Published online 9 May 2023

Nucleic Acids Research, 2023, Vol. 51, No. 12 ***6143–6155***
https://doi.org/10.1093/nar/gkad361

Forkhead box protein D2 suppresses colorectal cancer by reprogramming enhancer interactions

Hyo-Min Kim [1], Byunghee Kang [1], Sohyun Park [1], Hyorim Park [1], Chan Johng Kim [1], Hyeonji Lee [1], Mijoung Yoo [1], Mi-Na Kweon [2], Sin-Hyeog Im [1,3,4], Tae Il Kim [5] and Tae-Young Roh [6,7,*]

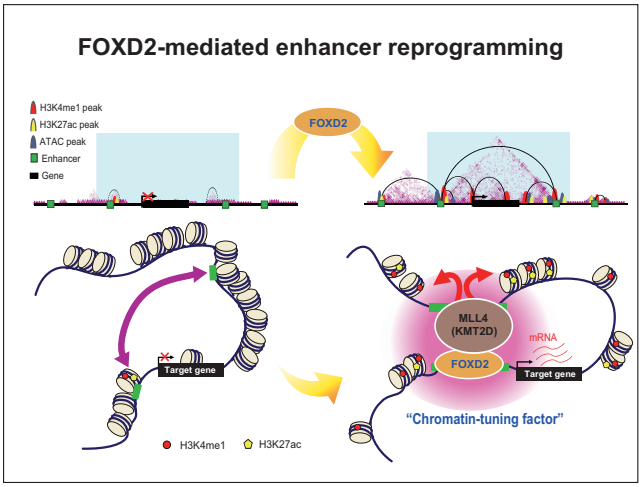
[1]Department of Life Sciences, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea, [2]Department of Convergence Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul 05505, Republic of Korea, [3]Institute of Convergence Science, Yonsei University, Seoul 03722, Republic of Korea, [4]ImmunoBiome, Inc., Pohang 37666, Republic of Korea, [5]Department of Internal Medicine, Institute of Gastroenterology, Yonsei University College of Medicine, Seoul 03722, Republic of Korea, [6]Department of Life Sciences, Ewha Womans University, Seoul 03760, Republic of Korea and [7]Sysgenlab Inc., Pohang 37673, Republic of Korea

Received March 15, 2022; Revised April 19, 2023; Editorial Decision April 20, 2023; Accepted April 26, 2023

ABSTRACT

Somatic stem cells contribute to normal tissue homeostasis, and their epigenomic features play an important role in regulating tissue identities or developing disease states. Enhancers are one of the key players controlling chromatin context-specific gene expression in a spatial and temporal manner while maintaining tissue homeostasis, and their dysregulation leads to tumorigenesis. Here, epigenomic and transcriptomic analyses reveal that forkhead box protein D2 (FOXD2) is a hub for the gene regulatory network exclusive to large intestinal stem cells, and its overexpression plays a significant role in colon cancer regression. FOXD2 is positioned at the closed chromatin and facilitates mixed-lineage leukemia protein-4 (MLL4/KMT2D) binding to deposit H3K4 monomethylation. *De novo* FOXD2-mediated chromatin interactions rewire the regulation of p53-responsive genes and induction of apoptosis. Taken together, our findings illustrate the novel mechanistic details of FOXD2 in suppressing colorectal cancer growth and suggest its function as a chromatin-tuning factor and a potential therapeutic target for colorectal cancer.

GRAPHICAL ABSTRACT



INTRODUCTION

Evolutionarily conserved forkhead box (FOX) proteins are transcription factors (TFs) characterized by a common DNA-binding domain, the so-called winged-helix domain. The FOX superfamily proteins are also known as pioneer factors binding to condensed chromatin and play important roles in diverse biological processes such as cell growth, development and differentiation (1–3). Moreover, the various modes of regulation and function of Fox proteins depend on the types of cells and tissues. In consequence, recent studies show that the somatic misregulation or mutation of Fox genes changes cell fate and is associated with cancer development (4).

*To whom correspondence should be addressed. Tel: +82 2 3277 3890; Email: tyroh@ewha.ac.kr

Cancers can be characterized by dysregulation of tissue-specific gene expression programs and epigenetic alterations, which results in cancer-promoting events without variations in DNA sequences (5). The epigenetic regulation mediated by the patterns of DNA methylation and histone modifications, for example, specify the transcriptional machinery to maintain cellular identity and respond to internal and external stimuli. Histone modifications such as H3K4me1 and H3K27ac are enriched in enhancer regions that typically serve as TF-binding sites to induce transcriptional activation. Moreover, H3K4me1 and H3K27ac can separate active enhancers from poised enhancers (6). However, these classifications and features are not generally applied to explain all types of tissues and states of diseases.

Since tissue-specific transcriptional machinery is essential to maintain the cellular fate, and the loss of cellular characteristics leads to cancer development, we investigated the possibility that TFs necessary for the determination of tissue fate might play a functional role in preventing cancer progression in a tissue-specific manner. An intestinal master TF, caudal type homeobox 2 (CDX2), plays a crucial role not only in intestinal stem cell (ISC) differentiation but also in cancer progression or regression (7–9). The involvement of CDX2 during cancer development demonstrates that differentiated cells are innately resistant to malignant transformation and that cellular reprogramming is required to promote malignancy. Considering the compartmentalization in the intestinal epithelium and the high prevalence of large intestine cancers, it is important to elucidate large intestine-specific TFs for the determination of tissue fate and tumor suppression in addition to CDX2 (10).

Accordingly, we analyzed the differential transcription machinery between small intestinal stem cells (SISCs) and large intestinal stem cells (LISCs). Then a cell-specific gene regulatory network (GRN) was established, which showed multiple tissue-specific TFs important for tissue homeostasis. FOXD2 was identified as a colon-specific TF hub in our GRN. FOXD2 is known to be functionally redundant in renal development and kidney hypoplasia, although its loss is not related to embryonic lethality (11). FOXD2 is also recognized as a functional TF regulating cAMP-dependent proteins and promoters in T lymphocytes (12). However, the precise molecular function and epigenetic mechanism of FOXD2 in the large intestine for maintaining a tissue identity and during cancer development remain elusive. Moreover, the TF-mediated enhancer landscape in colorectal cancer (CRC) is still unexplored. In this study, we show that colonic stem cell-enriched FOXD2 transactivates p53-responsive genes and induces apoptosis in human CRC cells by recruiting the mixed-lineage leukemia protein-4 (MLL4/KMT2D) gene specifically to reprogram enhancer and transcriptional machinery.

MATERIALS AND METHODS

Mice

Lgr5-EGFP-Ires-CreERT2$^{+/-}$ mice were used to study adult ISCs. All animal experiments were approved by the Animal Experimentation Committee of the Pohang University of Science and Technology (POSTECH).

Small and large intestinal crypt isolation

Crypt isolation was performed following previously described methods (13). The tissues were chopped into 2–3 cm pieces. Tissue fragments were suspended with 10 ml of cold ethylenediaminetetraacetate (EDTA) chelation buffer and placed on ice for 30 min. After incubation, the tissues were resuspended vigorously ∼10–15 times using a 10 ml pipette. The supernatant was collected and filtered with a 70 μm cell strainer to remove large fragments from the tissue. The remaining tissues were suspended several times with 10 ml of cold EDTA chelation buffer. All supernatants enriched with crypts were centrifuged at 200 *g* for 5 min and washed with cold 1× phosphate-buffered saline (PBS) buffer.

Intestinal stem cell preparation and flow cytometry

Small and large intestinal crypts were dissociated with TrypLE Express (Thermo Fisher Scientific, #12605028) containing Y-27632 (1 μM; Sigma-Aldrich, #Y0503) at 37°C. Dissociated cells were then filtered with a 70 μm cell strainer to remove cell clumps and washed twice with 10 ml of cold PBS buffer. Then, cells were centrifuged at 720 *g* for 5 min and the wash step with cold PBS buffer was repeated twice. After removal of the supernatant, cell pellets were resuspended with basal media and filtered in a 20 μm strainer (Falcon, #38030). Finally, small and large intestinal Lgr5-EGFP+/high cells were sorted by flow cytometry (MoFlo, Beckman Coulter and SONY-SH800S) for further analysis.

Cloning and lentivirus production for the stable cell line

For the overexpression of human FOXD2 transcript, the full-length coding sequence (CDS) from the *Foxd2* (NM_004474) human tagged open reading frame (ORF) clone (Origene, RC222086) was inserted in the pSin-EF2 vector (Addgene, #116877).

For lentivirus production, HEK293T cells were maintained at 70% confluency in 162 cm^2 culture plate in culture medium [Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin]. Then, DNA transfection solution containing 40 μg of total plasmid DNA was prepared by adding lentiviral packaging vectors (7 μg of pMD2.G; 13 μg of pCMV.dR 8.74) and 20 μg of lentiviral plasmid encoding the gene. Moreover, a polyethylenimine (PEI) transfection solution was prepared by adding 80 μl of 1 mg/ml PEI to 920 μl of opti-MEM (Gibco, 31985062) and incubating for 5 min at room temperature. HEK293T cells was treated with a total of 2 ml of the DNA transfection solution for 4 h. After 2 days of incubation, the culture medium was collected by centrifugation for 5 min at 1000 *g*. The culture medium was additionally collected a day later for maximal production of lentivirus. Finally, the supernatant was filtered with a 0.45 μm filter.

