progression [8-10]. Prior study revealed reorganizations of chromatin structure in colorectal cancers compared with normal colons. And the topological changes of chromatins restrain malignant progression of colorectal cancers [11]. In addition, changes of higher-order chromatin structures in colorectal cancer cells are accompanied by epigenetic changes like alterations of DNA methylation [11, 12]. These epigenetic changes are associated with changes of transcription factors' binding

and activation of some gene regulatory elements like enhancers. Some enhancers are activated in cancer cells while not in normal cells [13]. These “cancer specific enhancers” may regulate several target genes through higher-order chromatin structures and play roles in cancer progression. These cancer specific enhancers are potential therapeutic targets of cancer. However, how colorectal cancer specific enhancers functions in cancer initiation and progression is still not clear.

In this study, by combination analysis of ChIP-seq, HiChIP and Hi-C data, we show profile of colorectal cancer specific enhancers and gene regulation network of these enhancers. Results in this study suggest that colorectal cancer specific enhancers could regulate multiple target genes including cancer driver genes, and genes regulated by these colorectal cancer specific enhancers are associated with occurrence and development of colorectal cancer.

2. Results

2.1 H3K27ac profiles define colorectal cancer specific enhancers.

H3K27ac is a well-characterized histone modification marker of active enhancers[14]. To identify colorectal cancer-specific enhancers, we analyzed H3K27ac CHIP-seq datasets from 7 published colorectal datasets and 10 normal tissues from the NIH Roadmap database [15]. Unsupervised hierarchical clustering was performed on H3K27ac enrichment loci which identified 10 clusters from the 59 samples (Fig. 1A) and show distribution pattern of enhancers was highly tissue specific, which is consistent with previous studies [13]. 21132 normal colon tissue specific enhancers and 11463 colorectal cancer specific enhancers were identified. And the *t*-distributed stochastic neighbor embedding (*t*-SNE) analysis indicated colorectal cancer enhancer profiles were distinct from normal colon tissue enhancers (Fig. 1B). In addition, we identified 763 super enhancers among the colorectal cancer specific enhancers (Fig. 1C). Some of these super enhancers are associated with oncogenes, like *EPHA2*, *LIF*, *ID1*, *SMAD7*, *BMP4*, *FOXA1*, *NOTCH1*.

2.2 HiChIP identifies chromatin interactions containing colorectal cancer specific enhancers.

Positions of colorectal cancer specific enhancers are shown in Figure 2A, these colorectal cancer specific enhancers are evenly distributed on each chromosome (Fig. 2A). To identify the regulation network of colorectal cancer specific enhancers, HiChIP for H3K27ac was performed in colorectal cancer cell line HCT116. HiChIP data indicate interaction loops between enhancers/super enhancers and other chromatin fragments (Fig. 2B, C, interaction loops in chromosome 1 were shown as an example). And colorectal cancer specific enhancers/super enhancers also interact with several chromatin fragments including some long-range chromatin regions (Fig. 2D, E, interaction loops in chromosome 1 were shown as an example). Most of the interaction loops that contain colorectal cancer specific enhancers are less than 100kb in length and include more short-range interactions than random interaction loops from the HiChIP

data (Fig. 2F). In addition, for the interaction loops that are more than 100kb in length, random longer interaction loops (>100kb) have a lower proportion (Fig. 2G), but longer interaction loops (>100kb) that contain colorectal cancer specific enhancers didn't have a significant lower proportion (Fig. 2G), which suggests more long-range interactions are associated with colorectal cancer specific enhancers.

2.3 Change of TADs' boundaries in colorectal cancer cells compared with normal colon cells

Based on analysis of Hi-C data in colorectal cancer cells and normal colon cells [11], about 65% boundaries of TADs changed in colorectal cancer cells compared with normal colon cells (Fig. 3A-F). The boundary-changed TADs were divided into six categories: expand (Fig. 3A), narrow (Fig. 3B), fusion (Fig. 3C), split (Fig. 3D), appear (Fig. 3E) and disappear (Fig. 3F), that respectively represent six kinds of change patterns of TADs' boundaries in colorectal cancer cells compared with normal colon cells. Previous studies have indicated that gene transcription is correlated with insulation at TAD boundaries [6, 11]. Expand, fusion and appear of TADs in colorectal cancer cells are associated with more interactions between enhancers and genes, while split, narrow and disappear are associated with less interactions between some enhancers and genes in colorectal cancer cells. 65% of changed TADs in colorectal cancer cells compared with normal colon cells suggest global alterations of gene regulations in colorectal cancer cells. We noticed that more than 50% (51%-86%, except for the "disappear" group, which is 24%) of these changed TADs containing colorectal cancer specific enhancers (Fig. 3A-F), which suggests important roles of colorectal cancer specific enhancers in the alterations of the gene regulation network.

2.4 Transcriptome change in colorectal cancer cells is associated with colorectal specific enhancers.

To figure out whether transcription change of genes in colorectal cancer cells were associated with colorectal specific enhancers, we analyzed the transcriptome data of colon tissues (n=349) and colorectal cancer tissues (n=275) from TCGA. 4714 genes were significantly upregulated (fold change>1.5, q<0.01) and 4924 genes were significantly downregulated (fold change>1.5, q<0.01) in colorectal cancer tissue (Fig. 4A). Combination analysis of RNA-seq and HiChIP data identified 152 genes are significant upregulated (fold change>1.5, q<0.01) in colorectal cancer tissue and have comparatively high interaction signal (interaction counts ≥ 5) with colorectal cancer specific enhancers (Fig. 4B). These genes are the potential target genes that are directly regulated by colorectal cancer specific enhancers. About 79% of the interactions between colorectal cancer specific enhancer and target genes are long-range (≥ 20 kb) interactions (Fig. 4C). Pathway and process enrichment analysis by Megaspaces [16] showed potential target genes of colorectal cancer specific enhancers are enriched in RNA catabolic process, $\alpha 6\beta 1\alpha 6\beta 4$ integrin pathway and regulation of cyclin-dependent protein serine/threonine kinase activity (Fig. 4D). These pathways or biological processes are reported associated with colorectal cancer progression [17-21]. *ITGB4* encodes integrin subunit $\beta 4$ and was involved in the $\alpha 6\beta 1\alpha 6\beta 4$ integrin pathway.

