



## OPEN ACCESS

## EDITED BY

Olga V. Anatskaya,  
Russian Academy of Sciences (RAS), Russia

## REVIEWED BY

Alexander Vinogradov,  
Institute of Cytology, Russia  
Elena Tolkunova,  
Institute of Cytology (RAS), Russia

## \*CORRESPONDENCE

Sanjeev Galande,  
✉ sanjeev.galande@snu.edu.in  
Soumyadeep Nandi,  
✉ soumyadeep.nandi@gmail.com

## †PRESENT ADDRESS

Soumyadeep Nandi,  
Umeå Plant Science Centre, Department of  
Plant Physiology, Umeå University, Sweden

†These authors have contributed equally to  
this work

RECEIVED 30 December 2024

ACCEPTED 24 February 2025

PUBLISHED 28 March 2025

## CITATION

Saw AK, Madhok A, Bhattacharya A, Nandi S and  
Galande S (2025) Integrated promoter-capture  
Hi-C and Hi-C analysis reveals fine-tuned  
regulation of the 3D chromatin architecture in  
colorectal cancer.  
*Front. Genet.* 16:1553469.  
doi: 10.3389/fgene.2025.1553469

## COPYRIGHT

© 2025 Saw, Madhok, Bhattacharya, Nandi and  
Galande. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other forums is  
permitted, provided the original author(s) and  
the copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Integrated promoter-capture Hi-C and Hi-C analysis reveals fine-tuned regulation of the 3D chromatin architecture in colorectal cancer

Ajay Kumar Saw<sup>1</sup>, Ayush Madhok<sup>1†</sup>, Anupam Bhattacharya<sup>2,3†</sup>,  
Soumyadeep Nandi<sup>4\*†</sup> and Sanjeev Galande<sup>1,5\*</sup>

<sup>1</sup>Laboratory of Chromatin Biology and Epigenetics, Department of Biology, Indian Institute of Science Education and Research, Pune, India, <sup>2</sup>Division of Life Sciences, Institute of Advanced Study in Science and Technology, Vigyan Path, Paschim Boragaon, Garchuk, Guwahati, Assam, India, <sup>3</sup>Department of Molecular Biology and Biotechnology, Cotton University, Panbazar, Guwahati, Assam, India, <sup>4</sup>Data Sciences and Computational Biology Centre, Amity Institute of Integrative Sciences and Health, Amity University Haryana, Gurugram, Manesar, Haryana, India, <sup>5</sup>Center of Excellence in Epigenetics, Department of Life Sciences, Shiv Nadar University, Gautam Buddha Nagar, Uttar Pradesh, India

**Introduction:** Hi-C is a widely used technique for mapping chromosomal interactions within a 3D genomic framework, however, its resolution is often constrained by sequencing depth, making it challenging to detect fine-scale interactions. To overcome this limitation, Promoter-Capture Hi-C (PChI-C), as it selectively enriches for promoter-associated interactions, was employed. This study integrates PChI-C and Hi-C datasets from colorectal cancer (CRC) models investigate chromosomal interaction dynamics across various regulatory levels, from cis-regulatory elements to topologically associated domains (TADs). The primary goal is to examine how genomic structural alterations shape the epigenomic landscape in CRC and to assess their potential role in colorectal cancer susceptibility.

**Methods:** PChI-C and Hi-C datasets from multiple colorectal cancer (CRC) studies were integrated to enhance the resolution of chromatin interaction mapping. The analysis focused on identifying fine-scale interactions within topologically associated domains (TADs) while incorporating histone modification landscapes (H3K27ac, H3K4me3) and transcriptomic signatures from CRC cell lines and the TCGA database. For experimental validation, ChIP-quantitative PCR was performed at the promoters of target genes using the highly malignant colorectal cell line HT29 and compared it to an embryonic cell line NT2D1.

**Results:** Our integrated analysis revealed significant genomic structural instability in CRC cells, closely associated with tumor-suppressive transcriptional programs. We identified nine dysregulated genes, including long non-coding RNAs (MALAT1, NEAT1, FTX, and PVT1), small nucleolar RNAs (SNORA26 and SNORA71A), and protein-coding genes (TMPRSS11D, TSPEAR, and DSG4), all of which exhibited a substantial increase in expression in CRC cell lines compared to human embryonic stem cells (hESCs). Additionally, we observed enriched activation-associated histone modifications (H3K27ac and H3K4me3) at the potential enhancer regions of these genes, indicating possible transcriptional activation. ChIP-quantitative PCRs conducted using in the highly malignant CRC

cell line HT29, compared to the embryonic cell line NT2D1, further validated these findings, reinforcing the link between altered chromosomal interactions and gene dysregulation in CRC.

**Discussion:** This study sheds light on the dynamic 3D genome organization in CRC, highlighting critical structural changes associated with disease-associated loci. The identification of nine dysregulated genes points to potential biomarkers for colorectal cancer, with implications for diagnostic and therapeutic strategies. The combination of Hi-C and PCHi-C offers a refined approach for detecting chromosomal interactions at a higher resolution, laying the foundation for future studies on cancer-associated chromatin architecture.

#### KEYWORDS

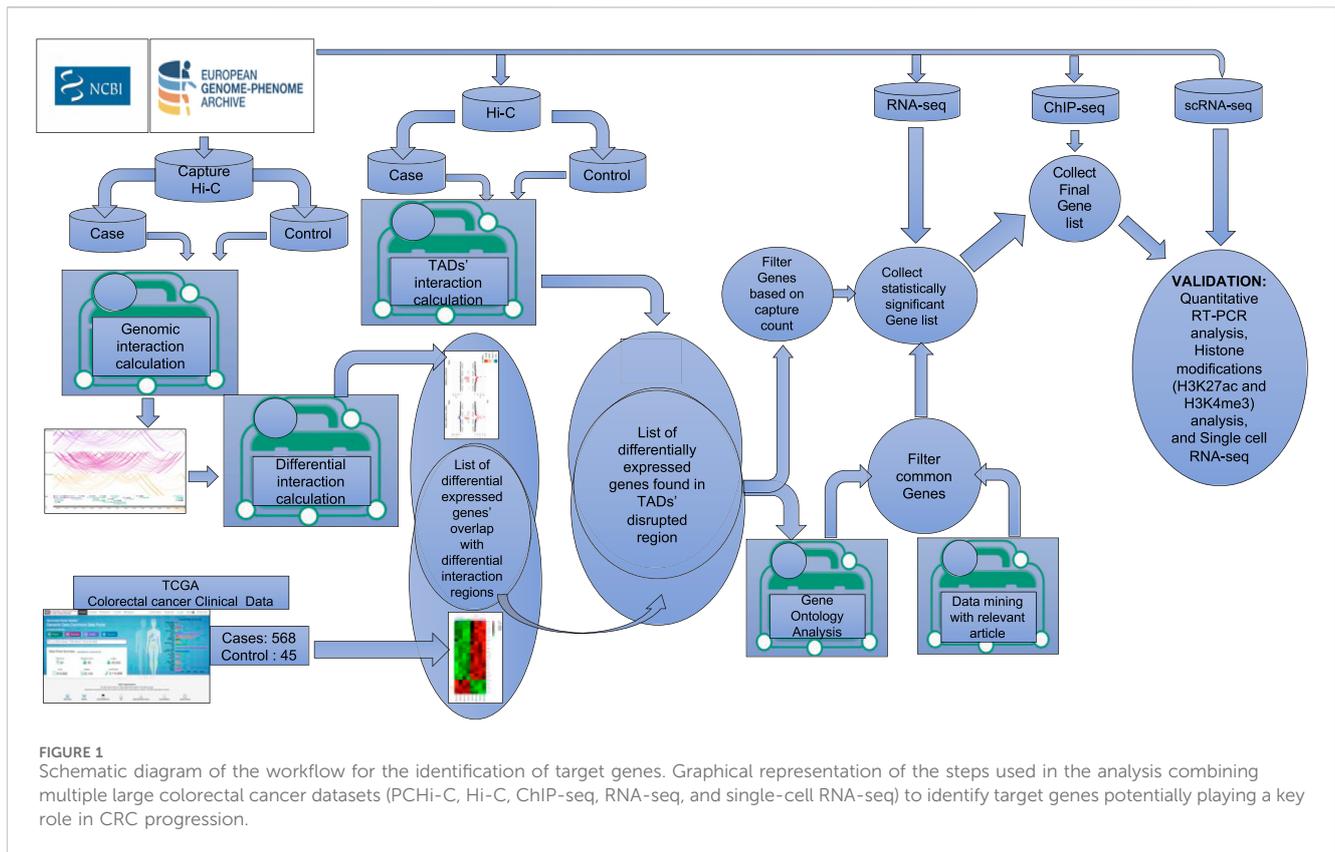
colorectal cancer, topologically associated domains, genome organization, epigenome, gene expression, biomarker

## 1 Introduction

Until recently, studies on the impact of three-dimensional changes on gene expression regulation were restricted due to the limited tools available to investigate the three-dimensional genome structure. Cutting-edge techniques for unraveling chromatin organization have enabled explorations into the gene expression mechanisms in a new dimension. Techniques such as fluorescence *in situ* hybridization (FISH) (Bienko et al., 2013; Beliveau et al., 2012) and chromosome conformation capture (3C, 4C, 5C, Hi-C, and promoter-capture Hi-C) have been developed for studying the mechanisms of gene expression. Using these methods, several studies (Dekker et al., 2002; Lieberman-Aiden et al., 2009; Zhao et al., 2006; Rao et al., 2014; Freire-Pritchett et al., 2021) have revealed hierarchical layers of spatial organization of genomes, further unveiling that the 3D genome architecture is highly dynamic throughout the course of an animal's development. Chromosome conformation capture methods, such as Hi-C and promoter-capture Hi-C (PCHi-C), provide averaged 3D genomic interaction information derived from their application to millions of cells. Hi-C maps have been generated across a variety of human cell categories, encompassing embryonic stem cells and early embryonic lineages (Dixon et al., 2012; Dixon et al., 2015; Pelham-Webb et al., 2020; Djeghloul et al., 2020; Du et al., 2020; Collombet et al., 2020), immune cells (Rao et al., 2014), fibroblasts (Jin et al., 2013), and other primary tissue types (Schmitt et al., 2016). Hi-C<sup>4</sup> enables whole genome-wide mapping of long-range chromatin interactions and, therefore, represents a powerful approach for predicting the distal gene targets of disease-associated loci. Hi-C-derived methodologies are advancing our understanding of TAD-level organization, A/B compartments, and loop formations (Klein et al., 2021; Fukuda et al., 2021; Spracklin et al., 2021; Hildebrand and Dekker, 2020; van Schaik et al., 2020; Su et al., 2020). However, effectively identifying intra-TAD interactions, such as regulatory loops, from Hi-C data remains a challenging task due to the complexity of the Hi-C libraries and the considerable expense associated with deep sequencing to obtain statistically significant interactions. On the other hand, targeted chromatin-capture techniques such as PCHi-C provide cis-regulatory insights for a specific subset of clinically relevant genomic regions at a substantially lower cost with fine-scale genomic interactions. High-resolution maps of clinically relevant loci enable more accurate predictions of the impacts of structural changes and alterations that may lead to human diseases or developmental abnormalities (Lupiáñez

et al., 2016). PCHi-C significantly enhances the capability to identify interactions that encompass promoter sequences. PCHi-C analysis across distinct cell types recognized numerous enhancer-promoter interactions and unveiled significant differences in promoter architecture between cell types and during differentiation (Schoenfelder et al., 2015; Mifsud et al., 2015; Javierre et al., 2016; Freire-Pritchett et al., 2017; Rubin et al., 2017; Siersbæk et al., 2017). These investigations demonstrated that the organization of the genome reflects cellular identity, underscoring the importance of disease-relevant cell types toward deciphering the gene regulatory mechanisms associated with disease loci. To support this idea, several recent studies have utilized high-resolution promoter interaction maps to identify tissue-specific target genes associated with GWAS findings. An examination of promoter-capture Hi-C data across 17 primary human blood cell types captured 2,604 potentially significant genes related to immune and blood-related disorders, including a considerable number of genes with roles yet to be annotated in those diseases (Javierre et al., 2016). Montefiori et al. generated high-resolution PCHi-C mapping in human-induced pluripotent stem cells (iPSCs), and iPSC-derived cardiomyocytes (CMs) unveiled 1,999 single-nucleotide polymorphisms (SNPs) associated with cardiovascular diseases, which were linked to 347 target genes. This highlights the significance of incorporating long-range chromatin interactions into the interpretation of functional targets associated with disease loci (Montefiori et al., 2018). Philip et al. performed a GWAS meta-analysis to explore the gene regulatory mechanisms underpinning all GWAS risk loci. They achieved this by analyzing PCHi-C datasets and capturing chromatin interactions between predisposition loci and target genes. They also scrutinized gene expression data and integrated these findings with chromatin immunoprecipitation-sequencing (ChIP-seq) data for 31 newly identified loci, in addition to the previously known loci, to investigate the heritable risk of colorectal cancer susceptibility (Law et al., 2019).