For establishing stable cell lines, pSin and pSin-FOXD2 lentivirus was infected into HCT116 cells. A total of 0.2 × 10^6 cells were split onto 6-well plates a day before the lentivirus treatment and incubated for 48–72 h for complete transduction. Then cells were selected and expanded in a culture medium containing puromycin (5 μg/ml).

Apoptosis and wound healing analysis

The control and FOXD2-expressing HCT116 cancer cells were washed twice with cold 1× PBS and analyzed with the FITC Annexin V Apoptosis Detection Kit with PI (BioLegend, 640914) according to the manufacturer's instructions. The scratch assay was performed to assess the cell migration of control and FOXD2-overexpressing (OE) HCT116 cells. The wound area was measured every 12 h for 2 days following the previously described method (14).

Immunofluorescence staining

The samples were cultured in 8-well chamber slides (MERK, C7182) and fixed with 4% paraformaldehyde. After washing three times with cold PBS, the fixed cells were incubated in 0.4% Triton X-100/1× PBS for 30 min at 25°C, blocked in 5% bovine serum albumin (BSA)/1× PBS for 1 h at 25°C and incubated overnight at 4 °C with the following primary antibodies: Ki67 (1:50, Thermo Fisher Scientific #14–5698-82), Caspase 3 (1:50, Thermo Fisher Scientific #PA5-77887) and Caspase 9 (1:50, Thermo Fisher Scientific #PA5-16358). The chamber wells were washed five times with 1× PBS and incubated with 1:1000 Alexa Flour 488- or 647-conjugated secondary antibodies for 1 h at 25°C. After washing five times with 1× PBS, Hoechst 33342 (10 μg/ml in 1× PBS) was added for 20 min at room temperature.

Immunoprecipitation and western blot

Cells were harvested and lysed with IP buffer [150 mM NaCl, Tris pH 7.4, 1% Nonidet-P40 (NP-40), 0.1% Na-deoxycholate] supplemented with 1× protease inhibitor cocktail and 1 mM phenylmethylsulfonyl fluoride (PMSF). After incubation on ice for 30 min, the supernatant was collected by centrifugation at 16 000 *g* for 30 min at 4 °C. The supernatant was pre-cleared for 1 h with mouse or rabbit IgG (Santa Cruz, sc-2025 or sc-2027) on Dynabeads Protein G (Thermo Fisher Scientific, #10003D). The pre-cleared cell lysate was incubated with monoclonal anti-FLAG antibody (F1804, Sigma Aldrich) or polyclonal anti-KMT2D/MLL4 antibody (Santa Cruz Biotechnology, sc-293217) using Dynabeads Protein G at 4 °C overnight.

For the western blot, the following primary antibodies were used; Caspase 3 (1:1000, Thermo Fisher Scientific #PA5-77887), Caspase 9 (1:1000, Thermo Fisher Scientific #PA5-16358), CDKN1A (1:1000, Thermo Fisher Scientific #MA5-14949), TP53 (1:500, Thermo Fisher Scientific #MA1-7629), PRDM1 (1:1000, Cell Signaling Technology #9115), anti-Flag (1:1000, Sigma Aldrich #SAB4200071), polyclonal anti-KMT2D/MLL4 (1:700, Atlas Antibodies HPA035977), anti-Flag (1:1000, Sigma Aldrich #F1804) and anti-FOXD2 (1:1000, Sigma Aldrich #AV31709). The protein bands were detected using SuperSignal™ Western Blot Substrate Bundle, Pico PLUS (Thermo Fisher Scientific, A43840) on Amersham Imager 680 (GE Healthcare and Life Sciences).

Quantitative reverse transcription–polymerase chain reaction (RT–qPCR)

The total RNA was extracted using TRIzol™ (Invitrogen, 15596026) and cDNA was synthesized by the SuperiorScript III cDNA Synthesis kit (Enzynomics, EZ405S) with oligo(dT). RT–qPCR was performed by using TB Green® Premix Ex Taq™ II (Takara, RR820A) on the AriaMx Real-Time PCR system (Agilent Technologies). The primers used for RT–qPCR experiments are listed in Supplementary Table S1 and the gene expression levels were normalized to the housekeeping gene glyceraldehyde phosphate dehydrogenase (GAPDH).

Preparation of patient-derived CRC organoids and overexpression of FOXD2

The information on CRC organoids is listed in Supplementary Table S2. Patient-derived CRC organoids were grown for 1 or 2 weeks depending on the growth rate. The organoids were dissociated into single cells which were then treated with FOXD2 lentivirus supplemented with polybrene (10 mg/ml) for ∼6 h to enhance transduction efficiency. After culturing CRC organoids, 10 000 cells of FOXD2 and control groups were re-imbedded in Matrigel with full growth medium supplemented with puromycin (5 μg/ml) for selection.

Xenograft mouse model

Stably FOXD2-OE HCT116 cells (2.0×10^6 cells) were subcutaneously transplanted into the right posterior flank of the nude mice. A week after injection, the size of the transplanted tumor was measured every 2 days and a tumor growth curve was plotted. For histological examination, the standard hematoxylin & eosin (H&E) staining protocol was applied.

Single-cell RNA sequencing (scRNA-seq) and analysis

The libraries for scRNA-seq were prepared using the Chromium Next GEM Single-cell 3′ Reagent Kit v3.1 following the manufacturer's protocol on a Chromium Controller (10X Genomics). Feature-barcode matrices and clustering were processed using the Cell Ranger pipeline (ver. 4.0.0). Cells with <200 genes and zero unique molecular index (UMI) counts were ignored. Integrative multimodal analysis was performed by Seurat 4.0 (15). Uniform manifold approximation and projection (UMAP) was generated based on the first 20 principal components. FeaturePlot in Seurat was used for projection of the expression of individual genes.

Chromatin immunoprecipitation (ChIP)

ChIP was performed as previously described (16,17). In brief, sonicated or micronuclease (MNase)-treated chromatin was pre-cleared in RIPA buffer [20 mM Tris–HCl pH 8.0, 2 mM EDTA, 50 mM NaCl, 1% NP-40, 0.01% sodium dodecyl sulfate (SDS) and 0.5% sodium deoxycholate], 1 mM PMSF and 1× protease inhibitor cocktail (Roche, 05892970001) with mouse IgG (Santa Cruz, sc-2025) or rabbit IgG (Santa Cruz, sc-2027) on Protein G Magnetic Beads (Thermo Fisher Scientific, 88848). The pre-cleared chromatin was subjected to immunoprecipitation using antibodies against H3K4me3 (Abcam, ab8580),

H3K4me1 (Abcam, ab8895), H3K27ac (Abcam, ab4729), MLL4/KMT2D (Santa Cruz, sc-293217), Anti-6× His tag® [HIS.H8] (Abcam, ab18184) or monoclonal ANTI-FLAG® M2 (Sigma-Aldrich, F1804) overnight at 4 °C with rotation. The pulled chromatin on beads was reverse-cross-linked with proteinase K (20 mg/ml) and 10% SDS for 6 h at 65 °C. The DNA was purified by phenol/chloroform extraction and eluted in 20 μl of Tris-EDTA buffer.

Library preparation for sequencing

The mRNA-seq libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit (NEB, E7760S) for Illumina with the mRNA Magnetic Isolation Module (NEB, E7490S).

The library for ChIP-seq was prepared using the ACCEL-NGS 2S PLUS DNA LIBRARY Kit (Swift Bioscience, # 28096) or the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645S) following the manufacturer's instructions. In brief, purified immunoprecipitated DNA was subjected to end-repairing, adaptor ligation and 12 cycles of PCR amplification. Amplified DNA was size-selected and quantified using Bioanalyzer (Agilent).

Following the ATAC-seq (Assay for Transposase-Accessible Chromatin with sequencing) protocol with minor modifications (18), 50 000 cells were lysed in 50 μl of lysis buffer (10 mM Tris–HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40, 0.1% Tween-20 and 0.01% Triton X-100) and the isolated nuclear pellets were treated with 50 μl of transposition mix, 25 μl of 2× TD buffer (20 mM Tris–HCl pH 7.6, 10 mM $MgCl_2$ and 20% dimethylformamide in sterile water) and 2.5 μl of transposase (100 nM). The transposed DNA was purified by the DNA Clean & Concentrator-5 Kit (Zymo Research) and then amplified using the NEBNext High-Fidelity 2x PCR Master Mix for 10 cycles.

Chromatin conformation capture with immunoprecipitation (HiChIP) for H3K4me1, H3K4me3 and H3K27ac was performed using control and FOXD2-OE HCT116 cell lines. All HiChIP libraries were generated using the Dovetail® HiChIP MNase Kit (Dovetail, #21007) according to the manufacturer's instructions.

Computational analysis

All duplicated sequencing data were obtained on the Illumina Hiseq 2500 or NovaSeq 6000 platform. The sequencing quality was examined by FastQC-0.11.5 (19). Illumina adapters were trimmed by cutadapt (20).