HiChIP data show *ITGB4* is regulated by a colorectal cancer specific enhancer (Fig. 4E) and *ITGB4* was upregulated in colorectal cancer tissue based on RNA-seq data, suggestive of regulation of *ITGB4* by colorectal cancer specific enhancer. Mutations and aberrant expression of cancer driver genes affect key cellular functions and induce cancer occurrence. *RECQL4* is one of the cancer driver genes [22] and potential target genes that are regulated by colorectal cancer specific enhancer (Fig. 4F). In addition, *RECQL4* is one of the most often mutated genes in colorectal cancer [23]. The colorectal cancer specific enhancer is about 15kb away from the *RECQL4* promoter and HiChIP data show interactions between colorectal cancer specific enhancer and *RECQL4* (Fig. 4F).

3. Discussion

This study shows colorectal specific enhancers interact with several potential target genes that are upregulated in colorectal cancers compared with normal colon. HiChIP data indicate higher-order chromatin structure induce interactions between colorectal cancer specific enhancers and target genes. Most altered TADs in colorectal cancer genome contain colorectal specific enhancers. But it is not clear whether changes of chromatin topological structures or the activation of colorectal specific enhancers is the primary event in the initiation of colorectal cancer. Previous study showed that chromosomal topological changes repress stemness and invasion programs and may restrain malignant progression of colorectal cancer and tumor-associated epigenomic changes are primarily oncogenic [11]. HiChIP data show that some of target genes of colorectal cancer specific enhancers are cancer driver genes, including *RECQL4* that is showed in Figure 4F, and analysis of Hi-C data indicate most of boundary-changed TADs contain colorectal specific enhancers (Fig. 3). It is possible that the activation of colorectal specific enhancers is before alteration of higher-order chromatin structure during occurrence and progression of colorectal cancer. Further studies are needed to verify this hypothesis.

With the development of chromatin conformation capture techniques [24-27], analysis of higher-order chromatin structures in cancer cells revealed complicated long-range gene networks of enhancers and indicated important functions of these enhancers or gene networks in cancer progression [8-10, 13]. A recent study reported that mutation in enhancer loci leads to activity change of the enhancer and affects expression of long-range target genes of this enhancer and further improve the progression of prostate cancer [8]. Prior study showed that ependymoma enhancer profiles were distinct from other tissues and most super enhancers were tumor-specific and enriched with cancer-associated genes [13], suggestive of important functions of these tumour-specific enhancers in the initiation of tumor. Our study profiles the colorectal cancer specific enhancers and presents regulation network of these enhancers based on HiChIP data and RNA-seq data, further unveils the mechanism of initiation and progression of colorectal cancer.

Combination analysis of HiChIP data and RNA-seq data indicated several potential target genes of colorectal specific enhancer. *ITGB4* is one of these potential target genes. *ITGB4* is one of integrin (ITG) molecules that are associated with cell migration, proliferation and cancer development [28-30]. Aberrant *ITGB4* expression was reported in several cancers including colorectal cancers [18, 31]. RNA-seq analysis also showed significant upregulation of *ITGB4* in colorectal cancer tissue. A recent study indicated that *ITGB4* might be a prognosis marker for individual therapy of colon cancer [18]. Data in our study suggests upregulation of *ITGB4* in colorectal cancer cells may due to the regulation of colorectal cancer specific enhancer. In addition, we noticed that *GALK1*, which is beside *ITGB4* (Fig. 4E), may also be one of the potential target genes of the colorectal cancer specific enhancer which is showed in Figure 4E. HiChIP data indicate interactions between the colorectal cancer specific enhancer and *GALK1*, and RNA-seq analysis showed upregulation (fold change=1.97, $q=7.76 \times 10^{-49}$) of *GALK1* in colorectal cancer tissue compared with normal colon tissue. However, the interaction signal is not very robust in the HiChIP data (count=3). This may due to the long-range interactions between the colorectal cancer specific enhancer and *GALK1* (~45kb). Our study suggests *GALK1* may be regulated by colorectal cancer specific enhancer and associated with progression of colorectal cancer. However, few studies reported function of *GALK1* in colorectal cancer for now. *RECQL4* is an example of cancer driver genes that are regulated by colorectal cancer specific enhancers. *RECQL4* was reported as DNA helicase and function in DNA replication, DNA repair and recombination [32], which play important role in maintaining the genomic stability [33]. Several studies indicated *RECQL4* is associated with cancer progression and could be diagnostic marker for cancer [33-37]. Examples include overexpression of *RECQL4* is associated with poor prognosis of gastric cancer [35] and predicts poor prognosis in hepatocellular carcinoma [36]. Investigation of mutation spectrum of cancer-associated genes in patients with early onset of colorectal cancer showed that *RECQL4* is one of the most often mutated genes [23]. And analysis of RNA-seq data from 275 colon cancer tissues and 349 normal colon tissues show that *RECQL4* is upregulated in colorectal cancer cells. Based on literatures and data analysis in the present study, it suggests *RECQL4* is an important target gene of colorectal cancer specific enhancers and involved in the progression of colorectal cancer. Although further studies are needed to pinpoint whether there are key driver genes in the colorectal cancer specific enhancer regulation network or the progression of colorectal cancer a combined result of these target genes, our study suggests important functions of colorectal cancer specific enhancers in colorectal cancer progression, and these enhancers could be potential therapeutic targets of colorectal cancer.

In summary, our study shows that colorectal cancer specific enhancers regulate several target genes including cancer driver genes and may play important roles in occurrence and progression of colorectal cancer.