Colorectal adenocarcinoma, the fourth most common epithelial tumor, continues to be the predominant cause of mortality worldwide (Bray et al., 2018; Alzahrani et al., 2021). In order to acquire detailed genome architectural insights into the progression of colorectal tumors, it is essential to establish a comprehensive gene regulatory map of human colon cells. This will facilitate an understanding of how TAD disruptions influence gene regulation. In this study, we present an integrative approach to



comprehensively detect structural changes in cancer genomes by combining Hi-C, PCHi-C, RNA-seq, ChIP-seq, and single-cell RNA-seq (scRNA-seq) analyses for the identification of the multiple target genes that may potentially serve as biomarkers for CRC susceptibility (the strategy is depicted in Figure 1). We further demonstrate that the expression of these genes, along with the associated activatory chromatin modifications, such as H3K27ac and H3K4me3, at the target regions is upregulated in CRC. In summary, this study illustrates the dynamic interplay between global and local chromatin architecture. Furthermore, by integrating chromatin architecture with gene expression and chromatin modification profiles, we identified novel regions of colorectal cancer susceptibility.

## 2 Materials and methods

### 2.1 Promoter-capture Hi-C data processing and analysis

Raw PCHi-C sequencing reads were aligned to the hg38 human reference genome and filtered using HiCUP (S et al., 2015) and Bowtie2 (Langmead and Salzberg, 2012). To identify regions that interacted with promoters, referred to as promoter-interacting regions (PIRs), we utilized the CHiCAGO tools (Cairns et al., 2016). Promoter-interacting regions were identified by comparing the number of promoter-ligated regions within each genomic bin to the expected reads, according to the generated model. Genomic bins with CHiCAGO scores  $\geq 5$  were considered to be PIRs. BAM files were converted into a

CHiCAGO-compatible format (.chinput) using bam2chicago script. Chinput files are tab-delimited text files that contain essential information of all pairs of promoter-“other end” regions for which at least one ligation region was detected. Each line in the .chinput file provides details about one promoter-“other end” pair, including their respective genomic coordinates.

### 2.2 Differential chromosomal interaction analysis using PCHi-C data

Detecting differential signals in sequencing data is a fundamental and frequently performed task in genomic analyses. We used Chicdiff (Cairns et al., 2019) for the differential analysis of significant interactions identified by CHiCAGO between different cell types. This tool integrates moderated differential testing for count data using negative binomial generalized linear models implemented in DESeq2 (Love et al., 2014), with signal normalization informed by CHiCAGO and non-uniform multiple testing correction.

### 2.3 Hi-C data processing and analysis

Reads were aligned to the human reference genome hg38 using Bowtie2 (Langmead and Salzberg, 2012), and then SAMtools (Li et al., 2009) was used to convert the reads to the BAM format. We used the HiCExplorer package (Wolff et al., 2020; Ramírez et al., 2018) for Hi-C data processing. The “hicBuildMatrix” command was used to construct the matrix of read counts over the genomic

bins, taking into account sites around the given restriction site (AAGCTT). The matrix was corrected using the “hicCorrectMatrix” command to eliminate biases related to GC content, open chromatin, and, most importantly, normalize the number of restriction sites per bin. Topologically associated domains (TADs) were identified using the “hicFindTADs” command with specific parameters, including “-minDepth 60,000,” “-maxDepth 120,000,” and “-step 20,000.” The visualization of the interaction matrix and other signal tracks was performed using the “hicplotsTADs” command. The positions of TAD boundaries were used to analyze the localization of identified genomic interactions relative to these TAD boundaries.

## 2.4 ChIP-seq data processing and analysis

ChIP-seq reads were trimmed using the Trim Galore tool for adapter removal. Filtered reads were mapped to the human reference genome hg38 using Bowtie2 (Langmead and Salzberg, 2012), and then SAMtools (Li et al., 2009) was used to convert the reads to the BAM format. macs2\_callpeak was used for peak calling with  $-tsize = 26$ ,  $-gsize = 2.7e9$ , and  $-extsize 200$ , with the remaining settings left as default. The MACS2 tool is designed to account for the impact of genome complexity when assessing the significance of enriched ChIP regions. It enhances the spatial resolution of binding sites by incorporating information from both the position and orientation of sequencing tags.

## 2.5 RNA-seq data processing and analysis

### 2.5.1 RNA-seq pipeline 1

The gene expression abundance from paired-end RNA-seq datasets was downloaded from The Cancer Genome Atlas (TCGA) data portal. For the retrieval of colon cancer datasets, primary sites “colon” and “rectum,” project IDs “TCGA-COAD” and “TCGA-READ,” and sample types “primary tumor,” “tumor,” and “metastatic” were selected. For the control datasets, “blood-derived normal” and “solid tissue normal” parameters were selected during the data retrieval process. The sample manifest file was prepared, and GDC client (Morgan and Davis, 2017) was used for downloading the datasets.

### 2.5.2 RNA-seq pipeline 2

Quality control of the dataset and mapping of processed reads were carried out as follows:

Raw files were downloaded and converted using the SRA toolkit. After that, the QC program was used to filter low-quality reads and remove any sequencing errors. The processed reads were aligned using the RNA-seq aligner STAR (v2.5)<sup>42</sup> with the Ensembl human reference annotation (GRCh38).

### 2.5.3 Quantifying transcript abundances

RSEM (Li and Dewey, 2011) was employed to quantify transcript abundance from the paired-end RNA-seq datasets. The human reference genome was built using the rsem-prepare-reference script. We used STAR to perform transcriptome-based mapping, and gene expression levels were calculated from STAR-generated BAM files by rsem-calculate-expression scripts.

## 2.5.4 Identification of differentially expressed genes

The edgeR package (v3.30.0)<sup>44</sup>, which relies upon count-based expression data for determining differential expression in R, was used for differential gene expression analysis. Transcripts with zero expression values were removed. Normalization of the counts was performed using the calcNormFactors function of edgeR, which normalizes for differences in RNA composition by finding a set of scaling factors for the library sizes that minimize the log-fold changes between samples for most genes. The TMM method of normalization was used for the datasets.

### 2.5.5 RNA-seq pipeline 3

The preprocessed RNA transcripts were mapped to *Homo sapiens* (Gencode v29) transcriptomes using Salmon (Patro et al., 2017). The resulting quant files were imported into R (v4.0.3) for exploratory data analysis. R package tximeta (Love et al., 2020) was used to import and summarize transcript-level data to the gene level. DESeq2 (Love et al., 2014) was used for differential expression analysis with the default setting. Further details of the used pipeline are available at the following link: (<https://bioconductor.org/packages/release/workflows/vignettes/rnaseqGene/inst/doc/rnaseqGene.html>).

## 2.6 GO analysis, network analysis, and mutational analysis details

The STRING database was used for the functional annotation of genes. During the analysis, GO biological process libraries 2021 and molecular function libraries 2021 were selected. The statistically significant results ( $p\text{-value} \leq 0.05$ ) were further considered for downstream analysis as a reference organism for the annotation of genes. The enriched gene clusters were visualized in the STRING network. We curated the function of genes based on functional enrichment analysis associated with key functions such as chromatin remodeling, chromatin organization, and cellular organization, among others. For mutational analysis, the Colon Cancer Atlas and the COSMIC database were used as reference databases, containing data from 13,711 CRC tissues and more than 165 CRC cell lines.

## 2.7 Single-cell RNA-seq data processing and analysis

We used the Seurat package (Stuart et al., 2019) (v4.0.6) for this integrative multimodal analysis. Genes detected in fewer than 100 cells, cells exhibiting expression of less than 200 detected genes, and cells expressing >15% mitochondrial genes were removed for downstream analysis. We adopted the general protocol described by Stuart et al. (2019) for grouping single cells into different cell subsets. We employed the following steps: data normalization, identification of highly variable features, scaling the data, clustering the cells, reordering the clusters based on their similarity, running non-linear dimensional reduction (tSNE), and labeling the cell types. Principal component analysis (PCA) was carried out on the scaled data of highly variable genes. The first 30 principal components (PCs) were used to cluster the cells and perform a subtype analysis by non-linear dimensionality reduction

(t-SNE) (Van Der Maaten, 2014). Additionally, we utilized the SingleR package (v1.8.1) for labeling cell types (Aran et al., 2019).

## 2.8 Cell lines and culture

NT2D1 embryonic cells (gifted by Dr. Peter Andrews) were cultured in DMEM (Invitrogen), with 20% FBS (Invitrogen), 1X antibiotics (Antibiotic-Antimycotic, Invitrogen) and NEAA supplements (Lonza) at 37°C with 5% CO<sub>2</sub>. The cells were split after every 3 days and maintained at 70% confluency to avoid any differentiation. HT29 colorectal carcinoma cell line (ATCC, HTB-38) was cultured in DMEM, with 10% FBS and antibiotics. Cells were split after every 1.5 days. Similar cell culture regime was followed with a change to RPMI-1640 media (Thermo Fisher Scientific) for other colon cancer lines used including DLD1 (ATCC, CCL-221) COLO-205 (ATCC, CCL-222) and HCT-115 (ATCC, CCL-225). FHC (fetal human colon) and hESC (human embryonic stem cells) were used as control cell lines, while HT29 and LoVo served as treated cell lines. We compared FHC and hESC with the cancer cell lines (HT29 and LoVo) to evaluate their similarity in terms of changes in Topologically Associated Domains (TADs). The primary objective was to assess whether hESC could function as a control for further downstream analyses. Two or 10 million cells were harvested for RNA extraction and chromatin immunoprecipitation assays, respectively.

## 2.9 RNA extraction and quantitative RT-PCR

The cells harvested ( $2 \times 10^6$ ) were washed twice with cold 1X PBS, and the pellets were re-suspended in TRIzol (Invitrogen), followed by the isolation of total RNA. Following DNase I (Promega) digestion, RNA (260/280 ratio ~2) was subjected to cDNA synthesis using a high-capacity cDNA synthesis kit (Applied Biosystems), as per the manufacturer's protocol. Target oligos were designed using the UCSC Genome Browser and Primer3 software. The expression level of h18s RNA was used as the internal control for the normalization of target transcripts. Quantitative RT-PCR analyses were performed using TB Green II qPCR Master Mix (TaKaRa) with the following PCR conditions: step 1, 95°C for 5 min; step 2, 95°C for 45 s, 60°C for 45 s, and 72°C for 1 min, repeated for 40 cycles, using the ViiA 7 Real-Time PCR system (Applied Biosystems). The change in gene expression was calculated using the formula  $\Delta Ct = Ct_{\text{target}} - Ct_{\text{control}}$ . Normalized transcript expression was calculated using the equation;  $2^{-(\Delta Ct)}$ , where  $\Delta Ct = Ct_{\text{target}} - Ct_{\text{control}}$ . The oligonucleotide primer sequences used for qRT-PCR analyses are listed in [Supplementary Table S1](#). Statistical analysis was performed using one-way ANOVA (GraphPad v9.1) on three biological replicates.