After filtering the mRNA-seq data, high quality reads were mapped to the mouse reference genome and transcriptome (GRCm38_ensGene_100) using STAR-2.7.2b aligner with the suggested parameters (21). For the quantification of gene and isoform expression, RSEM-1.2.31 was used with the following parameters: –paired-end –estimate-rspd. Using gene annotation in Ensembl, the transcripts were annotated and normalized by counts per million (CPM) and transcripts per million (TPM) (22). Only the genes with a total read count >10 were used for further analysis. Differential expression analysis between different conditions was performed using the Bioconductor package DESeq2-1.30.1, and differentially expressed genes were defined as those with a fold change (FC) >2 (log_2FC >1) and a corrected *P*-value (Benjamini–Hochberg) <0.01 (23). The enrichment analysis of Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was done using Metascape and Enrichr (24,25). The gene set enrichment analysis (GSEA) was performed using the Molecular Signatures Database (MSigDB) v7.2 and GSEA-4.1 with the following options: -metric log2_Ratio_of_Classes -permute gene_set (26,27).

For the data analysis from ATAC-seq and ChIP-seq, all reads were aligned to hg38/GRCh38 or mm10/GRCm38 using Bowtie2 with parameters for ATAC-seq: '-k 4 –end-to–end' and ChIP-seq: '–end-to-end' (28). Low-quality reads (duplicates, unmapped or mate unmapped reads, not primary alignment, reads failing platform and mapped to ENCODE blacklist regions version 2) were excluded for further analysis. Reads mapped onto mitochondrial DNA were additionally excluded from the ATAC-seq analysis. ATAC-seq reads were offset by +4 bp for the + strand and –5 bp for the – strand (29). Model-based analysis of ChIP-seq 2 (MACS2) was used to call peaks for each sample with parameters '-q 0.01 –nomodel –shift -100 –extsize 200' for ATAC-seq and '-q 0.01'. Differentially accessible regions and differential histone modification analysis for merged peaks were performed by edgeR-3.32.1 (30,31). Thresholds for differential accessibility and histone modifications were >2-fold and >4-fold changes, respectively, with false discovery rate (FDR) <0.01. Peak annotation and motif enrichment were performed using Homer and SEA-5.4.0 (32,33). Heatmap and average profiles were generated using deepTools-3.5.1 (34). For ATAC-seq, footprints were found to identify transcription regulatory sites, and TF activity scores were calculated using a framework called HINT (Hidden Markov Model-based IdeNtification of Transcription factor footprints) from the Regulatory Genomics Toolbox (RGT-0.13.2) (35,36).

To build the LISC-specific and SISC-specific GRNs, ATAC-seq footprints common in LISC and SISC were excluded, and TFs with significantly (*P*-value < 0.05) different activity scores were selected. Also, footprints on transcription start sites (TSSs) around 1 kb regions were considered to build promoters targeting the TF network. Lastly, only differentially expressed genes were selected to construct GRNs. The networks were visualized using Cytoscape software (37).

For HiChIP data analysis, all read pairs were aligned to hg38/GRCh38 using BWA with parameters for '-5SP -T0' (38). Reads with mapping quality (MAPQ) <40 and duplicated pairs were not considered, and reads with ligation junctions were rescued if the alignment gap was <30 bp using pairtools-0.3.0. FitHiChIP was used to call loops for each sample with the options of peak to all background, FDR < 0.01 and multiple bin sizes (5000, 10 000 and 25 000) (39). Read pairs with distance >2 Mbp or <100 bp were excluded from the loop calling. Loops with different bin sizes were merged using HiCExplorer-3.7.2 (40). The loops anchoring on two adjacent bins were excluded from further analysis.

All sequencing metrics were summarized to show sequencing depth, library complexity, FRiP score for

ChIP-seq, reproducibility of ChIP-seq peaks, HiChIP interactions and RNA-seq, quality of scRNA-seq and reproducibility of ATAC-seq peaks (Supplementary Table S3). To understand the landscape of chromatin interactions, we performed biological replicates of HiChIP-seq on H3K4me1, along with technical replicates on H3K27ac and H3K4me3. The quality of all sequencing data was confirmed by conducting Pearson's analysis, which involved calculating the RPM (reads per million) values in the replicates (Supplementary Figure S1).

Survival analysis with The Cancer Genome Atlas (TCGA) datasets

The clinical information on 449 primary colon adenocarcinoma (COAD) datasets was downloaded from the National Cancer Institute Genomic Data Portal (https://portal.gdc.cancer.gov/). The groups in the upper and lower 20th percentile of the average *Foxd2* expression levels were considered. The Kaplan–Meier analysis was used to estimate survival probability. Survival curve and log-rank test were generated using R packages such as survival, survminer and ggplot2.

RESULTS

Gene regulatory network analysis revealed FOXD2 as a core TF regulating murine LISCs

To determine the transcriptional machinery operating during the differentiation of gut epithelium delineating and compartmentalizing small and large intestines, murine Lgr5+ SISCs and LISCs were isolated by fluorescence-activated cell sorting (FACS) analysis and the data from RNA-seq and ATAC-seq were analyzed (Figure 1A). The TF-mediated chromatin structural dynamics were examined by analyzing chromatin accessibility within DNA footprints of TF-binding motifs and differences in gene expression levels. The GRN for SISCs and LISCs was built, considering conjectured TF-binding sites and their target genes. The SISC-specific TFs were from the *Klf10*, *Sp5*, *Gata4* and *Egr* families, where *Gata4* is a well-known cofactor of *Cdx2* for the regulation of intestinal homeostasis (Supplementary Figure S2A) (41–43). *Cdx2* enriched in both types of ISCs regulates Hox gene expression essential for lineage specification in the intestinal epithelium during early embryogenesis (10). In particular, the posterior HOX code, *Hoxa* and *Hoxb* families, were enriched in the LISC GRN. Among them, *Foxa1* is associated with the regulation of the ISC niche-supporting factor. *Foxd2* was identified in the hub of the LISC GRN and linked with multiple target genes specifically regulated in the large intestine (Figure 1B, C).

Next, we integrated the scRNA-seq data of LISCs and SISCs to confirm the specific enrichment of the *Foxd2* transcript in LISCs at single-cell resolution (44). *Cdx2* was equally expressed in LISCs and SISCs, but *Olfm4* was enriched explicitly in SISCs (Figure 1D; Supplementary Figure S2B, C). Moreover, consistent with bulk RNA-seq analysis, the *Foxd2* transcript was highly enriched in LISCs (Figure 1E). To further examine whether *Foxd2* is related to determining colonic identity, we investigated RNA-seq data from *Foxd2* knock-out colonic organoids (45). In the *Foxd2*-loss colonic organoids, 256 genes were significantly down-regulated (FDR < 0.05 and FC < -1.0). GSEA indicated that ablation of Foxd2 led to the loss of the large intestinal identity (Supplementary Figure S2D). Taken together, these results suggest that *Foxd2* is specifically enriched in murine colonic stem cells and is an essential factor in regulating the homeostasis of the colonic epithelium.

FOXD2 induces cancer cell apoptosis by transactivating p53-responsive genes in CRC cells

Somatic stem cell and tissue-specific enriched factors are frequently misregulated in tumors, which contributes to cancer inhibition or development (9,46). Given that the loss of *Foxd2* might impair murine colonic identity, we speculated that *Foxd2* expression would be related to the inhibition of colon cancer cell growth. The TCGA data showed that *Foxd2* expression was significantly down-regulated in COAD tissue compared with normal tissue (Supplementary Figure S3A). Considering the gene expression data in the upper and lower 20th percentiles, a total of 473 COAD datasets from TCGA showed that the reduced *Foxd2* expression was significantly related to the poor survival in COAD patients (P-value $= 0.03049$, Figure 2A). To confirm the poor prognosis of survival at the molecular level, the full-length *Foxd2* gene was overexpressed in HCT116 cells. Strikingly, ectopic expression of *Foxd2* regressed cancer cell growth, and significantly inhibited cell migration ability, with 29.3% for wound confluence of FOXD2-OE cells versus 77.8% for the control group (Figure 2B–D). FOXD2 overexpression increased early and late apoptotic cell populations and showed significantly reduced G_0/G_1-phase ratios and elevated S- and G_2/M-phase ratios at the same time. These results illustrate cancer cell apoptosis and cell cycle arrest (Supplementary Figure S3B, C). Differentially expressed genes between the control group and FOXD2-OE cells in RNA-seq analysis were used for the GSEA, and were significantly related to cell apoptosis and the activated p53 pathway (Figure 2E). Although the protein level of p53 was not directly regulated by FOXD2 overexpression, p53-responsive CDKN1A/p21 was highly transactivated (Supplementary Figure S3D). We also observed the induction of caspases 3 and 9, and the repression of Ki67 (Supplementary Figure S3E, F). These findings suggested that the colon-enriched factor FOXD2 activated the p53 pathway and apoptosis to inhibit colon cancer cell development in human CRC cells.