4. Materials and methods

4.1 Cell Culture

The colorectal cancer cell line HCT116 was obtained from the ATCC. HCT116 cells were incubated at 37 °C with 5% CO₂ and cultured in Dulbecco's Modified Eagle Medium (GIBCO) medium, supplemented with 10% fetal bovine serum (FBS) (BI) and 1% penicillin-streptomycin (GIBCO).

4.2 HiChIP

HiChIP was performed as described previously with modifications [26]. Briefly, cells were crosslinked with 1% formaldehyde and lysed. Then the chromatin was digested using *Mbo*I (NEB), and the restricted ends were religated by T4 ligase (NEB). Pelleted nuclei were dissolved in nuclear lysis buffer (50 mM Tris-HCl, pH 7.5, 10 mM EDTA, 1% SDS and protease inhibitors) and were sonicated and diluted in ChIP Dilution Buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 7.5, and 167 mM NaCl). Then immunoprecipitation was performed overnight at 4 °C by incubating H3K27ac antibody (Abcam) precoated on protein A-coated magnetic beads (Thermo Fisher Scientific). Immunocomplexes were washed three times each with low-salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 7.5 and 150 mM NaCl), high-salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 7.5 and 500 mM NaCl), LiCl buffer (10 mM Tris-HCl, pH 7.5, 250 mM LiCl, 1% NP-40, 1% Na-Doc and 1 mM EDTA). Beads were resuspended in DNA elution Buffer (50 mM NaHCO₃ and 1% SDS). After elution, ChIP sample was incubated with 10mg/mL proteinase K 4 h at 55 °C. Then DNA was purified using AMPure XP Beads (Beckman). Streptavidin C1 beads were used to capture biotinylated DNA. QIAseq FX DNA Library Kits were used to generate the sequencing library and HiChIP libraries were size-selected to 300–700 bp using AMPure XP beads (Beckman) and subjected to 2 × 50-bp paired-end sequencing on HiSeq XTen (Illumina).

4.3 ChIP-seq data processing and super enhancer analysis

Mapping of ChIP-seq data to hg19 was performed using bowtie2[38]. H3K27ac peak calling was performed using MACS1.4 with default parameter[39]. Peak calling for each sample was performed separately. Peaks that could not be identified in each colorectal cancer sample and peaks appeared within the region surrounding ±2.5 kb of transcriptional start sites were excluded from any further analysis. Afterwards, the H3K27ac peaks of the 7 individual samples were merged into a single set of peaks. Super enhancers were identified using the rank ordering of super enhancers (ROSE) algorithm[14] In the case of colorectal cancer-specific enhancers, all peak regions were removed that contained any overlap with a peak detected in each normal colon region.

4.4 Unsupervised hierarchical clustering analysis.

A matrix of the normalized H3K27ac density was generated using HOMER[40]. Variant enhancer loci (VELs) were defined as enhancers, which exhibited the greatest median absolute deviation (MAD) across all samples used for clustering. In the case of

unsupervised hierarchical clustering between colorectal cancer and Roadmap Epigenomics samples data, 11,463 VELs were retained. These enhancers were used for unsupervised hierarchical clustering using a Pearson correlation as a distance metric.

4.5 *t*-SNE analysis

For clustering of H3K27ac ChIP-seq data from the colorectal cancer and normal colon cohorts together, we generated both cohorts normalized H3K27ac density matrix. Distance between samples was calculated by using 1-Spearman correlation coefficient as the distance measure. The resulting distance matrix was used to perform the *t*-SNE analysis (Rtsne package)

4.6 HiChIP data processing

Valid fragment pairs and interaction matrix was generated using the HiC-Pro[41]. Then the interaction loop calling was performed using the hichipper[42] and interaction loops in Figure 2B-E were shown as circos plot [43]. In the case of colorectal cancer specific enhancer loops detecting, the loops were picked that contained anchors overlapped with cancer specific enhancers.

4.7 HiC data analysis

Two HiC data sets in GSE133928, BRD3179 and BRD3179N [11], representing colorectal cancer tissue and normal colon tissue respectively, was analyzed to detect cancer and normal TADs. For the unchanged-boundary cancer TADs, the cancer TADs boundary was extremely overlapped with normal TADs. For the expand cancer TADs, the cancer TAD region contained only one complete normal TAD. For fusion cancer TADs, each cancer TAD region contained more than one complete normal TADs. For narrow cancer TADs, each cancer TAD was included in one normal TAD. For the split cancer TADs, more than one cancer TADs were included in one normal TADs. For the appear or disappear cancer TAD, the TADs existed only in cancer sample or normal sample respectively.

Acknowledgments

This work was supported by the National Key R&D Program of China (NO.2017YFA0102600) and the Fundamental Research Funds for the Central Universities, Nankai University (63201087).

Conflict of interest

The authors declare that they have no competing interests.

Authors Contributions

B.C.,L.Z, conception and design, drafting or revising the article; B.C., L.Z., acquisition of data, analysis and interpretation of data; B.C., Y.M., A.H., J.B., W.W., and G.S. acquisition of data; B. C., analysis of all High-through sequencing data; B. C., W.W., acquisition of data and production of all bioinformatics figures; J.S, Z.Z, L.Z., revising the article.

Availability of supporting data

Published data used in this study

The following published datasets were used in our analysis: GSE36204 for H3K27ac ChIP-seq analyses[15], GSE133928 for Hi-C data in colorectal cancer cells and normal colon cells [11] and RNA-seq data was gained from TCGA and GEPIA database[44].

Data Availability

HiChIP data are deposited in the NCBI Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) under the accession number GSE173699.