## 2.10 Chromatin immunoprecipitation q-PCR

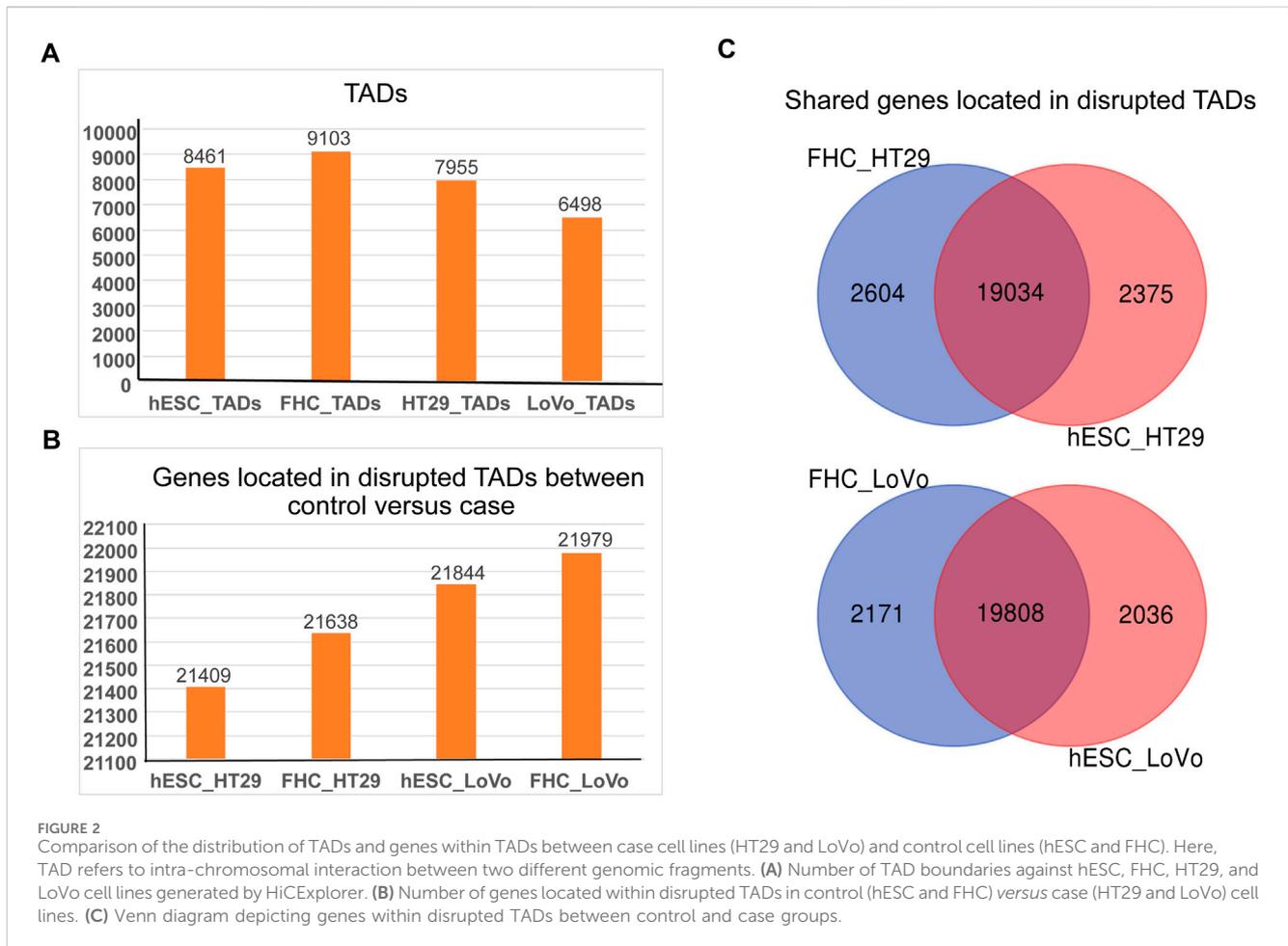
Cells obtained were cross-linked using 1.25% formaldehyde (Sigma), followed by quenching with 150 mM glycine. Cross-

linked cells were washed twice with PBS and subjected to chromatin isolation and shearing, as described by Patta et al. (2020), with a few modifications. In brief, nuclei were isolated using the hypotonic buffer (25 mM Tris-HCl at pH 7.9, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 0.1% NP-40, 1 mM DTT, 0.5 mM PMSF, and 1× protease inhibitor cocktail) (Roche). Pelleted nuclei were lysed using the sonication buffer (50 mM Tris-HCl at pH 7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.5% SDS, 0.5 mM PMSF, and 1× protease inhibitor cocktail) (Roche), and chromatin was subjected to sonication using the Covaris M220 sonicator with the following parameters: peak power 75, duty factor 20, burst 300, and duration 10 min. We obtained the chromatin fragment size of 200–400 bp. After pre-clearing, chromatin was subjected to immunoprecipitation using anti-H3k27ac (Abcam) and anti-H3K4me3 (Abcam) antibodies overnight at 4°C. Similarly, normal IgG was used as a control. Immunoprecipitated complexes were pulled down by adding protein A/G beads (Thermo Fisher Scientific), and the cocktail was incubated at 4°C for 4 h. The immunoprecipitated bead-bound chromatin was washed thoroughly using low-salt, high-salt, and lithium chloride buffers and subjected to elution using the elution buffer (1% SDS, 0.1 M NaHCO<sub>3</sub>). The eluted chromatin was de-crosslinked, and protein and RNA contamination were removed by treating with proteinase K (Sigma-Aldrich) and RNase A (Sigma-Aldrich), respectively. Furthermore, the immunoprecipitated chromatin was purified and subjected to the quantitative PCR analysis using the formula  $\Delta Ct = Ct_{\text{Target}} - Ct_{\text{Input}}$ . Fold differences in enrichment were calculated using equation  $2^{-(\Delta Ct)}$ , where  $\Delta Ct = Ct_{\text{Target}} - Ct_{\text{Input}}$ , for both IgG- and TCF1-immunoprecipitated DNA. The primer sequences used for ChIP-PCR analysis are listed in [Supplementary Table S2](#). Statistical analysis was performed using one-way ANOVA (GraphPad v9.1) with three biological replicates.

## 3 Results

### 3.1 hESC and FHC cell lines share high percentage of common genes with HT29 and LoVo CRC cell lines in TADs disruption comparison

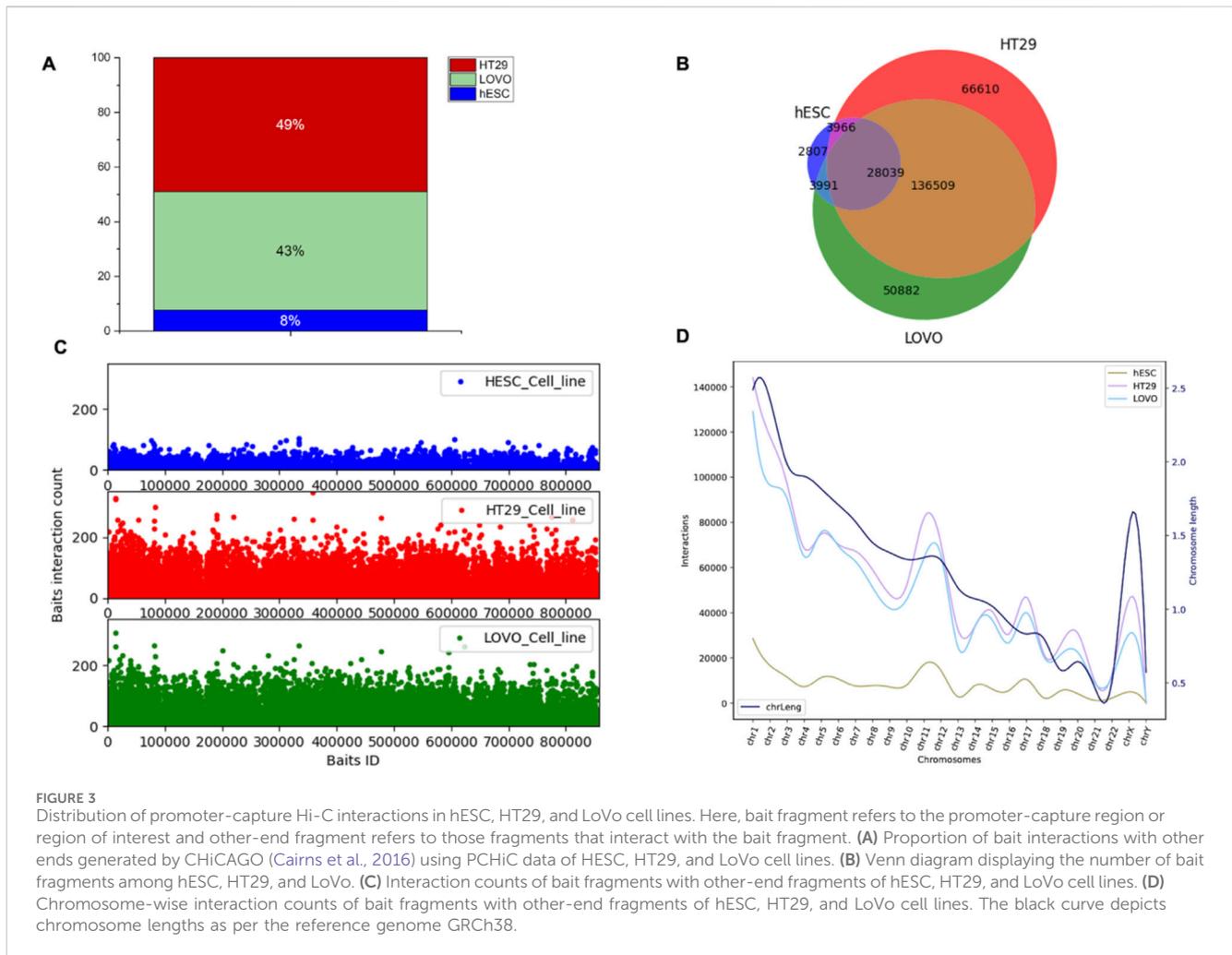
We performed comparative study of FHC (fetal human colon) and hESC (human embryonic stem cell) lines to understand the chromatin architectural differences between FHC and hESC versus the CRC cell lines (HT29 and LoVo). The basis for selecting conserved Topologically Associated Domains (TADs) was ensuring that the genes within the overlapping TADs matched those in their respective TADs. Consequently, overlapping TADs that did not meet this criterion were regarded as disrupted TADs. Towards this, we filtered affected genes which have been found in TADs disrupted regions in FHC & hESC versus CRC (HT29 & LoVo) cell lines. HiCExplorer39 has been used to search for topological associated domains (TADs) ([Figure 2A](#)). [Figure 2B](#) depicts the number of genes that have been found in TADs disruption regions in FHC & hESC versus CRC (HT29 & LoVo) cell lines. Further comparative analyses revealed 79% common genes within the TADs



disruption region in hESC versus HT29 and FHC versus HT29 comparison and 83% common genes lies in the TADs disruption region in hESC versus LoVo and FHC versus LoVo cells. This indicates that TADs generated using FHC and hESC datasets have maximum gene similarity with the CRC cell in the TADs disruption region. However, due to the limited depth of sequencing of Hi-C datasets, we were unable to capture fine scale interactions within TADs and therefore it is difficult to integrate fine-scale interaction with histone modification marks and transcriptional signature. In such a scenario, promoter-capture Hi-C (PChI-C) data is often used for mapping fine scale interactions. Currently, PChI-C data is available only for hESC, HT29 and LoVo, not FHC cells in publicly available databases. We performed a comparative study of hESC and FHC against CRC cell lines (Figure 2C) to capture and compare the subset of common genes in the TADs disruption region. We obtained 79% and 83% common genes in TADs disruption loci respectively (Figure 2C), showing good similarity between hESC and FHC cell lines while comparing with CRC cell lines in the context of TADs disruption. Since the PChI-C data is not available for FHC, we therefore chose to use hESC instead FHC for further comparative study. Thus, we used hESC as a control and HT29 and LoVo as a case for capturing the fine-scale interactions and association with epigenomic marks for novel biomarker identification in CRC progression.