FOXD2 opens chromatin and recruits MLL4/KMT2D to deposit H3K4 monomethylation on its target regions

To elucidate the functional role of FOXD2 in human CRC cells, the FOXD2-binding sites and the epigenetic modification patterns were examined by ChIP-seq and ATAC-seq. Upon FOXD2 overexpression, H3K4me1 ChIP-seq analysis revealed 17 031 altered regions, where H3K4me1 levels in 13 689 regions were substantially increased (denoted by GAIN) and 3342 regions were decreased (denoted by LOSS). These regions were located mainly at the intronic and intergenic regions (Figure 3A, B). GO analysis showed that H3K4me1 GAIN regions were significantly associated with the regulation of apoptotic process

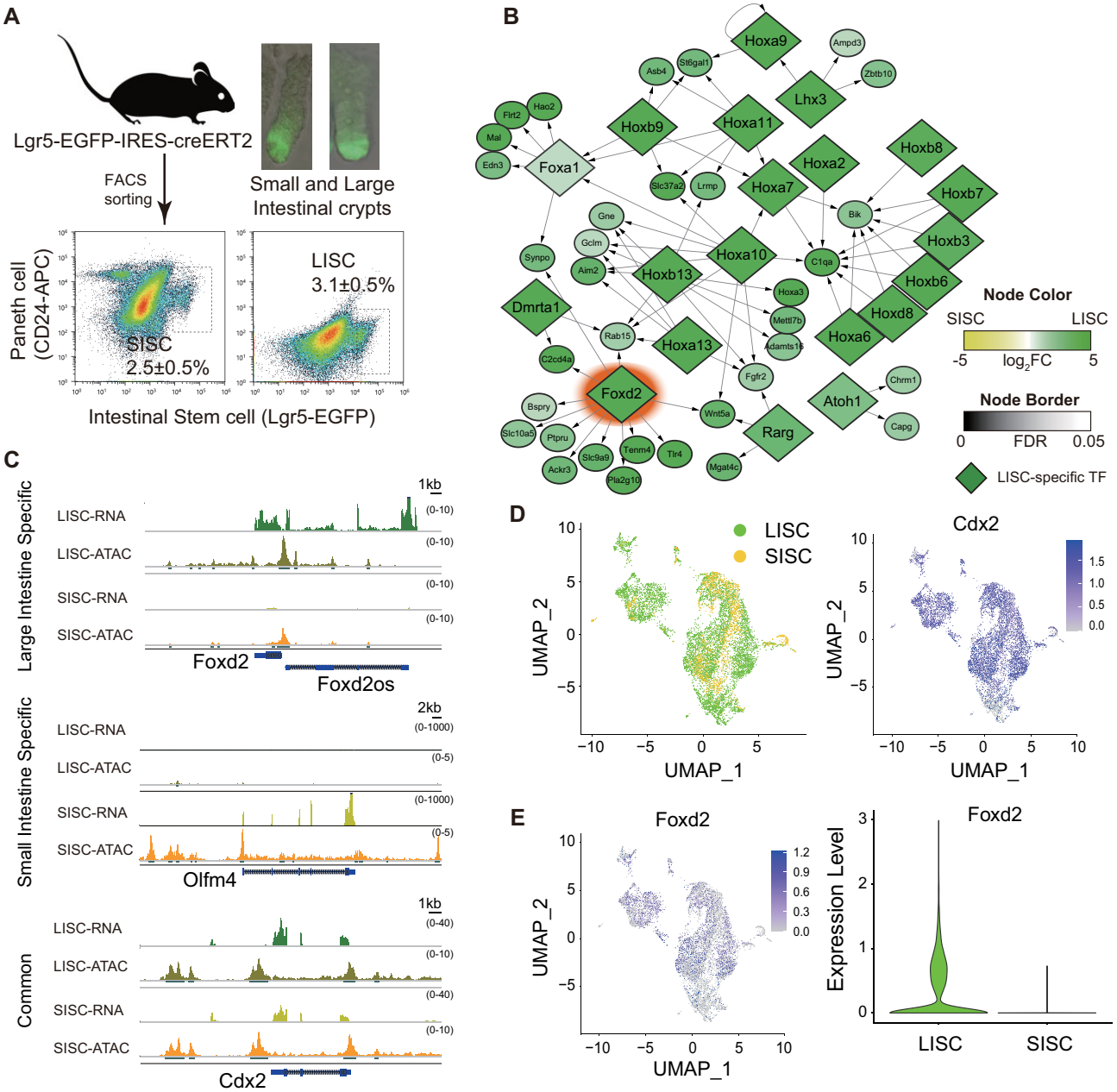


Figure 1. GRN analysis on ISCs revealed that *Foxd2* has regional specificity on LISCs. (**A**) Experimental scheme for isolating ISCs in the murine intestinal crypt. EGFP+/high stem cells were isolated from Lgr5-EGFP-Ires-CreERT2+/– mice by FACS and subject to RNA-seq and ATAC-seq. (**B**) GRN for LISCs (FDR < 0.01 and Outdegree ≥ 10). (**C**) RNA-seq and ATAC-seq profiles of ISCs at the loci of *Foxd2*, *Olfm4* and *Cdx2*. (**D**) UMAP plot for single-cell RNA-seq data of LISCs and SISCs (left). The expression of *Cdx2* at the single-cell level is shown on the right. (**E**) The expression level of *Foxd2* at single-cell resolution (left) and the overall *Foxd2* expression level in LISCs and SISCs (*P*-value = 1.24E-221, two-tailed *t*-test) (right).

and programmed cell death (Supplementary Figure S4A) (47). Moreover, motif enrichment analysis of H3K4me1 GAIN regions showed that FOXD2, CREB1, CREM and TP53 were highly ranked TFs. The enrichment of the TP53-binding motif (q-value = 1.37e-9, Supplementary Figure S4B) was implicated in cancer cell apoptosis, which is consistent with the result in Figure 2E and Supplementary Figure S3C. Although the global level of H3K27ac, another enhancer mark, was not significantly altered, the average signal of H3K27ac in H3K4me1 GAIN regions was elevated in FOXD2-OE cells (Figure 3C). These results support that FOXD2 could change the enhancer landscape in human CRC cells.

An integrated analysis of H3K4me1 GAIN or LOSS regions with changes in chromatin accessibility was carried out to further assess the epigenomic changes caused by FOXD2 overexpression in HCT116 cells. A total of 2299 regions became more accessible in FOXD2-OE cells, and 574 domains were overlapped with H3K4me1 regions (Supplementary Figure S4C, D). The changes in the level of

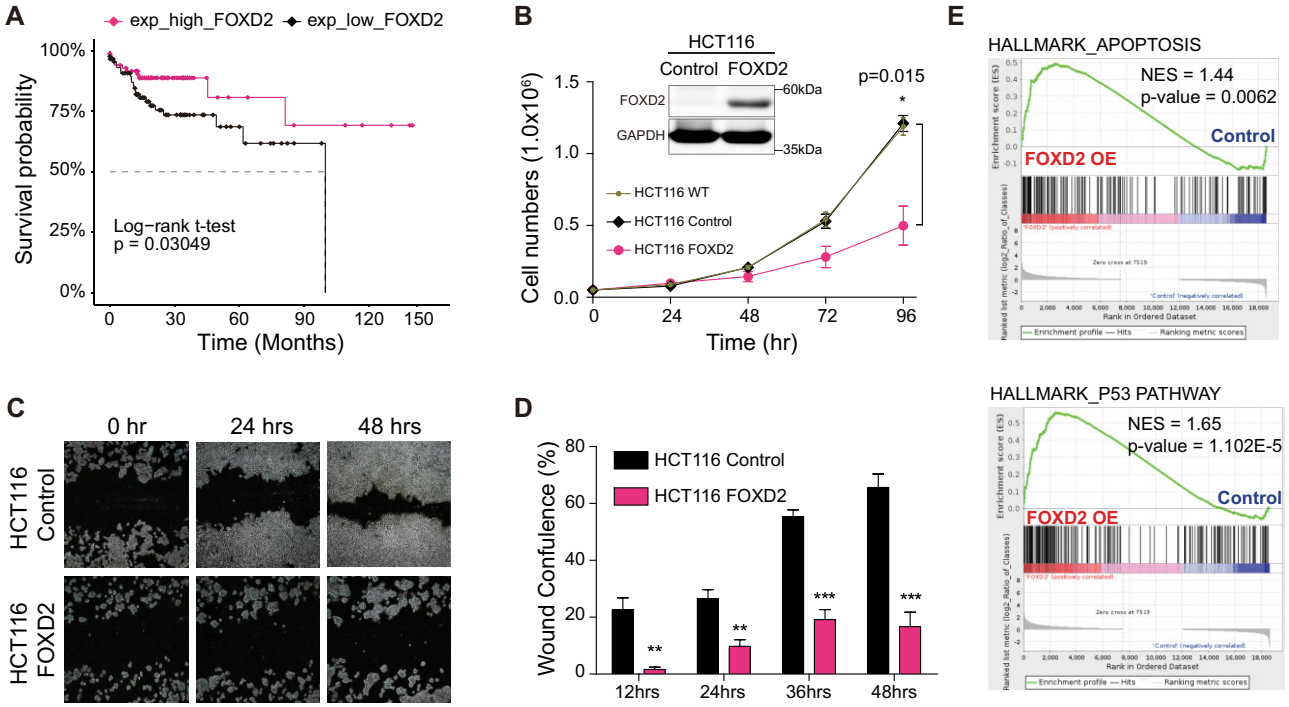


Figure 2. Ectopic expression of FOXD2 in human CRC cells inhibits growth *in vitro*. (**A**) TCGA survival analysis of COAD patients (High, top 20%; Low, bottom 20% by *Foxd2* expression level). (**B**) HCT116 cell growth in wild-type (WT), control (vehicle) and FOXD2 (full length of FOXD2 OE). The overexpression of FOXD2 was confirmed by western blot (inset). (**C**) Wound healing assay in control versus FOXD2 OE. (**D**) Wound confluence in FOXD2-OE HCT116 cells. Two-tailed *t*-test; *P*-values $*P < 0.05$, $**P < 0.001$ and $***P < 0.0001$. (**E**) GSEA for up-regulated genes in FOXD2-OE HCT116 cells. Up-regulated apoptosis (upper) and p53 pathway (lower).