References:

1. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet*. 2019;394(10207):1467-80. doi: 10.1016/S0140-6736(19)32319-0. PubMed PMID: 31631858.
2. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487(7407):330-7. doi: 10.1038/nature11252. PubMed PMID: 22810696; PubMed Central PMCID: PMC3401966.
3. Fearon ER. Molecular genetics of colorectal cancer. *Annual review of pathology*. 2011;6:479-507. doi: 10.1146/annurev-pathol-011110-130235. PubMed PMID: 21090969.
4. Wei Z, Gao F, Kim S, Yang H, Lyu J, An W, et al. Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. *Cell stem cell*. 2013;13(1):36-47. doi: 10.1016/j.stem.2013.05.010. PubMed PMID: 23747203.
5. Su G, Guo D, Chen J, Liu M, Zheng J, Wang W, et al. A distal enhancer maintaining Hoxa1 expression orchestrates retinoic acid-induced early ESCs differentiation. *Nucleic acids research*. 2019;47(13):6737-52. doi: 10.1093/nar/gkz482. PubMed PMID: 31147716; PubMed Central PMCID: PMC6649716.
6. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, et al. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*. 2017;171(3):557-72 e24. doi: 10.1016/j.cell.2017.09.043. PubMed PMID: 29053968; PubMed Central PMCID: PMC5651218.
7. Zhang L, He A, Chen B, Bi J, Chen J, Guo D, et al. A HOTAIR regulatory element modulates glioma cell sensitivity to temozolomide through long-range regulation of multiple target genes. *Genome research*. 2020;30(2):155-63. doi: 10.1101/gr.251058.119. PubMed PMID: 31953347.
8. Qian Y, Zhang L, Cai M, Li H, Xu H, Yang H, et al. The prostate cancer risk variant rs55958994 regulates multiple gene expression through extreme long-range chromatin interaction to control

- tumor progression. *Science advances*. 2019;5(7):eaaw6710. doi: 10.1126/sciadv.aaw6710. PubMed PMID: 31328168; PubMed Central PMCID: PMC6636982.
9. Kloetgen A, Thandapani P, Ntziachristos P, Ghebrecristos Y, Nomikou S, Lazaris C, et al. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nature genetics*. 2020;52(4):388-400. doi: 10.1038/s41588-020-0602-9. PubMed PMID: 32203470; PubMed Central PMCID: PMC7138649.
 10. Rhie SK, Perez AA, Lay FD, Schreiner S, Shi J, Polin J, et al. A high-resolution 3D epigenomic map reveals insights into the creation of the prostate cancer transcriptome. *Nature communications*. 2019;10(1):4154. doi: 10.1038/s41467-019-12079-8. PubMed PMID: 31515496; PubMed Central PMCID: PMC6742760.
 11. Johnstone SE, Reyes A, Qi Y, Adriaens C, Hegazi E, Pelka K, et al. Large-Scale Topological Changes Restrict Malignant Progression in Colorectal Cancer. *Cell*. 2020;182(6):1474-89 e23. doi: 10.1016/j.cell.2020.07.030. PubMed PMID: 32841603; PubMed Central PMCID: PMC7575124.
 12. Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nature genetics*. 2011;44(1):40-6. doi: 10.1038/ng.969. PubMed PMID: 22120008; PubMed Central PMCID: PMC4309644.
 13. Mack SC, Pajtler KW, Chavez L, Okonechnikov K, Bertrand KC, Wang X, et al. Therapeutic targeting of ependymoma as informed by oncogenic enhancer profiling. *Nature*. 2018;553(7686):101-5. doi: 10.1038/nature25169. PubMed PMID: 29258295; PubMed Central PMCID: PMC5993422.
 14. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153(2):307-19. doi: 10.1016/j.cell.2013.03.035. PubMed PMID: 23582322; PubMed Central PMCID: PMC3653129.
 15. Akhtar-Zaidi B, Cowper-Sal-lari R, Corradin O, Saiakhova A, Bartels CF, Balasubramanian D, et al. Epigenomic enhancer profiling defines a signature of colon cancer. *Science*. 2012;336(6082):736-9. doi: 10.1126/science.1217277. PubMed PMID: 22499810; PubMed Central PMCID: PMC3711120.
 16. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature communications*. 2019;10(1):1523. doi: 10.1038/s41467-019-09234-6. PubMed PMID: 30944313; PubMed Central PMCID: PMC6447622.
 17. Mercurio AM, Bachelder RE, Rabinovitz I, O'Connor KL, Tani T, Shaw LM. The metastatic odyssey: the integrin connection. *Surgical oncology clinics of North America*. 2001;10(2):313-28, viii-ix. PubMed PMID: 11382589.
 18. Li M, Jiang X, Wang G, Zhai C, Liu Y, Li H, et al. ITGB4 is a novel prognostic factor in colon cancer. *Journal of Cancer*. 2019;10(21):5223-33. doi: 10.7150/jca.29269. PubMed PMID: 31602273; PubMed Central PMCID: PMC6775604.
 19. Goodall GJ, Wickramasinghe VO. RNA in cancer. *Nature reviews Cancer*. 2021;21(1):22-36. doi: 10.1038/s41568-020-00306-0. PubMed PMID: 33082563.
 20. Shi XN, Li H, Yao H, Liu X, Li L, Leung KS, et al. Adapalene inhibits the activity of cyclin-dependent kinase 2 in colorectal carcinoma. *Molecular medicine reports*. 2015;12(5):6501-8. doi: 10.3892/mmr.2015.4310. PubMed PMID: 26398439; PubMed Central PMCID: PMC4626183.