### 3.2 Enhancer–promoter interactions are enriched in HT29 and LoVo cell lines compared to hESC, which corresponds to changes in gene expression dynamics

We used PChI-C data to examine genomic interaction in CRC. The captured regions (genomic region of interest or diseases-associated loci) are referred to as “baits” throughout the manuscript. In this study, we selected two colon cancer cell lines (HT29 and LoVo) and compared them with human embryonic stem cells (hESCs). We used ChIAGO (Cairns et al., 2016) to identify significant interactions. We investigated the interactions that were separated by a distance of at least 10 Kb. We identified bait interactions with other genomic loci, and the results are shown in Figure 3. In total, we identified 1,283,056 (49%) and 1,135,609 (43%) bait interactions in HT29 and LoVo cell lines, respectively, and 201,356 (8%) interactions in the hESC cell line (Figure 3A). Most interactions were found in cancer cell lines (HT29 and LoVo) compared to those in hESCs. A significant proportion of bait interactions was shared between the cancer cell lines compared to those in hESCs (Figure 3B). We also calculated the counts of interactions with respect to each bait fragment and chromosome-wise across all the cell lines (Figures 3C,D). We observed similar interaction patterns in HT29 and LoVo cell lines

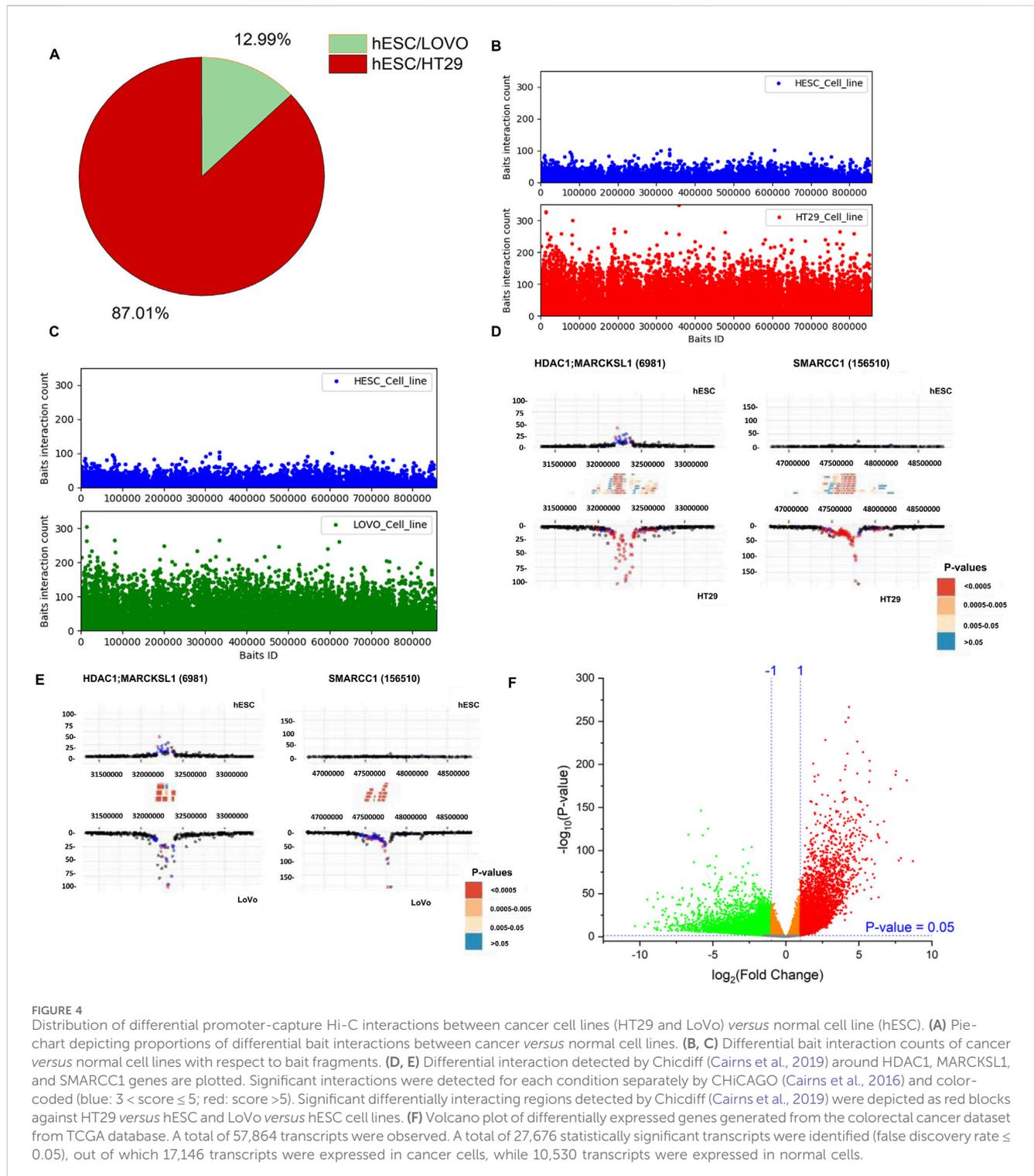


compared to those in hESCs (Figure 3D). We used Chicdiff (Cairns et al., 2019) for differential bait interaction between cancer and normal cell lines. We plotted the distribution of differential bait interactions (Figures 4B,C) and observed 3,895,502 (~87%) and 582,084 (~13%) differential interactions of HT29 *versus* hESC and LoVo *versus* hESC cell lines, respectively (Figure 4A). Differential interactions between two bait fragments (Figures 4D,E) were plotted with interactions within 500 Kb upstream and downstream from the bait, passing the threshold of 5 (threshold >5). The threshold criteria >5 was set up by the Chicdiff (Cairns et al., 2019) tool for significant differential interactions. Similar to the differential interaction analysis, we performed differential RNA-seq analysis using CRC datasets obtained from TCGA databases (The Cancer Genome Atlas Network, et al., 2012). The identified differentially expressed (DE) genes are depicted in Figure 4F. We found approximately twenty-eight thousand transcripts that were differentially expressed (false discovery rate  $\leq 0.05$ ). Nearly 62% of transcripts were expressed in cancer cells compared to those in normal cells. In cancer cells, 35% were protein-coding genes, and the remaining were non-coding genes (lincRNA, pseudogenes, antisense, miRNA, etc.). All the transcripts, along with their annotations, are provided in Supplementary Table S3. Gene expressions related to colorectal cancer (primary tumor and

metastatic) were compared with blood-derived normal, solid tissue to determine the DE genes using edgeR (Robinson et al., 2010). For the extraction of DE genes that lie within the differential interaction region, we overlapped the genomic coordinates of differential bait fragments with the genomic coordinates of DE genes. In hESC/HT29 and hESC/LoVo case studies, we found 24,438 and 17,385 DE transcripts, respectively, after overlapping. The abovementioned steps enabled finding the repertoire of DE genes due to structural changes at the chromatin loop level.

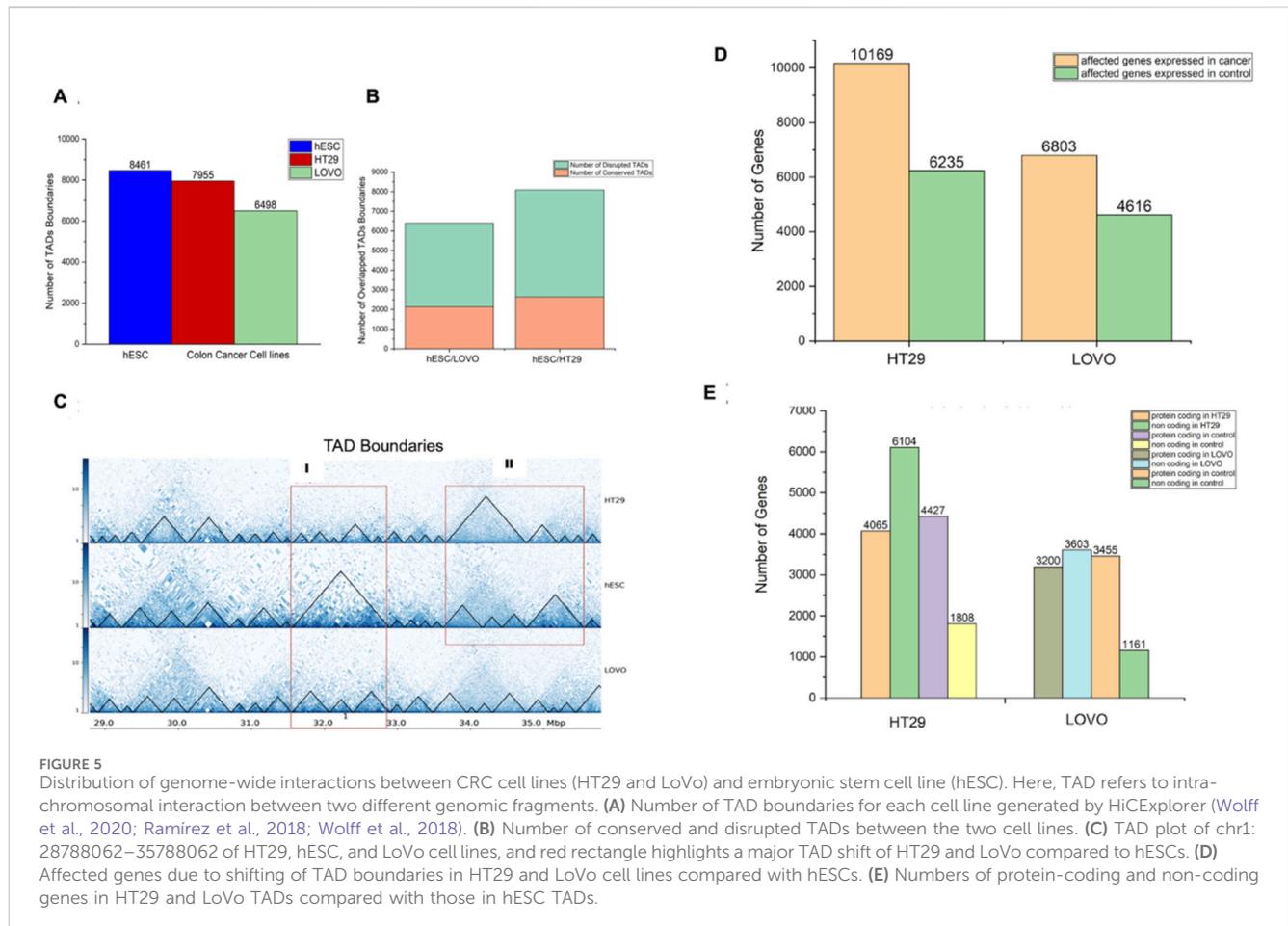
### 3.3 Disruption in TAD dynamics of HT29 and LoVo cell lines, compared to that in hESCs, involves a subset of colorectal-specific genes

In the previous section, we overlaid differentially expressed genes with the bait fragment's coordinates of the differential interaction region. To understand the distribution of TAD boundary regions with respect to bait fragments, we further overlapped the bait fragments in such a way that the starting bait should lie within the range of TAD boundary regions. Using HiCExplorer, we plotted the TADs with 10 Kb resolution and compared the TAD boundaries among the three cell lines (hESC, HT29, and LoVo). We found 8,461, 7,955, and 6,498 TAD boundaries,



respectively, in hESC, HT29, and LoVo cells (Figure 5A). Next, the overlapped TAD boundaries between two cell lines containing at least one DE gene were classified into two categories: conserved TADs and disrupted TADs. The total number of conserved and disrupted TADs (Figure 5B) of hESC/HT29 and hESC/LoVo are 2,637 and 5,448 and 2,138 and 4,255, respectively. The criterion for selecting conserved TADs is that the genes within the overlapping TADs are the same as those in their corresponding TADs. Hence, overlapping TADs that are not

conserved were considered disrupted TADs. We then investigated nearby genes in detail around the TAD disrupted regions. We filtered genes based on their position in the TAD boundary shift loci between normal and cancer cells (Figure 5C). We mapped and filtered the disrupted genes in TAD shifting boundary regions with the TCGA database (Figure 5D). We found that 47% of the disrupted genes were protein-coding and 53% were non-coding genes, which included lincRNAs, miRNAs, pseudogenes, and other non-coding variants in