H3K4me1 were moderately correlated with the degree of chromatin accessibility, suggesting that the deposition of H3K4me1 is related to the openness of chromatin (Figure 3D). Moreover, the FOXD2-binding sites were more overlapped with ATAC GAIN/LOSS regions than H3K4me1 GAIN/LOSS regions (Figure 3E; Supplementary Figure S4E). As a result of the FOXD2 overexpression in HCT116 cells, GAIN regions in both ATAC and H3K4me1 showed *de novo* chromatin openness and FOXD2 occupancy (Figure 3E, center panel). The level of FOXD2 binding showed a higher correlation with chromatin accessibility than the deposition of H3K4me1 (Figure 3E, left and right panel). In addition, the FOXD2 motif was detected most frequently in both H3K4me1 and ATAC GAIN regions in FOXD2-OE HCT116 cells (Supplementary Figure S4F). These results indicate that the deposition of H3K4me1 itself does not intensely affect global chromatin openness but changes in the chromatin structure are tuned when FOXD2 is co-localized. GO analysis showed that 574 H3K4me1/ATAC GAIN regions were significantly associated with the regulation of cell death and the apoptotic process (Figure 3F). The enriched TF-binding motifs at both GAIN regions were those of Jun, KLF4, FOXD2, ZNF460, PRDM1 and CREB1 (Figure 3G).

Since the ectopic expression of FOXD2 in human CRC cells inhibited cancer cell growth by dynamic deposition of H3K4me1 on *de novo* open chromatin regions, we tested whether FOXD2 would recruit a major mammalian H3K4 mono- and di-methyltransferase, MLL4, to deposit H3K4me1 in an enhancer- and/or gene-specific manner (48). Surprisingly, MLL4 peaks were also highly positioned in *de novo* H3K4me1/ATAC GAIN regions and MLL4 physically interacted with FOXD2 [Figure 3E (center panel), 3H]. Overall, these findings demonstrate that FOXD2 tunes chromatin structure and deposition of H3K4me1 by recruiting MLL4 to inhibit colon cancer development.

FOXD2 reprograms enhancer interactions and coordinates the gene expression profile enriched for cancer cell apoptotic pathways

To speculate how the chromatin structure modulates the epigenetic transcription machinery during cancer regression, the chromatin looping formation based on histone modifications was measured by the HiChIP method using anti-H3K4me3, H3K27ac and H3K4me1 antibodies. H3K4me3 is localized predominantly on promoters, whereas H3K27ac and H3K4me1 are mostly positioned at active enhancers. Significant chromatin interactions specific to FOXD2-OE CRC cells occupied 17.0% (2393 out of 14 067), 24.7% (3353 out of 13 558) and 12.4% (2323 out of 18 795) of chromatin loops anchored on the histone modifications of H3K4me3, H3K27ac and H3K4me1, respectively (Supplementary Figure S5A, B). As expected, the chromatin loops formed between enhancer–enhancer, enhancer–promoter or promoter–promoter pairs. HiChIP loops for H3K4me3 were primarily associated with promoters (control, 83%; FOXD2, 84%), for H3K27ac with enhancers (control, 71%; FOXD2, 71%) and for

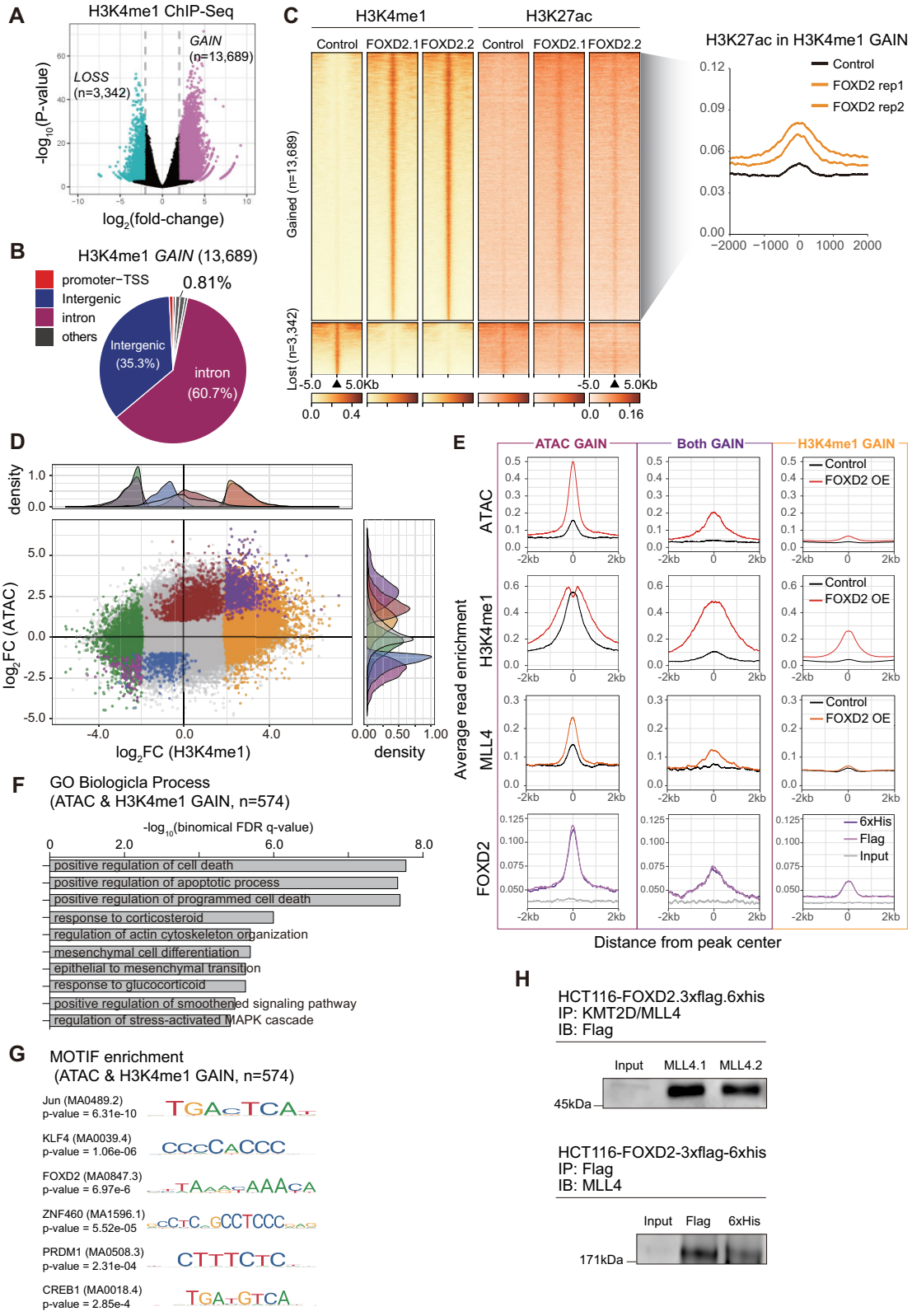


Figure 3. FOXD2 interacts with MLL4 to deposit H3K4 monomethylation in the *de novo* open chromatin region. (**A**) Volcano plots for differential H3K4me1 levels between control and FOXD2-OE HCT116 cells (FC > 4, FDR < 0.05). (**B**) Genome-wide distribution of H3K4me1 GAIN regions. (**C**) Heatmap analysis in gained and lost H3K27ac loci with two duplicated experiments. Levels of H3K27ac1 centered on H3K4me1 GAIN regions (right). (**D**) Comparison of H3K4me1 and ATAC signals at 574 overlapping regions. Upper and right peak plots are the relative density of differential levels of H3K4me1 or chromatin accessibility corresponding to dots with a different color. (**E**) Average read enrichment around the peak (right, ATAC GAIN regions; middle, both H3K4me1 and ATAC GAIN regions; left, H3K4me1 only GAIN regions). (**F**) GO analysis for biological processes associated with genes within 1 kb of H3K4me1 and ATAC, both GAIN genes ($n = 574$). (**G**) Highly enriched TF sequence motifs identified by H3K4me1 and ATAC GAIN sites. (**H**) Co-immunoprecipitation (Co-IP) of FOXD2 and MLL4. Duplicated experiments are denoted by MLL4.1 and MLL4.2.