21. Cam WR, Masaki T, Shiratori TY, Kato N, Okamoto M, Yamaji Y, et al. Activation of cyclin E-dependent kinase activity in colorectal cancer. *Digestive diseases and sciences*. 2001;46(10):2187-98. doi: 10.1023/a:1011962915280. PubMed PMID: 11680595.
22. Martinez-Jimenez F, Muinos F, Sentis I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, et al. A compendium of mutational cancer driver genes. *Nature reviews Cancer*. 2020;20(10):555-72. doi: 10.1038/s41568-020-0290-x. PubMed PMID: 32778778.
23. Zhunussova G, Afonin G, Abdikerim S, Jumanov A, Perfilyeva A, Kaidarova D, et al. Mutation Spectrum of Cancer-Associated Genes in Patients With Early Onset of Colorectal Cancer. *Frontiers in oncology*. 2019;9:673. doi: 10.3389/fonc.2019.00673. PubMed PMID: 31428572; PubMed Central PMCID: PMC6688539.
24. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*. 2002;295(5558):1306-11. doi: 10.1126/science.1067799. PubMed PMID: 11847345.
25. Belaghzal H, Dekker J, Gibcus JH. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods*. 2017;123:56-65. doi: 10.1016/j.ymeth.2017.04.004. PubMed PMID: 28435001; PubMed Central PMCID: PMC5522765.
26. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nature methods*. 2016;13(11):919-22. doi: 10.1038/nmeth.3999. PubMed PMID: 27643841; PubMed Central PMCID: PMC5501173.
27. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozcy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289-93. doi: 10.1126/science.1181369. PubMed PMID: 19815776; PubMed Central PMCID: PMC2858594.
28. Moreno-Layseca P, Streuli CH. Signalling pathways linking integrins with cell cycle progression. *Matrix biology : journal of the International Society for Matrix Biology*. 2014;34:144-53. doi: 10.1016/j.matbio.2013.10.011. PubMed PMID: 24184828.
29. Wilkinson EJ, Woodworth AM, Parker M, Phillips JL, Malley RC, Dickinson JL, et al. Epigenetic regulation of the ITGB4 gene in prostate cancer. *Experimental cell research*. 2020;392(2):112055. doi: 10.1016/j.yexcr.2020.112055. PubMed PMID: 32376286.
30. Stewart RL, O'Connor KL. Clinical significance of the integrin alpha6beta4 in human malignancies. *Laboratory investigation; a journal of technical methods and pathology*. 2015;95(9):976-86. doi: 10.1038/labinvest.2015.82. PubMed PMID: 26121317; PubMed Central PMCID: PMC4554527.
31. Ni H, Dydensborg AB, Herring FE, Basora N, Gagne D, Vachon PH, et al. Upregulation of a functional form of the beta4 integrin subunit in colorectal cancers correlates with c-Myc expression. *Oncogene*. 2005;24(45):6820-9. doi: 10.1038/sj.onc.1208848. PubMed PMID: 16007143.
32. Croteau DL, Popuri V, Opresko PL, Bohr VA. Human RecQ helicases in DNA repair, recombination, and replication. *Annual review of biochemistry*. 2014;83:519-52. doi: 10.1146/annurev-biochem-060713-035428. PubMed PMID: 24606147; PubMed Central PMCID: PMC4586249.
33. Mo D, Zhao Y, Balajee AS. Human RecQL4 helicase plays multifaceted roles in the genomic stability of normal and cancer cells. *Cancer letters*. 2018;413:1-10. doi: 10.1016/j.canlet.2017.10.021. PubMed PMID: 29080750.
34. Guo L, Li Y, Zhao C, Peng J, Song K, Chen L, et al. RECQL4, Negatively Regulated by miR-10a-5p, Facilitates Cell Proliferation and Invasion via MAFB in Ovarian Cancer. *Frontiers in oncology*.

2020;10:524128. doi: 10.3389/fonc.2020.524128. PubMed PMID: 33014878; PubMed Central PMCID: PMC7500455.

35. Chen H, Yuan K, Wang X, Wang H, Wu Q, Wu X, et al. Overexpression of RECQL4 is associated with poor prognosis in patients with gastric cancer. *Oncology letters*. 2018;16(4):5419-25. doi: 10.3892/ol.2018.9318. PubMed PMID: 30250613; PubMed Central PMCID: PMC6144911.
36. Li J, Jin J, Liao M, Dang W, Chen X, Wu Y, et al. Upregulation of RECQL4 expression predicts poor prognosis in hepatocellular carcinoma. *Oncology letters*. 2018;15(4):4248-54. doi: 10.3892/ol.2018.7860. PubMed PMID: 29541191; PubMed Central PMCID: PMC5835851.
37. Arora A, Agarwal D, Abdel-Fatah TM, Lu H, Croteau DL, Moseley P, et al. RECQL4 helicase has oncogenic potential in sporadic breast cancers. *The Journal of pathology*. 2016;238(4):495-501. doi: 10.1002/path.4681. PubMed PMID: 26690729; PubMed Central PMCID: PMC5576893.
38. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-9. doi: 10.1038/nmeth.1923. PubMed PMID: 22388286; PubMed Central PMCID: PMC3322381.
39. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying CHIP-seq enrichment using MACS. *Nat Protoc*. 2012;7(9):1728-40. doi: 10.1038/nprot.2012.101. PubMed PMID: 22936215; PubMed Central PMCID: PMC3386217.
40. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576-89. doi: 10.1016/j.molcel.2010.05.004. PubMed PMID: 20513432; PubMed Central PMCID: PMC2898526.
41. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol*. 2015;16:259. doi: 10.1186/s13059-015-0831-x. PubMed PMID: 26619908; PubMed Central PMCID: PMC4665391.
42. Lareau CA, Aryee MJ. hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat Methods*. 2018;15(3):155-6. doi: 10.1038/nmeth.4583. PubMed PMID: 29489746.
43. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome research*. 2009;19(9):1639-45. doi: 10.1101/gr.092759.109. PubMed PMID: 19541911; PubMed Central PMCID: PMC2752132.
44. 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*. 2019;47(W1):W556-W60. doi: 10.1093/nar/gkz430. PubMed PMID: 31114875; PubMed Central PMCID: PMC6602440.