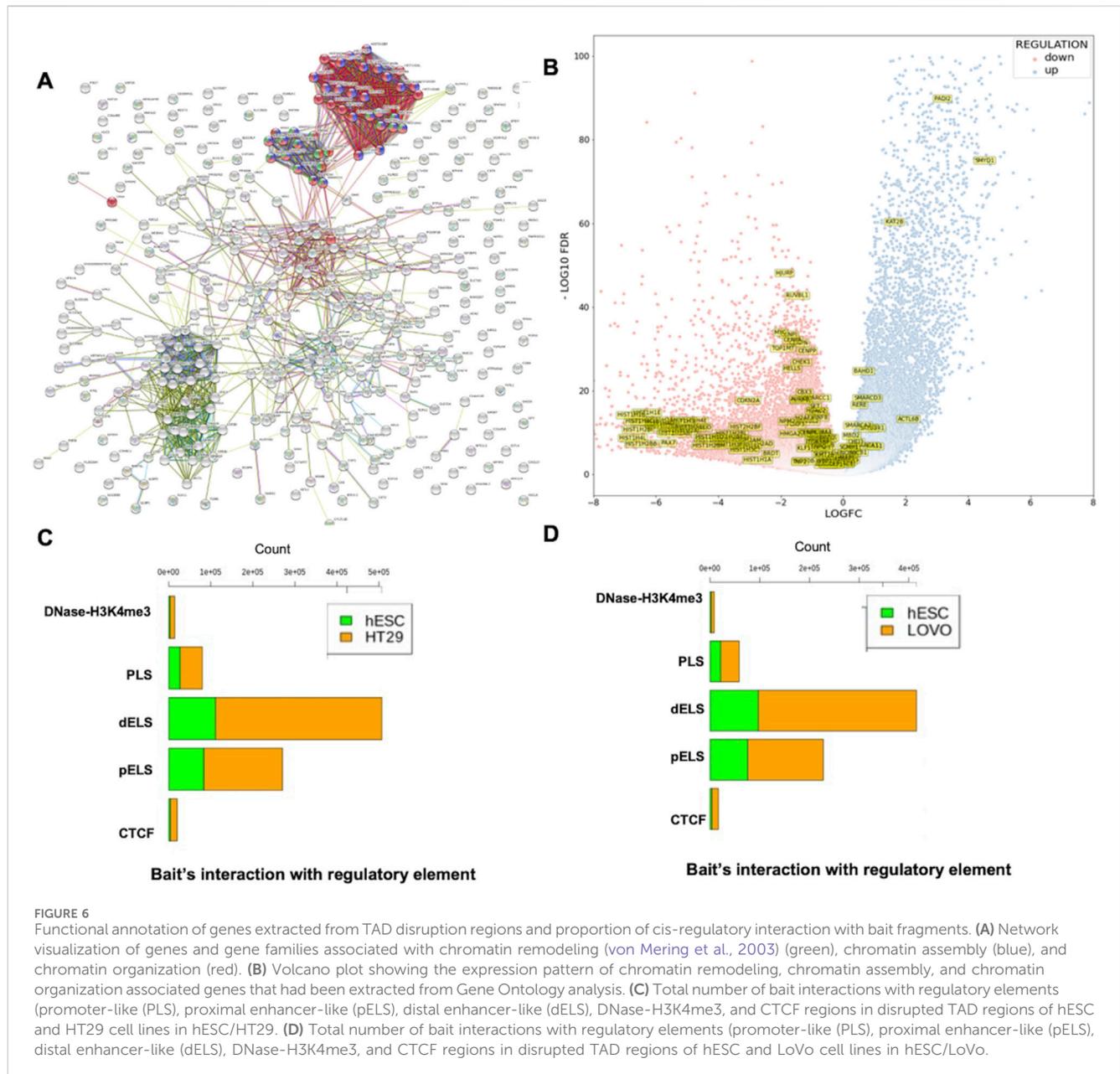


**TABLE 1** List of genes associated with chromosome organizations and assembly.

Function	Name of the genes
Chromatin assembly	Histone H1 HIST1H1A, HIST1H1B, HIST1H1D, and HIST1H1E Histone H2 HIST1H2BB, HIST1H2BE, HIST1H2BF, HIST1H2BI, HIST1H2BL, HIST1H2BM, and HIST1H2BO Histone H3 HIST1H3A, HIST1H3B, HIST1H3C, HIST1H3F, HIST1H3I, and HIST1H3J Histone H4 HIST1H4A, HIST1H4B, HIST1H4C, HIST1H4D, HIST1H4E, HIST1H4L, and HIST2H2BF
Regulation of chromatin remodeling	CDKN2A, HIST1H4A, HIST1H4B, HIST1H4C, HIST1H4D, HIST1H4E, and HIST1H4L INO80, SMARCA1, SMARCA2, SMARCA4, SMARCA5, SMARCB1, SMARCC1, SMARCC2, SMARCD1, SMARCD2, and SMARCD3
Chromatin organization and nucleosome organization	HIST1H2AB, HIST1H2AD, HIST1H2AH, HIST1H2AJ, HIST1H2AL, HIST1H2AM, HIST1H3A, HIST1H3B, HIST1H3C, HIST1H3F, HIST1H3I, HIST1H3J, HIST1H4A, HIST1H4B, HIST1H4C, HIST1H4D, HIST1H4E, HIST1H4L, and HIST2H2AB CDKN2A, CENPN, CENPU, CENPT, HDAC2, CENPW, CENPQ, CENPI, H2AFX, CHEK1, CBX3, TNP2, CASC5, AURKB, ITGB3BP, RBBP7, SET, CHRAC1, SUPT16H, ATRX, HJURP, ACTB, ACTL6B, ARID1A, BAHD1, BAZ1B, BAZ2A, BRD4, BRDT, CBX3, CENPA, CENPK, CENPO, CENPP, CHD4, CHRAC1, DMAP1, ESRI, FOXP3, HDAC1, HDAC2, HDAC5, HELLS, HILS1, HMGA2, HNRNPC, INO80E, INO80C, INO80B, INO80B-WBP1, KAT2B, KLF1, KMT2B, MBD2, MEN1, MIS18A, MIS18BP1, MYC, NASP, NPM2, OIP5, PADI2, PAX7, PBRM1, PSME4, RBBP7, RERE, RNF8, RUVBL1, RUVBL2, SATB1, SATB2, SCMH1, SIRT1, SMYD1, TOP1MT, and TTF1

the hESC/LoVo cells. Similarly, 40% expressed genes were protein-coding and 60% were non-coding genes in hESC/HT29 cells (Figure 5E). This approach allowed us to monitor the distribution of nearby genes across

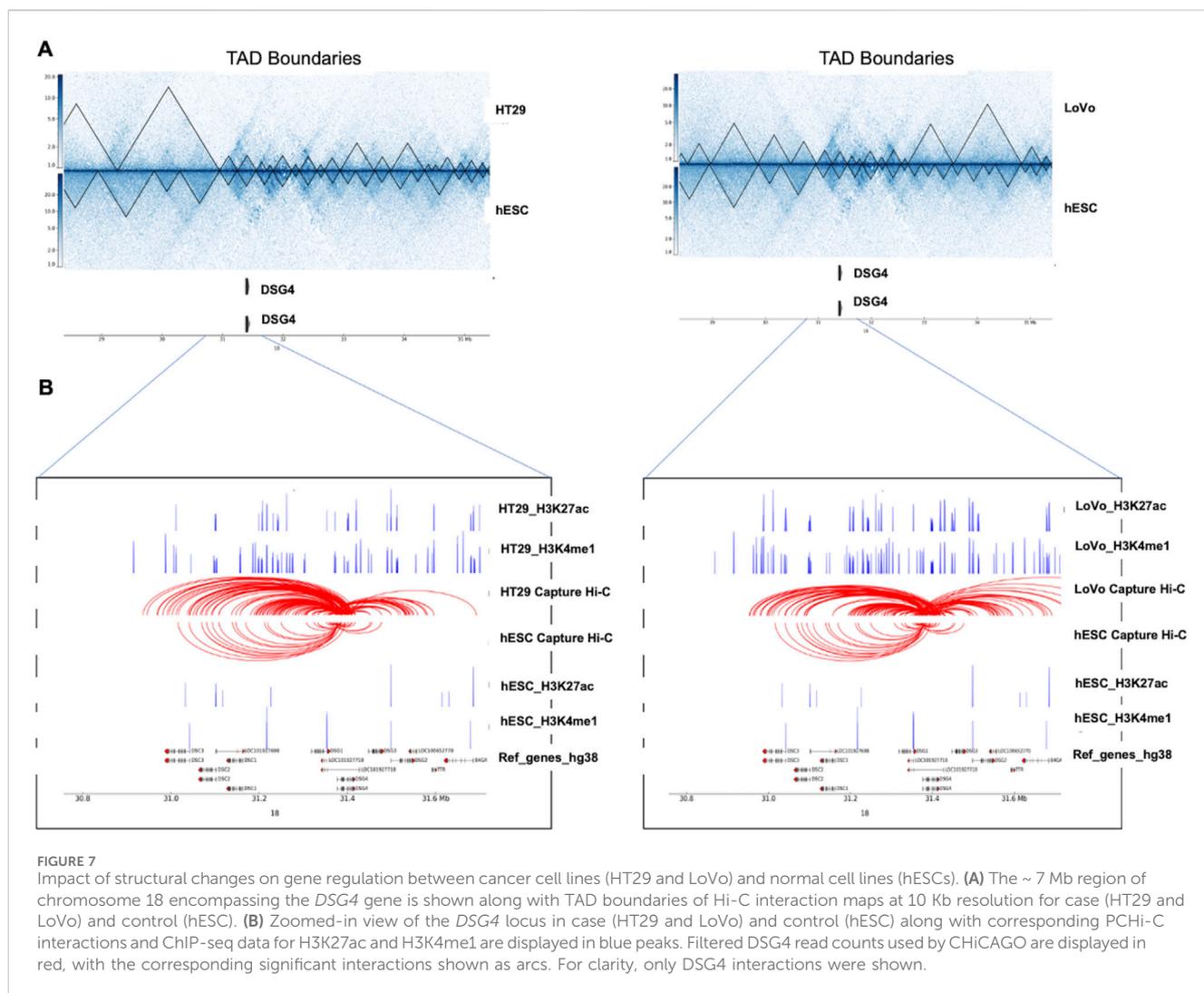
TAD boundaries. Taken together, these findings provided insights into the role of regulatory mechanisms in structural instability in CRC compared to normal cells.



### 3.4 Gene ontology, network analysis, and mutational study of colorectal-specific genes in the vicinity of disrupted TAD boundaries

Functional analysis revealed that most of the genes found at the disrupted TAD boundaries were involved in chromatin remodeling, chromatin assembly, and chromatin organization (Table 1). Network analysis revealed that SWI/SNF regulatory networks and histone family clusters were significantly associated with chromatin remodeling, chromatin assembly, and chromatin organization (Figure 6A). The results of enriched Gene Ontology and pathway analyses for both HT29 and LoVo cell lines are provided in Supplementary Table S4. Figure 6B depicts a volcano plot showing the list of the affected genes that were involved in chromatin remodeling, chromatin assembly, and chromatin organization expressed in cancer cells. The heatmap and

normalized count of the list of affected genes are listed in Supplementary Figure S1. Furthermore, we integrated the mutational datasets with the affected genes due to TAD changes and examined those affected genes that were flagged as oncogenes in the CRC and COSMIC databases (Tate et al., 2019; Chisanga et al., 2016). Mutational analysis revealed that ARID1A, ATRX, centromere protein complex, histone complex, chromatin organizers SATB1 and SATB2, SWI1/SNF1 complex, and histone deacetylases were frequently mutated in this dataset. Approximately 585 genes exhibited mutations, specifically in LoVo and HT29 cells, in which nonsynonymous and missense mutations were prominent. Supplementary Figures S2A, B provide the frequency of mutations in terms of CDs and nucleotide changes in genes associated with chromatin-associated genes collected from Gene Ontology analysis, which lie in TAD disruption regions according to Tate et al. (2019) and Chisanga et al. (2016). Supplementary Figures S2C, E list the various types of mutations in genes collected from the



database (Chisanga et al., 2016), which lie in the TAD disruption regions. We performed a comparative analysis with the list of chromatin remodeling-associated genes with relevant literature related to CRC (Chisanga et al., 2016; Chubb et al., 2016; Rudd et al., 2005). We obtained 33, 41, and 106 common genes when comparing our study with the previous CRC studies (Chisanga et al., 2016; Chubb et al., 2016; Rudd et al., 2005), respectively (Supplementary Figure S3). Finally, we obtained 10 unique genes that were commonly affected in the cited CRC studies (Chisanga et al., 2016; Chubb et al., 2016; Rudd et al., 2005) and in our study. These genes are TOP1MT, HELLS, AURKB, CHEK1, SMARCC1, TTF1, HDAC1, SCM1, RERE, and PADI2, out of which HDAC1, RERE, and PADI2 are known oncogenes (Liu et al., 2017) (Supplementary Figure S3). These findings suggest that the shifting of TAD boundaries affects gene regulation in CRC.