H3K4me1 with enhancers (control, 76%; FOXD2, 75%) (Supplementary Figure S5C). Significantly, the interaction loops only for H3K4me1 showed positive correlation with up-regulated genes, H3K4me1 enrichment and chromatin accessibility in FOXD2-OE CRC cells, indicating that FOXD2-mediated enhancer interactions were reorganized at the regions where the changes in the levels of H3K4me1 and chromatin accessibility occurred (Figure 4A–C; Supplementary Figure S5D–F). The HiChIP loops for H3K4me3 and H3K27ac were generally associated with the global deposition of H3K4me1 (Supplementary Figure S5E, F). Additionally, chromatin interactions of enhancers and promoters were considerably augmented on open chromatin regions but had little effect on closed chromatin (Figure 4C; Supplementary Figure S5F). According to the categorized chromatin regions in Figure 3E, both H3K4me1/ATAC GAIN regions were highly correlated with chromatin loop formation of H3K4me3, H3K4me1 and H3K27ac, whereas H3K4me1/ATAC LOSS regions had a moderate correlation (Figure 4D–F). Transactivated genes bearing all HiChIP loops associated with H3K4me1/ATAC GAIN regions were enriched in the FOXD2-OE CRC cells; P-value = 0.00097, normalized enrichment score (NES) = 1.28 (Figure 4G). Also, down-regulated target genes were correlated with chromatin interactions (Supplementary Figure S5G, H). GO analysis of leading-edge genes showed significantly enriched biological processes such as positive regulation of programmed cell death and negative regulation of cell population proliferation, highlighting that enhancer interactions were transformed in a gene-specific manner to induce the apoptotic process in FOXD2-OE CRC cells (leading edge genes n = 219, Figure 4H; Supplementary Table S4). As exemplified in Figure 4I, new chromatin interactions were formed at enhancer regions near OASL (oligoadenylate synthase-like), PRDM1, TNFAIP3 (TNF Alpha Induced Protein 3) and GADD45A (Growth Arrest and DNA Damage Inducible Alpha) in accordance with *de novo* H3K4me1/ATAC GAIN. Thus, FOXD2-mediated reprogramming caused changes in enhancer–promoter interactions at the GAIN regions of H3K4me1/ATAC and transactivation of the target genes to suppress human CRC development.

FOXD2 inhibits the growth of patient-derived CRC organoid and xenografted cancer *in vivo*

To further evaluate the tumor-suppressing role of FOXD2, patient-derived CRC organoids from two human surgical specimens were grown as illustrated in Figure 5A. The overexpression of FOXD2 substantially repressed CRC organoid growth, and its size was much smaller than that of the control group (Figure 5B, C). FOXD2 overexpression impaired the proliferation of CRC organoids and induced apoptosis, as confirmed by immunostaining using antibodies against Ki67, and caspases 3 and 9 (Figure 5D). The expression of PR domain zinc finger protein 1 (PRDM1, also known as Blimp-1), a tumor suppressor and the p53-responsive element, was increased as an effect of FOXD2 overexpression in HCT116 cells and CRC organoids (Supplementary Figure S6A, B) (49,50).

FOXD2-OE HCT116 cells were transplanted subcutaneously into the back of nude mice. A week after injection, the xenografted tumor group with FOXD2-OE HCT116 cells grew more slowly, and the tumor size was significantly smaller than in the control group (Figure 5E, F). The reduced tumor weights measured after sacrifice substantiated that FOXD2 suppressed the tumor development (Figure 5G). In addition, the level of PRDM1 expression was increased and the abundant apoptotic vacuoles in FOXD2-OE xenografted tumor were observed by H&E staining analysis (Supplementary Figure S6C, D). These results indicate a novel role for FOXD2 associated with the suppressive effect of tumor development as a chromatin-tuning factor.

DISCUSSION

The FOX family members function as diverse tissue- and cell type-specific TFs (51). They are generally known as important regulators in embryogenesis and maintenance of cellular and tissue homeostasis (52). Also, they can act as transcriptional activators and repressors, although it remains unclear how they function under the epigenetic mechanisms through DNA methylation, histone modifications and chromatin structure. In this study, the TF-mediated chromatin structural dynamics were investigated by comparing changes in chromatin accessibility within DNA footprints of TF-binding motifs and differences in gene expression levels between SISCs and LISCs. Interestingly, *Foxd2* was identified as an LISC-specific hub in the GRN (Figure 1). In addition, the expression level of FOXD2 in CRC was negatively correlated with survival rate by TCGA data analysis (Figure 2). To understand the molecular role of FOXD2, an integrative analysis of the epigenome and transcriptome was performed in cellular and mouse models. FOXD2 overexpression transformed the chromatin dynamics, especially the *cis*-regulatory landscape between promoters and enhancers, and induced an extensive cellular reprogramming to alter the cell cycle and tumor growth (Figures 3 and 4).

Further investigation on the changes in H3K4me1 patterns and chromatin openness suggested that FOXD2 opened the compact chromatin domain and recruited MLL4, an H3K4 mono-methyltransferase introducing *de novo* histone modification on the target regions (Figure 3). The chromatin loops identified by HiChIP analysis revealed that the regulatory elements such as enhancers and promoters were typically marked by histone modifications and formed differential chromatin loops with each other when FOXD2 was overexpressed. Overall, FOXD2 reprogrammed enhancer–promoter interactions on *de novo* GAIN regions of H3K4me1/ATAC to suppress human CRC development in a cell type- and gene-specific manner (Figure 4). The tumor-suppressing role of FOXD2 was validated such that ectopic expression of FOXD2 in patient-derived CRC organoids and xenograft mouse models substantially repressed the growth due to induction of the apoptotic pathway (Figure 5).

The functional effect of many FOX family proteins can be either transactivation or inhibition of gene transcription. However, this study delineates the novel role of FOXD2 as a chromatin-tuning factor by providing evidence of

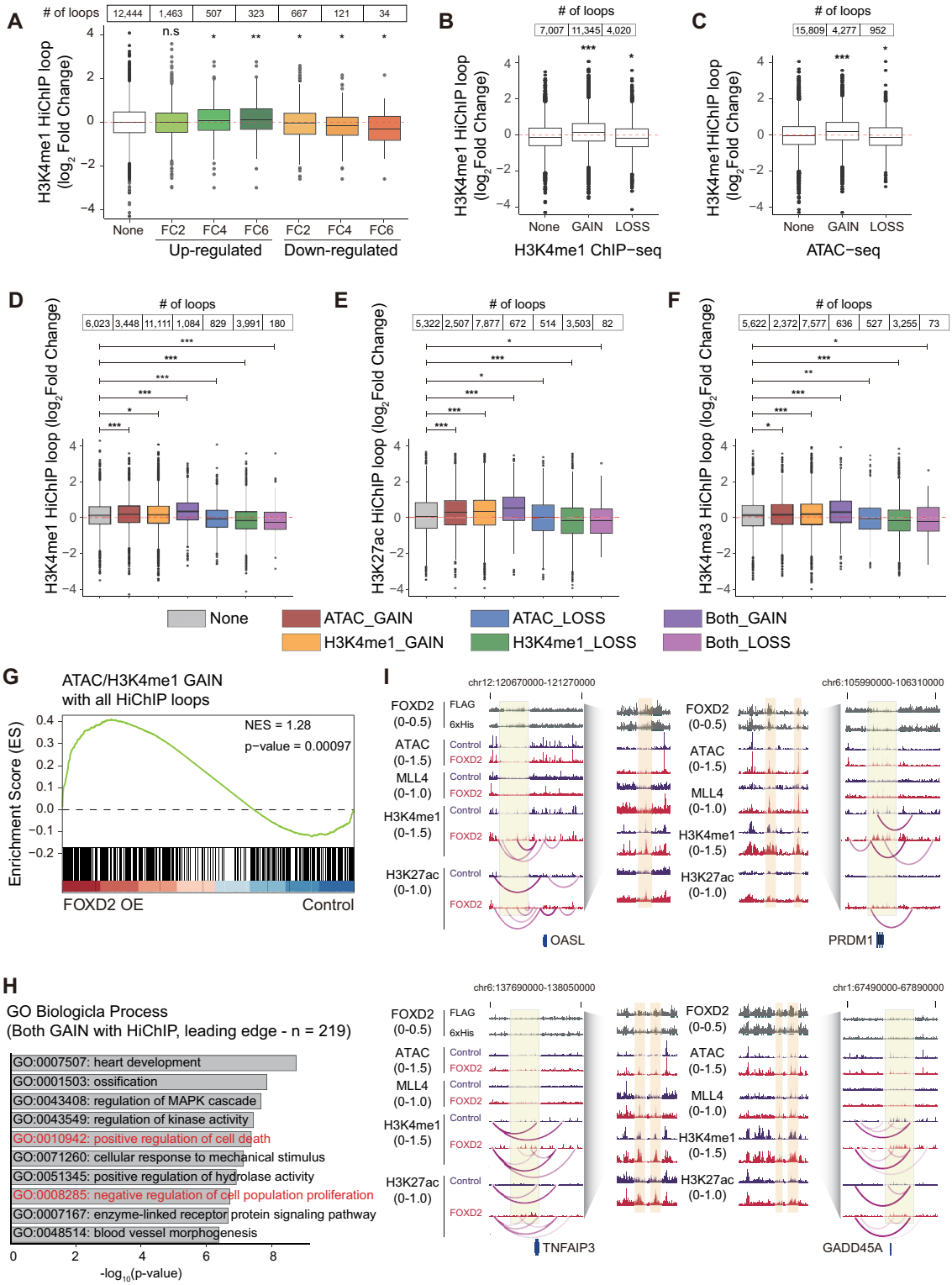


Figure 4. FOXD2-mediated enhancer interactions transactivate genes toward the apoptotic process. The *P*-values for Wilcoxon signed-rank test are $*P < 0.05$, $**P < 0.001$, and $***P < 0.0001$. (**A**) H3K4me1-anchored HiChIP loops and genes up- or down-regulated by FOXD2 overexpression. FC2, FC4, and FC6 denote 2-, 4- and 6-fold change in gene expression. (**B**) H3K4me1-anchored loops versus changes of H3K4me1 levels. (**C**) H3K4me1-anchored loops versus changes of chromatin accessibility (ATAC-seq). (**D**) Relationship of H3K4me1 HiChIP loops to changes of H3K4me1 and ATAC. (**E**) Relationship of H3K27ac HiChIP loops to changes of H3K4me1 and ATAC. (**F**) Relationship of H3K4me3 HiChIP loops to changes of H3K4me1 and ATAC. (**G**) GSEA for target genes with H3K4me1/ATAC GAIN regions overlapped with all three HiChIP loops in FOXD2-OE CRC cells. (**H**) GO analysis for biological processes associated with leading edge genes ($n = 219$). (**I**) Representative profiles of ChIP-seq for FOXD2, MLL4, H3K4me1, and H3K27ac, ATAC-seq for chromatin accessibility and HiChIP for H3K4me1 and H3K27ac (loops). The *P*-values for Wilcoxon signed-rank test are $*P < 0.05$, $**P < 0.001$, and $***P < 0.0001$. Zoom-in snapshots of the boxed region in yellow are shown in the middle.