Figures

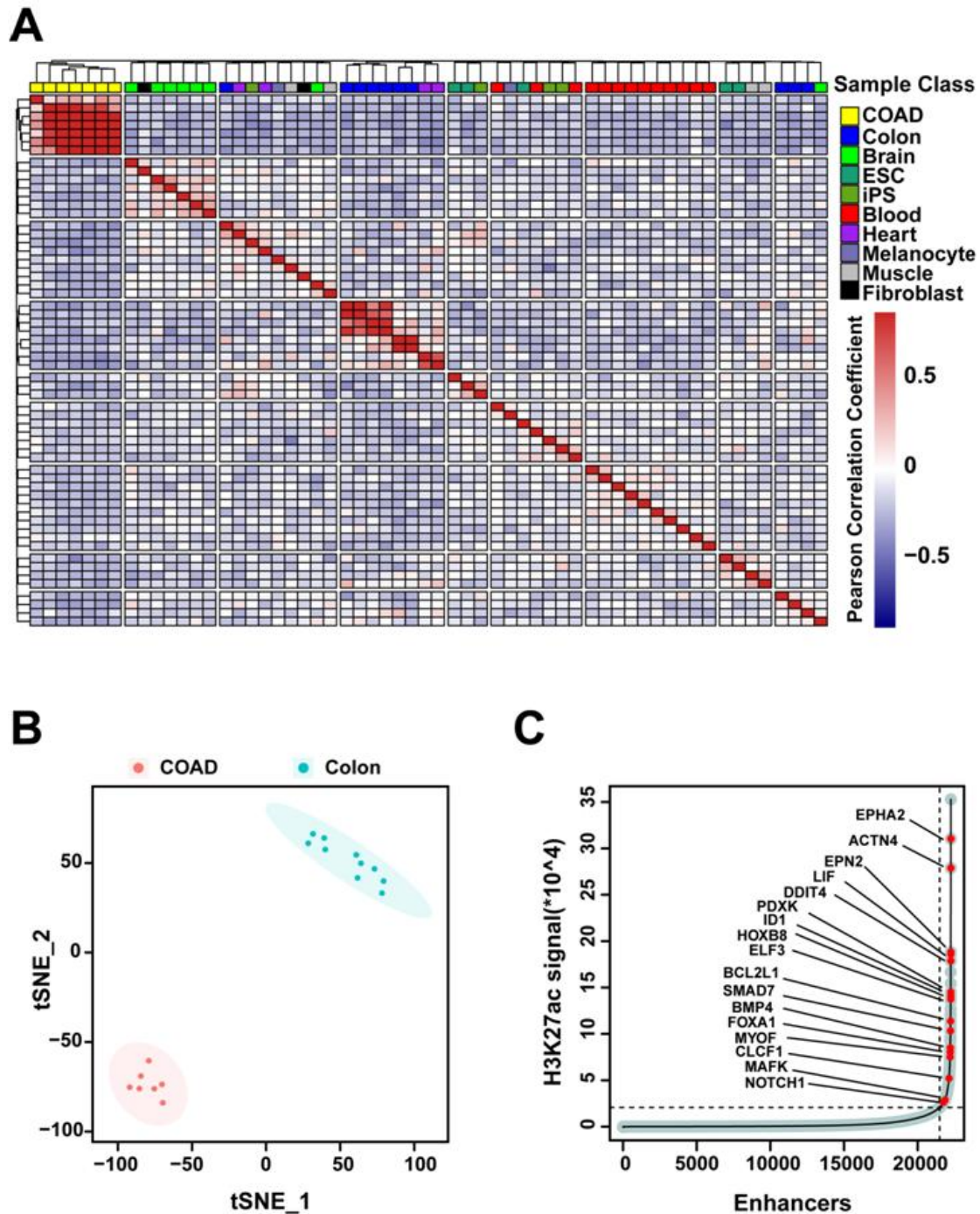


Fig 1. H3K27ac profiles define colorectal cancer specific enhancers. A) Unsupervised hierarchical clustering of the 11463 enhancer loci detected in colorectal cancer (n=7) compared to the normal tissue samples (n=10). B) *t*-distributed stochastic neighbour embedding (*t*-SNE) analysis of normal colon tissue specific enhancers and colorectal cancer specific enhancers. C) Inflection plot showing identified super enhancers among the cancer-specific enhancers.