### 3.5 Distribution of the captured promoters' interactions with candidate cis-regulatory elements derived from ENCODE data

As mentioned above, we used ChIcAGO, a PChI-C interaction tool, to identify bait interactions with other-end

genomic loci using PChI-C data (Figure 3). We only selected bait IDs having a non-zero bait count and bait count differences greater than or equal to 10. Following the abovementioned criteria, we obtained 115,198 and 630,095 bait interactions with other end (OE) regions of hESC and HT29 in the hESC/HT29 sample (Supplementary Figure S4A). Similarly, 112,383 and 554,536 bait interactions with OE regions of hESC and LoVo in the hESC/LoVo sample (Supplementary Figure S4B). For the sake of convenience, we referred to the above process as STAGE 1. Again, we filtered the bait IDs from STAGE1 that had common bait IDs generated in the previous section (Figure 4A). We referred to this filtered interaction as STAGE 2. Finally, we filtered bait IDs of STAGE 2 that have common bait IDs with those of the affected genes due to TAD disruption (Figure 5B). The bait IDs of the list of affected genes can be filtered from bait map IDs. The final bait interactions file is referred to as STAGE 3. We overlapped the OE region of STAGE 3 interaction files with regulatory elements such as promoter-like (PLS), proximal enhancer-like (pELS), distal enhancer-like (dELS), DNase-H3K4me3, and CTCF regions (<https://screen-v2.wenglab.org>) (ENCODE Project Consortium et al., 2020). The stage-wise details of the total number of bait interactions with regulatory elements of hESC/HT29 and hESC/

LoVo are tabulated in [Supplementary Table S5](#). It is clearly evident that there are high proximal enhancer-like (pELS) and distal enhancer-like (dELS) interactions with promoter-capture counts, compared to those with the remaining cis-regulatory interactions ([Figures 6C,D](#)).

### 3.6 Role of enhancer–promoter interactions within TAD disruption regions in identifying disease-relevant target genes in CRC

The integration of PChi-C with Hi-C was performed to explore the genomic landscape from the loop level to the TAD-level. This was then combined with differentially expressed genes from the TCGA database and from HT29 (case) *versus* NT2D1 (control) cell lines was to understand gene regulation. We also overlapped these data with histone modification marks (H3K4me1 and H3K27ac) to verify the promoter interaction with the enhancer, which overall mimics the impact of TAD disruption on a gene's expression. These could serve as key genes for potential therapeutic targets. Based on the capture count and log2fold value, we selected a few statistically significant DE genes that lie in the TAD disruption regions. The statistically significant gene list includes long non-coding genes (MALAT1, NEAT1, FTX, and PVT1), small nucleolar genes (SNORA26 and SNORA71A), and protein-coding genes (TMPRSS11D, TSPEAR, and DSG4). In this study, statistical significance refers to selecting the top-hit genes based on high capture counts, which have enhancer–promoter loops, lie within TAD disruption regions, and are upregulated genes in CRC. For the visualization of Hi-C and PChi-C interaction maps in the context of gene regulation, we highlighted one of the statistically significant upregulated gene loci, *DSG4*, in [Figures 7A,B](#), with the remaining loci in [Supplementary Figure S5-11](#). The *DSG4* gene lies within the TAD disruption region in case cells (HT29 and LoVo) compared to control cells (hESCs), as identified by Hi-C analysis. Similarly, PChi-C detected a higher frequency of interactions between the *DSG4* promoter and histone modification marks H3K27ac and H3K4me1 in case cells (HT29 and LoVo) compared to the control cells (hESCs). TAD-based analysis helps in defining a gene's cis-regulatory landscape, while high-resolution promoter interaction data provide the necessary precision for accurately mapping the enhancer–promoter interactions. There were more active enhancer marks and enhancer–promoter loop formations in the case cells than in the control cells in the vicinity of the *DSG4* gene locus, suggesting its upregulation ([Figures 7A,B](#)). PChi-C techniques employed across various cell types detected numerous enhancer–promoter interactions, uncovering significant differences in promoter structure among these cell types and during differentiation ([Schoenfelder et al., 2015; Mifsud et al., 2015; Javierre et al., 2016; Freire-Pritchett et al., 2017; Rubin et al., 2017; Siersbæk et al., 2017; Montefiori et al., 2018](#)). The variation in enhancer–promoter interactions within TAD disruption regions ([Figures 7A,B](#)) suggests that disease-relevant cell types (such as HT29 and LoVo) serve as tractable models for effectively probing the gene regulatory mechanisms associated with disease loci.

### 3.7 Validation of the dysregulated genes revealed putative biomarkers

As discussed above, using our combined Hi-C, promoter-capture Hi-C, RNA-seq, and ChIP-seq analysis, we have uncovered new potential targets that exhibit a high degree of dysregulation in CRC datasets with respect to the 3D structure. To validate whether the predictions of our analysis hold true, we next sought to monitor the expression profiles of a few of the statistically significant hits from our dataset in CRC models, some of which were previously reported to be dysregulated in CRC. lncRNA PVT1 has been implicated in the progression of colorectal cancer via the VEGFA-AKT axis ([Wu et al., 2020](#)). lncRNA NEAT1 modulates chromatin accessibility in CRC and directly regulates Myc and ALDH1 ([Zhu et al., 2020](#)). Similarly, recent studies ([Zhao et al., 2020; Yang et al., 2018; Jia et al., 2022; Guo et al., 2015](#)) have reported that lncRNA FTX plays a crucial role in the initiation and progression of CRC. We used a highly malignant colorectal cell line HT29 and compared it to an embryonic cell line NT2D1 in our experimental validations to closely represent the primary datasets used for the bioinformatics analysis. We selected our targets, in which the contact frequency was at least two-fold. We observed that expression levels of genes belonging to long non-coding genes (MALAT1, NEAT1, FTX, and PVT1), small nucleolar genes (SNORA26 and SNORA71A), and protein-coding families (TMPRSS11D, TSPEAR, and DSG4) were significantly higher in HT29 and various other colorectal model cell lines than in NT2D1 embryonic cells ([Figure 8A](#)). Next, we monitored the relative levels of transcriptional activation-associated chromatin modifications H3K4me3 and H3K27ac at the promoters of these genes by performing ChIP-quantitative PCRs. We observed a significant enrichment of both these modifications at the promoters of all selected target genes ([Figures 8B,C](#)), suggesting a hierarchical modification of the chromatin following TAD disruption. Collectively, these analyses suggest that changes in transcription-associated chromatin modifications correlated with the 3D chromatin changes shown previously, leading to the activation of the transcription machinery.

The integrative analysis reported above is based on bulk-seq data. In bulk-seq studies, samples are treated as a homogeneous population. However, the cell populations in the human body are heterogeneous in nature, and each cell reflects unique activity. Due to recent breakthroughs, it is now possible to analyze the transcriptome at the single-cell level for millions of cells ([Stuart and Satija, 2019; Khaliq et al., 2022](#)). This enabled us to differentiate, characterize, and classify each cell at the transcriptome level, which leads to the prediction of rare cell populations. Taking advantage of the single-cell human colon cancer atlas database ([Pelka et al., 2021](#)), we analyzed transcriptionally profiled 371,223 tumor and adjacent normal cells generated based on scRNA-seq. A detailed analysis is provided in [Supplementary Material M1](#). The purpose of the single-cell RNA-seq analysis is to assess the enrichment of the potential targeted genes (MALAT1, NEAT1, FTX, PVT1, SNORA26, SNORA71A, TMPRSS11D, TSPEAR, and DSG4) at the single-cell level. We did not observe a significant enrichment score at the single-cell level despite these genes exhibiting a high degree of dysregulation at the bulk-seq level.

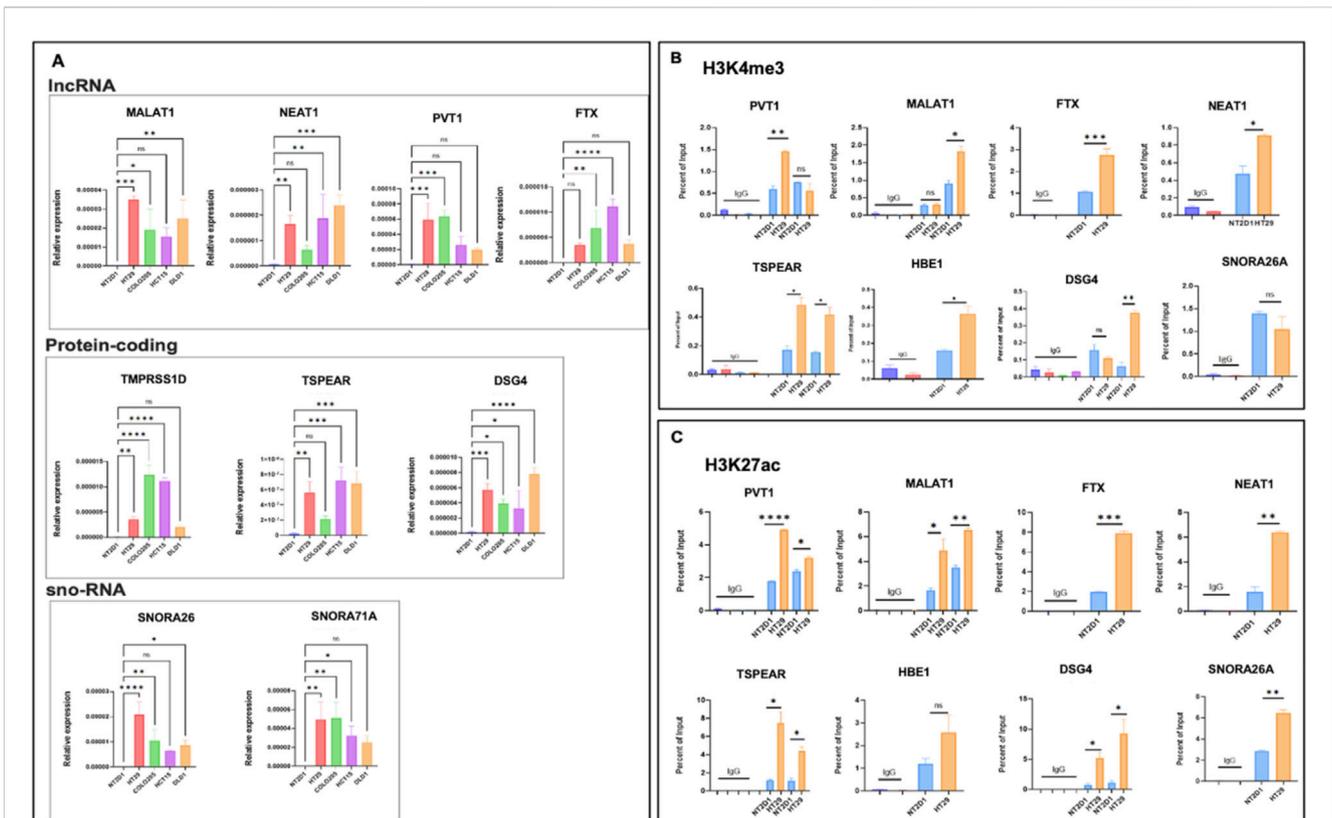


FIGURE 8

Validation of target genes extracted from integrative analysis. **(A)** qRT-PCR analysis of target genes in NT2D1, HT29, Colo205, HCT15, and DLD1 cell lines, showing relative enrichment of gene expression in the CRC cell lines compared to NT2D1 cells. **(B, C)** ChIP-qPCR for H3K4me3 and H3K27ac modifications, respectively, on the transcription start site (TSS) upstream regions of target genes. The cell lines used are indicated below the bars. Three biological replicates were used for statistical analysis using GraphPad v9.1. A one-way ANOVA or unpaired t-test was used as the significance test.