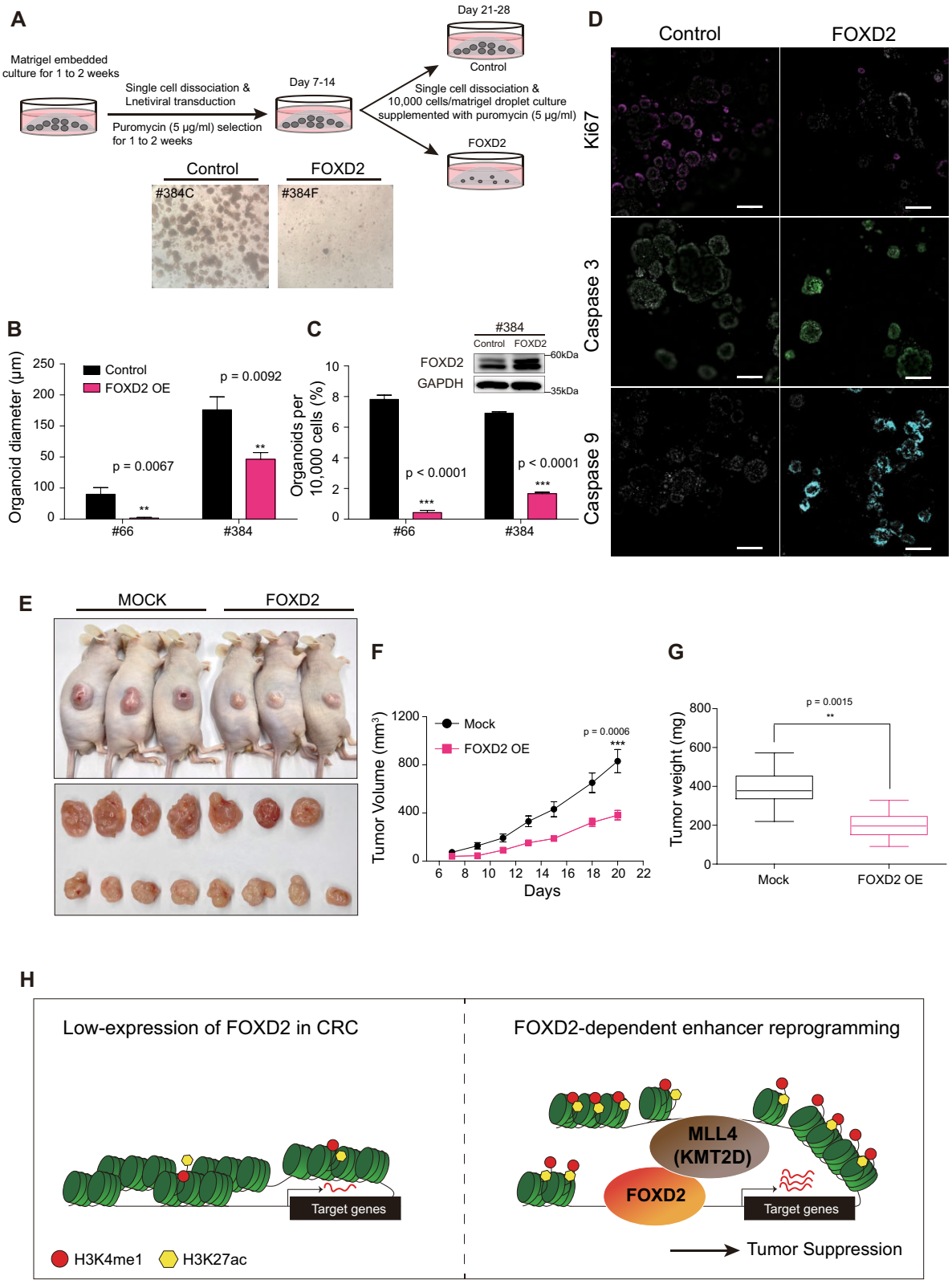


Figure 5. FOXD2 inhibited CRC organoid growth and cancer development *in vivo*. (**A**) Schematic of the tumor organoid-forming assay in FOXD2-OE cells and control. (**B**) CRC organoid size in FOXD2-OE cells and control (biological replicates $n = 2$). (**C**) Occurrence of organoids per 10 000 cells at CRC organoid establishment (***$P < 0.0001$, biological replicates $n = 2$, triplicated). The expression levels of FOXD2 were verified by western blot (inset). (**D**) Immunostaining of the proliferation marker (Ki67) and apoptosis markers (caspases 3 and 9) in a CRC organoid (#384, scale bar 200 μm). (**E**) Colon cancer growth in transplanted mice. (**F**) Tumor volumes of the FOXD2-OE and control groups. (**G**) Tumor weight of the FOXD2-OE and control groups. The *P*-values for the two-tailed *t*-test are *$P < 0.05$, **$P < 0.001$, and ***$P < 0.0001$. (**H**) Schematic model for the role of FOXD2 as a chromatin-tuning factor in tumor development.

dynamic alterations in epigenomic features during suppression of CRC development. In addition, we identified 76 LOSS regions of ATAC and H3K4me1 bound by FOXD2, which implicates that FOXD2 would be marginally associated with closed chromatin (Supplementary Figure S4C). In a future study, it would be interesting to test whether these FOXD2-linked closed chromatin domains are related with the polycomb repressive complex (PRC) to repress transcription in a chromatin context-dependent manner, in contrast to the FOXD2–MLL4 axis. As a chromatin-tuning factor, FOXD2 would exert a diverse function in a different chromatin milieu so that cells could respond effectively to the environmental signals. As shown in this study, FOXD2 seems to be associated with regulatory regions in sequence- and chromatin feature-specific manners and to play a role in fine-tuning the gene expression program of a limited set of genes to suppress CRC growth. Overall, our data support a model where FOXD2 binding induces the closed chromatin domain to change to the open structure, and MLL4 is recruited to implement H3K4 monomethylation. Eventually, variations in chromatin interactions transactivate target genes such as tumor suppressors in colon cancer (Figure 5H). Our FOXD2-associated multi-omics databases could be used for further analysis to identify another aspect of chromatin activity in intestinal stem cells and CRC cells.

DATA AVAILABILITY

The raw sequencing data and the processed data are available in the Gene Expression Omnibus (GEO) repository at https://www.ncbi.nlm.nih.gov/geo/ with the accession numbers GSE196922 and GSE196916 (ATAC-seq, GSE196913 and GSE196918; RNA-seq, GSE196914, and GSE196921; ChIP-seq, GSE196919; HiChIP-seq, GSE196920; scRNA-seq of LISCs, GSE196915). The raw sequencing data are also available on the Korean Nucleotide Archive (KoNA) at https://www.kobic.re.kr/kona/ with the accession number PRJKA220170. The publicly available GEO dataset for scRNA-seq (GSE99457) was used.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Dr. C.-S. Hwang and Mr. S.-H. Mun for providing pcDNA 3.1–3×flag-6×his plasmid, S.-H. Shin for constructing sequencing libraries. Dr. D.-K. Kim for providing basic information of patient-derived colon cancer organoids, and Dr. D.-H. Kim for advising on TCGA analysis. We thank all the lab members for their valuable comments and discussions.

FUNDING

The National Research Foundation of Korea [NRF-2014M3C9A3064548, 2019M3A9H1103711, 2022M3A9B6082674 and 2022R1A2C3011663].
Conflict of interest statement. None declared.