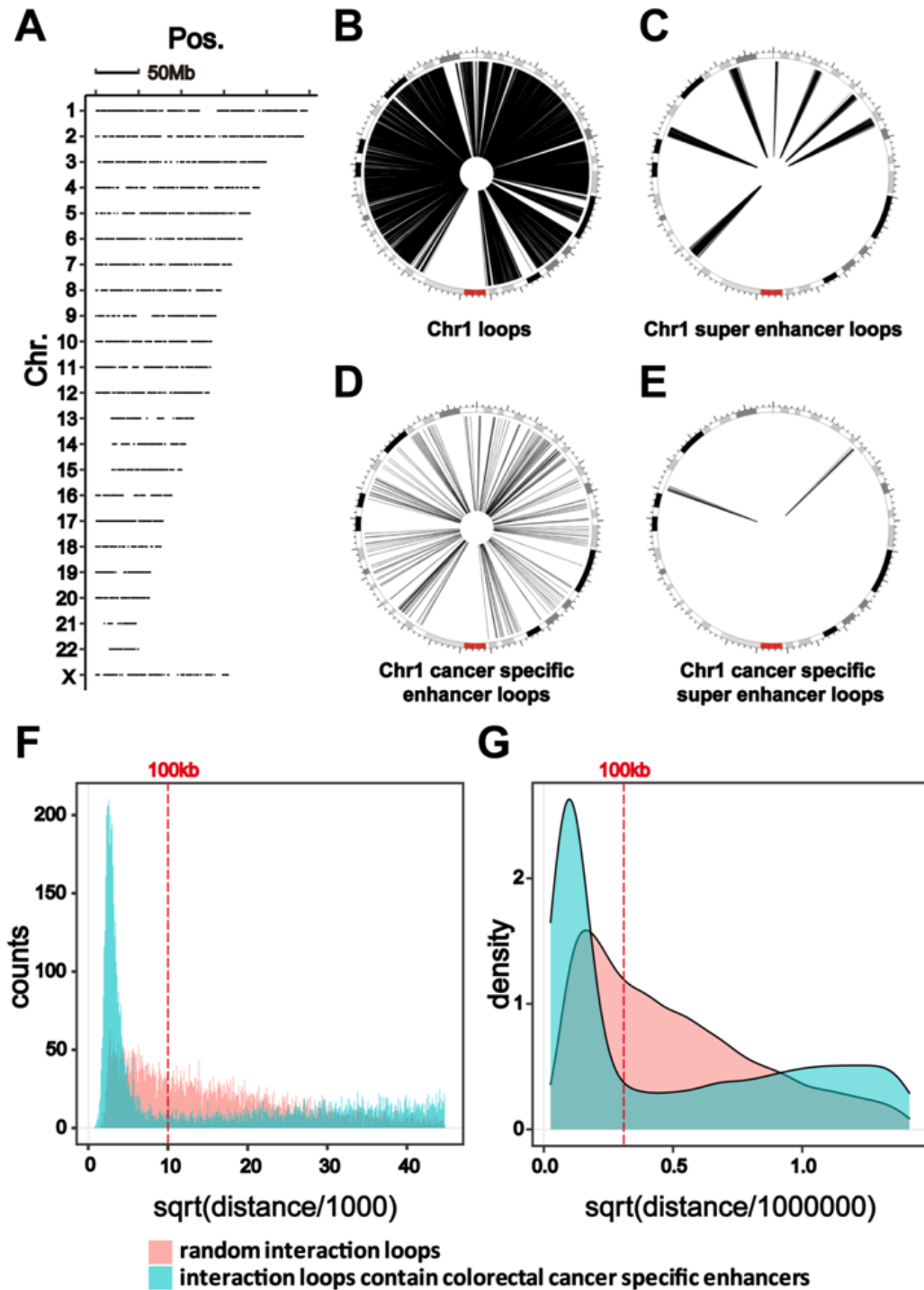


Fig 2. HiChIP identifies chromatin interactions containing colorectal cancer specific enhancers. A) Distributions of colorectal cancer specific enhancers on human genome (hg19). B and C) Circos plot showing interactions in Chr1 indicated by curves extending from enhancers and super enhancers in colorectal cancer cells. D and E) Circos plot showing interactions in Chr1 indicated by curves extending from colorectal cancer specific enhancers and super enhancers in colorectal cancer cells. F and G) Analysis of length distribution of random interaction loops from HiChIP data (Red) and interaction loops that contain colorectal cancer specific enhancers (Blue).

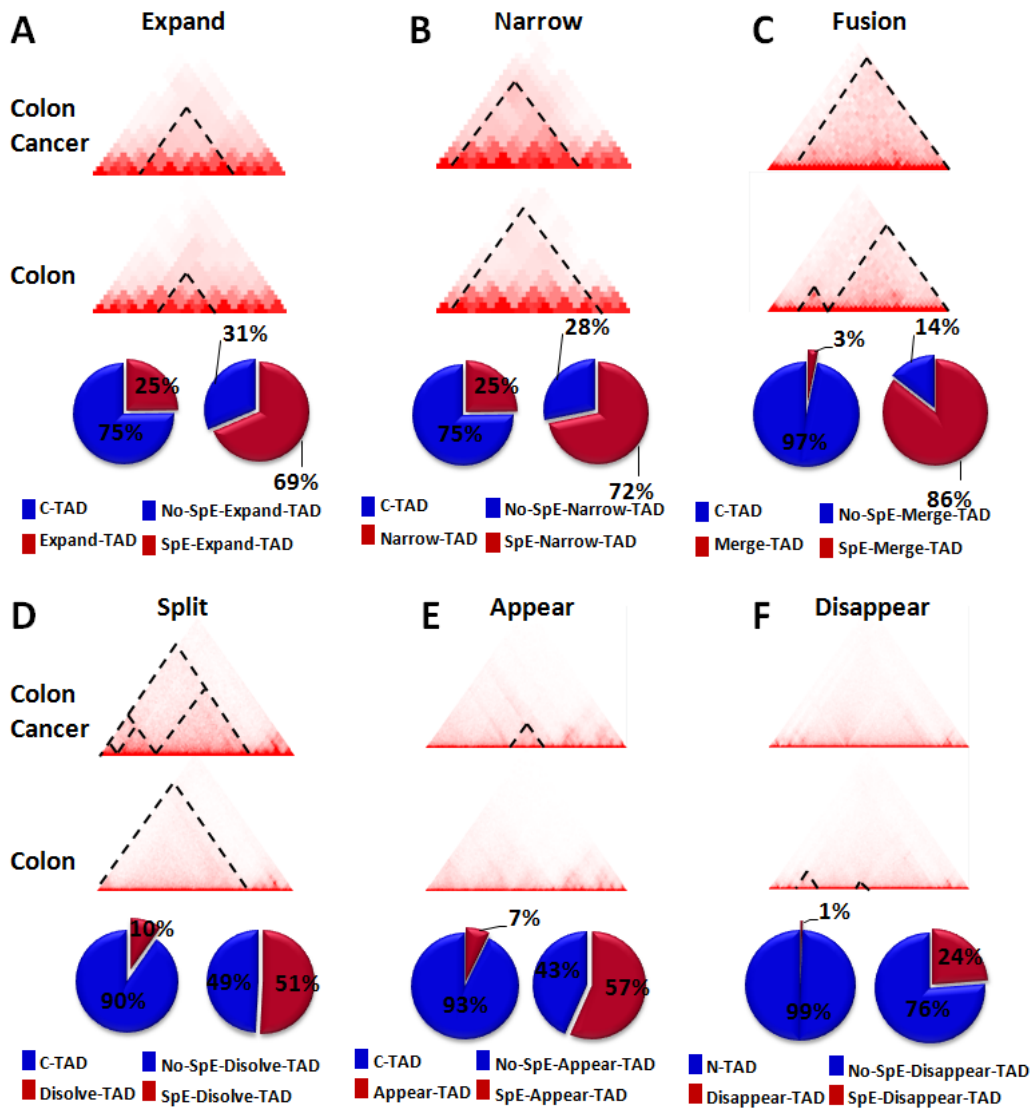


Fig 3. Change of TADs' boundaries in colorectal cancer cells compared with normal colon cells. A) “Expanded TADs” in colorectal cancer cells compared with normal colon cells. B) “Narrowed TADs” in colorectal cancer cells compared with normal colon cells. C) “Fusion TADs” in colorectal cancer cells, two or more TADs in normal colon cells fused to one TAD in colorectal cancer cells. D) “Split TADs” in colorectal cancer cells, new insulation boundaries were formed in one TAD in colorectal cancer cells. E) “Appeared TADs” in colorectal cancer cells that are not exist in normal colon cells. F) “Disappeared TADs”, TADs in normal colon cells that are disappeared in colorectal cancer cells. Pie charts in A-F show the percentages of specific category of boundary-changed TADs (category-TAD) in total changed TADs (C-TAD) in colorectal cancer cells and the percentages of specific category of boundary-changed TADs contain colorectal specific enhancers (SpE-category-TAD) in total changed TADs of this category.