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ , and ns, non-significant.

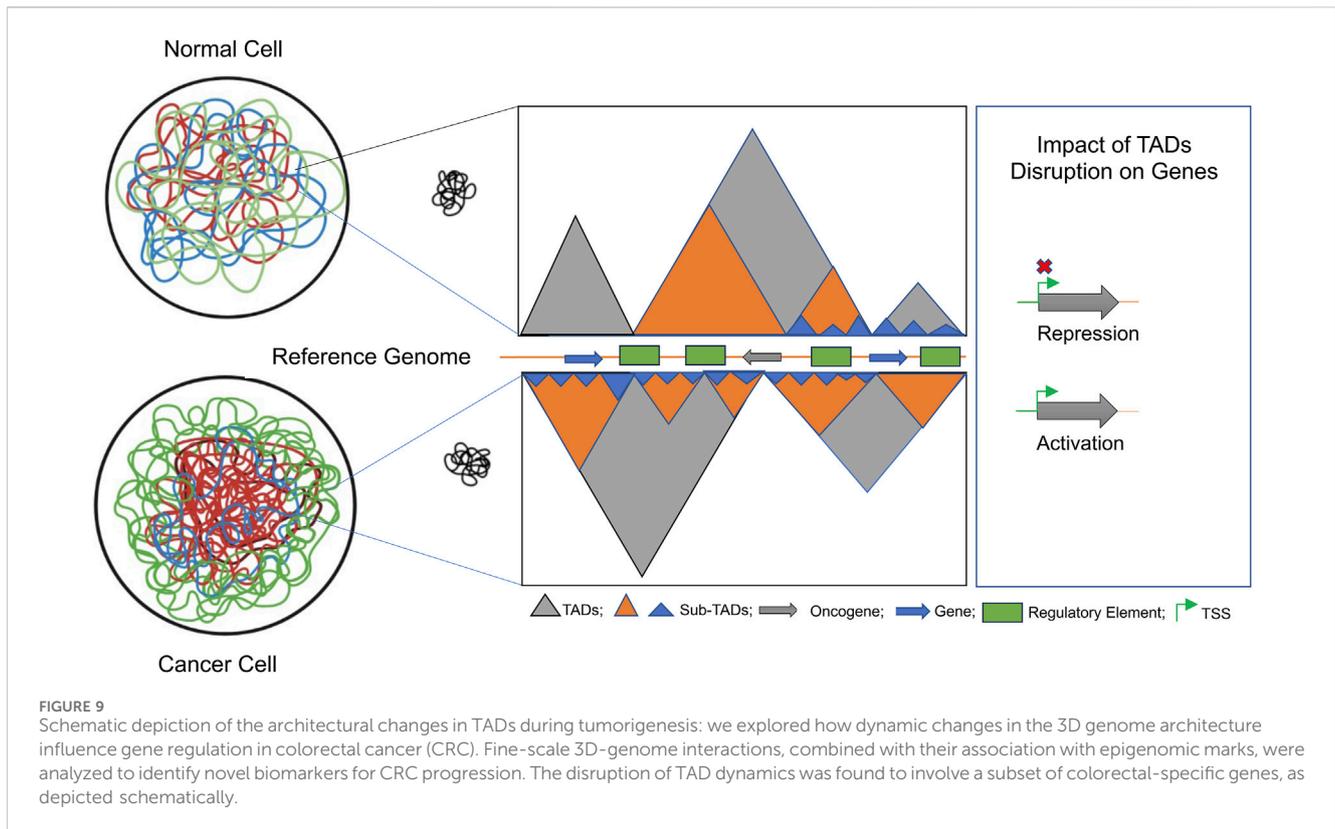
## 4 Discussion

TADs are increasingly recognized as pivotal elements of 3D chromosome organization and gene regulation. Our objective was to understand the communication between enhancer-promoter interactions and impact on gene expression. In this study, we investigated how the abnormalities in 3D chromosome folding affect gene expression patterns in colorectal cancer. The instability of cancer genomes often results in a higher frequency of TAD disruption in cancer cells compared to their normal counterparts. Such disruptions could trigger significant alterations in gene expression, potentially contributing to the initiation and progression of tumorigenesis. Our comprehensive analysis attempted to explore the molecular basis of the genetic risk for CRC. The configurations of TADs in cancer cells could thus offer valuable insights into the genetic underpinnings of gene expression in these cells, potentially identifying specific targets for therapeutic interventions. In our study, we focused on identifying multiple potential therapeutic targets for colorectal cancer by systematically integrating 3D datasets with 1D datasets. We employed publicly available datasets including Hi-C, PChI-C, ChIP-seq, RNA-seq, and single-cell RNA-seq data towards this comprehensive analysis. Generating high-resolution Hi-C data through Hi-C experiments requires the use of millions of

mammalian cells, resulting in the generation of billions of paired-end reads. Because of the substantial costs of sequencing, most Hi-C datasets exhibit a relatively low resolution, which might not be suitable for examining detailed interactions in the vicinity of regions of interest, such as disease-associated loci.

To overcome this limitation, we integrated PChI-C data with Hi-C to understand the dynamics of regulatory chromosomal interactions from local-to-TAD level. This approach enabled us to investigate the structural organization and changes in chromatin interactions across different cellular contexts. During our analysis, we observed that most interactions were found in cancer cell lines (HT29, LoVo) compared to hESCs. This finding suggests a higher frequency of structural alterations in tumor cells, highlighting the extensive chromatin remodeling and genomic instability associated with tumor cell population in contrast to human embryonic stem cells. Such modifications may contribute to cancer-specific transcriptional programs by reshaping regulatory interactions.

TADs exhibit evolutionary conservation and serve as crucial elements in governing and facilitating long-range regulation of gene expression<sub>67</sub>. They act as self-contained domains that confine enhancer-promoter interactions, ensuring precise gene expression control. When a TAD boundary is deleted, it can lead to the fusion of two adjacent TADs, resulting in the misregulation of genes previously restricted to separate regulatory environments.



Genomic rearrangements that fragment existing TADs and create novel regulatory domains, while preserving the original TAD boundaries. The disruption of TAD boundaries can lead to abnormal gene expression patterns by exposing genes to inappropriate regulatory elements. This disruption is frequently observed in cancer cells, where alterations in the 3D chromatin architecture drive oncogenic transcriptional programs and plays a crucial role in tumorigenesis<sub>68–70</sub> (Figure 9). Comprehending the significance of TAD disruption and long-range chromatin interactions is essential for gaining insights into the broader mechanisms of gene regulation. Additionally, it sheds light on how genomic rearrangements and mutations in cancer genomes can result in the atypical expression of oncogenes and tumor suppressors. This knowledge contributes not only to our understanding of gene regulation in a general context but also to our comprehension of the molecular underpinnings of cancer progression.

For gene expression analysis, we performed differential RNA-seq analysis using colorectal cancer samples (primary tumor, metastatic) compared with normal samples (blood-derived normal, solid tissue) obtained from the TCGA databases. We filtered differentially expressed genes found around variably interacting regions to check the distribution of DGEs across TADs boundaries. We used H3K27ac and H3K4me1 histone modification marks for understanding the gene's cis-regulatory landscape to precisely map the enhancer-promoter interactions. The functional study revealed that most of the genes upregulated in cancer were found around TAD disruption regions, and are involved in chromatin remodeling, chromatin assembly and chromatin organization. We also integrated the mutational

datasets collected from databases like cosmic database<sub>52,53</sub> and human oncogene database<sub>56</sub> with the affected genes due to TAD disruptions. Results of such analyses revealed that genes encoding key chromatin associated regulators including AR1D1A, ATRX, centromere protein complex, histone complex, SATB1, SATB2, SWI1/SNF1 complex, histone deacetylases were frequently mutated in these datasets.

Finally for experimental validation, we filtered a statistically significant gene-list based on high capture count and log<sub>2</sub> fold change in transcript expression; and with relevant literature related to colorectal cancer. We found expression of long non-coding genes (MALAT1, NEAT1, FTX and PVT1), small nucleolar (SNORA26 and SNORA71A) and protein-coding (TMPRSS11D, TSPEAR and DSG4) were significantly much higher in HT29 compared to NT2D1; and enriched with activation-associated histone modifications H3K4me3 and H3K27ac. Our data, while confirming the expression patterns, strongly argues in favor of a priori involvement of epigenomic mechanisms in form of dynamic histone modifications as well as the chromatin interaction changes that modulate the gene expression.

Additionally, for understanding the expression patterns of target genes across diverse cell types, we analyzed 371,223 tumor and adjacent normal cells taken from the single cell human colon cancer atlas database. We found expression of FTX and PVT1 is contributed from the rare cell population while MALAT1 and NEAT1 were expressed almost everywhere in diverse cell populations. However, expression of TSPEAR and DSG4 was not observed in the scRNA-seq analysis. Collectively, our study suggests that changes in 3-D genomic architecture affect transcriptome signatures which might be associated with tumor-suppressive transcriptional programs (Figure 9). Such integrative analysis

resulted in identification of multiple target genes which may potentially serve as biomarkers and could be used to better understand CRC progression.

## Data availability statement

The dataset analyzed in the study is publicly available. The accession number can be found in the article/Supplementary Material.

## Author contributions

AS: conceptualization, data curation, formal analysis, investigation, methodology, writing–original draft, and writing–review and editing. AM: data curation, formal analysis, investigation, methodology, validation, and writing–review and editing. AB: data curation, formal analysis, investigation, methodology, and writing–review and editing. SN: conceptualization, formal analysis, investigation, methodology, supervision, and writing–original draft. SG: conceptualization, funding acquisition, methodology, project administration, resources, supervision, writing–original draft, and writing–review and editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. AKS was funded by the Science and Engineering Research Board (SERB), Government of India, through the National Postdoctoral Fellowship (N-PDF) (PDF/2019/000503). SG was funded by SERB under the JC Bose Fellowship (JCB/2019/000013).

## Acknowledgments

AKS acknowledges SERB for National Postdoctoral Fellowship (N-PDF). AM acknowledges fellowship from the CSIR, Government

## References

- Alzahrani, S. M., Al Doghaither, H. A., and Al-Ghafari, A. B. (2021). General insight into cancer: An overview of colorectal cancer (Review). *Molecular and Clinical Oncology*, *Mol. Clin. Oncol.* 15, 271. doi:10.3892/mco.2021.2433
- Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* 20, 163–172. doi:10.1038/s41590-018-0276-y
- Beliveau, B. J., Joyce, E. F., Apostolopoulos, N., Yilmaz, F., Fonseka, C. Y., McCole, R. B., et al. (2012). Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci.* 109, 21301–21306. doi:10.1073/pnas.1213818110
- Bienko, M., Crossetto, N., Teytelman, L., Klemm, S., Itzkovitz, S., and van Oudenaarden, A. (2013). A versatile genome-scale PCR-based pipeline for high-definition DNA FISH. *Nat. Methods* 10, 122–124. doi:10.1038/nmeth.2306
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 68, 394–424. doi:10.3322/caac.21492
- Cairns, J., Freire-Pritchett, P., Wingett, S. W., Várnai, C., Dimond, A., Plagnol, V., et al. (2016). CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* 17, 127. doi:10.1186/s13059-016-0992-2
- Cairns, J., Orchard, W. R., Malysheva, V., and Spivakov, M. (2019). Chicdiff: a computational pipeline for detecting differential chromosomal interactions in Capture Hi-C data. *Bioinformatics* 35, 4764–4766. doi:10.1093/bioinformatics/btz450
- Chisanga, D., Keerthikumar, S., Pathan, M., Ariyaratne, D., Kalra, H., Boukouris, S., et al. (2016). Colorectal cancer atlas: an integrative resource for genomic and proteomic annotations from colorectal cancer cell lines and tissues. *Nucleic Acids Res.* 44, D969–D974. doi:10.1093/nar/gkv1097
- Chubb, D., Broderick, P., Dobbins, S. E., Frampton, M., Kinnersley, B., Penegar, S., et al. (2016). Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. *Nat. Commun.* 7, 11883. doi:10.1038/ncomms11883
- Collombet, S., Ranisavljevic, N., Nagano, T., Varnai, C., Shisode, T., Leung, W., et al. (2020). Parental-to-embryo switch of chromosome organization in early embryogenesis. *Nature* 580, 142–146. doi:10.1038/s41586-020-2125-z
- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311. doi:10.1126/science.1067799
- Di Stefano, M., Stadhouders, R., Farabella, I., Castillo, D., Serra, F., Graf, T., et al. (2020). Transcriptional activation during cell reprogramming correlates with the formation of 3D open chromatin hubs. *Nat. Commun.* 11, 2564. doi:10.1038/s41467-020-16396-1

of India. SN acknowledges ICR/MPG-ICR, EGA for sharing the CRC data. The support and the resources provided by ‘PARAM Brahma Facility’ under the National Supercomputing Mission, Government of India at the Indian Institute of Science Education and Research (IISER) Pune are gratefully acknowledged. SG acknowledges SERB, Government of India for the JC Bose Fellowship (JCB/2019/000013).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1553469/full#supplementary-material>

- Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature* 518, 331–336. doi:10.1038/nature14222
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. doi:10.1038/nature11082
- Djehoul, D., Patel, B., Kramer, H., Dimond, A., Whilding, C., Brown, K., et al. (2020). Identifying proteins bound to native mitotic ESC chromosomes reveals chromatin repressors are important for compaction. *Nat. Commun.* 11, 4118. doi:10.1038/s41467-020-17823-z
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* 29, 15–21. doi:10.1093/bioinformatics/bts635
- Du, Z., Zheng, H., Kawamura, Y. K., Zhang, K., Gassler, J., Powell, S., et al. (2020). Polycomb group proteins regulate chromatin architecture in mouse oocytes and early embryos. *Mol. Cell* 77, 825–839. doi:10.1016/j.molcel.2019.11.011
- ENCODE Project Consortium Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shores, N., Adrian, J., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 583, 699–710. doi:10.1038/s41586-020-2493-4
- Freire-Pritchett, P., Ray-Jones, H., Della Rosa, M., Eijsbouts, C. Q., Orchard, W. R., Wingett, S. W., et al. (2021). Detecting chromosomal interactions in Capture Hi-C data with CHiCAGO and companion tools. *Nat. Protoc.* 16, 4144–4176. doi:10.1038/s41596-021-00567-5
- Freire-Pritchett, P., Schoenfelder, S., Várnai, C., Wingett, S. W., Cairns, J., Collier, A. J., et al. (2017). Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. *eLife* 6, e21926. doi:10.7554/eLife.21926
- Fukuda, K., Shimura, C., Miura, H., Tanigawa, A., Suzuki, T., Dohmae, N., et al. (2021). Regulation of mammalian 3D genome organization and histone H3K9 dimethylation by H3K9 methyltransferases. *Commun. Biol.* 4, 571. doi:10.1038/s42003-021-02089-y
- Guo, X.-B., Hua, Z., Li, C., Peng, L. P., Wang, J. S., Wang, B., et al. (2015). Biological significance of long non-coding RNA FTX expression in human colorectal cancer. *International journal of clinical and experimental medicine*, 8(9), 15591–15600.
- Hildebrand, E. M., and Dekker, J. (2020). Mechanisms and functions of chromosome compartmentalization. *Trends biochem. Sci.* 45, 385–396. doi:10.1016/j.tibs.2020.01.002
- Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., et al. (2016). Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* 167, 1369–1384.e19. doi:10.1016/j.cell.2016.09.037
- Jia, R., Song, L., Fei, Z., Qin, C., and Zhao, Q. (2022). Long noncoding RNA Ftx regulates the protein expression profile in HCT116 human colon cancer cells. *Proteome Sci.* 20, 7. doi:10.1186/s12953-022-00187-1
- Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., et al. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 503, 290–294. doi:10.1038/nature12644
- Khalid, A. M., Erdogan, C., Kurt, Z., Turgut, S. S., Grunvald, M. W., Rand, T., et al. (2022). Refining colorectal cancer classification and clinical stratification through a single-cell atlas. *Genome Biol.* 23, 113. doi:10.1186/s13059-022-02677-z
- Klein, K. N., Zhao, P. A., Lyu, X., Sasaki, T., Bartlett, D. A., Singh, A. M., et al. (2021). Replication timing maintains the global epigenetic state in human cells. *Science* 372, 371–378. doi:10.1126/science.aba5545
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923
- Law, P. J., Timofeeva, M., Fernandez-Rozadilla, C., Broderick, P., Studd, J., Fernandez-Tajes, J., et al. (2019). Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat. Commun.* 10, 2154. doi:10.1038/s41467-019-09775-w
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinforma.* 12, 323. doi:10.1186/1471-2105-12-323
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinforma. Oxf. Engl.* 25, 2078–2079. doi:10.1093/bioinformatics/btp352
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi:10.1126/science.1181369
- Liu, Y., Sun, J., and Zhao, M. (2017). ONGene: a literature-based database for human oncogenes. *J. Genet. Genomics* 44, 119–121. doi:10.1016/j.jgg.2016.12.004
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8
- Love, M. I., Soneson, C., Hickey, P. F., Johnson, L. K., Pierce, N. T., Shepherd, L., et al. (2020). Tximeta: reference sequence checkups for provenance identification in RNA-seq. *PLoS Comput. Biol.* 16, e1007664. doi:10.1371/journal.pcbi.1007664
- Lupiáñez, D. G., Spielmann, M., and Mundlos, S. (2016). Breaking TADs: how alterations of chromatin domains result in disease. *Trends Genet. TIG* 32, 225–237. doi:10.1016/j.tig.2016.01.003
- Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* 47, 598–606. doi:10.1038/ng.3286
- Montefiori, L. E., Sobreira, D. R., Sakabe, N. J., Aneas, I., Joslin, A. C., Hansen, G. T., et al. (2018). A promoter interaction map for cardiovascular disease genetics. *eLife* 7, e35788. doi:10.7554/eLife.35788
- Morgan, M. T., and Davis, S. R. (2017). GenomicDataCommons: a bioconductor interface to the NCI genomic data commons. 117200 preprint at. doi:10.1101/117200
- Muzny, D. M., et al. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330–337. doi:10.1038/nature11252
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi:10.1038/nmeth.4197
- Patta, I., Madhok, A., Khare, S., Gottimukkala, K. P., Verma, A., Giri, S., et al. (2020). Dynamic regulation of chromatin organizer SATB1 via TCR-induced alternative promoter switch during T-cell development. *Nucleic Acids Res.* 48, 5873–5890. doi:10.1093/nar/gkaa321
- Pelham-Webb, B., Murphy, D., and Apostolou, E. (2020). Dynamic 3D chromatin reorganization during establishment and maintenance of pluripotency. *Stem Cell Rep.* 15, 1176–1195. doi:10.1016/j.stemcr.2020.10.012
- Pelka, K., Hofree, M., Chen, J. H., Sarkizova, S., Pirl, J. D., Jorgji, V., et al. (2021). Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* 184, 4734–4752.e20. doi:10.1016/j.cell.2021.08.003
- Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K. C., Grüning, B. A., Villaveces, J., et al. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* 9, 189. doi:10.1038/s41467-017-02525-w
- Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. doi:10.1016/j.cell.2014.11.021
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi:10.1093/bioinformatics/btp616
- Rubin, A. J., Barajas, B. C., Furlan-Magaril, M., Lopez-Pajares, V., Mumbach, M. R., Howard, I., et al. (2017). Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* 49, 1522–1528. doi:10.1038/ng.3935
- Rudd, M. F., Williams, R. D., Webb, E. L., Schmidt, S., Sellick, G. S., and Houlston, R. S. (2005). The predicted impact of coding single nucleotide polymorphisms database. *Cancer Epidemiol. Biomarkers Prev.* 14, 2598–2604. doi:10.1158/1055-9965.EPI-05-0469
- Sati, S., Bonev, B., Szabo, Q., Jost, D., Bensedoun, P., Serra, F., et al. (2020). 4D genome rewiring during oncogene-induced and replicative senescence. *Mol. Cell* 78, 522–538. doi:10.1016/j.molcel.2020.03.007
- Schmitt, A. D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C. L., et al. (2016). A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell reports*. 17 (8), 2042–2059. doi:10.1016/j.celrep.2016.10.061
- Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B. M., et al. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* 25, 582–597. doi:10.1101/gr.185272.114
- Siersbæk, R., Madsen, J. G. S., Javierre, B. M., Nielsen, R., Bagge, E. K., Cairns, J., et al. (2017). Dynamic rewiring of promoter-anchored chromatin loops during adipocyte differentiation. *Mol. Cell* 66, 420–435. doi:10.1016/j.molcel.2017.04.010
- Spracklin, G., Abdennur, N., Imakaev, M., Chowdhury, N., Pradhan, S., Mirny, L., et al. (2021). Heterochromatin diversity modulates genome compartmentalization and loop extrusion barriers. doi:10.1101/2021.08.05.455340
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., et al. (2019). Comprehensive integration of single-cell data. *Cell* 177, 1888–1902. doi:10.1016/j.cell.2019.05.031
- Stuart, T., and Satija, R. (2019). Integrative single-cell analysis. *Nat. Rev. Genet.* 20, 257–272. doi:10.1038/s41576-019-0093-7
- Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B., and Zhuang, X. (2020). Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell* 182, 1641–1659. doi:10.1016/j.cell.2020.07.032
- S, W., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., et al. (2015). HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research* 4, 1310. doi:10.12688/f1000research.7334.1
- Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., et al. (2019). COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 47, D941–D947. doi:10.1093/nar/gky1015

- Valton, A.-L., and Dekker, J. (2016). TAD disruption as oncogenic driver. *Curr. Opin. Genet. Dev.* 36, 34–40. doi:10.1016/j.gde.2016.03.008
- Van Der Maaten, L. (2014). Accelerating t-SNE using tree-based algorithms. *The journal of machine learning research.* 15(1), 3221–3245.
- van Schaik, T., Vos, M., Peric-Hupkes, D., Hn Celie, P., and van Steensel, B. (2020). Cell cycle dynamics of lamina-associated DNA. *EMBO Rep.* 21, e50636. doi:10.15252/embr.202050636
- Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D. T., Tanay, A., et al. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* 10, 1297–1309. doi:10.1016/j.celrep.2015.02.004
- von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003). STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* 31, 258–261. doi:10.1093/nar/gkg034
- Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., et al. (2018). Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* 46, W11–W16. doi:10.1093/nar/gky504
- Wolff, J., Rabbani, L., Gilsbach, R., Richard, G., Manke, T., Backofen, R., et al. (2020). 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. doi:10.1093/nar/gkaa220
- Wu, H., Wei, M., Jiang, X., Tan, J., Xu, W., Fan, X., et al. (2020). lncRNA PVT1 promotes tumorigenesis of colorectal cancer by stabilizing miR-16-5p and interacting with the VEGFA/VEGFR1/AKT Axis. *Mol. Ther. Nucleic Acids* 20, 438–450. doi:10.1016/j.omtn.2020.03.006
- Yang, Y., Zhang, J., Chen, X., Xu, X., Cao, G., Li, H., et al. (2018). lncRNA FTX sponges miR-215 and inhibits phosphorylation of vimentin for promoting colorectal cancer progression. *Gene Ther.* 25, 321–330. doi:10.1038/s41434-018-0026-7
- Zhao, K., Ye, Z., Li, Y., Li, C., Yang, X., Chen, Q., et al. (2020). lncRNA FTX contributes to the progression of colorectal cancer through regulating miR-192-5p/eif5a2 Axis. *OncoTargets Ther.* 13, 2677–2688. doi:10.2147/OTT.S241011
- Zhao, Z., Tavoosidana, G., Sjölander, M., Göndör, A., Mariano, P., Wang, S., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* 38, 1341–1347. doi:10.1038/ng1891
- Zhu, Y., Hu, H., Yuan, Z., Zhang, Q., Xiong, H., Hu, Z., et al. (2020). lncRNA NEAT1 remodels chromatin to promote the 5-Fu resistance by maintaining colorectal cancer stemness. *Cell Death Dis.* 11, 962. doi:10.1038/s41419-020-03164-8