REFERENCES

1. Iwafuchi-Doi,M., Donahue,G., Kakumanu,A., Watts,J.A., Mahony,S., Pugh,B.F., Lee,D., Kaestner,K.H. and Zaret,K.S. (2016) The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell*, **62**, 79–91.
2. Newman,E.A., Kim,D.W., Wan,J., Wang,J., Qian,J. and Blackshaw,S. (2018) Foxd1 is required for terminal differentiation of anterior hypothalamic neuronal subtypes. *Dev. Biol.*, **439**, 102–111.
3. Yoshida,M., Hata,K., Takashima,R., Ono,K., Nakamura,E., Takahata,Y., Murakami,T., Iseki,S., Takano-Yamamoto,T., Nishimura,R. *et al.* (2015) The transcription factor Foxc1 is necessary for Ihh-Gli2-regulated endochondral ossification. *Nat. Commun.*, **6**, 6653.
4. Myatt,S.S. and Lam,E.W. (2007) The emerging roles of forkhead box (Fox) proteins in cancer. *Nat. Rev. Cancer*, **7**, 847–859.
5. Dawson,M.A. and Kouzarides,T. (2012) Cancer epigenetics: from mechanism to therapy. *Cell*, **150**, 12–27.
6. Creyghton,M.P., Cheng,A.W., Welstead,G.G., Kooistra,T., Carey,B.W., Steine,E.J., Hanna,J., Lodato,M.A., Frampton,G.M., Sharp,P.A. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA*, **107**, 21931–21936.
7. Dalerba,P., Sahoo,D., Paik,S., Guo,X., Yothers,G., Song,N., Wilcox-Fogel,N., Forgo,E., Rajendran,P.S., Miranda,S.P. *et al.* (2016) CDX2 as a prognostic biomarker in stage II and stage III colon cancer. *N. Engl. J. Med.*, **374**, 211–222.
8. Stringer,E.J., Duluc,I., Saandi,T., Davidson,I., Bialecka,M., Sato,T., Barker,N., Clevers,H., Pritchard,C.A., Winton,D.J. *et al.* (2012) Cdx2 determines the fate of postnatal intestinal endoderm. *Development*, **139**, 465–474.
9. Yu,J., Liu,D., Sun,X., Yang,K., Yao,J., Cheng,C., Wang,C. and Zheng,J. (2019) CDX2 inhibits the proliferation and tumor formation of colon cancer cells by suppressing Wnt/beta-catenin signaling via transactivation of GSK-3beta and Axin2 expression. *Cell Death Dis.*, **10**, 26.
10. Gao,N., White,P. and Kaestner,K.H. (2009) Establishment of intestinal identity and epithelial–mesenchymal signaling by Cdx2. *Dev. Cell*, **16**, 588–599.
11. Kume,T., Deng,K. and Hogan,B.L. (2000) Minimal phenotype of mice homozygous for a null mutation in the forkhead/winged helix gene, Mf2. *Mol. Cell. Biol.*, **20**, 1419–1425.
12. Johansson,C.C., Dahle,M.K., Blomqvist,S.R., Gronning,L.M., Aandahl,E.M., Enerback,S. and Tasken,K. (2003) A winged helix forkhead (FOXD2) tunes sensitivity to cAMP in T lymphocytes through regulation of cAMP-dependent protein kinase RIalpha. *J. Biol. Chem.*, **278**, 17573–17579.
13. Sato,T., Vries,R.G., Snippert,H.J., van de Wetering,M., Barker,N., Stange,D.E., van Es,J.H., Abo,A., Kujala,P., Peters,P.J. *et al.* (2009) Single Lgr5 stem cells build crypt–villus structures in vitro without a mesenchymal niche. *Nature*, **459**, 262–265.
14. Liang,C.C., Park,A.Y. and Guan,J.L. (2007) In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. *Nat. Protoc.*, **2**, 329–333.
15. Hao,Y., Hao,S., Andersen-Nissen,E., Mauck,W.M. 3rd, Zheng,S., Butler,A., Lee,M.J., Wilk,A.J., Darby,C., Zager,M. *et al.* (2021) Integrated analysis of multimodal single-cell data. *Cell*, **184**, 3573–3587.
16. Choe,M.K., Hong,C.P., Park,J., Seo,S.H. and Roh,T.Y. (2012) Functional elements demarcated by histone modifications in breast cancer cells. *Biochem. Biophys. Res. Commun.*, **418**, 475–482.
17. Brind'Amour,J., Liu,S., Hudson,M., Chen,C., Karimi,M.M. and Lorincz,M.C. (2015) An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat. Commun.*, **6**, 6033.
18. Buenrostro,J.D., Wu,B., Chang,H.Y. and Greenleaf,W.J. (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.*, **109**, 21.29.1–21.29.9.
19. Patel,R.K. and Jain,M. (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*, **7**, e30619.
20. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.
21. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

22. Li,B. and Dewey,C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.*, **12**, 323.
23. 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.
24. Xie,Z., Bailey,A., Kuleshov,M.V., Clarke,D.J.B., Evangelista,J.E., Jenkins,S.L., Lachmann,A., Wojciechowicz,M.L., Kropiwnicki,E., Jagodnik,K.M. *et al.* (2021) Gene set knowledge discovery with Enrichr. *Curr. Protoc.*, **1**, e90.
25. Zhou,Y., Zhou,B., Pache,L., Chang,M., Khodabakhshi,A.H., Tanaseichuk,O., Benner,C. and Chanda,S.K. (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.*, **10**, 1523.
26. Liberzon,A., Birger,C., Thorvaldsdóttir,H., Ghandi,M., Mesirov,J.P. and Tamayo,P. (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.*, **1**, 417–425.
27. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.
28. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
29. Buenrostro,J.D., Giresi,P.G., Zaba,L.C., Chang,H.Y. and Greenleaf,W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, **10**, 1213–1218.
30. 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.
31. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
32. Bailey,T.L. and Grant,C.E. (2021) SEA: simple enrichment analysis of motifs. bioRxiv doi: https://doi.org/10.1101/2021.08.23.457422, 24 August 2021, Preprint: not peer reviewed.
33. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
34. Ramirez,F., Ryan,D.P., Gruning,B., Bhardwaj,V., Kilpert,F., Richter,A.S., Heyne,S., Dundar,F. and Manke,T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*, **44**, W160–W165.
35. Castro-Mondragon,J.A., Riudavets-Puig,R., Rauluseviciute,I., Lemma,R.B., Turchi,L., Blanc-Mathieu,R., Lucas,J., Boddie,P., Khan,A., Manosalva Perez,N. *et al.* (2022) JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **50**, D165–D173.
36. Li,Z., Schulz,M.H., Look,T., Begemann,M., Zenke,M. and Costa,I.G. (2019) Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.*, **20**, 45.
37. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
38. Li,H. and Durbin,R. (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, **26**, 589–595.
39. Bhattacharyya,S., Chandra,V., Vijayanand,P. and Ay,F. (2019) Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nat. Commun.*, **10**, 4221.
40. Wolff,J., Backofen,R. and Grüning,B. (2021) Robust and efficient single-cell Hi-C clustering with approximate k-nearest neighbor graphs. *Bioinformatics*, **37**, 4006–4013.
41. Bosse,T., Piaseckyj,C.M., Burghard,E., Fialkovich,J.J., Rajagopal,S., Pu,W.T. and Krasinski,S.D. (2006) Gata4 is essential for the maintenance of jejunal–ileal identities in the adult mouse small intestine. *Mol. Cell Biol.*, **26**, 9060–9070.
42. San Roman,A.K., Aronson,B.E., Krasinski,S.D., Shivdasani,R.A. and Verzi,M.P. (2015) Transcription factors GATA4 and HNF4A control distinct aspects of intestinal homeostasis in conjunction with transcription factor CDX2. *J. Biol. Chem.*, **290**, 1850–1860.
43. Thompson,C.A., Wojta,K., Pulakanti,K., Rao,S., Dawson,P. and Battle,M.A. (2017) GATA4 is sufficient to establish jejunal versus ileal identity in the small intestine. *Cell. Mol. Gastroenterol. Hepatol.*, **3**, 422–446.
44. Yan,K.S., Gevaert,O., Zheng,G.X.Y., Anchang,B., Probert,C.S., Larkin,K.A., Davies,P.S., Cheng,Z.F., Kaddis,J.S., Han,A. *et al.* (2017) Intestinal enteroendocrine lineage cells possess homeostatic and injury-inducible stem cell activity. *Cell Stem Cell*, **21**, 78–90.
45. Gu,W., Wang,H., Huang,X., Kraiczy,J., Singh,P.N.P., Ng,C., Dagdeviren,S., Houghton,S., Pellon-Cardenas,O., Lan,Y. *et al.* (2022) SATB2 preserves colon stem cell identity and mediates ileum–colon conversion via enhancer remodeling. *Cell Stem Cell*, **29**, 101–115.
46. Sarkar,A., Huebner,A.J., Sulahian,R., Anselmo,A., Xu,X., Flattery,K., Desai,N., Sebastian,C., Yram,M.A., Arnold,K. *et al.* (2016) Sox2 suppresses gastric tumorigenesis in mice. *Cell Rep.*, **16**, 1929–1941.
47. McLean,C.Y., Bristor,D., Hiller,M., Clarke,S.L., Schaar,B.T., Lowe,C.B., Wenger,A.M. and Bejerano,G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
48. Hu,D., Gao,X., Morgan,M.A., Herz,H.M., Smith,E.R. and Shilatifard,A. (2013) The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol. Cell. Biol.*, **33**, 4745–4754.
49. Kucuk,C., Iqbal,J., Hu,X., Gaulard,P., De Leval,L., Srivastava,G., Au,W.Y., McKeithan,T.W. and Chan,W.C. (2011) PRDM1 is a tumor suppressor gene in natural killer cell malignancies. *Proc. Natl Acad. Sci. USA*, **108**, 20119–20124.
50. Liu,C., Banister,C.E., Weige,C.C., Altomare,D., Richardson,J.H., Contreras,C.M. and Buckhaults,P.J. (2018) PRDM1 silences stem cell-related genes and inhibits proliferation of human colon tumor organoids. *Proc. Natl Acad. Sci. USA*, **115**, E5066–E5075.
51. Lam,E.W., Brosens,J.J., Gomes,A.R. and Koo,C.Y. (2013) Forkhead box proteins: tuning forks for transcriptional harmony. *Nat. Rev. Cancer*, **13**, 482–495.
52. Golson,M.L. and Kaestner,K.H. (2016) Fox transcription factors: from development to disease. *Development*, **143**, 4558–4570.