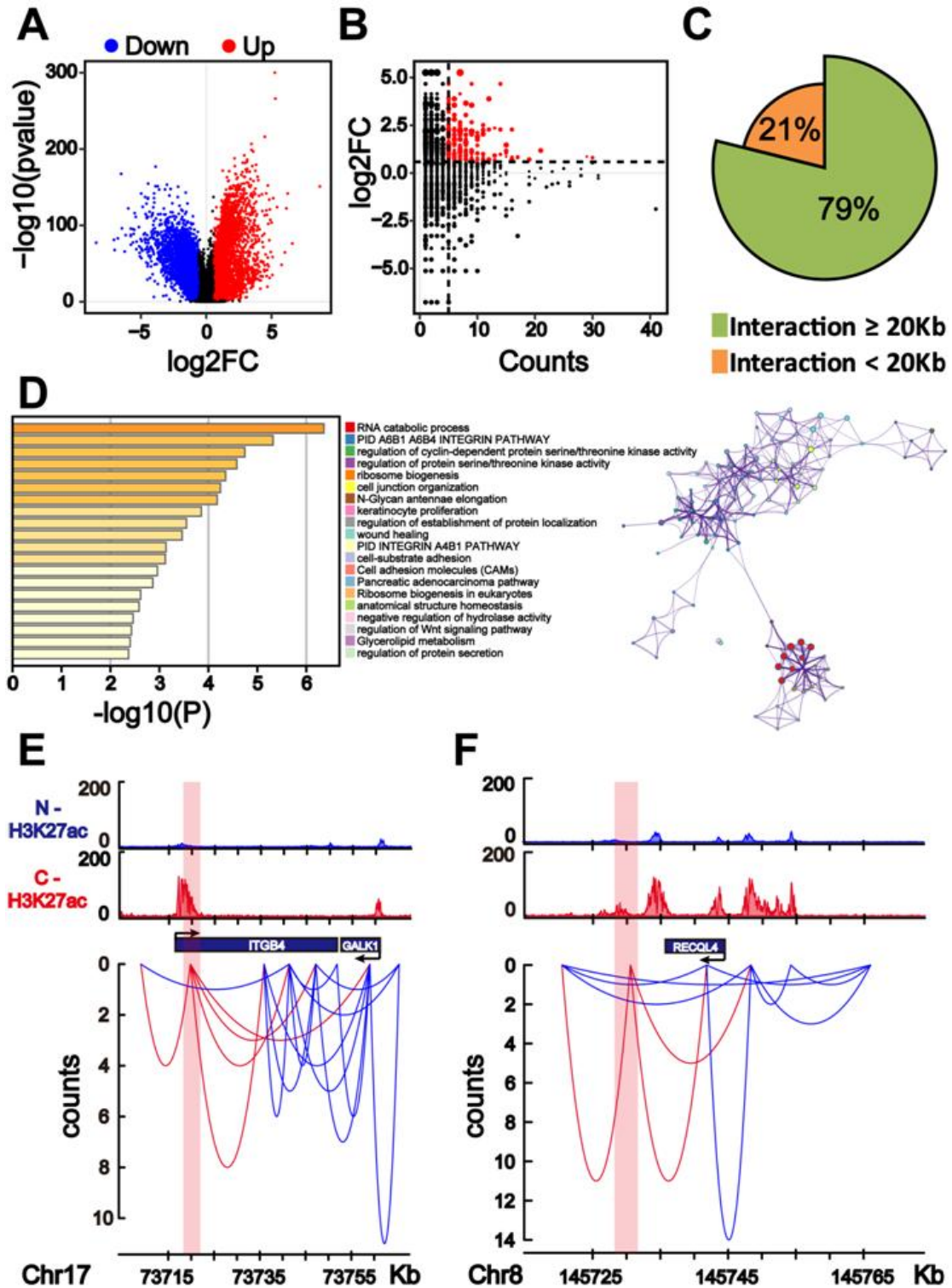


Fig 4. Transcriptome change in colorectal cancer cells is associated with colorectal specific enhancers. A) Volcano plot depicting gene expression changes in normal colon tissue (n=349) and colorectal cancer tissue (n=275). B) Comparison of fold-changes in gene expression based on RNA-seq and HiChIP signal. 735 target genes are shown on the plot. Red plots indicate target genes that significantly changed (fold change>1.5, q<0.01) in colorectal cancer tissue compared with normal colon tissue and have significant HiChIP signal (counts ≥ 5). C) Percentage of long-range

(>20kb) interaction loops between colorectal cancer specific enhancers and target genes. D) Functions of genes as analyzed by Megaspacer. E) H3K27ac enrichment for normal colon (N, blue) and colorectal cancer (C, red) and HiChIP interaction loops (red curves) between target gene *ITGB4* and colorectal cancer specific enhancer. Colorectal cancer specific enhancer is shaded in red. Blue curves indicate other interaction loops in this area. F) H3K27ac enrichment for normal colon (N, blue) and colorectal cancer (C, red) and HiChIP interaction loops (red curves) between target gene *RECQL4* and colorectal cancer specific enhancer. Colorectal cancer specific enhancer is shaded in red. Blue curves indicate other interaction loops in this area.